IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications
Volume 6 Issue 11
November 2015
ISSN 2156-5570(Online)
ISSN 2158-107X(Print)
SAI
www.ijacsa.thesai.org

Editorial Preface

From the Desk of Managing Editor…

It may be difficult to imagine that almost half a century ago we used computers far less sophisticated than current home desktop computers to put a man on the moon. In that 50 year span, the field of computer science has exploded.

Computer science has opened new avenues for thought and experimentation. What began as a way to simplify the calculation process has given birth to technology once only imagined by the human mind. The ability to communicate and share ideas even though collaborators are half a world away and exploration of not just the stars above but the internal workings of the human genome are some of the ways that this field has moved at an exponential pace.

At the International Journal of Advanced Computer Science and Applications it is our mission to provide an outlet for quality research. We want to promote universal access and opportunities for the international scientific community to share and disseminate scientific and technical information.

We believe in spreading knowledge of computer science and its applications to all classes of audiences. That is why we deliver up-to-date, authoritative coverage and offer open access of all our articles. Our archives have served as a place to provoke philosophical, theoretical, and empirical ideas from some of the finest minds in the field.

We utilize the talents and experience of editor and reviewers working at Universities and Institutions from around the world. We would like to express our gratitude to all authors, whose research results have been published in our journal, as well as our referees for their in-depth evaluations. Our high standards are maintained through a double blind review process.

We hope that this edition of IJACSA inspires and entices you to submit your own contributions in upcoming issues. Thank you for sharing wisdom.

Thank you for Sharing Wisdom!

Managing Editor
IJACSA
Volume 6 Issue 11 November 2015
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

Editorial Board

Reviewer Board Members

CONTENTS

Distance Prediction for Commercial Serial Crime Cases Using Time Delay Neural Networks

Anahita Ghazvini
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia (UKM)
43600 Bangi, Selangor, MALAYSIA

Mohammed Ariff Abdullah
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia (UKM)
43600 Bangi, Selangor, MALAYSIA

Siti Norul Huda Sheikh Abdullah
Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia (UKM)
43600 Bangi, Selangor, MALAYSIA

Md Nawawi Junoh
upt. at Royal Malaysia Police (RMP), Dept. of
Crime Prevention & Community Safety
Royal Malaysia Police Headquarters, 50560
Bukit Aman, Kuala Lumpur, MALAYSIA

Zainal Abidin bin Kasim
SAC at Royal Malaysia Police,
Inspector General Secretariat, R&D
Royal Malaysia Police Headquarters, 50560
Bukit Aman, Kuala Lumpur, MALAYSIA

***Abstract*—The prediction of the next serial criminal time is important in the field of criminology for preventing the recurring actions of serial criminals. In the associated dynamic systems, one of the main sources of instability and poor performances is the time delay, which is commonly predicted based on nonlinear methods. The aim of this study is to introduce a dynamic neural network model by using nonlinear autoregressive time series with exogenous (external) input (NARX) and Back Propagation Through Time (BPTT), which is verified intensively with MATLAB to predict and model the crime times for the next distance of serial cases. Recurrent neural networks have been extensively used for modeling of nonlinear dynamic systems. There are different types of recurrent neural networks such as Time Delay Neural Networks (TDNN), layer recurrent networks, NARX, and BPTT. The NARX model for the two cases of input-output modeling of dynamic systems and time series prediction draw more attention. In this study, a comparison of two models of NARX and BPTT used for the prediction of the next serial criminal time illustrates that the NARX model exhibits better performance for the prediction of serial cases than the BPTT model. Our future work aims to improve the NARX model by combining objective functions.**

Keywords*—Criminology and Computational Criminology; Neural Network; modeling; NARX; BPTT; Quantum GIS

I. Introduction

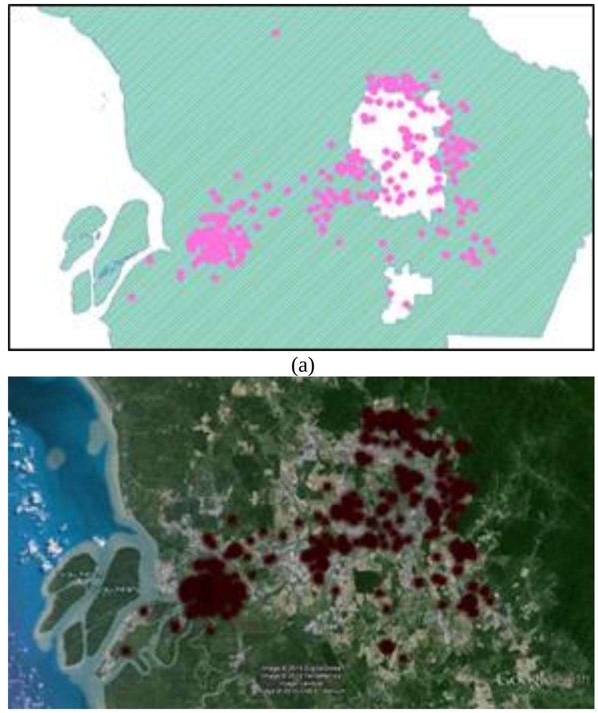
Crime is a long-standing concern of everyday people, governments, and researchers, and an education on crime and criminals has become a requirement for many careers. Criminology, the study of crime, is a complicated field that examines and combines all the connections and disagreements of law, social science and psychology to solve criminal cases. The implementation of criminal behavioral analysis on a database of behavioral patterns can be used to recognize crime patterns; however, a combined modeling technique can be applied to lead to a viewpoint that can state the criminal behavior analysis in a precise, official and computer-supported manner. Computational Criminology uses the power of computers to recognize different types of crimes, such as patterns and developing patterns, crime producers and attractors, and cybercrime. For development of algorithms, Computational topology and dynamic information systems analysis are used. Crime does not appear randomly; it follows rules that are similar to those in various types of non-criminal behavior. Thus, improvement of the understanding of these rules for the development of suitable algorithms for identifying risky parts is the continuing focus of computational criminology [1]. According to computational criminology research, the next crime time can be predicted based on previous data that has been interconnected to serial criminals. Thus, in this study, an appropriate improved method for the future values of time series is selected to create modeling distance crime predictions for serial cases using a time delay neural network, which could help the police to reduce their workload in the investigation of serial criminals.

A neural network is a structure that represents the biological actions of neurons, and it can integrate with a learning or training process such that the known data can be used to alter the coefficients of the neural network. The neural network can function as a way to present the actual output for all of the known inputs in the case of a successful training process. If the units' connections to one another and their weights are set correctly, artificial neural network can perform a particular task. The connections show one unit can influence another unit and the strength of this influence is being identified by weight. In many fields of science and engineering, artificial neural networks are widely used for extensive classification and regression tasks [2]. The time delay in process control can destabilize or degrade the performance of the system. The existent control action component cannot alter the current output unless the time delay elapsed. Therefore, for large time delay the predictor is used. To design the predictor, identification of the time delay parameter that has already been used by the neural network is crucial. Static neural networks are widely used to process modeling and control because of their approximation capabilities, but they usually cannot give satisfactory results in the case of dynamical or time delay systems [3], [4].

Quantitative forecasting methods, based on mathematical modeling, are classified into two models: causal and time series. A linear relationship is used in a casual model based on predicted and independent variables. The time series models have been categorized into two models, namely linear and nonlinear, and are used for the estimation of historical data. The linear models are mainly statistical models; however, the nonlinear models are comprised of both statistical and non-statistical models. Artificial Neural Networks (ANN) and Support Vector Regression (SVR) are worthy instances of nonlinear models. Crime forecasting is one of the principal areas of criminology. Regression and econometric models can explain the fundamental connection of crime variables and other descriptive variable factors to forecast the rate of crime. Time series models are used as an alternative method to forecast crime rates, as some theoretical rules are needed to describe the connectivity of descriptive variables. The normal time series models need a considerable number of observations, but, due to the lack of information in the crime area, these models are not considered appropriate to forecast the rate of crime. Thus, a novel model that outfits a small dataset is required to improve the performance of crime rate forecasting. Time series models are divided into two groups, linear and non-linear models, in which linear and nonlinear relationships can be modeled, respectively.

The effects of three primary factors on the forecasting of nonlinear time series have been investigated by computer simulation-based tests, and the findings showed that neural networks are suitable for both modeling and forecasting of nonlinear time series, whereas traditional linear methods are not [5]. Artificial neural network (ANN) can be simulated via MATLAB to predict maximum and minimum temperatures (two important weather parameters). Based on findings of this research, multilayer neural network predicts weather as a powerful method that correctly provides mapping between input and output by using historical data [6].

Geographical information systems (GIS) can be used to interpret physical components in the real world and their arrangements is shown, operated and analyzed. The two types of data models that can be used by GIS for the interpretation of information are vector and raster information. The basic components of spatial data, such as points, lines, and polygons, can be signified by the vector data, whereas raster data are images, i.e., a grid formed by pixels, and can be a photograph provided by aerial means, satellite or remote sensing. Fig.1 (a) and Fig.1 (b) shows a Malaysian map of crime hot spots.



(a)

(b)

Fig. 1. Malaysian Map for serial crime hotspots in Selangor and Federal District

A raster-type image can be larger or smaller depending on its resolution. GIS analysis is improved by using these two data models. Collection, editing and validation of data can be carried out via spatial analysis to assess the Relations and configuration between them, as well as those with the space surrounding them [7]. A massive number of studies, including criminology ones, used Geographical Information Systems (GIS) and considered a significant development of GIS to have been observed for the analysis of crime, police precinct repositioning, and schemes for the decrease of crime. When police records become computerized, this material can be useful for the recognition of crime and intelligence analyses. [8]. Predictive models, which are geo-statistics tools, can predict the occurrence of incidents by allowance of simple descriptions of phenomena.

This study includes Dynamic Neural Networks using nonlinear autoregressive time series with exogenous (external) input (NARX) and Back Propagation Through Time (BPTT), which is verified intensively with MATLAB to predict and model crime time for serial cases.

The present study is structured as follows: a literature review in Section II, materials and methods in Section III, experimental results in Section IV, the conclusion in Section V, And, finally, discussion and future work in Section VI.

II. Literature Review

A. Neural Network Modelling

Neural Networks are an influential data modeling method that can signify complex input/output relationships to learn to identify patterns in any data, which is a division of the field recognized as "Artificial Intelligence." The primary advantage of Neural Networks is to solve complex problems for conventional technologies that cannot be solved by algorithm, or for which an algorithmic solution is too difficult to define for it. Neural networks are able to solve problems that can be easily solved by people but not by computers, such as pattern recognition and forecasting that the tendency of the data are essential.

Artificial Neural Network implementations have three different neurons, input, hidden and output. Neural networks can use linear and nonlinear relationships directly from the data that is modeled, but, for modeled data with nonlinear characteristics, the traditional linear models are not appropriate. The multi-layer perceptron (MLP), which is known as a superintended network because it requires a specific output to learn, uses the historical data to map the input and output correctly to produce the output when the anticipated output is unknown. An MLP contains nodes in various layers in a fixed graph, for which each layer is totally connected to the following layer. A superintended learning method, error back propagation, was used to train the network [9], [10]. The multilayer perceptron (MLP) mainly uses a feed-forward topology that has an architecture of a layered type with entirely interrelated layers. The back propagation through time (BPTT), a feed-forward time-delay network that is used instead in the case of time-related problems, mostly uses the back propagation variant in the case of recurrent topologies, which is a suitable algorithm for recurrent topologies. The design of time delay networks is similar to MLP design, which is a particular topology of back propagation algorithms.

B. Time Series Neural Network

The precision of pure time delay measurements by traditional methods is not accurate or reliable. For a linear system, the weight can be distributed in place of sampling to place the correct value of the time delay by the neural network, but identification of the time delay of the nonlinear system still remains a difficult problem [3]. A new branch of computing systems is inspired by the human brain as a biological structure for the identification of parameters to provide a new identification method for the time delay. The nonlinear system problems can be identified for theorem of approximation by NN methods due to their capabilities for process modeling and control [11].

For system modeling and forecasting, a recurrent neural network with internal feedback was used [12], [13]. The weakness of static networks can be overcome with dynamic recurrent networks, which have boundaries to recognize the dynamical systems. The nonlinear dynamics method and chaos theory, which are often related with continuous differential equations (but not for discrete variables), are, together, one of the effective techniques for the prediction of difficult natural procedures and variables. The nonlinear dynamics and chaos theory are data driven methods, as they are totally based on the observation data analysis [14].

III. Materials and Methodology

Dataset descriptions and attribute information are followed by pre-processing on the attributes of datasets, pseudo code, pre-processing steps, the NARX method and the BPTT Neural Network.

A. Dataset Descriptions

In this study, two types of datasets were used: (1) Raw Commercial Criminal (SCC) datasets and (2) Crime Plotting (CP) datasets using the SPBS system [15]. SCC provides information about ‘Suspect & Capture,’ while CP shares information about Longitude and Latitude for each matching criminal case from SCC. We obtained ethical approval and used SCC datasets from “PDRM” for the year of 2013 covering only Selangor and Wilayah Persekutuan states to map with CP datasets obtained from “PDRM.” The “Suspect & Capture” information is one of the large datasets, and we combined its plotting from CP datasets to predict the distance for commercial serial criminals. The descriptions of the “Suspect & Capture’’ and “Crime Plotting” datasets are shown in Table 1 and Table 2, respectively.

TABLE I. “Suspect & Capture” Dataset Description

Name of Dataset	Suspect & Capture
Data set characteristics	Multivariate
Attribute characteristics	Categorical, Integer
Number of instances	105934
Number of attributes	14
Missing values	Yes

TABLE II. “Crime Plotting” Dataset Description

Name of Dataset	Crime Plotting
Data set characteristics	Multivariate
Attribute characteristics	Categorical, Integer
Number of instances	136407
Number of attributes	14
Missing values	Yes

In this study, the total number of instances for the “Suspect & Capture” and “Crime Plotting” datasets are 105,934 and 136,407, respectively. The “Suspect & Capture” and “Crime Plotting” datasets both contain fourteen attributes; the Name, No ID and NO Case attributes from “Suspect & Capture” and Latitude & Longitude and NO Case from “Crime Plotting” were used, while No of Case in both datasets plays a key role in combining Latitude & Longitude from one dataset with another for use in the prediction of distance crimes for serial cases.

Based on the number of duplicates, the sum of data for repetition of three and above after the combination of the two above-mentioned datasets comes to 8256. The serial criminal cases of repetitions three and four at the hidden layers of 10, 11, 12, 13, 14 and 15 at the default epoch of 1000 after removing the outliers were examined.

B. Merging into the Serial Commercial Crime & Plotting (SCCP) Dataset

The specific information is needed for every attribute instead of the total description of the dataset. Continues values and String values are two main groups of attributes, which are shown in Table 3.

TABLE III. ATTRIBUTE INFORMATION

NO	Attribute
1	Name
2	No ID
3	No Case
4	Latitude
5	Longitude

C. Pseudo code

***Pseudo code* of the Pre-Processing Steps**

Start

a. **Step 1:** Input attributes.
b. **Step 2:** Sort the suspect names from the "Suspect & Capture "and "Crime Plotting" datasets in ascending order.
c. **Step 3:** Sort the numbers of IDs.
d. **Step 4:** Find the duplicate values from among the suspect names and numbers of IDs.
e. **Step 5:** Find the frequency of each duplicate.
f. **Step 6:** Keep the duplicates with frequencies of three or higher.
g. **Step 7:** Match the "Suspect & Capture" dataset entries to those of "Crime Plotting" by the Number of Cases.
h. **Step 8:** Find the distance between the serial events of A (Latitude1, Longitude2)-B (Latitude1, Longitude2) by:

$Distance(i,j) = \sum_{i,j}^{n,m} \cos^{-1}((\sin(\pi - \theta^{La1}) \times \sin(\pi - \theta^{La2}) + \cos(\pi - \theta^{La1}) \times \cos(\pi - \theta^{La2})) \times \cos(\theta^{Lo2} - \theta^{Lo1})) \times R$, [16].

where i = the number of SCCP criminal records between [1, n], n = the number of records, j = the number of frequencies between [1, m], m = the number of repetitions, La and Lo are the latitude and longitude of two serial criminal events, and R = 6371 Km (the Earth's radius).

End

D. Pre-Processing Steps

The name and the No ID attributes in the "Suspect & Capture "dataset were sorted to find the duplicate values of the above-mentioned attributes. Table 4 illustrates the duplicate values of the name (Sample Name) and No ID (Sample No ID) attributes.

TABLE IV. DATA SORTING AND OBSERVATION OF DUPLICATE VALUES

Name	No ID
ABCD	123456789
ABCD	123456789
ABCD	123456789
XDFT	908765443
XDFT	908765443
XDFT	908765443
XDFT	908765443

Combination results of two "Suspect & Capture" and "Crime Plotting" datasets by using a primary key (No Case) are illustrated in Table 5.

TABLE V. COMBINATION RESULTS OF "SUSPECT & CAPTURE" AND "CRIME PLOTTING" DATASETS BASED ON A PRIMARY KEY (NO CASE)

Name	No ID	No Case	Latitude	Longitude
ABCD	123456789	Pkl88888009	3.106197	101.5467
ABCD	123456789	Pkl88888009	3.2060718432205	101.6068
ABCD	123456789	Pkl88888009	3.0524875306253	101.5377
XDFT	908765443	Lkmp777885433	3.148755792	101.7566
XDFT	908765443	Lkmp777885433	3.148755792	101.7566
XDFT	908765443	Lkmp777885433	3.148755792	101.7566
XDFT	908765443	Lkmp777885433	3.148755792	101.7566
…….	…………	…………….	……………..	…………..

The distances were calculated based on step eight of the pseudo code of the pre-processing steps for the attributes of Name and No ID with three or more repetitions. Table 6. Illustrates the Distance Values.

TABLE VI. DISTANCE VALUES

Name	Number of ID	No Case	Latitude	Longitude	Distance
ABCD	123456789	Pkl88888009	3.106197	101.546686	11.11168
ABCD	123456789	Pkl88888009	3.206071843	101.6068298	17.08294
ABCD	123456789	Pkl88888009	3.052487531	101.5377287	Will be predicted by best model
XDFT	908765443	Lkmp777885433	3.1487558	101.789299	0.071799
XDFT	908765443	Lkmp777885433	3.1487568	101.777543	0.127581
XDFT	908765443	Lkmp777885433	3.14875679	101.756655	9.49E-05
XDFT	908765443	Lkmp777885433	3.14875578	101.756656	Will be predicted by best model
…….	……	……….	………	……	…….

Normalization is mainly suitable for the classification of algorithms concerning neural networks. In this study, because the data are serial and the min and max values of the data in repetitions three and four are unknown, the Z-score normalization method was used. In z-score normalization, the values for the attribute D_V are normalized based on the mean and standard deviation of D_V. The data value of D_V is normalized to D_V' by the computation of (1):

$$D_V' = \frac{D_V - M}{STN\_D} \quad (1)$$

where M and STN_D are the mean and the standard deviation, respectively, of the attribute D_V. This method of normalization is useful when the actual minimum and maximum of the attribute D_V are unknown.

The calculated values of the distance based on the Latitude and Longitude showed that some values of distances are outliers, which are defined as data points that are statistically inconsistent with the rest of the data.

Therefore, to determine the outliers, the percentiles function in Microsoft Excel was used. By using this statistical method, the percentile function calculated the 30th percentile and the 90th percentile of the data to remove the outliers. Fig.2 and Fig.3 show the data of repetitions of three and four (a) with and (b) without the outliers.

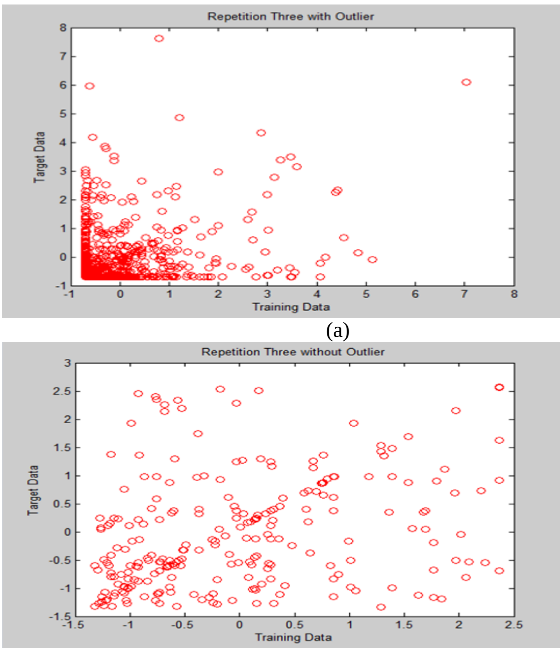


Fig. 2. Data of repetitions of three (a) with and (b) without outliers

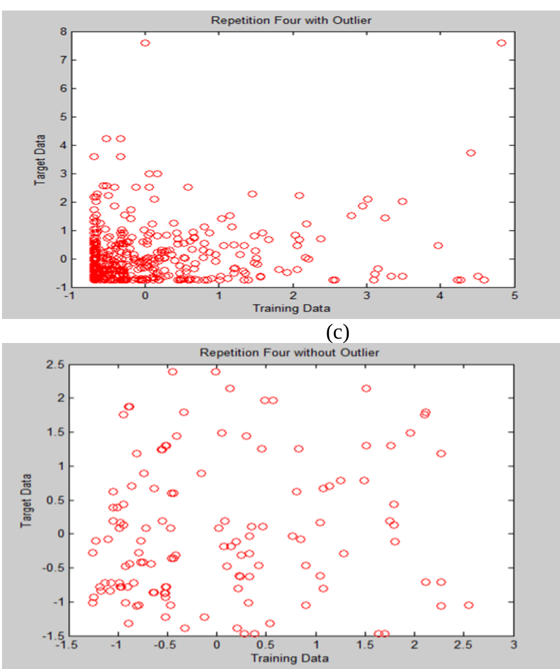


Fig. 3. Data of repetitions of four (c) with and (d) without antlers

E. NARX Neural Network

Recurrent NNs have been widely used for the modeling of nonlinear dynamical systems [17], [18]. The NARX model, which is based on the model of linear ARX, is used to predict the serial criminal time, and, for the classification stage, the back propagation algorithm is used to find the best repetition experimental results. NARX consists of input, hidden, and output layers, where the current and historical inputs and outputs that are provided for the hidden layer are held by the input.

One or more neurons in the hidden layer enable it to produce a nonlinear charting of the affine weighted mixture of the values from the input, whereas the output contains an affine combination of the values from the hidden layer. The dynamical order of inputs and outputs, as well as the number of neurons in each layer of this network, are specified. For determination of a valid training set, some pre- and post-processing should be performed on the input and target values [19]. These procedures record the input and target data in the range of $[-1, 1]$, and normalization of the inputs and targets will be performed to achieve zero mean and variance. Finally, constant inputs and outputs are eliminated for the processing of unknown inputs. (2), which describes the input-output recurrent model of NARX, is as follows:

$$y(t+1) = F(y(t), \ldots., y(t-1), D(n), \ldots., D(t-1)) \qquad (2)$$

This model contains one single input and one single output for repetitions of three and two inputs and one output for repetitions of four. The single input is applied to a tapped delay line memory, whereas the single output is returned to the input through another tapped delay line memory for the purpose of storing the previous values. The input layer of the multilayer perceptron is fed with the content of the two above-mentioned tapped delay line memories, whereas the value of the input of the model is shown as D(n) and the value of the output of the model is presented as y(t+1).

The signal vector provided for the input layer of the multilayer perceptron contains current and past values of the input (exogenous inputs) and delayed values of the output (regressed). Fig.4 shows the NARX model.

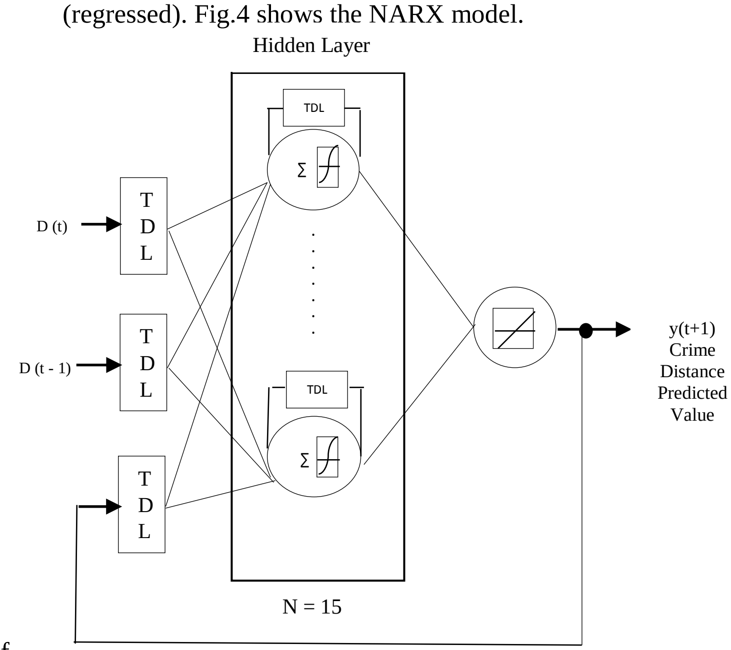


Fig. 4. Nonlinear autoregressive NARX model with exogenous inputs

F. Back Propagation Through Time Neural Network

The back propagation through time (BPTT) algorithm is an extension of the standard Back propagation algorithm for recurrent neural networks [20]. The first step of the BPTT algorithm contains the unfolding process of the network which involves replicating times of the recurrent network to obtain a corresponding feed-forward network. In this procedure, each replicated connection shares its value in all times, which can be trained by using the Back propagation algorithm. In the forward process of Back propagation, the calculation of output y_i at time t of each neuron can be defined in (3) to (7) as follows:

$$Y_{i(t)} = f(D_i\ (t)) \quad (3)$$

$$D_i(t) = \sum_{j\in H} Y_i(t)W_{ij} + \sum_{j\in i} D_j^{in}\ (t)W_{ij} + \sum_{J\in M} y_j\big(t - T_{ij}\big)W_{ij} \quad (4)$$

Where f = the neuron activation function, H = the hidden layer indexes, i = the input neuron indexes, and $D_j^{\ in}$ is the j^{th} input neuron; however, M indicates the indexes of neurons that supply information regarding the previous network stages, and T_{ij} is an integer value representing the displacement in recurrent connections through time. Regarding the error of the back propagation process, each neuron has an error size, and, for the output layer, this error value is considered as the variation between the predictable and the obtained value (4), while, for the hidden layers, it can be gained by taking into account the error in the successor layers (5).

$$\delta_j(t) = (d_j - y_j)y_j(1 - y_j) \quad (5)$$

$$\delta_j(t) = y_j\big(1 - y_j\big)\sum_{i\in suc}(J)W_{ij}\delta_j \quad (6)$$

The weights are calculated with regard to (6), where the necessary change in weights is computed based on the learning rate α:

$$\Delta\, W_{ij}^{e+1} = \alpha\delta_j y \quad (7)$$

where e = the order of the weight updates in the learning process. The second phase of the BPTT algorithm continues to process (the folding process) when the back propagation algorithm is used to obtain the FFN. The original recurrent network is obtained in this stage, where the connection weights are the combination of the corresponding connections on each fold. “Fig.4” shows the BPTT with its Architecture Graph of two neurons, and the Single flow graph of network 15 unfolds in time where D is the actual distances value. Fig.5 shows the BPTT model.

G. Evaluation Metrics

There are various statistical metrics to validate the performance of the models, but a more commonly used benchmark metric is to use the regression coefficient (R) and mean square error (MSE), which are used in this paper. The R criterion is given by (8) as follows:

$$\mathrm{R} = \frac{R_{yt}}{std(y)std(t)} \quad (8)$$

where *y* and *t* are the predicted and actual distance of the serial criminal, *Ryt* = the covariance between *y* and *t*, and *std*(y) and *std*(*t*) represent the standard deviation of *y* and *t*, respectively. The value of the regression coefficient varies between 0 and 1.

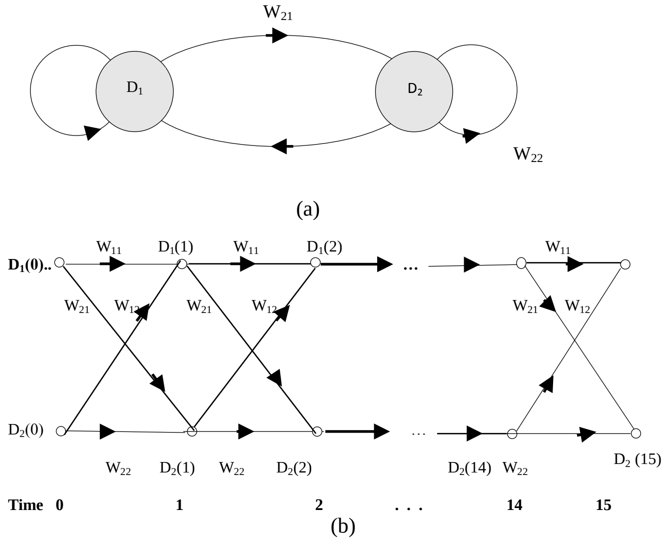


Fig. 5. Back Propagation Through Time model (a) Architecture Graph of two neurons and (b) Single flow graph of network 15 unfolded in time

The value being near to 0 indicates that the output is deviating and near to 1 that the output matches the target.

The mean square error (MSE) criterion is widely used to find the difference between the values gained by the given model and the actual values that are expected. As the MSE is less, the model is more accurate and reliable.

The MSE criterion is given by (9) as follows:

$$MSE = \frac{1}{N}\ \sum_{I=1}^{N}(t_i - y_i)^2 \quad (9)$$

where *N* is the number of predicted sample.

IV. EXPERIMENTAL RESULTS

This research aims to solve the problem of next crime time prediction for serial cases based on time delay neural networks. To predict the time series data, the NARX and BPTT methods were used to create a model for Distance Crime Prediction for serial cases using Time Delay Neural Networks. The NARX and BPTT models are examined and compared for each dataset for repetitions of three and four with 558 and 198 instances (after removing the outliers with the percentile method) of time series criminal data, respectively. The performances of serial criminal cases of repetitions of three and four were examined at the default epoch of 1000. The NARX network is a feed-forward network, for which its hidden layer contains a tan-sigmoid transfer function and its output layer contains a linear transfer function as its default. NARX has two inputs, an external input and a feedback connection from the network output that can be closed after training of the network. There is a tapped delay line to stock previous values of each of the above mentioned inputs to allocate the network architecture for a NARX network, and it must choose the delays related with each tapped delay line. The input delays and the feedback delays can be set in a range from 1 to 4, and, as the delay number increases, more calculation is required and over-fitting of the data can occur. However, an overly high delay number can allow the network to solve complex problems. In this study, the delay number is set to 2, and the number of hidden layers of 10, 11, 12, 13, 14 and 15 were chosen to examine the performance of the NARX and BPTT models.

This study is considered a one-step-ahead prediction for the performance of all the experiments. Tables 7 and 8 show the performance measures of the NARX model, and Tables 9 and 10 present the performance of the BPTT model with various hidden layer selections for repetitions of three and four. It was decided to vary the hidden layers as it has been proven that, a single hidden layer with sufficient neurons in a neural network will give the desired accuracy for any nonlinear function [21], [22], [23]. Fig.6 and Fig.7 show comparisons of the accuracy versus hidden layers for the NARX and BPTT Models for repetitions of three and four.

TABLE VII. PERFORMANCE METRIC COMPARATIVE RESULTS FOR REPETITION OF THREE WITH THE NARX MODEL

Hidden Layer	R	MSE	RMSE	R^2
10	0.4334	0.8192	0.905	0.1878
11	0.4186	0.8479	0.9208	0.1752
12	0.4272	0.8223	0.9068	0.1825
13	0.4227	0.8238	0.9086	0.1787
14	0.4008	0.8532	0.9237	0.1606
15	**0.4570**	**0.8149**	**0.9027**	**0.2088**

TABLE VIII. PERFORMANCE METRIC COMPARATIVE RESULTS FOR REPETITION OF FOUR WITH THE NARX MODEL

Hidden Layer	R	MSE	RMSE	R^2
10	0.4571	0.8598	0.9273	0.2089
11	**0.7441**	**0.5154**	**0.7179**	**0.5537**
12	0.7420	0.5292	0.7274	0.5506
13	0.6344	0.6114	0.7819	0.4025
14	0.5013	0.7494	0.8657	0.2513
15	0.7340	0.5701	0.7551	0.5388

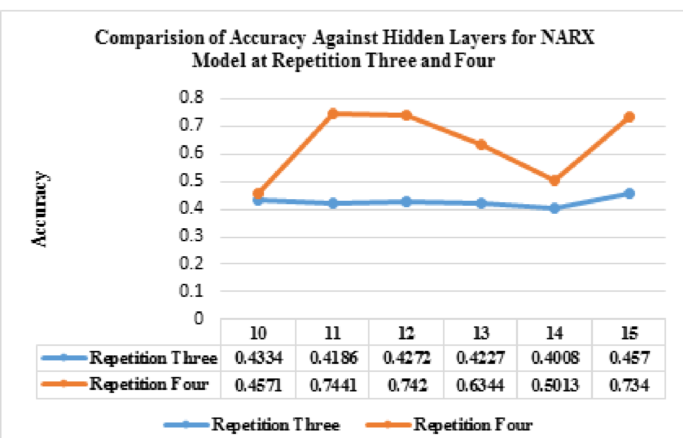


Fig. 6. Comparison of the accuracy versus the hidden layers for the NARX model with repetitions of three and four

Tables 9 and 10 present the NARX model performance measurements with different hidden layers for repetitions of three and four. The observation of results showed that the NARX model with 15 hidden layers for repetition of three and 11 hidden layers for repetition of four yields good accuracy for the prediction with regard to lower error measures (0.8149 and 0.5154) and a high accuracy coefficient (0.4570 and 0.7441) for repetitions of three and four, comparatively.

TABLE IX. PERFORMANCE METRIC COMPARATIVE RESULTS FOR REPETITION OF THREE WITH THE BPTT MODEL

Hidden Layer	R	MSE	RMSE	R^2
10	**0.4462**	**0.8151**	**0.9029**	**0.1991**
11	0.4188	0.823	0.9072	0.1754
12	0.4108	0.8504	0.9222	0.1688
13	0.4057	0.8512	0.9387	0.1646
14	0.4398	0.8196	0.9053	0.1934
15	0.4340	0.8240	0.9078	0.1884

TABLE X. PERFORMANCE METRIC COMPARATIVE RESULTS FOR REPETITION OF FOUR WITH THE BPTT MODEL

Hidden Layer	R	MSE	RMSE	R^2
10	0.4533	0.8867	0.9416	0.2055
11	0.4853	0.8165	0.9036	0.2355
12	0.5042	0.783	0.8849	0.2542
13	0.4347	0.9019	0.9497	0.1890
14	0.5600	0.7378	0.8595	0.3136
15	**0.5604**	**0.7005**	**0.8370**	**0.3140**

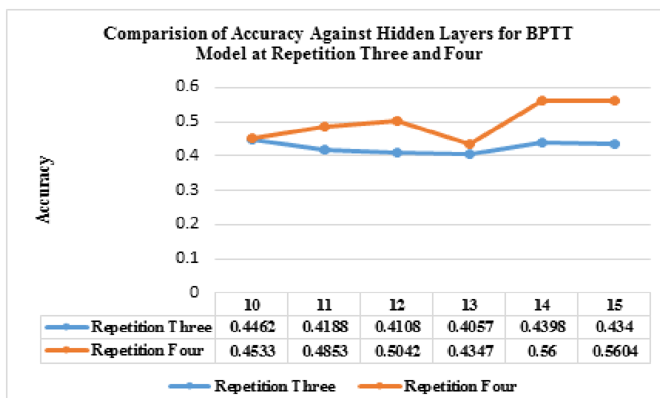


Fig. 7. Comparison of the accuracy versus the hidden layers for the BPTT model with repetitions of three and four

Tables 9 and 10 present the BPTT model performance measurements with different numbers of hidden layers for repetitions of three and four. The observation of results showed that the BPTT model with 10 hidden layers for repetitions of three and 15 hidden layers for repetitions of four yields good accuracy for prediction with regard to lower error measures (0.8151 and 0.7005) and a high regression coefficient (0.4462 and 0.5604) for repetitions of three and four, comparatively.

The comparison of the NARX and BPTT models indicates that the NARX model yields more accurate results than the BPTT model regarding the mean and standard deviation; thus, to evaluate the performance of the NARX and BPTT models, two repetitions were used for the training, validation and testing phases of both the NARX and BPTT models; the performance metrics with the mean and STDV are presented in Table 11. Fig.8 shows a comparison of the accuracy values of the NARX and BPTT models.

TABLE XI. PERFORMANCE COMPARISON FOR THE NARX AND BPTT MODELS

Repetition	NARX MODEL				BPTT MODEL			
	R	***MSE***	***RMSE***	***R^2***	***R***	***MSE***	***RMSE***	***R^2***
3	**0.4570**	0.8149	0.9027	0.2089	**0.4462**	0.8151	0.9029	0.1991
4	**0.7441**	0.5154	0.7179	0.5537	**0.5604**	0.7005	0.8370	0.3140
Mean	0.6006				0.5033			
STDV	0.2030				0.0808			

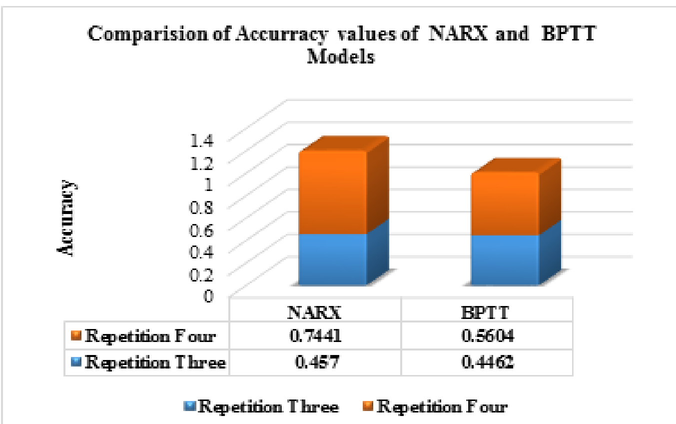
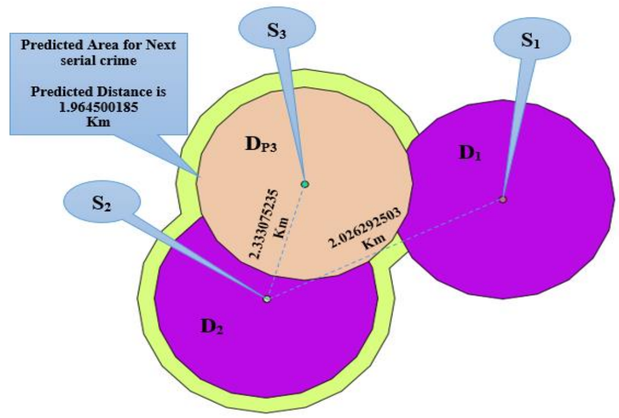


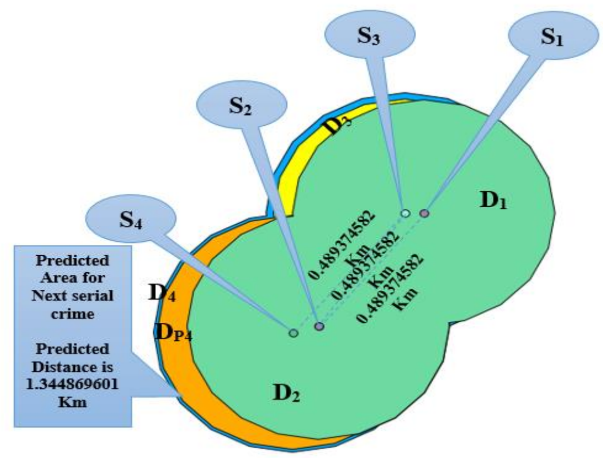
Fig. 8. Comparison of the Accuracy values of the NARX and BPTT Models

H. Quantum Geographical Information System (QGIS)

In this study, the actual and predicted values have been demonstrated by Google Earth. From the results of this study, the NARX model showed better performance, so the predicted value from this model was used. To plot the differences of distance between the actual and predicted values for repetitions of three and four, one case of a repetition of three and one case of four were selected randomly. The repetitions of three and four contain three and four distance values, respectively; the first two distance values in the case of the repetition of three and the first three for the repetition of four were calculated based on the formula mentioned in step eight of the pseudo code from the pre-processing steps, and then the third and fourth distance values of the repetitions of three and four are predicted by the NARX model. To present the differences between the first two distance (D_1, D_2) values with the predicted distance (D_{P3}) value in the repetition of three and the first three distance (D_1, D_2, D_3) values with the predicted distance (D_{P4}) value in the repetition of four, buffering was created around the latitude and longitude of each serial event for the repetitions of three and four by a Quantum Geographical Information System (QGIS). The QGIS was used for the buffering of each point that was calculated from the distance values for the repetitions of three and four to be used by Google Earth to show the exact area of each serial event and the differences of the actual from the predicted distance to prevent the next serial event. Fig.9 shows buffering for the repetitions of three and four to present the differences of distance between the actual and predicted values by the QGIS, and Fig.10 shows the differences between the first two distance values with the predicted distance value for (a) the repetition of three and (b) the repetition of four serial criminal events at Selangor and Wilayah Persekutuan by Google Earth.



(a)



(b)

Fig. 9. Differences between the first two distance values with the predicted distance value for (a) the repetition of three and (b) the repetition of four serial criminal event at Selangor and Wilayah Persekutuan by the QGIS

First distance (D_1) between S_1 and S_2 of the serial events for the repetition of three at a zoom of 11.29 mi
(a)

First distance (D_1) between S_1 and S_2 of the serial events for the repetition of four at a zoom of 17101 ft
(b)

Second distance (D_2) between S_2 and S_3 of the serial events for the repetition of three at a zoom of 11.29 mi
(a)

Second distance (D_2) between S_2 and S_3 of the serial events for the repetition of four at a zoom of 17101 ft
(b)

Third distance (D_3) between S_3 and S_{P4} of the serial events for the repetition of three at a zoom of 11.29 mi
(a)

Third distance (D_2) between S_3 and S_4 of the serial events for the repetition of four at a zoom of 17101 ft
(b)

Third distance (D_2) between S_3 and S_{P4} of the serial events for the repetition of four at a zoom of 17101 ft

(b)

Fig. 10. Differences between the first two distance values with the predicted distance value for (a) the repetition of three and (b) the repetition of four serial criminal events at Selangor and Wilayah Persekutuan

V. Conclusion

For prediction of next crime distances NARX and BPTT models were chosen due to their capabilities. In this research, the two models were tested for modeling distance crime prediction of serial cases by using time delay neural networks for repetitions of three and four serial cases after removing the outliers. Serial criminal cases of repetitions three and four were examined with hidden layers of 10, 11, 12, 13, 14 and 15 for both models. The best performances of NARX is observed with hidden layers of 15 for repetition three and 11 for repetition four, while this is at hidden layers of 10 and 15 for BPTT model. The results of this study have illustrated that the accuracy of the NARX model for repetitions three and four are 0.4570 and 0.7441, respectively, whereas, for the BPTT model, they are 0.4462 and 0.5604, respectively. Based on mean and standard deviation measurements, the results stated that NARX model has better performance than the BPTT model.

VI. Discussion and Future Work

NARX is a well-known model for dynamic systems, not only for being an accurate model but also for its simplicity of training. The use of the Lavenberg algorithm to create the NARX model has some advantages and disadvantages. As a disadvantage, there are various complications in the training process in NARX relating to weights or the number of connections, which can cause overtraining of the data and the wrong fit, which then cannot yield a better prediction. According to Levenberg-Marquardt optimization, modification of the training functions of NARX can update the weight and bias, including a regularization technique to minimize the squared errors and weights to present a correct combination for a network that is well generalized. Additionally, for a network with a few hundred weights, faster convergence occurs to solve the problem of function approximation, and this advantage in the case of an accurate training requirement is more obvious in the NARX model. Regarding the disadvantages of this network, it can be noted that, as the weights in the network rise, the advantage of the Levenberg algorithm will be reduced [24]. BPTT is, in fact, a back propagation algorithm with memory of the previous outputs, which is considerably faster for the training of recurrent neural networks than general purpose optimization techniques. A low speed of convergence and the possibility of becoming trapped in local minima of the error surface are disadvantages of BPTT [25].

The results illustrate that the NARX network shows better results for distance crime time prediction than Back Propagation Through Time (BPTT). Additionally, using the Quantum Geographical Information System (QGIS) can show the difference between the predicted values of the next crime distance and our actual values in the NARX model and give a better understanding of the entire concept. Because the NARX approach is a more suitable model for the prediction of the next crime distance, in our future work, the NARX model will be combined with objective functions that are directly minimized during training to improve the NARX model by getting more accurate results of prediction with a lower mean square error, causing the predicted value of the next crime distance to be much closer to the actual value. As a function of the weights, the total error or the average error for the training set is usually minimized by the NN training performance, which can lead to overfitting and poor generalization when the number of training cases is small in comparison to the complexity of the network, so regularization (a function of the weights) can be used as a method to improve generalization in a way that minimizes an objective function by adding the total error function and a regularization function.

Acknowledgement

This research is based on two fundamental research grants from Exploration Research Grant (ERGS) "ERGS/1/2013/ICT02/UKM/02/4 from Ministry of Education, entitled "GEO-TEMPORAL CRIME NAVIGATION BASED ON MULTI-OBJECTIVE TIME DELAY NEURAL NETWORK". I also would like to thank previous the CAIT researchers such as Associate Prof. Dr. Siti Norul Huda Sheikh Abdullah, PHD Candidate Mohd Ariff bin Abdullah, Datuk' Md Nawawi Junoh. Supt police, Dato' Zainal Abidin bin Kasim.

References

[1] P. L. Brantingham, and S. Fraser, "Computational Criminology", European Intelligence and Security Informatics Conference, pp, 3-5, IEEE Computer Society, DOI: 10.1109/EISIC.2011.79, 2011.

[2] S. Haykin, "A Comprehensive Foundation", 2^{nd} edition, Prentice Hall PTR Upper Saddle River, NJ, USA, 1999.

[3] A. Yazdizadeh, K. Khorasani , and R. V. Patel "Identification of a Two-Link Flexible Manipulator Using Adaptive Time Delay Neural Network" in IEEE Trans. On System, Man and Cybernetics Part B: Cybernetics. Vol.30, No. l, pp. 165-172, 2000.

[4] H. Karlene, A. Sinzinger, D. Eric, and M. J., "Improvements in the Predictive Capability of Neural Networks" In Journal of Process Control, Vol. 12, pp. 193-202, 2002.

[5] G. P. Zhang, B. E. Patuwo, M. Y. Hu, "A simulation study of artificial neural networks for nonlinear time-Research 28, 381-396, 2001.

[6] N. Kumar, G. Kumar Jha, "A Time Series ANN Approach for Weather Forecasting" in International Journal Of Control Theory and Computer Modeling (IJCTCM) Vol.3, No.1, January 2013.

[7] J.H., Ratcliffe, "Strategic Thinking in Criminal Intelligence", Sydney: Federation Press.2004.

[8] J. Ferreira, P. João, and J. Martins, "GIS for Crime Analysis: Geography for Predictive Models" in The Electronic Journal Information Systems Evaluation Volume 15 Issue 1, pp36-49, 2012.

[9] Rosenblatt, Frank. X., "Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms" in Spartan Books, Washington DC, 1961.

[10] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning Internal Representations by Error Propagation", D. E. Rumelhart, J. L. McClelland, and the PDP research group. (Editors), Parallel distributed Processing: Explorations in the microstructure of cognition, Volume 1: Foundations. MIT Press, 1986.

[11] P. J. Werbos, "the Roots of Backpropagation: From Ordered Derivation to Neural Networks and Political Forecasting, February 1994.

[12] A. Balestrino, F. B., and A. Landi, "On-line Process estimation by ANNs and Smith Controller design" in IEE Proc.-Control Theory Appl., Vol.145, No.2, 1998.

[13] K. S. Narendra, and K. Parthasarathy, "Identification and Control of Dynamical Systems Using Neural Networks" in IEEE Trans. Neural Networks, Vol. 1, No. 1, pp. 4-27, 1990.

[14] Z. Hongyu, Z. jian, and C. LiMin, "Elman Network-Based Implicit Generalized Predictive Control of Superheated Steam Temperature", Control Theory and Applications, Vol.22, No.7, pp l3-16, 2003.

[15] Siow Suan Neo, "Safe City Monitoring System", GIS Web Based Applications in Crime Monitoring in Malaysia, Geospatial World Forum, Geneva, Switzerland, May 2014.

[16] Deriving the haversine formula, Ask Dr. Math, Apr. 20–21, 1999.

[17] Siek, M. and D. P. Solomatine, "Nonlinear chaotic model for predicting storm surges", Nonlin. Processes Geophys. 17, 405–420, 2010.

[18] A. Waibel, T. Hanazawa, G. Hinton, G., and S. Kiyohiro, "Phoneme Recognition Using Time-Delay Neural Networks", in IEEE Transactions on Acoustics, Speech, and Processing, Vol. . 37, NO. 3, March 1989.

[19] L. Ljung, System Identification: Theory for the User, 2nd ed. Prentice Hall PTR, 1998.

[20] P. J. Werbos, "Backpropagation through Time: What it does and How to do it", Proceedings of the IEEE, Vol. 78 NO. 10, pp. 1550–1560, 1990.

[21] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feed forward network are universal approximators", Neural Networks, Vol. 2 NO. 5, pp. 359–366, 1989.

[22] G. Cybenko, "Approximations by Superpositions of a Sigmoidal Function", Mathematics of Control, Signals and Systems, Vol. 2, pp. 303–314, 1989.

[23] G. Zhang, B. E. Patuwo, and M. Y. Hu, "Forecasting with artificial neural networks: The state of the art", International Journal of Forecasting", Vol. 14, pp. 35–62, 1998.

[24] Diaconescu, E., "The use of narx neural networks to predict chaotic time series, WEEAS Transactions on Computer Research", issue 3, volume 3, March 2008.

[25] Mastorocostas, P., Varsamis, D., Mastorocoastas, C., Rekanos, I., "An Accelerating Learning Algorithm for Block-Diagonal Recurrent Neural Networks", IEEE Proceedings of the International Conference on Computational Intelligence for Modelling, Control and Automation, and International Conference Intelligent Agents, 2005.

Using the Sub-Game Perfect Nash Equilibrium to Deduce the Effect of Government Subsidy on Consumption Rates and Prices

Dr. Magdi Amer
Department of Electrical Engineering
Faculty of Engineering, Fayoum University
Fayoum, Egypt

Dr. Ahmed Kattan
Department of Computer Science
Faculty of Computer and IS, Umm Al-Qura University
Makkah, Saudi Arabia

***Abstract*—Governments are interested in inducing positive habits and behaviors in its citizens and discouraging ones that are harmful to the individual or to the society. Taxation and legislation are usually used to discourage negative behaviors. Subsidy seems the politically correct way to encourage positive behaviors. In this paper, the Subgame Perfect Nash Equilibrium is used to deduce the effect of the government subsidy on the user consumption, prices and producer and distributor profits.**

Keywords—game-theory; subsidy

I. Introduction

Governments all over the world have been trying to use taxation and subsidies to modify the behavior of its citizens, providing them with incentives to pursue actions that will benefit the society and discouraging them from those actions that are harmful to the individual and the society.

To fight smoking, governments had to go up against smoking companies in order to acquire the public approval needed to implement policies discouraging smoking as described in McKenzie [1]. The publicity war was won to the level that even heavy smokers were supportive for policies aiming to prevent youth from smoking, as proven in Lazuras [2]. Naiman [3] has proved that a full smoking ban was also associated in a decrease in the number of smokers and on secondhand smoke exposure in private settings.

Governments have also tried to solve traffic congestion by applying a congestion pricing on high-traffic roads. Small [4] and Ze-bin [5] have studied the effect of applying a congestion pricing on roads, both of which have shown, along with actual experience, that congestion-pricing car affects citizens' behavior and reduce traffic congestion. This was achieved as a congestion pricing encourages people commuting in a non-congested time, using non-congested roads and changing their transportation manners to use carpooling and public transportations [4]. The challenge was to acquire the political approval for such a system, as the citizens need to feel that it is just another added tax. Nevertheless, people gaining middle and lower incomes do not benefit from such a system, and when the saved time by the policy is not obvious, the high-income group and businesses do not benefit either [5].

Obesity and chronic diseases associated with poor dietary choices have become the leading cause of death [6]. When it comes to dissuade un-healthy eating habits, taxes or imposing higher service prices on over-weight citizens (fat-taxes) does not seem a valid alternative due to the high political opposition that such laws may generate and governments need to start to gather support for fat-taxes to make such policies acceptable by the voters [7]. Imposing taxes on unhealthy food is also unfair as they are imposed on both thin and fat citizens and are usually ineffective, as wealthy consumers are not affected by food prices while consumers with low incomes have the tendency to buy cheap energy-dense food [8].

There have been some attempts to impose taxes on un-healthy foods, but, in some cases, the government had to abandon such attempts due to the public opposition [9]. On the other hand, governments' attempts to encourage healthy eating habits through subsidies and incentives seem to be the politically correct choice. Such approaches are costly, as the cost of a statistical life saved has been estimated to be 1.29 million $ in the USA [9].

In our research, we study the government incentives effect to increase healthy eating habits. Our work is based on Stackelberg [10], who has built a model in 1934 for the market behavior when there is a leader firm. Chen [11] and Simaan [12] have built the mathematical model for Stackelberg problems in the field of game theory. Shubik [13] has written a review on the application of game theory in economics. De Fraja [14] has established the mathematical framework for the research in mixed oligopoly market inspired from a survey of research in this area.

A mixed oligopoly market is a market where public and private firms compete where private firms aim to maximize their profit and public firms are interested in optimizing social targets [15]. Pal [16] has studied the effect of privatization in a mixed oligopoly with private, public and foreign firms when the government maximizes welfare through subsidy or through imposing tariffs on the foreign firms. He obtained interesting results such as privatization always improves welfare when using subsidies and it improves welfare when using tariffs if there are more than one public and one private company in the market and k is large, where k is the constant of the cost function c(q) = k q2, and q is the production level. In all cases, privatization when using subsidies decreases the consumption and increases prices. Welfare is defined as the profits of local companies, public and private, plus the amount of market surplus minus the actual subsidy paid by the government plus the tariff revenue. Pal [16] has also proven that subsidy increases welfare more than tariff regardless of the privatization if and only if k is small. Poyago-Theotoky [17] has studied the effect of subsidy on the production levels of companies in a mixed oligopoly. He has proven that the optimal subsidy that maximize welfare is the same whether all companies, including the public company, act as a profit maximizers or the public company act as a Stackelberg leader or the public company moves simultaneously with the other companies. Q. Gu [18] has studied the effect of subsidy for home-appliance replacement on the price of appliance.

II. BUILBING THE MATHEMATICAL MODEL

The goal of our research is to provide the Saudi government with a plan to manage the governmental product subsidy to increase the consumption of more healthy food. To model the market, we assume that there is a clear leader of the market for every subsidized product, as defined by Stackelberg [10]. We also assume that the production cost and the distribution cost are fixed, independent on the consumption levels and external parameters. We also assume that all players are competing in a free market and their goal is to maximize their profits. We also assume that all the producers and all the retailers form a cartel, thus behave like a single producer and a single retailer, as proven by Sertel [19].

In this research, the government is not interested in increasing welfare function, defined as the profit of local companies, public and private, plus the amount of market surplus plus the revenue that the government may gain from tariff and taxes minus any expenses the government pays such as subsidies. Instead, the government is only interested in increasing consumption, similar to the approach taken by Merrill [15]. This makes sense as the increase in the consumption of healthy food would reflect on the general health of the citizen, which benefits surpluses any short-term financial gain that the government would have obtain from maximizing the welfare function.

In this paper, we are going to discuss the effect of a product subsidy to the producer and the distributor or to the consumer both in case of free market and in a price-controlled market.

A. The Market Without a Government Subsidy

Let us first examine the market before any government intervention. Let us assume that the producer's unit sale price is P_{1p} and the unit sale of the retailer is P_{1r}. Let us also assume that the production cost of a unit product is C_p and the retailer operation cost of carrying a unit product is C_r.

Let us also assume that the demand for a product is a decreasing function of the sale price of the unit product, represented by D_1. $D_1 = K - \alpha P_{1r}$, where K is a positive parameter representing the market demand the product while α is a positive parameter representing the sensitivity coefficient of the end-customer for the retail price of the product unit.

The profit of the producer is $\Pi_{1p} = (P_{1p} - Cp)(K - \alpha P_{1r})$

The profit of the retailer is $\Pi_{1r} = (P_{1r} - Cr - P_{1p})(K - \alpha P_{1r})$

Using the assumption that the producer has enough market power to act as the market leader, we apply the Sub-game Perfect Nash Equilibrium (SPNE) [20 and 21] to calculate the retailer maximum profit given the sale price of the producer. This is given by differentiating Π1r with respect to P1r and setting it to zero for maximization.

$\partial\Pi_{1r}/\partial P_{1r} = K - 2\alpha P_{1r} + \alpha C_r + \alpha P_{1p} = 0$

$P_{1r} = (K + \alpha C_r + \alpha P_{1p}) / 2\alpha$, which represents the best response function of the retailer in response to the price of the market leader, the producer.

Given the best response function of the retailer, the output that maximizes the producer's profit is found by differentiating Π_{1p} with respect to P_{1p} and setting it to zero.

$\partial\Pi_{1r}/\partial P_{1r}$

$= \partial/\partial P_{1r}\,[(P_{1p} - C_p)(K - \alpha[(K + \alpha C_r + \alpha P_{1p}) / 2\alpha])]$

$= 0$

This results is $P_{1p} = (K + \alpha C_p - \alpha C_r) / 2\alpha$.

Consequently, $P_{1r} = (3K + \alpha C_p + \alpha C_r) / 4\alpha$.

The consumption in this case is D_1

$= (K - \alpha[(3K + \alpha C_p + \alpha C_r) / 4\alpha])$

$= ¼ K - ¼ \alpha (C_p + C_r)$

The maximum profit of the producer and the retailer will be

$\Pi_{1p\,max} = (P_{1p\,max} - Cp)(K - \alpha P_{1r\,max})$

$= [K - \alpha (C_p + C_r)]^2 / 8\alpha$

$\Pi_{1r\,max} = (P_{1r\,max} - Cr - P_{1p\,max})(K - \alpha P_{1r\,max})$

$= [K - \alpha (C_p + C_r)]^2 / 16\alpha$

The previous calculation allowed us to express the price of the producer and the retailer that maximize their profits, the consumption and the maximum profit of the producer and of the retailer as a function of the market demand for the product and the sensitivity of the end-customer consumption to the retail price of the product without any governmental intervention.

To simplify the calculations, let's assume that K'

$= K - \alpha (C_p + C_r)$

$\Pi_{1p\,max} = [K']^2 / 8\alpha$

$\Pi_{1r\,max} = [K']^2 / 16\alpha$

B. The Market in Case of a Government Subsidy to the Producer and Retailer

Now, let us assume that in order to encourage the increase of a product consumption, the government decides to subsidies the producer and the retailer of this product with βp and βr respectively.

Thus, the profit of the producer becomes $\Pi_{2p} =$

$(P_{2p} - (C_p - \beta_p))(K - \alpha P_{2r})$.

The profit of the retailer becomes Π_{2r} =

$(P_{2r} - (C_r - \beta_r) - P_{2p}) (K - \alpha P_{2r})$.

Similar to the previous calculations, the prices of the producer and the retailer that maximize their profits when the producer is the effective market leader are:

$$P_{2p} = (K + \alpha (C_p - \beta_p) - \alpha (C_r - \beta_r)) / 2\alpha$$
$$= ½ k/\alpha + ½ (C_p - C_r) - ½ (\beta_p - \beta_r)$$
$$= P_{1p} + ½ \beta_r - ½ \beta_p$$

$$P_{2r} = (3K + \alpha (C_p - \beta_p) + \alpha (C_r - \beta_r)) / 4\alpha$$
$$= ¾ k/\alpha + ¼ (C_p + C_r) - ¼ (\beta_p + \beta_r)$$
$$= P_{1r} - ¼ (\beta_p + \beta_r)$$

The consumption in this case is

$$D_2 = (K - \alpha [¾ k/\alpha + ¼ (C_p + C_r) - ¼ (\beta_p + \beta_r)])$$
$$= ¼K - ¼ \alpha (C_p + C_r) + ¼ \alpha (\beta_p + \beta_r)$$

Knowing that the consumption that maximize the profit of the producer and retailer without government subsidy is D_1 = $¼K - ¼ \alpha (C_p + C_r)$

$$D_2 = D_1 + ¼ \alpha (\beta_p + \beta_r)$$

The maximum profit of the producer and the retailer are:

$$\Pi_{2p\,max} = [K - \alpha (C_p + C_r) + \alpha (\beta_p + \beta_r)]^2 / 8\alpha$$
$$\Pi_{2r\,max} = [K - \alpha (C_p + C_r) + \alpha (\beta_p + \beta_r)]^2 / 16\alpha$$

Using $K' = K - \alpha (C_p + C_r)$

$$\Pi_{2p\,max} = [K' + \alpha (\beta_p + \beta_r)]^2 / 8\alpha$$
$$= [(K')^2 + 2\alpha (\beta_p + \beta_r) K' + \alpha^2 (\beta_p + \beta_r)^2] / 8\alpha$$
$$= \Pi_{1p\,max} + ¼ (\beta_p + \beta_r) K' + ⅛ \alpha (\beta_p + \beta_r)^2$$
$$\Pi_{2r\,max} = \Pi_{1p\,max} + ⅛ (\beta_p + \beta_r) K' + (\alpha (\beta_p + \beta_r)^2) / 16$$

C. The Market in Case of a Government Subsidy to the Consumer

Let us assume that the government decides to subsidies the consumer with an amount β_c for every product unit.

Thus, the profit of the producer becomes Π_{3p} =

$(P_{3p} - C_p) (K - \alpha (P_{3r} - \beta_c))$.

The profit of the retailer becomes Π_{3r}

$= (P_{3r} - C_r - P_{3p}) (K - \alpha (P_{3r} - \beta_c))$.

The prices of the producer and the retailer that maximize their profits are:

$$P_{3p} = (K + \alpha \beta_c + \alpha C_p - \alpha C_r) / 2\alpha$$
$$= ½ k/\alpha + ½ (C_p - C_r) + ½ \beta_c$$
$$P_{3r} = (3K + 3\alpha \beta_c + \alpha C_p + \alpha C_r) / 4\alpha$$
$$= ¾ k/\alpha + ¼ (C_p + C_r) + ¾ \beta_c$$

The consumption in this case is

$$D_3 = (K - \alpha ([¾ k/\alpha + ¼ (C_p + C_r) + ¾ \beta_c] - \beta_c))$$
$$= ¼K - ¼ \alpha (C_p + C_r) + ¼ \alpha \beta_c$$

Knowing that $D_1 = ¼K - ¼ \alpha (C_p + C_r)$

$$D_3 = D_1 + ¼ \alpha \beta_c$$

The profit of the producer and the retailer are:

$$\Pi_{3p\,max} = (P_{3p\,max} - C_p) (K - \alpha (P_{3r\,max} + \beta_c))$$
$$= [K + \alpha \beta_c - \alpha (C_p + C_r)]^2 / 8\alpha$$
$$\Pi_{3r\,max} = (P_{3r\,max} - {}_{Cr} - P_{3p\,max}) (K - \alpha P_{3r\,max})$$
$$= [K + \alpha \beta_c - \alpha (C_p + C_r)]^2 / 16\alpha$$

Using $K' = K - \alpha (C_p + C_r)$

$$\Pi_{3p\,max} = [K' + \alpha \beta_c]^2 / 8\alpha = \Pi_{1p\,max} + ¼ \beta_c K' + ⅛ \alpha \beta_c^{\,2}$$
$$\Pi_{3r\,max} = [K' + \alpha \beta_c]^2 / 16\alpha = \Pi_{1p\,max} + ⅛ \beta_c K' + (\alpha \beta_c^{\,2}) / 16$$

Comparing the results of product consumption in case of producer and retailer subsidies and in the case of consumer subsidy, we find that if the sum of the subsidy per product unit given to the retailer and the producer is the same as subsidy per product unit given to the consumer (i.e. if $\beta_c = \beta_p + \beta_r$), the increase in the product consumption is the same and the profit of the producer and the retailer is the same. Moreover, it does not matter how the government divides the profit between the retailer and producer, their profit are the same. This is an expected result due to the fact that it is the producer that is driving the market and controlling the profits.

D. The Market in Case of a Government Subsidy to Illegible Citizen

Assume that the government decides to subsidies only a percentage ε of the total citizens by β_c for every product unit.

In this case the consumption for un-subsidized citizen will be:

$$D_{4a} = K - \alpha P_{4r}$$

The consumption for subsidized citizen will be:

$$D_{4b} = K - \alpha (P_{4r} - \beta_c)$$

So the average consumption will be:

$$D_4 = (1-\varepsilon) D_{4a} + \varepsilon D_{4b}$$
$$= K - \alpha P_{4r} + \varepsilon \alpha \beta_c$$

The profit of the producer becomes Π_{4p}

$= (P_{3p} - C_p) (K - \alpha P_{4r} + \varepsilon \alpha \beta_c)$

The profit of the retailer becomes Π_{4r}

$= (P_{3r} - Cr - P_{3p}) (K - \alpha P_{4r} + \varepsilon \alpha \beta_c)$

Using SPNE, the price of the producer and the retailer that will maximize their profit is:

$$P_{4p} = ½ k/\alpha + ½ (C_p - C_r) + ½ \varepsilon \beta_c = P_{1p} + ½ \varepsilon \beta_c$$
$$P_{4r} = ¾ k/\alpha + ¼ (C_p + C_r) + ¾ \varepsilon \beta_c = P_{1r} + ¾ \varepsilon \beta_c$$

The consumption in this case is:

$$D_4 = K - \alpha [¾ k/\alpha + ¼ (C_p + C_r) + ¾ \varepsilon \beta_c] + \varepsilon \alpha \beta_c$$

$= ¼ K – ¼ α (C_p + C_r) + ¼ ε α β_c$

$= D_1 + ¼ ε α β_c$

The profit of the producer and the retailer are:

$Π_{4p\ max} = Π_{1p\ max} + ¼ ε β_c K' + ⅛ α ε^2 β_c^{\ 2}$

$Π_{4r\ max} = Π_{1p\ max} + ⅛ ε β_c K' + (α ε^2 β_c^{\ 2}) / 16$

It is important to notice that when the government subsidies a product in a free market environment, it will be in the best interest of the producer and the retailer to increase their prices to maximize their profits when the subsidy is paid to the consumer, or to reduce their prices by only a percentage of the subsidy when it is paid directly to the producer and the retailer. As a result, the consumption will increase only by $¼ α β_c$, $¼ ε α β_c$ or $¼ α (β_p + β_r)$ depending whether the subsidy is paid to all consumers, a sub-set of the consumers or to the producer and retailer. This means that only 25% of the government funding will actually be used to increase the consumption while the other 75% will be used to increase the profit of the producer and retailer.

E. Government Subsidy to the Producer and Retailer under Price Control

Let us consider the case where the government decides to provide product subsidy only if the producer and the retailer fix their prices during the subsidy period to be equal to their prices before the subsidy reduce by the amount of subsidy. In this case, the producer and retailer prices will be:

$P_{5p} = ((K + α C_p – α C_r) / 2 α) – β_p$

$P_{5r} = ((3K + α C_p + α C_r) / 4 α) – (β_p + β_r)$

Note that the retailer sale price is reduced by both the subsidy of the retailer and the producer, as the later was used to reduce the price the retailer paid to the producer.

The consumption in this case will be:

$D_5 = K- α (((3K + α C_p + α C_r) / 4 α) – (β_p + β_r))$

$= D_1 + α (β_p + β_r)$

In this case the producer profit is:

$Π_{5p} = (P_{5p} – C_p + β_p) (D_5)$

$= (((K + α C_p – α C_r) / 2 α - β_p) – Cp + βp) (¼ K – ¼ α (Cp + Cr) + α (βp + βr))$

Using the assumption $K' = K – α (C_p + C_r)$

$Π_{5p} = ⅛ (K')^2/ α + ½ K' (β_p + β_r)$

$= Π_{1p\ max} + ½ (K – α (C_p + C_r)) (β_p + β_r)$

The retailer profit is:

$Π_{5r} = (P_{5r} – C_r – P_{5p} + β_r) (D_5)$

$= [((3K + α C_p + α C_r) / 4 α) – (β_p + β_r) – C_r – [((K + α C_p – α C_r) / 2 α) – β_p] + β_r] * (¼ K – ¼ α (C_p + C_r) + α (β_p + β_r))$

Using the assumption $K' = K – α (C_p + C_r)$

$Π_{5r} = (K')^2/ (16α) + ¼ K' (β_p + β_r)$

$= Π_{1r\ max} + ¼ (K – α (C_p + C_r)) (β_p + β_r)$

F. Government Subsidy to the Consumer under Price Control

If the government decides to pass regulations that freeze the prices of the producer and retailer to what they were before the subsidy, and pay the subsidy directly to the consumer, the producer and retailer prices will be:

$P_{6p} = ((K + α C_p – α C_r) / 2 α)$

$P_{6r} = ((3K + α C_p + α C_r) / 4 α)$

The consumption in this case will be:

$D_6 = K- α (((3K + α C_p + α C_r) / 4 α) – (β_c))$

$= D_1 + α β_c$

In this case the producer profit is:

$Π_{6p} = (((K + α C_p – α C_r) / 2 α) – C_p) (¼ K – ¼ α (C_p + C_r) + α β_c)$

Using the assumption $K' = K – α (C_p + C_r)$

$Π_{6p} = ⅛ (K')^2/ α + ½ K' (β_c)$

$= Π_{1p\ max} + ½ (K – α (C_p + C_r)) (β_c)$

The retailer profit is:

$Π_{6r} = (P_{6r} – C_r – P_{6p}) (D_6)$

$= (K')^2/ (16α) + ¼ K' β_c$

$= Π_{1r\ max} + ¼ (K – α (C_p + C_r)) (β_c)$

G. Government Subsidy to Illegible Consumer under Price Control

If the government decides to pass regulations that freeze the prices of the producer and retailer to what they were before the subsidy, and pay the subsidy to a percentage ε of the consumers who are eligible to receive financial aid, the producer and retailer prices will be:

$P_{7p} = ((K + α C_p – α C_r) / 2 α)$

$P_{7r} = ((3K + α C_p + α C_r) / 4 α)$

The average consumption in this case will be:

$D_7 = K – α P_{7r} + ε α β_c$

$= K- α ((3K + α C_p + α C_r) / 4 α) + ε α β_c$

$= D_1 + ε α β_c$

In this case the producer profit is:

$Π_{7p} = (((K + α C_p – α C_r) / 2 α) – C_p) (¼ K – ¼ α (C_p + C_r) + α ε β_c)$

Using the assumption $K' = K – α (C_p + C_r)$

$Π_{7p} = ⅛ (K')^2/ α + ½ K' (ε β_c)$

$= Π_{1p\ max} + ½ (K – α (C_p + C_r)) (ε β_c)$

Using the assumption $K' = K – α (C_p + C_r)$, the retailer profit is:

$Π_{7p} = Π_{1r\ max} + ¼ (K – α (C_p + C_r)) (ε β_c)$

III. CONCLUSION

When a government wants to reduce bad habits of its citizens such as smoking, taxation seems the way to go. On the other hand, to increase positive behavior such as eating healthier food, taxation of those who do not follow this behavior is a very difficult choice [7, 8 and 9]. Subsidy seems a more politically correct approach.

In this paper, we applied the Stackelberg [10] model and the Sub-game Perfect Nash Equilibrium [20 and 21] to understand the effect of subsidy on product consumption. To simplify the calculations, we made a few assumptions that are realistic, such as the existence of a market leader as defined by Stackelberg for each subsidized product, that the production cost and the distribution cost are not affected by the production level and that the consumption is a linear function inversely proportional to the product sale price.

In a free market environment, no matter how the subsidy is paid, it will always be in the best interest of the producer and retailer to increase their prices to increase their profit. In fact, only 25% of the government-subsidized money will be used to increase the consumption, while the rest will go towards increasing the profit of the producer and the consumer.

On the other hand, if the government decides to freeze the product price of the producer and retailer to what they were before the start of the subsidy program, all the subsidized money will be used to decrease the sale price of the product. In this case, all the funds the government spends for product subsidy will go towards increasing the product consumption. The producer and retailer will still see an increase in their profit due to an increase in the consumption. It is in the best interest of the government to force a price freeze on subsidized product either through regulations or by playing the role of the actual market leader.

Another important finding was that it does not matter whether the government pays the subsidy to the citizens or to the producer and the retailer and it does not matter the percentage given to the retailer compared to the producer. In all cases, if there is a market leader as defined by Stackelberg [10], the consumption rate, the product sale price and the profit of the producer and the consumer will be the same for the same amount of subsidy per product unit. The government can choose the subsidy approach that is easier to manage and have the best marketing effect on the citizens.

REFERENCES

[1] R. B. McKenzie, "Smoking War", in Heavy!, Springer, pp. 149-161, 2012.

[2] L. Lazuras, A. Rodafinos, D. B. Panagiotakos, J. R. Thyrian, U. John, and E. Polychronopoulos, "Support for smoke-free policies in a pro-smoking culture: findings from the European survey on tobacco control attitudes and knowledge", in International Journal of Public Health, Volume 54, Number 6, pp. 403-408, 2009.

[3] A. B. Naiman, R. H. Glazier, and R. Moineddin, "Is there an impact of public smoking bans on self-reported smoking status and exposure to secondhand smoke?", in BMC Public Health, Volume 11, Number 1, 2011. On Line: http://www.biomedcentral.com/1471-2458/11/146/

[4] K.A. Small, and J.A. Gomez-Ibanez, "Road pricing for congestion management: Transition from theory to policy", In: Road Prices, Traffic Congestion and the Environment, Issues of Efficiency and Social Feasibility, K Button, and E Verhoet (Eds), Edward Elgar Publishing, pp 213-246, 1998. On-line: http://www.uctc.net/papers/391.pdf

[5] Z. Ze-bin, A. Shi, and W. Jian, "The Welfare Effects of Congested Road-use Pricing and Redistribution Way", International Conference on Management Science and Engineering (ICMSE), pp: 2235 – 2240, 2007.

[6] J M. McGinnis, M. Nestle, "The Surgeon General's report on nutrition and health: policy implications and implementation strategies" , American journal of clinical nutrition, volume 49 (1) pp:23-28, 1989. On-line: http://ajcn.nutrition.org/content/49/1/23.full.pdf+html

[7] R. B. McKenzie, "Fat Taxes, Bans, and Discrimination", in Heavy!, Springer, pp: 163-200, 2012.

[8] M. J. Salois, J. R. Richard, "The Impacts of Fat Taxes and Thin Subsidies on Nutrient Intakes", Agricultural Economics Society, 85th Annual Conference, April 18-20, 2011, Warwick University, Coventry, UK. On-line: http://purl.umn.edu/108789

[9] S. Cash, D. Sunding, D. Zilberman, "Fat taxes and thin subsidies: prices, dietand Healthy outcomes", American Agricultural Economics Association, 2004 Annual meeting, August 1-4, Denver, CO On-line: http://purl.umn.edu/19961

[10] H. V. Stackelberg, "Market Structure and Equilibrium: 1st Edition Translation into English", Springer, 134 pages, 2011.

[11] C.I. Chen, and J. B. Jr Cruz, Stackelberg Solution for Two-Person Games with Biased Information Patterns, in IEEE Transactions on Automatic Control, VoI. AC-17, No. 5, 1972.

[12] M. Simaan, and Jr. J.B. Cruz, "On the Stackelberg Strategy in Nonzero-Sum Games", in Journal of Optimization Theory and Applications, Vol. 11, No. 5, pp. 533-555, May 1973. On-Line: http://link.springer.com/content/pdf/10.1007%2FBF00935665

[13] M. Shubik, "Game theory models and methods in political economy", in Handbook of Mathematical Economics, Volume 1, Editors: K. J. Arrow, and M. D. Intriligator, Elsevier, pp 285–330, 1981.

[14] G. de Fraja, and F. Delbono, "Game theoretic models of mixed oligopoly", in Journal of Economic Surveys, Vol. 4, No. 1, pp: 1 – 17, 1990.

[15] W. C. Merrill; N. Schneider, "Government Firms in Oligopoly Industries: A Short-Run Analysis" in the Quarterly Journal of Economics, Volume 80, Issue 3, Oxford Journals, pp: 400-412, 1966

[16] D. Pal, and M. D. White, "Mixed Oligopoly, Privatization and Strategic Trade Policy", in the southern economic journal, Volume 65, Issue 2, pp 264-281

[17] J. Poyago-Theotoky, "Mixed oligopoly, subsidization and the order of firms' moves: an irrelevance result", in economics bulletin, 2001, volume 12. On-line: http://www.accessecon.com/includes/CountdownloadPDF.aspx? PaperID=EB-01L30001

[18] Q. Gu; J. Ji, and T. Gao, "Price decisions for new product with and without subsidy", in IEEE International Conference on Service Operations and Logistics and Informatics (SOLI), pp: 35 – 39, 2010.

[19] M. Sertel, "Regulation by Participation", in Journal of Economics, Volume 48, Issue 2, Springer, pp: 111-134, 1988.

[20] M.J. Osborne; A. Rubenstein, "A Course in Game Theory", Chapter 6, MIT Press, pp: 89-115, 1994.

[21] R. Gibbons, "A primer in game theory", Chapter 2, Harvester-Wheatsheaf, pp: 55-141, 1992.

SmartOrBAC

Enforcing security in the Internet of Things

Imane BOUIJ-PASQUIER
ENSA UCA
Marrakech, MOROCCO

Anas ABOU EL KALAM
ENSA UCA
Marrakech, MOROCCO

Abdellah AIT OUAHMAN
ENSA UCA
Marrakech, MOROCCO

***Abstract*—The emergence of the Internet of Things (IoT) paradigm, provides a huge scope for more streamlined living through an increase of smart services but this coincides with an increase in security and privacy concerns, therefore access control has been an important factor in the development of IoT.**

This work proposes an authorization access model called SmartOrBAC built around a set of security and performance requirements. This model enhances the existing OrBAC (Organization-based Access Control) model and adapts it to IoT environments. SmartOrBAC separates the problem into different functional layers and then distributes processing costs between constrained devices and less constrained ones and at the same time addresses the collaborative aspect with a specific solution. This paper also presents the application of SmartOrBAC on a real example of IoT and gives a complexity study demonstrating that even though this model is extensive, it does not add additional complexity regarding traditional access control models.

Keywords—internet of things; security; privacy; access control model; authorization process

I. Introduction

Today we are seeing a change in our perception of Internet towards a global network of “smart objects”, which we can call the Internet of Things (IoT). These advances are estimated to accelerate over the next few years [1, 2] in response to reduced hardware costs, internet’s technological maturity and the swift development of communication technology. This will lead to a smooth assimilation of these smart objects into the Internet, which will in turn enable mobile and widespread access. Areas that are expected to be directly affected include healthcare [3, 4], supply chain management [5], transport systems [6], agriculture and environmental monitoring [7, 8], life at home and more, as we move towards “smart homes” [9, 10, 11] and the next generation of “smarter cities”[12].

This extension and proliferation of technology will certainly change our life, but will also present security and privacy challenges [13, 14, 15], since unexpected information leaks and illegitimate access to data and physical systems could have a high impact on our lives. Moreover, malicious modifications or denial of service may also cause damage in the context of IoT. Subsequently, the implementation of an access control mechanism that respects both the character of and the constraints on, smart objects in the IoT environment, is imperative. In this paper addresses one of the most relevant security issues - authorization and access control - in the context of distributed, cross-domain systems that consist of resource constrained devices not directly operated by humans. Especially, the problem where a single constrained device is communicating with several other devices from different organizations or domains. Based on OrBAC [16] access control model, our “Smart OrBAC” proposal is specifically designed for IoT environments. It, in fact, takes the main features of IoT into account and facilitates a distributed-centralized approach where authorization decisions are based on local conditions, and in this way offers context-aware access control.

The main contributions of this work can be outlined as follows:

- Exhaustive study and deep analysis of IoT security requirements and needs regarding its specific characteristics.
- Abstraction layers design regarding the specificities of IoT devices.
- SmartOrBAC, our access Control Model for IoT.
- Applying SmartOrBAC to an IoT case study and showing that it does not present additional complexity.

The rest of the paper is organized as follows: Section 2 presents a healthcare case study that allows us to extrapolate the relevant security requirements that an access control mechanism must fulfill, and then, generalizes these requirements for wider use in Section 3. Afterward, Section 4 gives an overview of the literature and discusses the important access control models currently existing in the IoT environment. Then, Section 5 describes the background needed to understand our new work. The SmartOrBAC access control model is then detailed in Section 6. Section 7 presents a complexity study comparing SmartOrBAC with traditional models followed by a brief description of the implementation in Section 8. Finally, Section 9 presents our conclusions.

II. Case Study

Before going into technical details, let us first discuss a representative scenario [17, 18]. A number of security requirements will be derived from this scenario.

Assume that John, a man with a heart condition, has opted for an assisted living service that is provided by a medical center. John uses a device that monitors his heart rate and his position; his home is also equipped with multiple sensors and actuators (temperature sensor, humidity sensor, luminosity sensor...). In the case of a cardiac problem, the heart monitor alerts the emergency services, and informs of John's current location. Moreover, the device uses smart logic to identify its owner “John” and allows him to configure the device's settings, including access control. This mainly prevents situations where someone else wearing that device acts as the owner and changes the access control and security settings.

In addition, John can add additional people to be notified in case of emergency, such as members of his family. Furthermore, the device saves the collected data, in order to assist his physician with his analysis.

However, John is worried that one of these authorized people may use the device to monitor his location even in the absence of an emergency. Furthermore, he is reluctant to let his health insurance company have access to this stored data, due to the possibility that they may decide he is too big a risk and therefore refuse to insure him.

A doctor, who monitors John's health remotely from the medical center, receives an alarm that John has fainted. An ambulance is instructed to go to assist John. A smart driving application is used by the ambulance to reach John's home as quickly as possible.

The situation requires the interaction of the following key actors:

- **Smart home of John**, actuators and sensors located in the house are used to collect vital information about the patient and sent to the monitoring service in the medical center, which oversees the patient's condition.
- **The medical center,** for monitoring John's health and the environmental conditions in the smart home. It then initiates appropriate action, such as alerting emergency services and sending the smart ambulance.
- The **ambulance** requests information from Traffic Monitoring in the police department in order to find the best route to John's home and save valuable time.
- **The police department for traffic jams monitoring, which** receives data from the distributed platforms sensors in order to infer the traffic status in the city's streets.
- **The smart city**: which includes all the previous stakeholders as a sub-stakeholders and where are various types of sensors, which are connected through Wireless Sensor Network (WSN) platform using various access technologies and/or communication protocols (ZigBee, Bluetooth, Wi-Fi, etc.) sharing their data.

In fact, each actor can be considered as (or belongs to) an *organization* or a domain e.g. "the medical center".

Subsequently, each organization is structured by different *roles* e.g. "doctor in the monitoring service", several *activities,* e.g. "consult", several *views* (groups of objects), e.g." patient's medical history, received sensor's data from monitored patient" and finally, the *context,* e.g. "A medical emergency such as John's faint".

The scenario intends to demonstrate:

- The cross-application nature of smart objects in one IoT service by showing their ability to simultaneously connect multiple application sectors and, more specifically, smart health, smart home, smart living, smart transport, etc…
- John needs to have the option of configuring his preferences related to trusted people or groups who can access his data in case of emergency (e.g. heart rate, location…).
- He must be able to block access to specific persons or groups, if he mistrusts them.
- The security measures must not affect the device's battery lifetime significantly. More precisely, since physically accessing the implanted device is hard or even impossible, the security measures should not affect battery lifetime significantly and not require direct physical interaction.
- Easy and intuitive configuration of the device.

III. General Requirements of the IoT

This section presents the most important IoT requirements derived from the case study, and then, generalized for wider use.

- *Interoperability*: The access control model must be designed for multiple organizations. On the one hand, each organization set up its own policies. On the other hand, it must respect other collaborating organization's policies.
- *Context awareness*: In IoT environments, context is highly important [19, 20]. In fact, services and applications use knowledge from the context surrounding them in order to gain information about their users and the users' environment [21, 22, 23, 24]. Thus authorization decisions are inextricably linked to local contextual data available to the device.
- *Ergonomie*: Due to the high saturation level of smartobjects in everyday life, many non-expert users are pushed to define permissions on their devices. Therefore, an access control mechanism must be simple to use: easily administrated, expressed and modified. In addition, it must enable policy updates without re-provisioning individual devices, and it must be designed so as not to require manual intervention of the user in the access control process.
- *Heterogeneity*: A collaborative environment may combine several technologies [25, 26, 27, 28]. This heterogeneity results in interoperation challenges, such as devices from different producers that provide proprietary features used by several services implemented according to diverse standards and protocols in order to initiate multiple functions [29].
- *Fine grained Access control*: The access control mechanism must be able to apply different permissions for different requesting entities rather than being all-or-nothing. Consequently, there is a need for granularity in authorization decisions.

- *Lightweight solution*: Due to the constrained energy nature of the IoT component, access control may minimize resource usage on the constrained device.
- *Scalability*: is the way to scale while managing increasingly large volumes of users, applications and connected devices. An Access control mechanism should naturally be extensible in size, structure, and number of organizations [30].

IV. Related Work

Zhang and Gong proposed in [31] the UCON model taking into consideration flexibility and heterogeneity in an IoT distributed environment. However, UCON is a conceptual model only, and thus it does not give details on the implementation of the monitoring process. This approach is indeed still not practical.

The CAPBAC model is implemented in a centralized approach in [32] where the proposed framework is based on a central Policy Decision Point (PDP) which handles authorization decisions. Whereas the implementation of capability-based access control in IoT is considered in [33] with an entirely distributed approach without intervention of central entities. The limits of both a purely centralized approach and fully distributed approach will be detailed bellow later on in this paper (see V.B Main architectures for access control in the Internet of Things).

The Capability-based Context-Aware Access Control (CCAAC) [34] is a delegation model based on a federated vision of IoT [35], where a central entity in each domain is in charge of authorizing a delegation request from a delegator, and making the decision about granting it to the delegate. However, this vision does not make use of technologies specifically designed for constrained highly context dependent environments such as IoT. Furthermore, the technical requirements in the constrained environment of the different actors involved in the proposed delegation mechanism are missing from this study.

Seitz et al. present in [36] an authorization framework based on XACML [37]. Evaluating XACML policies is too heavy-weight for constrained devices; therefore most of the authorization process is externalized. In order to convey the authorization decision from the external point to the device, an assertion is encoded in JSON [38] and is sent to the end-device (i.e., sensor or constrained device). The end-device takes responsibility for local conditions verification. However, this study does not give information about the central component involved neither about its management within the organization. Also, this proposal is bound to the use of XACML, which is not specifically designed for use in constrained devices.

V. Towards Centralized-Distributed Access Control for the Internet of Things

The integration of resource constrained devices into the Internet requires specifically designed technology and protocol that respect the nature of these smart objects. Recently, several IETF Working Groups have been focused on the adaptation of existing Internet protocols to IoT scenarios. These rising protocols, such as CoAP [39] and 6LoWPAN [40, 41, 42] aim to enable a seamless integration of the constrained devices into the Internet. It is then necessary to develop security mechanisms to fully take advantage of the huge potential offered by these protocols and technologies.

Prior to the detailed presentation of SmartOrBAC, this section describes briefly some of the core concepts that make up the proposed scheme. First of all, an overview of the OrBAC access control model and its benefits over other commonly accepted models are given. Then an overview of the main approaches and trends to provide access control logic in IoT scenarios is presented based on the architecture taxonomy proposed in [43].

A. Organization-Based Access control model (OrBAC)

The OrBAC model introduces the concept of organization as a structured group of active entities, in which subjects play specific roles. An activity is a group of one or more actions, a view is a group of one or more objects, and a context is a specific situation.

Actually, the Role entity is used to structure the link between the subjects and the organizations .The Empower (*org*, *r*, *s*) relationship (or predicate) means that *org* employs subject *s* in role *r*. In the same way, the objects that satisfy a common property are specified through views, and activities are used to abstract actions.

In security rules, permissions are expressed as Permission (*org*, *r*, *v*, *a*, *c*), obligations and prohibitions are defined similarly. Such an expression is interpreted as: in the context *c*, organization *org* grants role *r* the permission to perform activity *a* on view *v*.

As rules are expressed only through abstract entities, OrBAC is able to specify the security policies of several collaborating and heterogeneous organizations.

In our context, OrBAC presents several benefits:

- *Rules expressiveness*: OrBAC defines permissions, interdictions and obligations.
- *Abstraction of the security policy*: OrBAC has a structured and an abstracted expression of the policy; it also separates the specification from the implementation of the policy.
- *Scalability*: OrBAC has no limitation in size or capacity. It can define an extensible policy. It is then easily applicable to large-scale environments such as IoT.
- *Loose coupling*: each organization is responsible for its assets and entities. Implementation details as well as private information are managed separately by each organization.
- *Evolvability*: a policy in OrBAC is evolvable. It easily handles changes in organizations.
- *User-friendliness*: specifying and updating an OrBAC security policy are rather intuitive.
- *Popularity*: OrBAC has a growing community. Many research studies are being conducted, based on OrBAC.

- *Context-aware*: OrBAC takes the context (e.g. specific situations, time and location constraints) into account.
- *Fine-grained access control*: thanks to the context ad to its abstract and concrete concepts, OrBAC enables security administrators to define, set, specify, implement dynamic security policies and control access to individual data items and attributes.

However, despite the several advantages of OrBAC, it is not completely adapted to IoT. In particular, OrBAC is not able to manage collaboration-related aspects. In fact, as OrBAC security rules have the Permission (*org*, *r*, *v*, *a*, *c*) form, it is not possible to represent rules that involve several independent organizations (e.g. when the ambulance's driver, in *the medical center* organization, requests information from *Traffic Monitoring in the police department* organization in order to find the best route to John's home and save valuable time), or even, autonomous sub-organizations of a particular collaborative system (e.g. when the *police department for traffic jams monitoring* which is a sub-organization of smart city accede to data from the distributed sensors nodes in *the smart city* organization, in order to infer the traffic status in the city's streets). Moreover, it is impossible (for the same reason) to associate permissions to entities belonging to other partner-organizations (or to sub-organizations). As a result, if we can assume that OrBAC provides a framework for expressing the security policies of several organizations, it is unfortunately only adapted to centralized structures and does not cover the distribution, collaboration and interoperability needs, and these aspects are very important in the IoT context.

In order to overcome the limitations listed above, on one hand, the OrBAC model will be extended to include collaboration-related and context aware concepts; and on the other hand, a new architecture articulated around four functional layers will be proposed. The resulting framework is called "SmartOrBAC".

B. Main architectures for access control in the Internet of Things

This section gives an overview of the most popular current architecture providing access control in IoT services highlighting their main advantages and drawbacks.

1) Centralized architecture

In order to relieve smart objects from processing a large amount of access control related tasks, these functionalities are externalized to a back-end server or gateway responsible for authorization processing and thus, the end component (e.g sensors and actuators…) have a limited part (see

).

The most pertinent advantage of the centralized approach is that the access control logic is located within a non-constrained entity. It follows that the use of standard security protocols normally used in the traditional Web is not restricted. XACML may for example, be used to express access control policies.

Nonetheless, this approach encounters a major problem. In IoT scenarios such as the healthcare case study seen above, contextual information is of great importance, while in a centralized architecture, authorization evaluation doesn't take into account local contextual information related to the end component. Thereby, this one single vulnerability may compromise sensitive information, and this context insensitive central entity becomes the main weakness of the centralized approach.

2) Distributed approach

In this architecture, the access control process is carried out by the end component. This means that each device must be capable of handling authorization processes and having adequate resources to do so (see Fig. 1). An advantage of this approach is that end-devices act smartly, and are autonomous. A second advantage is that this approach allows real time contextual information to become central to the authorization decision. Furthermore, in this approach, end-to-end security is more easily achieved, as there is no need for an intermediate entity.

However, the need to extend the constrained device with access control logic makes the implementation of this approach unfeasible in resource-constrained devices.

3) Centralized-distributed approach

In this approach, the end-devices partially participate in the access control decisions (see Fig. 1) enabling the authorization evaluation process to take into account contextual information. As seen in the case study, there are environments where access control is not possible without the including information from the end component at the precise time of access request (e.g. location, temperature, humidity, CO2 level, heartbeat rate etc…).

This hybrid (centralized-distributed) approach, as in the centralized approach, allows us to use standard technologies to operate access control and the transmission of contextual information request will then be operated by specific application protocols as the *Constrained Application Protocol* (CoAP).

The most obvious disadvantage of this approach are the delays caused by the transmission of the contextual information from the end component to a central entity when needed. Due to this limitation, the value acquired by the end component may be different at the time of making the authorization decision, and consequently end-to-end security is unattainable.

Each one of these three approaches has advantages and drawbacks that need to be considered while considering them for the design of the access control.

In our proposal, the design of access control is based on the centralized-distributed approach. But unlike other proposals that use this approach, each separate group of components will have a central authorization engine (rather than just having one of these engines centrally performing all the authorization processes).The selection process that determines which entity will act as this engine depend on the contextual properties of the nodes in its group. The aim of this is to make the access control mechanism more time efficient by facilitating a smoother exchange of information between the end device and the authorization engine.

This vision is made possible by the fact that in a constrained environment, not all the devices are at the same level of constraint. In almost every WSN, less constrained nodes exist, and thus the central authorization server in charge of an area can be implemented on one of them. For more understanding, the next section gives an overview of the different actors involved in the proposed architecture and their properties.

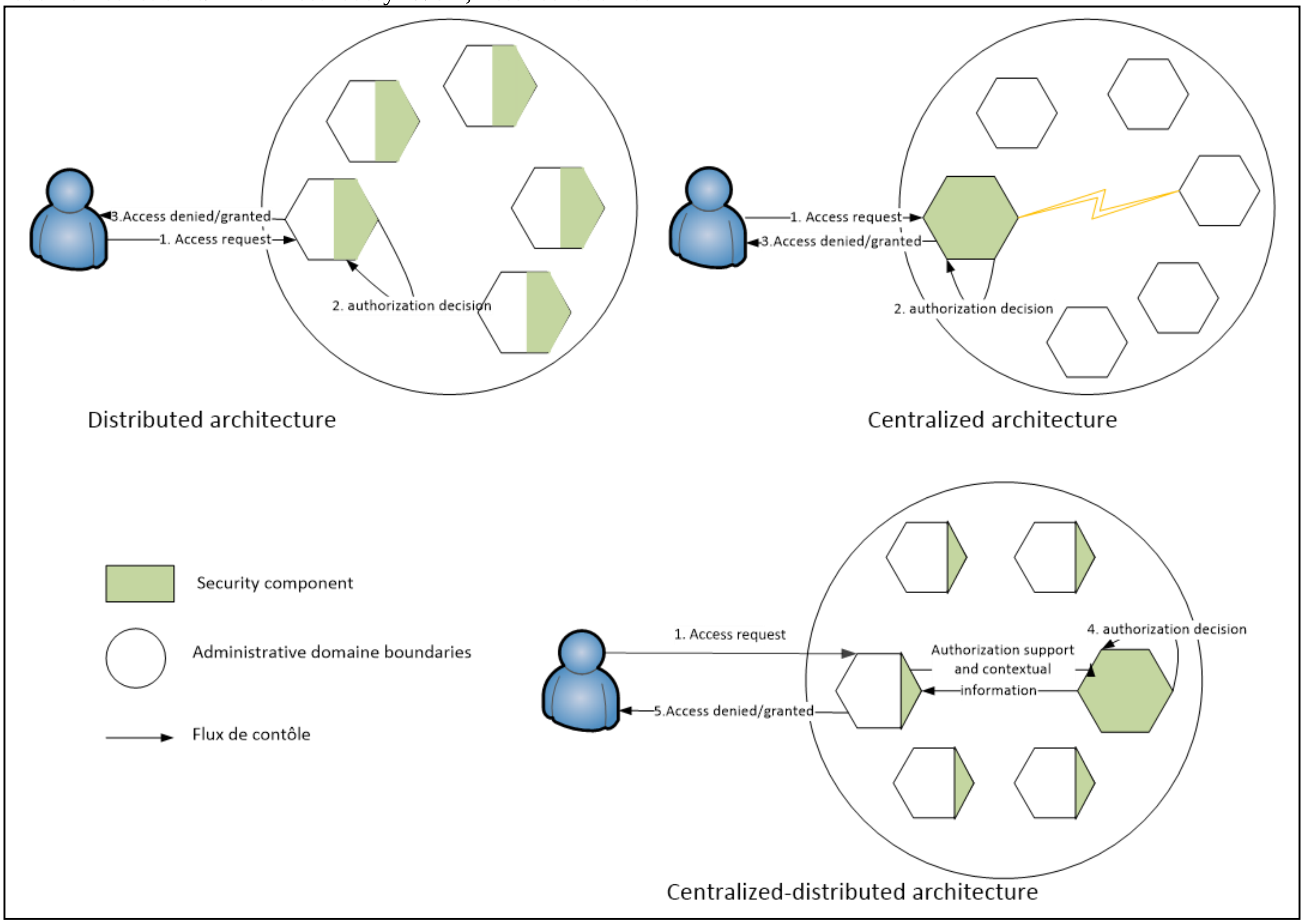


Fig. 1. Main architectures for IoT access control

C. Actors in SmartOrBAC

The main actors are the following [44]:

- **Resource Server (RS)**: An entity which hosts and represents a Resource that might contain sensor or actuator values or other information;
- **Resource Owner (RO)**: The principal that owns the resource and controls its access permissions;
- **Client (C)**: An entity which attempts to access a resource on a Resource Server;
- **Client Owner (CO)**: The principal that owns the Client and controls permissions concerning authorized representations of a Resource.

Consequently, in a basic scenario, *C* wants to access *R* located on *RS*. It follows logically that, *C* and / or *RS* are constrained.

VI. SMARTORBAC

The following paragraph contains description of the key aspects of our proposal. First, an explanation of the most relevant features of the abstraction layers design is given followed by a presentation of the collaborative solution. Then a structured expression of the *context* concept is presented. Finally the proposal is applied on the previous IoT scenario presented above.

A. SmartOrBAC abstraction layers

The SmartOrBAC architecture proposes, among others, a model based on a partitioning of the access control process into functional layers depending on the capabilities offered on each one. This approach is directly inspired by the fact that each device is constrained to a different level; they are in fact not all uniformly constrained. Note that the term “constrained node” is used according to the RFC 7228 [45]. While processing access control related tasks each layer assists the one below when needed. Note that the authentication process details are out the scope of this study. Only authorization aspects are treated. Four layers are introduced:

1) Constrained layer

One or both of *C* and *RS* are presumed to be located in a constrained node, but despite this, must perform access control related tasks. We thus consider that either of them may be unable to manage complex tasks while processing authorization requests. In addition, nodes do not always have permanent network connectivity. That's why both of *C* and *RS* are considered to be constrained layer actors. In order to address the limitations present in this layer, a less constrained device is associated to each area of constrained devices. This centric entity is defined by the upper layer called less-constrained layer (see Fig. 2).

2) Less constrained layer

To relieve constrained layer actors from conducting computationally intensive tasks, another layer is introduced. Each group of constrained layer actors is bound to a less constrained layer actor that belongs to the same security domain (see Fig. 2).

This link is configured by the entity in charge of the device (see Section VI.A.3) Organization layer). We call this central element the "*Client Authorization Engine*" (CAE), on the client side, and Resource Authorization Engine (RAE) on the resource side.

The Client Authorization Engine (CAE) belongs to the same security domain as *C*. It assists *C* in determining if *RS* is an authorized source for *R* by obtaining authorization information and supporting *C* in handling the authorization process.

The Resource Authorization Engine (RAE) belongs to the same security domain as *R* and *RS*. It assists *RS* in determining the correct permissions of *C* on the requested resource *R*. *RAE* obtains authorization information and supports *RS* in handling the authorization process.

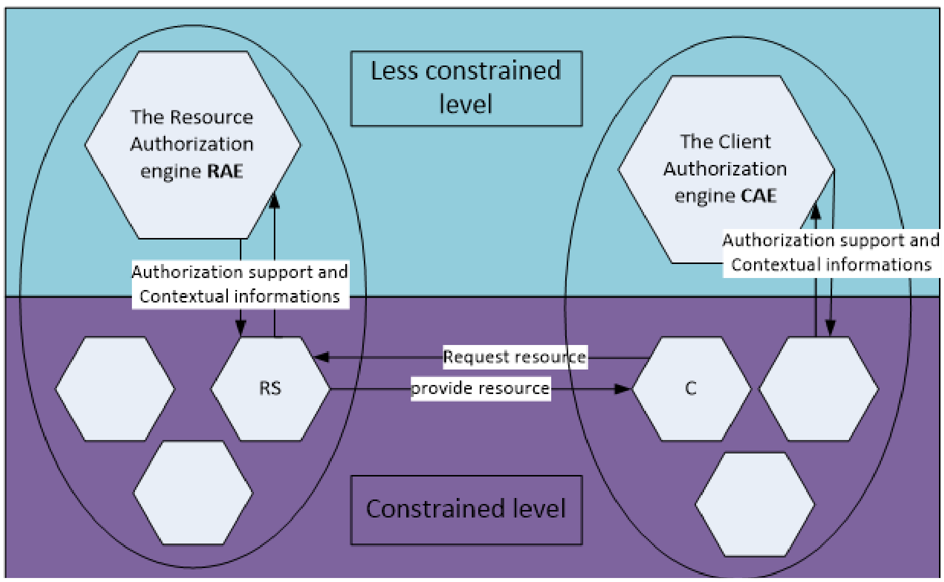


Fig. 2. Constrained and less constrained layers

3) Organization layer

In the real world, *C* and *R* are under the control of some physical entities. These entities are commonly called *ROr* (Resource Organisation) and *COr* (Client Organisation). In order to keep close to reality and to the OrBAC environment, this entity will be represented by Organisations (e.g. the police department, John's home, the medical center). Thus, each organization specifies the security policy for its devices and structures them in security domains.

The client organization *COr* is in charge of the entity proceeding to the resource request and thus, must specify security policies for *C*, including with whom *C* is allowed to communicate. This means that *COr* has to define authorized sources for a resource *R*. *COr* also configures *C* and *CAE* in order to make them belong to the same security domain.

The resource Organization *ROr* belongs to the same security domain as *R* and *RS*. *ROr* is in charge of *R* and *RS* and thus, must specify the authorization policies for *R* and decides with whom *RS* is allowed to communicate. That means that *ROr* has to configure if and how an entity with certain attributes is allowed to access *R*. *ROr* also configures *RS* and *RAE* in order to make them belong to the same security domain.

Subsequently, on the client side, *COr* defines authorized sources for *R*, and on the Resource side, *ROr* configures if and how an entity can access *R*.

In orders to do this, *ROr* and *COr* must have already agreed on the terms of such a service and on how to organize and structure this collaboration. An agreement is passed between the two entities before this interaction takes place (see Collaboration layer: a cross domain access control).

Note that an RS may in some cases be also the *RAE*. This holds in the same way for the *C* and the *CAE*.

4) Collaboration layer: a cross domain access control

As seen in the case study above, cross domain interaction is fundamental in the IoT. Furthermore, this characteristic represents the main difference between the Internet of Things paradigm and a simple sensor network based service that usually only deal with one domain. Note that, throughout this study, we define a domain as a structured independent organization.

Unfortunately, as seen above, the OrBAC access model does not handle the collaborative interaction aspects. To overcome this limitation, SmartOrBAC enhances OrBAC with new collaboration related concepts. This issue is addressed at the collaboration layer, by making a prior agreement between the involved organizations (as shown in Fig. 3) where the access rules to a given resource are jointly defined according to the OrBAC format by organizations that interact.

In order to manage this new agreement, the entity, located in the Organization layer, called *Principal Authorization Manager* "PAM" will be used. From the *RS* point of view, this agreement, which is interpreted in terms of access rules, will be treated just like all the other rules concerning local interactions. The complexity of the external interaction authorization management is hidden from the end constrained device, which keeps the same authorization processing no matter the nature of the client. This abstraction is made possible by the establishment of a fourth layer that manages the cooperation between different organizations.

Basically, SmartOrBAC begins with the publication and negotiation of collaboration rules as well as the corresponding access control rules. First, each organization determines which resources it will offer to external partners, and then references them into the *PAM*. At this point, other organizations can contact it to express their wish to use this specific referenced resource. To do that, the *COr* and the *ROr* negotiate and come to an agreement concerning the use of the resource *R*. Then, they establish a contract and jointly define security rules concerning access to *R*. The *COr*'s and *ROr*'s exchange format and the contract aspect will be discussed in a future paper. In the rest of this section, let us focus on access control rules. These rules are registered -according to an OrBAC format- in the *PAM* of both organizations. Parallel to this, *COr* creates locally a "*virtual resource*" called *R_image* which represents (the remote) *R*, and adds a rule in its OrBAC base to define which entities can invoke *R_image* to use *R* (see Fig. 3 and 4).

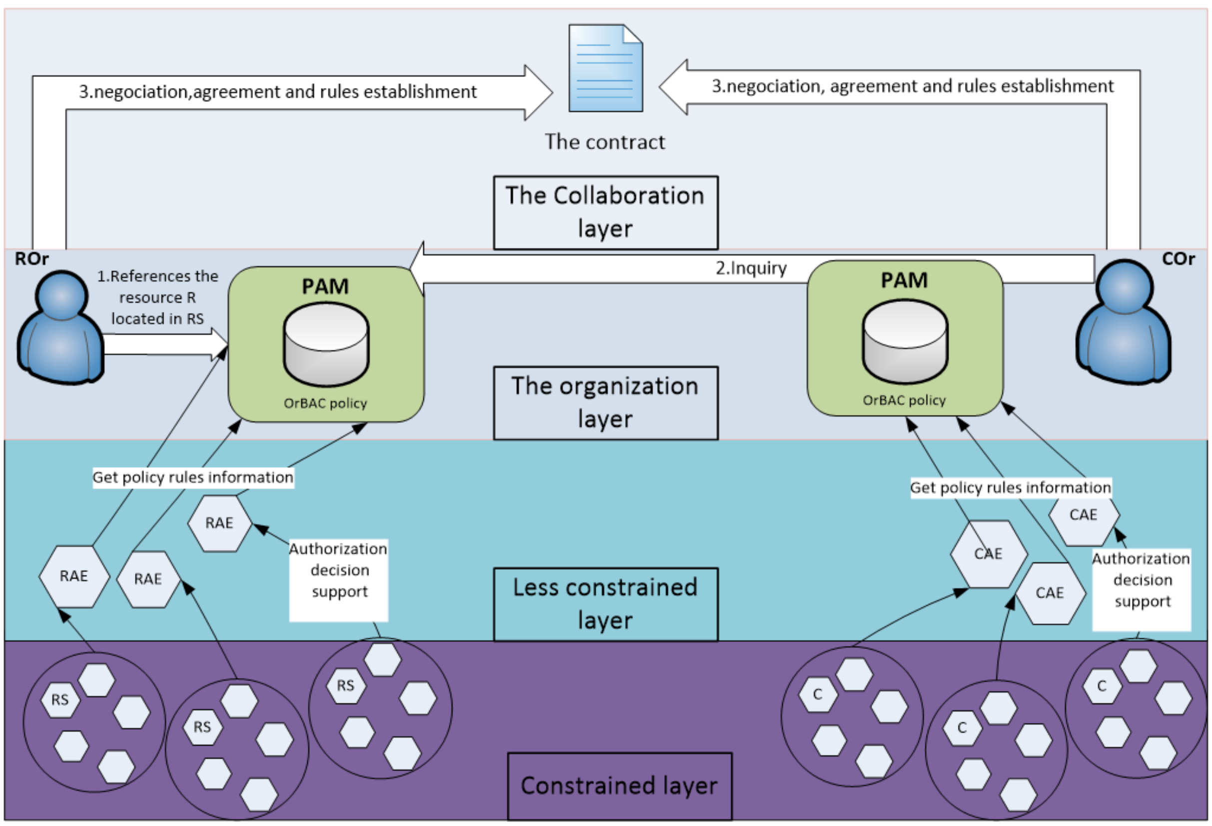


Fig. 3. Management of cross domain requirement in IoT environment

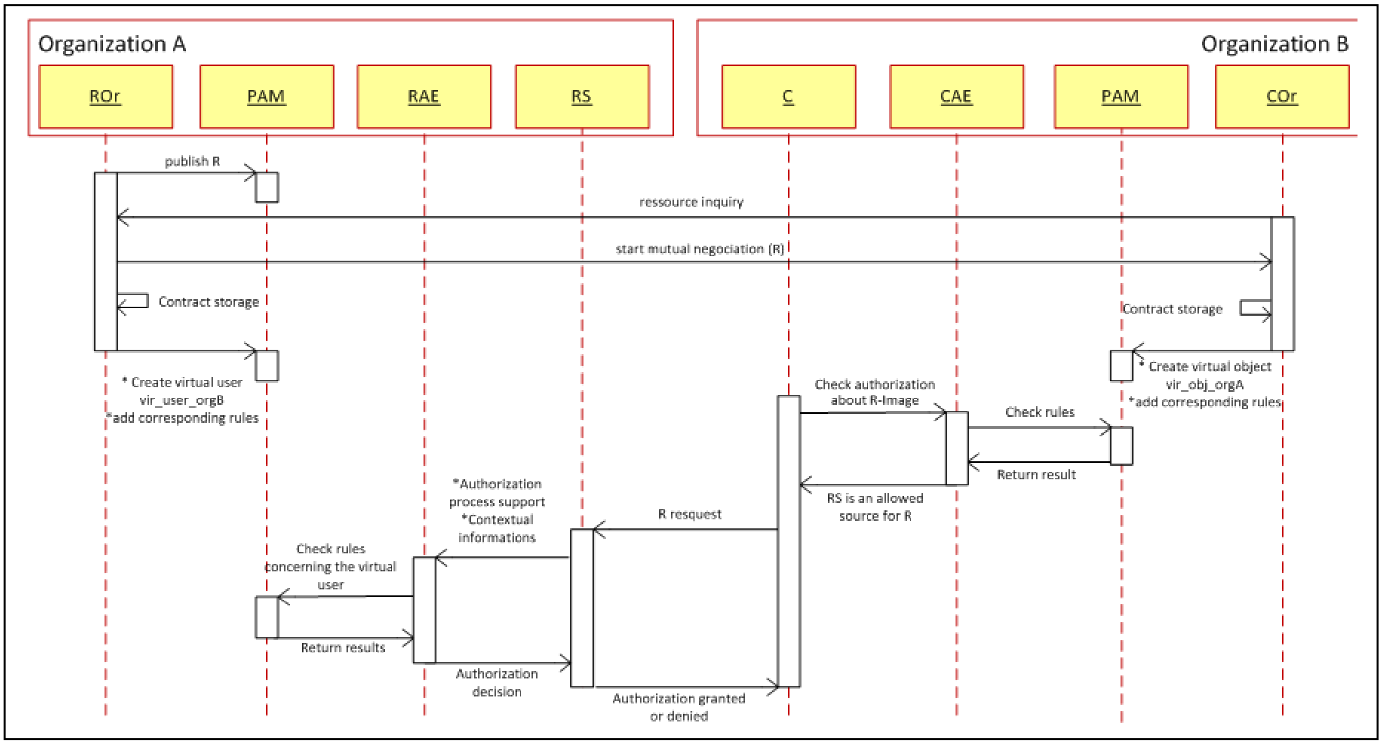


Fig. 4. Sequence diagram shows interactions between entities presented above when a Client from Organization B requests a resource from Organization A

B. Enhancing OrBAC for context awareness

Unlike traditional services where the concept of context is limited to a finite set of use cases, in the IoT environment, the concept is getting wider by taking on an ambient character in order to allow services taking into account the contextual information collected in real time by the different sensors [46]. The Context used in defining the SmartOrBAC rule is a group of contexts (C_{Set}) from several types (C_{Type}). The type of context represents a concrete characteristic such as location, temperature or time, but also security requirement such as trust level or risk level. Subsequently, to introduce the context in the access control decision, a value called (C_{Const}) is given to each C_{Type}. Thus the context definition in SmartOrBAC takes the following format:

$$C_{Type} = \{\text{authLevel, trustLevel, time, location,}\dots\} \quad (1)$$

$$C_{Set} = \{C_{Type(1)}, C_{Type(2)}, \dots , C_{Type(n)}\} \quad (2)$$

$$C_{Const} = <C_{Type}> <OP> <VALUE> \quad (3)$$

where OP is a logical operator, i.e. OP $\in \{>,<,\geq,\leq,=, \neq\}$, and VALUE is the estimated level of C_{Type}. Finally, C is expressed as a set of constraints C_{Const} as follows:

$$C = \{C_{Const(1)}, C_{Const(2)}, \dots, C_{Const(n)}\} \quad (4)$$

Typically in the previous use case, the emergency context would be defined by a set of constraints related to the patient movement, location and especially to his heartbeat measures.

C. Scenario

In order to illustrate SmartOrBAC, the different concepts detailed above are applied on the previous case study.

First of all, each organization determines which device's resources it will offer to external partners. At this stage, we find in the *PAM* of John's smarthome organization resources such as the heart monitor resource. The medical center organization makes an inquiry to the *PAM*. As soon as the target resource is found, the negotiation phase begins between the *ROr* of the smart home and the *COr* of the medical center. The resulting contract is then transcript in terms of authorization rules regarding the OrBAC format for both of the medical center and smart home of John. More precisely, if the agreement between the two organizations is: "*Assigned doctor from medical center have the permission to remotely actuate the implanted cardioverter defibrillator from the heart monitor device in the heart attack emergency context*", the *ROr* of Smart home should:

- have (or create) a rule that grants the permission to a certain role (e.g. *Doctor*) to actuate the heart monitor: *Permission(smart home, Doctor, vital equipement, Actuating,* $C_{heart\_attack\_Emergency}$*)*; Note that, from John's smart home's point of view, every user playing the "*Doctor*" role will have this permission;
- create a "virtual user" noted "v_user_doctor" that represents the medical center for its use of the implanted cardioverter defibrillator (see Fig. 5);
- add the following *Empower(smart home, v_user_doctor, Doctor)* association to its rule base. This rule grants the user medical center's doctor the right to play the Doctor role.

In parallel, the COr of the medical center creates locally a "virtual object" heart_monitor_image which represents the (remote) implanted device (the resource made available by

John's Smart Home), and adds a rule in its OrBAC base to define which of the medical center's roles can invoke heart_monitor_image to use the real heart monitor.

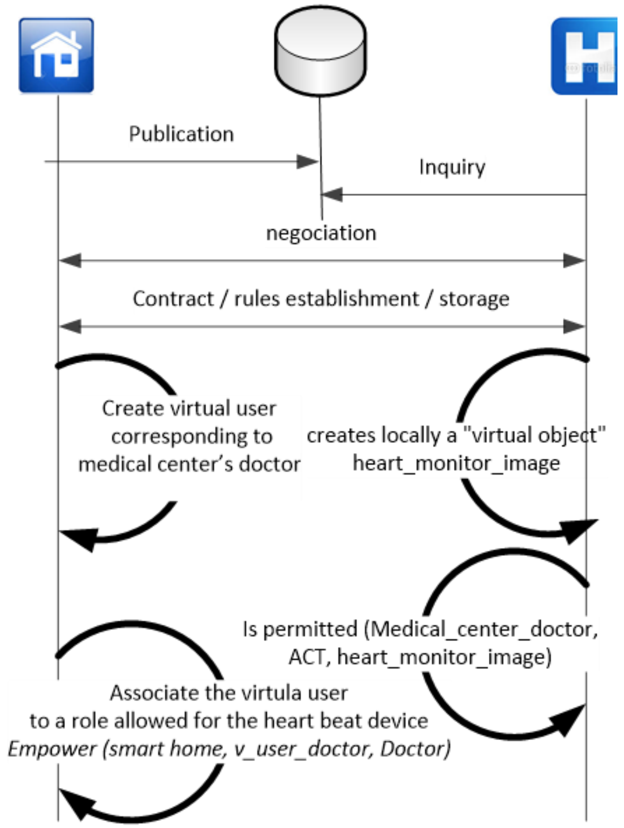


Fig. 5. Virtual user and virtual Object in SmartOrBAC

The derivation of the permission (i.e., instantiation of security rules) mentioned above can be formally expressed as follows:

Permission *(smart home, Doctor, vital equipement, Actuating, Cheart_attack_Emergency)* ∧
Empower *(smart home, v_user_doctor, Doctor)* ∧
Consider *(smart home, ACT, Actuating)* ∧
Use *(smart home, John's heartbeat, vital equipement)* ∧
Hold *(smart home, medical center's doctor, ACT, John's heartbeat, Cheart_attack_Emergency)*

→ ***Is permitted*** *(v_user_doctor, ACT, John's heartbeat)*

Fig. 6. Derivation of permissions in SmartOrBAC

Let's assume that the assisted living dispositive is a set of different devices (sensors and actuators) with different capabilities. We also assume that the specific device *RS* of heart monitoring that the medical center tries to access is located in the constrained layer, such as the client device *C* used by the doctor in the medical center. The link between the *RS* and its corresponding *RAE* located in the less constrained layer has already been configured by the *ROr* of John's smart home. The same applies for the *CAE* and *C* that have been already configured by the *COr* of medical center.

Before the doctor's device *C* in the medical center sends an actuating request to the heart monitoring device *RS*, it asks the corresponding CAE in the medical center for assistance in order to determine if the local image of *RS* (*heart_monitor_image*) is an authorized source.

At this moment, *CAE* starts evaluating the authorization policy rules, using as object the *heart_monitor_image*. Note that at this level, the external nature of the heart monitor device is unknown. Then, if information about policy rules are needed, a request is sent to the *PAM* of the medical center. Once this process is completed, if *RS* is an allowed source, an actuating request is directly sent to the heart monitoring device.

Once the request is received, the authorization decision process begins on the smart home organization side. For that, the device sends an authorization process request, with contextual information, to the corresponding *RAE* in john's smart home. The latter evaluates the authorization decision regarding authorization rules in John smart home's *PAM* - especially those detailed above where the subject is *v_user_doctor* -.The result is sent to *RS* which, in turn, sends an access response to the doctor's device.

VII. Complexity of SmartOrBAC

In this section, the complexity of the capabilities based models (frequently proposed for IoT) is compared to SmartOrBAC on the three following aspects of the access control process:

1) Operations related to the access control decision in an IoT environment.

2) Operations related to the security policy update.

3) The potential risk of errors during the access control policy administration.

Our aim (at this stage) is to demonstrate that even though SmartOrBAC is multifaceted, it is less complex than capabilities based models, it ameliorates the security policy management cost and it reduces the risks of errors.

B. Access control decision

In order to evaluate the complexity of the complete access control management process, we focused on two crucial parameters:

- Quantify the number of decisions required for the definition of the access control policy.
- Rate the complexity of each decision (each management operation).

It is clear that, the more complex the operations needed, the more management resources and processing times are required, and a higher probability of errors is observed which is even more prominent in an IoT environment. Furthermore the operations assigned to the security administrator are naturally more complex and more sensitive than those done by other actors. Therefore in our context, we identify two kinds of operations:

1) Sensitive operations assigned to the security administrator; we note D the cost of this kind of operations.

2) Secondary operations that may be executed by an operator not necessary aware of the security requirements; we note d the cost of such operations (e.g. the assignment of subject to roles in SmartOrBAC in a certain organization).

Next follows a comparison of the administration costs of the capability based models and the SmartOrBAC.

3) Core Capability based models

In the capability based models, the overall quantity of operations is |SUB|.|Op|.|OBJ|; where|.| is the the number of elements (the cardinal). As the quantity of operations is limited, |Op| = *constant*. Thus the number of operations is equal to |SUB|.|OBJ|.|*constant*|. However, this analysis is correct for traditional services but not for the IoT environment where the contextual information collected in real time by sensors has to be taken into account. As seen in "Enhancing OrBAC for context awareness", the concept of context is no more limited to a finite set of use cases. In order to better represent this reality, the above formula should be correlated to the multiplicity of existent contexts. Thus, the total number of operation would rather be |SUB|.|Op|.|OBJ|.|CONTEXT|.

Assuming that n designates the maximal value of |SUB|, |OBJ| and |CONTEXT|, the number of the operations is $O(n).O(n)$. $O(n).O(1) = O(n^3)$. Furthemore, all these operations require the administrator skills, consequently, the cost is equal to $D.O(n^3)$.

4) SmartOrBAC

Prior to the calculation of the administration cost, let us fisrt recall how the access decision is made in SmartOrBAC according to OrBAC:

$$\begin{aligned}&\forall org \in \text{ORG},\ \forall s \in \text{SUBJ},\ \forall \alpha \in \text{ACTION},\\ &\forall o \in \text{OBJ},\ \forall r \in \text{ROLE},\\ &\forall a \in \text{ACTIV},\ \forall v \in \text{VIEW},\ \forall c \in \text{CONT},\\ &\quad\quad \textit{Permission}\ (org, r, v, a, c) \wedge\\ &\quad\quad \textit{Empower}\ (org, s, r) \wedge\\ &\quad\quad \textit{Consider}\ (org, \alpha, a) \wedge\\ &\quad\quad \textit{Use}\ (org, o, v) \wedge\\ &\quad\quad \textit{Hold}\ (org, s, \alpha, o, c) \wedge\\ &\quad\quad\quad\quad \rightarrow \textit{Is permitted}(s, \alpha, o)\end{aligned}$$

$$Cost = D.\mathrm{C}(RULE) + d.[\ \mathrm{C}(Empower) + \mathrm{C}(Consider) + \mathrm{C}(Use) + \mathrm{C}(Hold)] \quad (5)$$

while:

$$\begin{aligned}\mathrm{C(RULE)} &= |\text{Access\_Mode}| + |\text{ORG}| + |\text{ROLE}| + |\text{VIEW}| + |\text{ACTIV}| + |\text{CONT}|\\ &= |\text{constant}| + |\text{ constant }| + |\text{ constant }| + |\text{ constant}|\\ &\quad + |\text{ constant }| + \mathrm{O}(n) \approx \mathrm{O}(n) \quad (6)\end{aligned}$$

$$\begin{aligned}\mathrm{C}(Empower) &= |\text{ORG}| + |\text{ROLE}| + |\text{SUBJ}|\\ &= |constant| + |\ constant\ | + O(n) \approx O(n) \quad (7)\end{aligned}$$

In the same way,

$$\mathrm{C}(Consider) \approx O(n),\ \mathrm{C}(Use) \approx O(n) \text{ and } \mathrm{C}(Hold) \approx O(n) \quad (8)$$

Hence,

$$\begin{aligned}Cost &\approx D.\ O(n) + \mathrm{d}\ .\ [O(n) + O(n) + O(n) + O(n)\ ]\\ &\approx D.\ O(n) + \mathrm{d}.O(n) \quad (9)\end{aligned}$$

TABLE I. THE COMPLEXITY OF DECISIONS

Capability based models	SmartOrBAC
D.O(n3)	D. O(n) + d.O(n)

In the capability based models, the complexity is a cubic function with a higher factor (D), while in SmartOrBAC it is a linear function with two factors D (major) and d (minor). Subsequently, SmartOrBAC leads to significant reduction of the management complexity.

C. The update of the security rules

In the capability based models, the subject's permission management (adding, updating or deleting) is in the order of $O(n)$. Moreover, all these operations require a major decision. The total cost is thus $D.O(n)$ correlated to contextual parameters it is upgraded to $D.O(n^2)$. Inversely, in SmartOrBAC the management of subjects' permissions corresponds to a change of its roles which involves only $d.O(1)$ operations. Now, if we concentrate our analysis on objects, and in particular, if we want to change the permissions associated to a certain object, the cost in the capability based models is $D.O(n^2)$. In SmartOrBAC, the cost is $d.O(1)$ as we only review the *Use* relationship.

TABLE II. THE COMPLEXITY OF ACCESS CONTROL CHANGES

Capability based models	SmartOrBAC
$D.O(n^2)$	$\mathrm{d}.O(1)$

D. Risk of errors

Generally, two indicators are used to evaluate the risks:

- The severity of the threat, estimated in term of its impact.
- The potentiality of the threat (estimated in term of frequency or probability if the cause is accidental, or in term of feasibility if the cause is deliberate).

Basically, the security administrator decisions are more sensitive than the operator or a secretary decision. In the first case, we note the severity by ***S***, while we denote it by ***s*** in the second case.

Besides, the potentiality (e.g. probability) that an administrator makes an error in the security policy definition is lower than the action of the operator or a secretary. Indeed, we note ***p*** the potentiality of the security administrator errors, while ***P*** denotes the potentiality of the operator/secretary errors.

We can thus conclude that:

- The assignments that could be done by a secretary or an operator has a factor of risk (of error) equivalent to s.P

- The factor is equivalent to S.p for Sensitive operations that are done by the security administrator.

Therefore, to calculate the risk of error for the two models (capability based and SmartOrBAC), we only replace *D* in table 1 by *S.p* and *d* by *s.P*. Table 3 summarizes the risk of errors in the three compared models.

TABLE III. THE RISK OF ADMINISTRATION ERRORS

Capability based models	SmartOrBAC
$S.p.O(n^3)$	$S.p.\ O(n) + s.P.O(n)$

The risk of access control management errors is thus reduced in SmartOrBAC.

Consequently, compared to the other models, it appears that not only SmartOrBAC gains simplicity and clarity in the IoT environment (e.g. by taking the context into account in the earlier stages), but it also greatly reduces the cost of administering access control policies as well as making the process less error prone while being clearly context aware.

VIII. IMPLEMENTATION

The transmissions between the different entities included in our Framework (C/RS, C/CAE, RS/RAE) are done via the CoAP [39] protocol (Constrained Application Protocol), which is a specialized web transfer protocol that is intended for use in resource-constrained internet devices. Like HTTP, CoAP is based on the wildly successful REST model: Servers make resources available under a URL, and clients access these resources using methods such as GET, PUT, POST, and DELETE.

Since the XML representation is too verbose for efficient transmission over limited channels, thus JSON-based notation is used for authorization requests and responses. In fact JSON [38] (JavaScript Object Notation) is a lightweight data-interchange format that efficiently reduces the size of the transmitted messages between *C* and *RS* devices and optimizes the processing time.

The device part of our framework (especially *C* and *RS*) was implemented on an example platform: The Arduino Mega 2560 board3. This board features a 16 MHz processor, 256 kB of Flash Memory, 8 kB of SRAM, and 4 kB of EEPROM. The choice of this board is made in order to test our approach on the lowest performance of the end constrained devices.

The board was programmed in JAVA using a custom implementation of the CoAP protocol stack and the assertions were wrapped in JSON format using the standard Java API (javax.json.*).

IX. CONCLUSION

Our SmartOrBAC access model is specifically designed for the IoT environment and it is conceived through an abstraction layer design that makes use of a deep understanding of the IoT paradigm as it is used in the real world. For these smart services, contextual information is a leading element in decision making therefore only a real-time consideration of this information will achieve smartness. For this reason, the "context" notion (originally present in OrBAC) is extended in order to fit the IoT requirements.

Understanding that users belonging to an organization need to dynamically access resources controlled by other organizations the proposed model is extended with specific collaborative mechanisms where the same OrBAC security policy can be used for local as well as external access. In this way, SmartOrBAC improves the management of the security policy and reduces considerably its complexity.

In our future work, the focus willl be laid on making the SmartOrBAC model more effective by going deeper in the study of the negotiation process and the e-contract format. Other possibilities include incorporating a secure authority delegation method based on OrBAC in order to control the link between the end device and the *RAE*/*CAE* in order to make it more dynamic.

Finally, another relevant research line related to this work is the consideration for additional privacy enhancement through techniques such as the use of pseudonyms or anonymous assertions.

REFERENCES

[1] G. D. Abowd, A. K. Dey, P. J. Brown, N. Davies, M. Smith, and P. Steggles, "Towards a better understanding of context and context-awareness" in *Proceedings of the 1st international symposium on Handheld and Ubiquitous Computing, ser. HUC '99*. London, UK: Springer-Verlag, 1999, pp. 304–307.

[2] European Commission, "Internet of things in 2020 road map for the future" *Working Group RFID of the ETP EPOSS, Tech. Rep*., May 2008, http://ec.europa.eu/information society/policy/rfid/ documents/iotprague2009.pdf [Accessed on: 2011-06-12].

[3] Istepanian, R.S.H, Jara, N. Philips and A. Sungoor. "Internet of things for M-health applications (IoMT)". In AMA IEEE Medical Technology Conference on Individualized Healthcare, Washington (2010).

[4] A. Jara, A. Skarmeta and M. Zamora : "An Internet of Things based personal device for diabetes therapy management in ambient assisted living (AAL)". Personal and Ubiquitous Computing, 15(4):431–440 (2011).

[5] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey" Computer Networks, vol.54, no.15, pp. 2787–2805, Oct. 2010.

[6] J. Santa, A. Skarmeta, A. Jara, and M. Zamora: "Telematic platform for integral management of agricultural/perishable goods in terrestrial logistics". In Computers and Electronics in Agriculture, 80:31–40 (2012).

[7] L. W. F. Chaves and C. Decker, "A survey on organic smart labels for the internet-of-things" in Networked Sensing Systems (INSS), 2010 Seventh International Conference on, 2010, pp. 161–164.

[8] J. Burrell, T. Brooke, and R. Beckwith, "Vineyard computing: sensor networks in agricultural production" Pervasive Computing, IEEE, vol. 3, no. 1, pp. 38 – 45, jan.-march 2004.

[9] V. Gungor, T. Kocak and al.: "Smart Grid and Smarthomes: Key Players and Pilot Projects". In Industrial Electronics Magazine, IEEE, 6(4):18–34 (2012).

[10] J.W. Hui and D.E. Culler. IP is dead, long live IP for wireless sensor networks. *In Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 15-28, Raleigh, NC, USA, 2008.

[11] Mathias Kovatsch, Markus Weiss, and Dominique Guinard. Embedding internet technology for home automation. *In Emerging Technologies and Factory Automation (ETFA), 2010 IEEE Conference* on, pages 1-8, Bilbao, Spain, September 2010. IEEE.

[12] M. Castro, A. Jara, and A. Skarmeta.Smart Lighting Solutions for Smart Cities. In Proc. of the 27th International Conference on Advanced Information Networking and Applications Workshops (WAINA'13), Barcelona, Spain, pages 1374–1379. IEEE, March 2013.

[13] L. Daigle and O. Kolkman. "The OAuth 2.0 Authorization Framework" Internet Engineering Task Force (IETF), Request For Comments (RFC) 6749, October 2012, http://www.ietf.org/rfc/rfc6749.txt.

[14] Xun Li and Sang Bong Yoo, "Extended Role-Based Security System using Context Information", *In: 2008 Second International Conference on Future Generation Communication and Networking*, pp. 7-12.

[15] Guillemin and P. Friess, "Internet of things strategic research roadmap" The Cluster of European Research Projects, Tech. Rep., September 2009.

[16] L. Daigle and O. Kolkman, "The OAuth 2.0 Authorization Framework," Internet Engineering Task Force (IETF), Request For Comments (RFC) 6749, October 2012, http://www.ietf.org/rfc/rfc6749.txt.

[17] M.Memon, S.Rahr Wagner, C. Pedersen, F. Beevi and F.Overgaard Hansen etc. Ambient Assisted Living Healthcare Frameworks, Platforms, Standards, and Quality Attributes. *Sensors* ***2014****, 14, 4312-4341; doi:10.3390/s140304312.*

[18] A. Dohr, R. Modre-Opsrian, M. Drobics, D. Hayn, and G. Schreier, "The internet of things for ambient assisted living," *in Information Technology: New Generations (ITNG), 2010 Seventh International Conference on, 2010*, pp. 804–809

[19] Xun Li and Sang Bong Yoo, "Extended Role-Based Security System using Context Information", *In: 2008 Second International Conference on Future Generation Communication and Networking*, pp. 7-12.

[20] H. Sundmaeker, P. Guillemin, P. Friess, and S. Woelffle, "Vision and challenges for realising the internet of things," *European Commission Information Society and Media, March 2010*.

[21] G.Zhang, J.Tian. An Extended Role Based Access Control Model for the Internet of Things. *In 2010 International Conference on Information, Networking and Automation (ICINA)*.

[22] A. Bernardos, P. Tarrio, and J. Casar, "A data fusion framework for context-aware mobile services," *in Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on, aug. 2008*, pp. 606 –613.

[23] L. Lin, "Application of the internet of thing in green agricultural products supply chain management," in Intelligent Computation Technology and Automation (ICICTA), 2011 International Conference on, vol. 1, 2011, pp. 1022–1025.

[24] D. Martin, C. Lamsfus, and A. Alzua, "Automatic context data life cycle management framework," in Pervasive Computing and Applications (ICPCA), 2010 5th International Conference on, dec. 2010, pp. 330 – 335.

[25] S. Helal, W. Mann, H. El-Zabadani, J. King, Y. Kaddoura, and E. Jansen. The Gator Tech Smart House: a programmable pervasive space. *Computer*, 38(3):50-60, 2005.

[26] C.Floerkemeier, M.Lampe. Facilitating RFID development with the accada prototyping platform. *In Fifth IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom '07)*, pages 495-500, New York, NY, USA, 2007. IEEE Computer Society.

[27] Christian Floerkemeier, Christof Roduner, and Matthias Lampe. RFID application development with the Accada middleware platform. *IEEE Systems Journal*, 1(2):82-94, December 2007.

[28] Deze Zeng, Song Guo, and Zixue Cheng. The Web of Things: A Survey. *Journal of communications*, vol.-, No.6, September 2011.

[29] F. Ramparany, R. Poortinga, M. Stikic, J. Schmalenstroer, and T. Prante, "An open context information management infrastructure the ist-amigo project," in *Intelligent Environments, 2007. IE 07. 3rd IET International Conference* on, sept. 2007, pp. 398 –403.

[30] Christian Floerkemeier, Christof Roduner, and Matthias Lampe. RFID application development with the Accada middleware platform. *IEEE Systems Journal*, 1(2):82-94, December 2007.

[31] G. Zhang and W. Gong. The Research of Access Control Based on UCON in the Internet of Things. Journalof Software, 6(4):724–731, April 2011.

[32] S. Gusmeroli, S. Piccione, and D. Rotondi. A capability-based security approach to manage access control in the internet of things. Mathematical and Computer Modelling, 58(5-6):1189–1205, Sept 2013.

[33] HERNÁNDEZ-RAMOS, J., JARA, Antonio J., MARIN, Leandro, etal. Distributed capability-based access control for the internet of things. J Internet ServInfSecur, 2013, vol. 3, no 3/4.

[34] B. Bayu, P. N. Mahalle, N. R. Prasad, and R. Prasad. Capability-based access control delegation model on the federated IoT network. In Proc. of the 15th International Symposium on Wirless Personal Multimedia Communications (WPMC'12), Taipei, China, pages 604–608. IEEE, September 2012.

[35] K. Hasebe and M. Mabuchi, "Capability-role-based delegation in workflow systems," in Embedded and Ubiquitous Computing (EUC), 2010 IEEE/IFIP 8th International Conference on, dec. 2010, pp. 711 – 717.

[36] L. Seitz, G. Selander, and C. Gehrmann. Authorization Framework for the Internet-of-Things. In Proc.of the 14th IEEE International Symposium and Workshops on a World of Wireless, Mobile and Multimedia Networks (WoWMoM'13), Madrid, Spain, pages 1–6. IEEE, June 2013.

[37] T. Moses. Extensible Access Control Markup Language (XACML) Version 2.0, 2005.

[38] D. Crockford. RFC 4627: The application/json Media Type for Javascript Object Notation (JSON). IETFRFC 4627, July 2006. http://www.ietf.org/rfc/rfc4627.txt.

[39] Z. Shelby, K. Hartke, and C. Bormann. Constrained Application Protocol (CoAP). IETF Internet-draft (work in progress), June 2013.http://tools.ietf.org/html/draft-ietf-core-coap-18.

[40] R. Prasad, My personal Adaptive Global NET (MAGNET), *ser. Signals and Communication Technology Book*. Springer Netherlands, 2010.

[41] J.W. Hui and D.E. Culler. Extending IP to low-power, wireless personal area networks. IEEE Internet Computing, 12(4):37-45, 2008.

[42] N. Kushalnagar, G. Montenegro, and C. Schumacher, Ipv6 Over Low-Power Wireless Personal Area Networks (6lowpans): Overview, Assumptions, Problem Statement, and Goals, RFC4919, August, 2007.

[43] R. Roman, J. Zhou, and J. Lopez. On the Features and Challenges of Security and Privacy in Distributed Internet of Things. Computer Networks, 57(10):2266–2279, July 2013.

[44] IEFT, "Constrained restful environments working group (core wg) - authentication and authorization for constrained environments (ace) mailing list".

[45] C. Bormann, M. Ersue, A. Keranen. Terminology for Constrained-Node Networks. Informational, IETF RFC 7228, May, 2014. http://tools.ietf.org/html/rfc7228.

[46] B. Bayu, P. N. Mahalle, N. R. Prasad, and R. Prasad. Capability-based access control delegation model on the federated IoT network. In Proc. of the 15th International Symposium on Wirless Personal Multimedia Communications (WPMC'12), Taipei, China, pages 604–608. IEEE, September 2012.

Implementation of Central Dogma Based Cryptographic Algorithm in Data Warehouse Architecture for Performance Enhancement

Rajdeep Chowdhury [1]
Assistant Professor, Department of Computer Application,
JIS College of Engineering,
Block–A, Phase–III, Kalyani,
Nadia–741235, West Bengal, India

Soupayan Datta [2]
Student, Department of Computer Science & Engineering,
JIS College of Engineering,
Block–A, Phase–III, Kalyani,
Nadia–741235, West Bengal, India

Saswata Dasgupta [3]
Student, Department of Computer Science & Engineering,
JIS College of Engineering,
Block–A, Phase–III, Kalyani,
Nadia–741235, West Bengal, India

Mallika De [4]
Professor, Department of Computer Science & Engineering,
Dr. Sudhir Chandra Sur Degree Engineering College, Surer Math, Melabagan Estate, Basak Bagan, Dum Dum,
Kolkata–700074, West Bengal, India

***Abstract*—Data warehouse is a set of integrated databases deliberated to expand decision-making and problem solving, espousing exceedingly condensed data. Data warehouse happens to be progressively more accepted theme for contemporary researchers with respect to contemporary inclination towards industry and executive purview. The crucial tip of the proposed work is integrated on delivering an enhanced and an exclusive innovative model based on the intention of enhancing security measures, which at times have been found wanting and also ensuring improved accessibility using Hashing modus operandi. An unsullied algorithm was engendered using the concept of protein synthesis, prevalently studied in Genetics, that is, in the field of Biotechnology, wherein three steps are observed, namely; DNA Replication, Translation and Transcription. In the proposed algorithm, the two latter steps, that is, Translation and Transcription have been taken into account and the concept have been used for competent encryption and proficient decryption of data. Central Dogma Model is the name of the explicit model that accounts for and elucidates the course of action for Protein Synthesis using the Codons which compose the RNA and the DNA and are implicated in numerous bio–chemical processes in living organisms. It could be observed that subsequently a dual stratum of encryption and decryption mechanism has been employed for optimal security. The formulation of the immaculate Hashing modus operandi ensure that there would be considerable diminution of access time, keeping in mind the apt retrieval of all indispensable data from the data vaults.**

The pertinent appliance of the proposed model with enhanced security might be in its significant service in a variety of organizations where accrual of protected data is of extreme magnitude. The variety of organizations might include educational organizations, corporate houses, medical establishments, private establishments and so on and so forth.

Keywords—Data Warehouse; Central Dogma; Replication; Translation; Transcription; Codon; Data Mart; Hashing

I. Introduction

Data warehouse is an entrenched depository of an organization's electronically summative data [1, 2, 5]. Data warehouse are designed with the objective to facilitate comprehensive reporting and proficient analysis. Endowing security for the warehouse is virtually an enormous vulnerability for any organization.

The proposed work is principally unforced, as it confers the formulation of the inventive architecture with the objective of enhancing security measures and data warehouse performance enhancement in the course of action [1, 2, 5, 6].

The proposed cryptographic algorithm introduces an intuitive as well as an innovative approach by employing a unanimously accepted concept in Genetics and Molecular Biology known as 'Central Dogma**.'** The Central Dogma introduced by Francis Crick endows with a sequential explanation of the flow of genetic information within a biological system. The general transfer principally consists of three activities, namely; DNA Replication, Transcription and Translation.

DNA Replication – DNA Replication is the process of engendering two identical replicas from one original DNA molecule. The biological process occurs in all living organisms and is the basis for biological inheritance.

Transcription – Transcription is the initial step of gene expression, in which a particular segment of DNA is copied into RNA by the enzyme RNA polymerase.

Translation – Translation is the process by which proteins are created. In the process, apiece codon codes for a specific amino acid and proteins are synthesized or rather they are found in the form of amino acids in the body.

Improved accessibility at the data warehouse using Hashing technique would ensure performance enhancement of the data warehouse in addition to the security measures adhered.

The entire paper deals with an interdisciplinary approach of employing the biological process by which protein is synthesized in living organisms, for development of an algorithm by which a plain text is converted to a cipher text.

A real world phenomenon is being employed in computation and its appliance on the data warehouse should endow the reader with an insight into implementation of bio–inspired algorithm and its prospective impact on the manner it is being perceived and professed, thereby solving computational problems.

II. Literature Survey

The section focuses on the interrelated work available in the similar genre, encouraging the formulation of the concerned paper [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14].

There are quite a few interrelated works on data warehouse and its security measures, which have formerly been carried out, but the precise design, implementation and incorporation of the innovative algorithm using protein synthesis is something inventive as well as appealing for contemporary researchers. Whenever the perception culminates in the mind about design, implementation and incorporation of security mechanism for data warehouse as well as architectural orientation, the extensive study of few papers need meticulous mention and the concise points are affirmed in details for ease in reference [1, 2].

In the paper titled “Design and Implementation of Proposed Drawer Model Based Data Warehouse Architecture Incorporating DNA Translation Cryptographic Algorithm for Security Enhancement,” the DNA Translation cryptographic algorithm is incorporated in the Proposed Drawer Model Based Data Warehouse Architecture in two distinguished tiers of security mechanism. Initially the algorithm is adhered during the transition from Operational Data Store to Data Vault and then it is adhered for the second time during the transition from Data Vault to Data Mart, thereby ensuring two-tier security enhancement for the proposed data warehouse model [1].

Through the formulation of the paper, discussions as well as Illustrations ensure how security could be implemented at distinguished tiers/levels using the innovative DNA Translation cryptographic algorithm through the proposed Drawer Model Based Data Warehouse Architecture as an Added measure over the existing Data Warehouse Architecture [1].

In the paper titled “Towards Data Security in Affordable Data Warehouse,” the data warehouse technique is based on clustering and the star schema is dispersed over the nodes of the cluster. Dimension table is replicated in apiece node of the cluster and fact table is distributed using strict round robin or hash partitioning. Security is assured by using signature in each column individually and in each row the verification is controlled by data warehouse middleware. Data warehouse middleware engender the signature for insert and update operations. In the concerned approach, encryption technique is applied on dimension table. Primary keys and foreign keys are not encrypted as they get filled with synthetic values. Encryption technique is not used in fact table due to performance issues. Fact table is fragmented into several clusters and fact data cannot be simply allied to the dimension data as they are encrypted [2, 6].

In the paper titled “An Integrated Conceptual Model for Temporal Data Warehouse Security,” it has been proposed that the first integrated conceptual model for addressing temporal data warehouse security requirements needs specification. The model is the first model which combines ETL model with temporal data warehouse model in one integrated model.

ETL processes are accountable for extraction of data from heterogeneous operational data sources, their transformation (conversion, cleaning, normalization, etc.) and their loading into data warehouses. ETL proposed model has six fragmented steps, namely; 1-Source authentication, 2-Extractraction, 3-Filter process, 4- Incorrect process, 5-Surrogate process and 6-Load process [9]. These are three of numerous papers which have had an untiring impact during the formulation of the compiled paper.

III. Proposed Work

For utmost simplicity in understanding, the flow chart of the entire proposed work has been designed in figure 1.

The proposed Central Dogma Based cryptographic algorithm is incorporated in the Data Warehouse Architecture for security as well as performance enhancement in distinguished tiers of security mechanism and improved accessibility using Hashing modus operandi. Initially the algorithm is adhered during the transition from Operational Data Store to Data Vault and then it is adhered for the second time during the transition from Data Vault/Storage Area to explicit segments of Data Mart, thereby ensuring two-tier security enhancement for the proposed data warehouse model. The interdisciplinary amalgamation of Biotechnology, especially genetics and Computer science, especially cryptography is an innovative approach, to say the least.

In the next fragment of the section, the detail of the proposed work has been specified and the commencement churns out with the notion about Central Dogma and the formulation of the proposed innovative Central Dogma Based cryptographic algorithm, as stated below for ease in reference:

The Central Dogma of Molecular Biology is an elucidation of the flow of genetic information within a biological system. It was initially stated by Francis Crick in 1956 and re-stated in a research paper published in ‘Nature’ in 1970. The Central Dogma of Molecular Biology deals with the exhaustive residue-by-residue transfer of sequential information. It states that such information cannot be transferred back from protein to protein or nucleic acid.

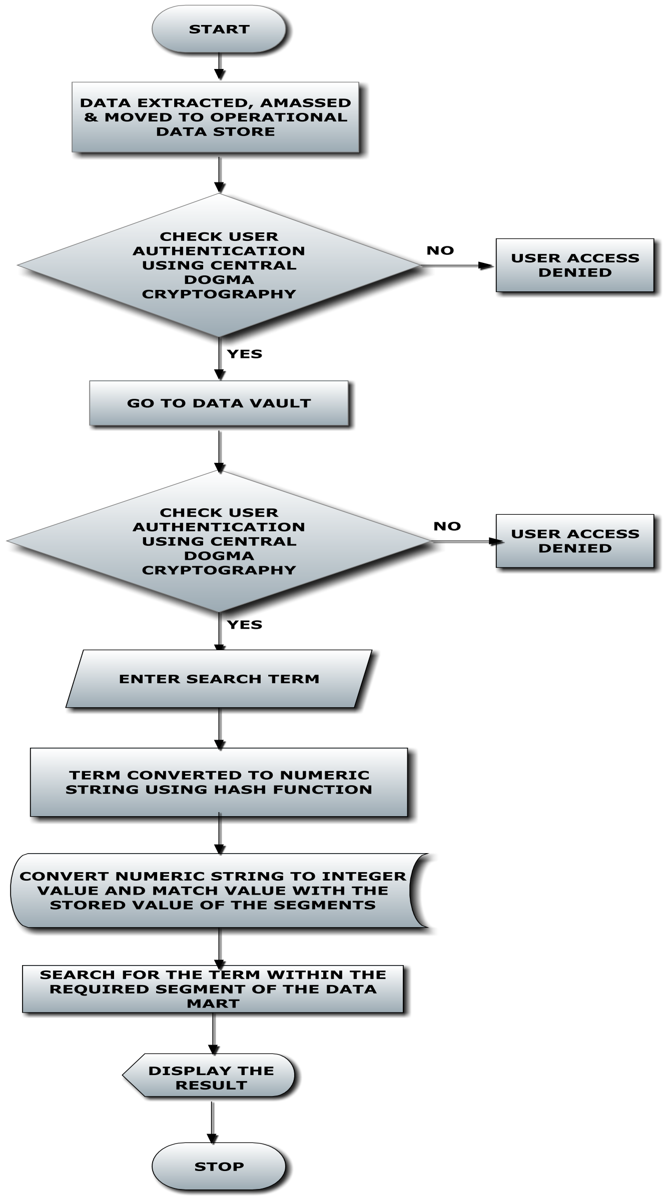


Fig. 1. Flow Chart of the Entire Proposed Work

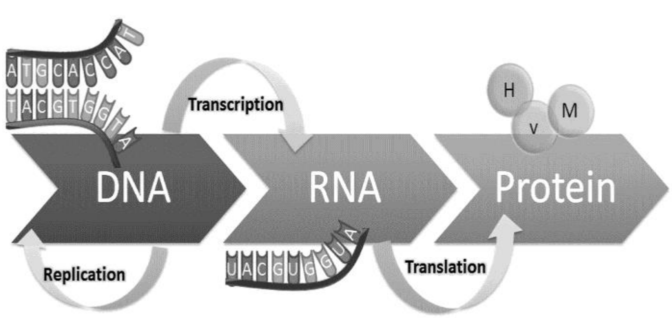


Fig. 2. Central Dogma Model

The Dogma is a framework for understanding the transfer of information sequence amid sequential information carrying biopolymers, in the most familiar or general case, in living organisms. There are three major classes of such biopolymers, namely; **DNA** and **RNA** (Both nucleic acids) and **Protein**. There are 3×3 = 9 conceivable unswerving transfers of information that could occur amid these. The Dogma classifies these into three groups of three: Three General Transfers, which are believed to transpire normally in most cells; Three Special Transfers, which are believed to transpire, only under specific conditions, in case of some viruses or in laboratories; Three Unknown Transfers, which are believed never to transpire.
The General Transfers illustrate the normal flow of biological information: DNA could be copied to DNA (DNA Replication); DNA information could be copied to mRNA (Transcription); Proteins could be synthesized using the information in mRNA as a template (Translation).

A. Proposed Central Dogma Based Cryptographic Algorithm

Proposed Central Dogma Based cryptographic algorithm employs the concept of protein synthesis which is an integral part of the 'Central Dogma' consisting of two steps, namely; Transcription and Translation. The procedure of Transcription engrosses transfer of information from DNA to RNA and the procedure of Translation engrosses the concluding step of protein synthesis, wherein specific group of nucleobases lead to specific protein synthesis.

ACGT stands for the four nucleic acid bases that make up DNA. 'A' stands for Adenine and pairs with 'T', which stands for Thymine. 'C' stands for Cytosine and pairs with 'G', which stands for Guanine. These four nucleic acids make up an organism's genetic code or DNA.

ACGU stands for the four nucleic acids that make up RNA. RNA pairs up in the similar manner as DNA, except that Thymine is replaced with Uracil.

B. Extended ASCII Code and Encryption Key

The Extended ASCII code has been employed for ease in working, taking into account all feasible characters possible during authentication process. Extended ASCII is an 8–bit or larger character encoding modus operandi that includes 7–bit ASCII characters as well as others. All the characters embodied within the Extended ASCII code could be represented by their equivalent 8–bit binary form.

Consequently for smooth functionality, the grouping mechanism for primary key has been employed, nevertheless to overlook the fact that this implicitly places the key into the category of public and symmetric encryption key.

Subsequently, as exemplified, the 8 bits have been clustered into 2 bits apiece, wherein apiece cluster of 2 bits embodies a specific nucleobase found in DNA (Ribonucleic Acid). After grouping in the mentioned manner, it is observed that there are 4 combinations of groups feasible, namely; 00, 01, 10, 11.

The feasible groups have been used to embody the specific nucleobase (Found in DNA) in the subsequent manner:
00 – A
01 – C
10 – G
11 – T

IV. Illustration and Example

A. Steps to obtain Cipher Text

Step 1 – Transcription Phase →

1. The binary representation of the character concerned from the Extended ASCII Table is referred and are clustered/grouped into 2–bits apiece. Subsequently, four groups are obtained.
2. The groups are reinstated by the nucleobases as mentioned above.
3. A text is obtained which would contain the four nucleobases, which might look like –> UUTC.
4. The code/text thus obtained in its DNA form, gets converted to its RNA form by the process of Transcription, following the rules mentioned below:
   T (Thymine) is replaced with U (Uracil), following the rule of complementary base pairing.
   For complementary base pairing, the following replacement is followed:

   A → U & Vice Versa

   G → C & Vice Versa

Step 2 – Translation Phase →

1. The Genetic Code Table is referred and it is observed that cluster/group of three nucleobases leads to the synthesis of a specific protein.
2. Employing the concept of Translation, the cipher text obtained by Transcription mechanism is further processed.

To illustrate the concept exclusively, it is assumed that the obtained cipher text from Transcription, for a random word of three letters is as mentioned below:

ACUGUCGACUAA

In order to translate the cipher text, the following rules are adhered:

1) Subsequent to the implementation of Transcription mechanism, for apiece character, a cipher text is obtained which consists of a combination of four nucleobases.

2) If the first nucleobase is left out, then the rest three nucleobases refer to a specific codon that code for a specific amino acid, as observed in the Genetic Code Table mentioned already.

3) It is observed that certain codon refers to the similar amino acid or rather a specific amino acid could be synthesized from multiple codons.

4) In order to avoid ambiguity, numbering mechanism should be employed, that is, '1' before the name of the amino acid (As visible in Genetic Code Table) should refer to the first codon from top, within a specific block, accountable for the synthesis of that amino acid.

*5) As there are three **STOP** codons, **UAA** would be referred to as **1sto**, **UAG** would be referred to as **2sto** and **UGA** would be referred to as **3sto**.*

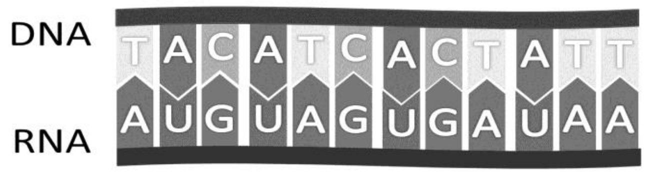
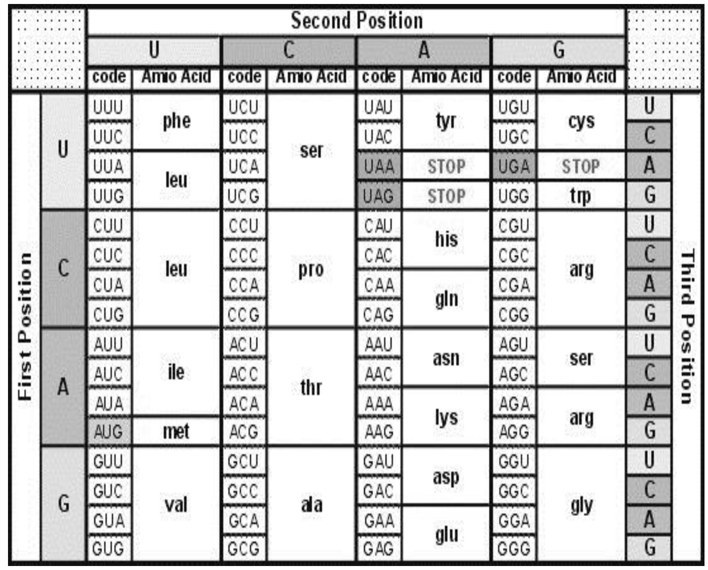


Fig. 3. Nucleobase Pairing



First Position	Second Position: U code	U Amio Acid	C code	C Amio Acid	A code	A Amio Acid	G code	G Amio Acid	Third Position
U	UUU	phe	UCU	ser	UAU	tyr	UGU	cys	U
U	UUC		UCC		UAC		UGC		C
U	UUA	leu	UCA		UAA	STOP	UGA	STOP	A
U	UUG		UCG		UAG	STOP	UGG	trp	G
C	CUU	leu	CCU	pro	CAU	his	CGU	arg	U
C	CUC		CCC		CAC		CGC		C
C	CUA		CCA		CAA	gln	CGA		A
C	CUG		CCG		CAG		CGG		G
A	AUU	ile	ACU	thr	AAU	asn	AGU	ser	U
A	AUC		ACC		AAC		AGC		C
A	AUA		ACA		AAA	lys	AGA	arg	A
A	AUG	met	ACG		AAG		AGG		G
G	GUU	val	GCU	ala	GAU	asp	GGU	gly	U
G	GUC		GCC		GAC		GGC		C
G	GUA		GCA		GAA	glu	GGA		A
G	GUG		GCG		GAG		GGG		G

Fig. 4. Genetic Code Table

*6) Hence, a random character, viz, **ACUG**, after implementation of Translation mechanism should be:*

A4leu

*7) After implementation of Transcription mechanism, the entire cipher text representing a random word of three letters, that is, **ACUGUCGACUAA,** enduring the Translation mechanism should be:*

A4leuU3proC1sto

Demonstration/Illustration of General Data Warehouse Architecture

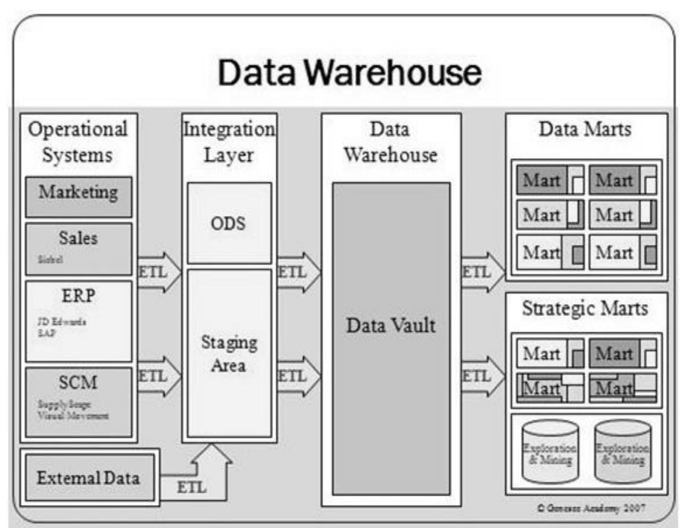


Fig. 5. General Data Warehouse Working Architecture

B. General Working of Data Warehouse

*1) The elucidation of the working exemplifies that data being extracted from various **Operational Systems**. The extracted data are of distinguished format.*

*2) After extraction, the data are amassed at **Staging Area**, which is an intermediary storage locale used for data processing.*

*3) From staging area, the data is moved to **Operational Data Store (ODS),** which assimilates all data and place them into a common format.*

*4) From ODS, data is moved to **Data Vault**, which is deliberated to endow with long term historical storage of data.*

*5) From data vault, the data is moved to different **Data Marts**, which is the access layer of data warehouse environment.*

C. Data Vault Working

The working of the data vault could be fragmented into three components/processes performed on data, namely;

a) OLAP Analysis

b) Adept Reporting

c) Data Mining

Proposed Hashing Technique Based Data Warehouse Architecture

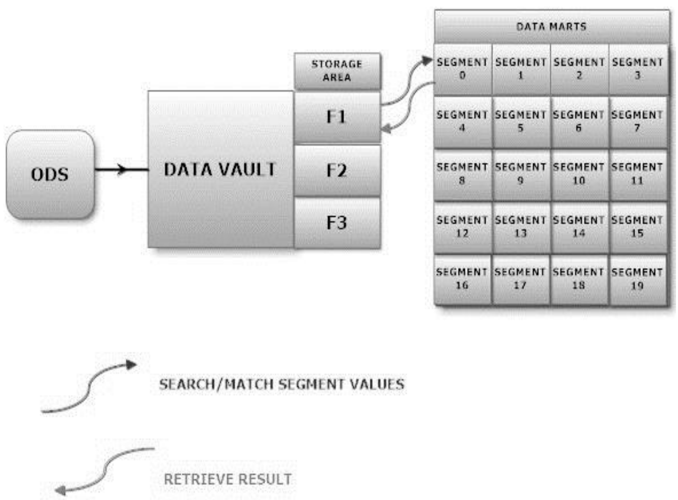


Fig. 6. Proposed Hashing Modus Operandi Based Data Warehouse Architecture

D. General Working of Proposed Hashing Modus Operandi Based Data Warehouse employing Central Dogma Based Cryptographic Algorithm for security enhancement

*1) Initially, data is extracted from various **Operational Systems**, then amassed in **Staging Area** and then moved to **Operational Data Store**.*

*2) From **ODS**, data is moved to **Data Vault,** deliberated to endow with long term historical storage.*

*3) From **Data Vault**, data are shifted to newly designed **Storage Area**, where data is separated and amassed in distinguished folders. Apiece folder is specified to accumulate data for explicit **Data Mart**.*

*4) All the **Data Marts** are initially empty, wherein data are transmitted through folders to the dedicated **Data Marts**.*

*5) At a particular time, only apiece folder could be transmitted from **Storage Area** to **Data Mart**.*

*6) After triumphant authentication employing **Central Dogma Based Cryptographic Algorithm**, the user is allowed to access the **Data Mart**, wherein the **Data Mart** is fragmented into **Segments** numbered from 0 to 19.*

*7) A database pertaining to a bank should contain Employee Unique ID, Name, Age, Sex, etc., which might be treated as **Segments**.*

*8) Gaining the access to the **Data Mart**, the search term concerning the requisite data is entered by the user, wherein the search term is converted into a numeric string employing the **Hashing function**:*

$$F(n) = \text{Sum of the (digits or alphabets)} \bmod 20 \quad (1)$$

Set of rules are followed to engender an integer value which would denote the segment number.

The rules followed are:

(i) If digits would be present, then they would be added as usual.

Example: If Unique ID is 120110, then Sum is 5.

(ii) If alphabets would be present, then the respective digit value would be considered.

Example: a = 1, b = 2, c = 3, etc.

(iii) If an amalgamation of digits and alphabets would be present, then they would be converted in respective manner described already, followed by addition to obtain the Sum.

*9) It is reasonably undemanding to search for an integer rather than an assortment of characters, ensuring diminution of accessibility time for retrieving the requisite data from the specific **Segment**.*

*10)Subsequent to finding a match with respect to the integer value obtained after **Hashing**, the search progresses by seeking the requisite and precise data within that specific integer **Segment**.*

V. CONCLUSION

Whenever the term safety comes in intellect and initiative, security is synonymous, but from time to time implementing security mechanism(s) like cryptographic techniques, biometric methodologies, genetic algorithm, quick response code mechanisms, etc. has not only been sturdy but cost constrained as well. On the other hand, design, implementation and incorporation of any security methodology at internal structure of the data warehouse are affable.

The accessibility is explicitly improved by employing Hashing Modus Operandi, thereby ensuring diminution of access time and validating performance enhancement for the data warehouse. The design, implementation and incorporation of the Central Dogma Based cryptographic algorithm is the core of the conferred security amid predicament at bay like malicious intrusions.

Through the formulation of the paper, deliberations as well as illustrations ensure how security could be implemented and incorporated at distinguished tiers/levels using the innovative Central Dogma Based cryptographic algorithm through the proposed Hashing Modus Operandi Based Data Warehouse Architecture, aided as an additional quantifier over the existing Data Warehouse Architecture.

References

[1] Chowdhury, R., Dey, K., S., Datta, S., Shaw, S., "Design and Implementation of Proposed Drawer Model Based Data Warehouse Architecture Incorporating DNA Translation Cryptographic Algorithm for Security Enhancement", Proceedings of International Conference on Contemporary Computing and Informatics, (IC3I 2014), Organized by Sri Jayachamarajendra College of Engineering, Mysore, Proceedings in USB: CFP14AWQ-USB, ISBN–978-1-4799-6628-8, INSPEC Accession Number–14881472, Published and Archived in IEEE Digital Xplore, ISBN–978-1-4799-6629-5, pp. 55–60.

[2] Chowdhury, R., Chatterjee, P., Mitra, P., Roy, O., "Design and Implementation of Security Mechanism for Data Warehouse Performance Enhancement using Two Tier User Authentication Techniques", International Journal of Innovative Research in Science, Engineering and Technology, An ISO 3297:2007 Certified Organization, 3 (6), 2014, Special Issue, pp. 441–449, ISSN (Online)–2319 8753, ISSN (Print)–2347 6710, Proceedings of National Conference on Emerging Technology and Applied Sciences–2014 (NCETAS 2014).

[3] Chowdhury, R., Pal, B., Ghosh, A., De, M., "A Data Warehouse Architectural Design using Proposed Pseudo Mesh Schema", Proceedings of First International Conference on Intelligent Infrastructure (CSI ICII 2012), 47th Annual National Convention of Computer Society of India, Science City Auditorium, Kolkata, Tata McGraw Hill Education Private Limited, 2012, pp. 138–141.

[4] Chowdhury, R., Pal, B., "Proposed Hybrid Data Warehouse Architecture based on Data Model", International Journal of Computer Science and Communication, 1 (2), 2010, pp. 211–213.

[5] Saurabh, A., K., Nagpal, B., "A Survey on Current Security Strategies in Data Warehouses", International Journal of Engineering Science and Technology, 3 (4), 2011, pp. 3484–3488.

[6] Vieira, M., Vieira, J., Madeira, H., "Towards Data Security in Affordable Data Warehouse", 7th European Dependable Computing Conference, (2008).

[7] Patel, A., Patel, J., M., "Data Modeling Techniques for Data Warehouse", International Journal of Multidisciplinary Research, 2 (2), 2012, pp. 240–246.

[8] Chaudhuri, S., Dayal, U., "An Overview of Data Warehousing and OLAP Technology", ACM SIGMOD Record, 26 (1), (1997).

[9] Farhan, M., S., Marie, M., E., El-Fangary, L., M., Helmy, Y., K., "An Integrated Conceptual Model for Temporal Data Warehouse Security", Computer and Information Science, 4 (4), (2011), pp. 46–57.

[10] Golfarelli, M., Maio, D., Rizzi, S., "The Dimensional Fact Model: A Conceptual Model for Data Warehouses", International Journal of Cooperative Information Systems, 7 (2–3), (1998), pp. 215–247.

[11] Golfarelli, M., Rizzi, S., "A Methodological Framework for Data Warehouse Design", Proceedings of ACM First International Workshop on Data Warehousing and OLAP, DOLAP, Washington, (1998), pp. 3–9.

[12] Chowdhury, R., Bose, R., Sengupta, N., De, M., "Logarithmic Formula Generated Seed based Cryptographic Technique using Proposed Alphanumeric Number System and Rubik Rotation Algorithm", Proceedings of IEEE 2012 International Conference on Communications, Devices and Intelligent Systems (CODIS 2012), Organized by Jadavpur University, Kolkata, Proceedings in CD: IEEE Catalog Number–CFP1207U-CDR, ISBN–978-1-4673-4698-6, Proceedings in Print: IEEE Catalog Number–CFP1207U-PRT, ISBN–978-1-4673-4697-9, INSPEC Accession Number–13285714, DOI–10.1109/CODIS.2012.6422265, Published and Archived in IEEE Digital Xplore, ISBN–978-1-4673-4700-6, pp. 564–567.

[13] Chowdhury, R., Ghosh, S., De, M., "String Graphixification based Asymmetric Key Cryptographic Algorithm using Proposed Concepts of GDC and S-Loop Matrix", Proceedings of IEEE/OSA/IAPR International Conference on Informatics, Electronics & Vision 2012 (ICIEV 2012), Organized by University of Dhaka, Dhaka, Bangladesh, Proceedings in CD: IEEE Catalog Number–CFP1244S-CDR, ISBN–978-1-4673-1152-6, Proceedings in Print: IEEE Catalog Number–CFP1244S-PRT, ISBN–978-1-4673-1151-9, Conference Proceedings: ISSN–2226 2105, INSPEC Accession Number–13058551, DOI–10.1109/ICIEV.2012.6317483, Published and Archived in IEEE Digital Xplore, ISBN–978-1-4673-1153-3, pp. 1152–1157.

[14] Chowdhury, R., Gupta, S., Saha, A., "Stochastic Seed based Cryptographic Technique [SSCT] using Dual Formula Key [DFK]", Proceedings of International Conference on Communication and Industrial Applications (ICCIA 2011), Organized by Narula Institute of Technology at Science City, Kolkata, Proceedings in CD: IEEE Catalog Number–CFP1135R-CDR, ISBN–978-1-4577-1916-5, Proceedings in Print: IEEE Catalog Number–CFP1135R-PRT, ISBN–978-1-4577-1915-8, INSPEC Accession Number–12540264, DOI–10.1109/ICCIndA.2011.6146660, Published and Archived in IEEE Digital Xplore, ISBN–978-1-4577-1915-8, pp. 1–5.

Embed Attitude from Student on Elearning Using Instructional Design with ADDIE Model

Ni Putu Linda Santiari
STMIK STIKOM Bali, High School of Management Information Technology and Computer
Denpasar, Bali, Indonesia

***Abstract*—Attitude is very important in an education, without a good attitude certainly education will not be able to run smoothly, even education can be said to fail if the output of the education did not have a good attitude in the community in the workplace. To determine the value of the attitude in elearning is not easy. In this study will try to create a method or means that can be used to determine the value of the attitude of a student in the learning system elearning. The method to be used is instructional design using ADDIE Model, where the latter begins by determining the parameters to be assessed from that attitude, the parameters used are each - each part of Affective Learning. After determining the parameters are then carried out the design and manufacture of questioner, before this questioner deployed then ever before will be testing the validity and reliability using SPSS. If questioner has valid and reliabl, then the next can be done questioner deployment and then be evaluated. Questioner from spreading to some of the students showed that students that the attitude of the students already Very Good with a total student getting very good value are 96 people with a percentage of 48%.**

Keywords—Instructional Design; eLearning; Affective Learning; Reliability; Validity

I. Introduction

Knowing the value of attitude in a learning system is not easy, there are currently very much a learning system and of course, using a variety of learning models. In Indonesia, the government has begun to allow and authorize an official to a learning system Elearning. But because it is a relatively new learning programs, such as the system still has many shortcomings. In this research, I want to try to find out the value attitude to the students in the learning system Elearning using the ADDIE Instructional Design models. The study began by determining the object and the media that will be used as the data in the study, the object to be used in this study were students STMIK STIKOM Bali, who are following Elearning classes. Then proceed by selecting one of several values in attitude, in this study I use Affective as parameter measurements [1]. Affective itself consists of the several parts or taxonomy [2], namely receiving, responding, valuing, organization and characterization.

Instructional Design used in this research is the ADDIE models. ADDIE model of instructional design is the most widely known and most widely used. ADDIE models begin with an analysis of the situation and conditions on the ground [3] [4], then proceed to create a design model or create a design questionnaires that will be distributed at the time of implementation to some object of research that students STMIK STIKOM Bali being Elearning classes. Making the design of questionnaires is done by making some questions where questions that already includes several questions about each taxonomy of Affective, then the result of the implementation or deployment of questionnaires were collected, and the data is stored as a result of implementation, and then the next is to evaluate the results implementation. With this research is expected eventually to be able to provide a solution of the problems faced by earlier about how do to know the affective value of the students who attend classes elearning. Because students generally get dross elearning little direct meetings between teachers and students, lack of emotional closeness between teachers and students will be able to be more difficult to provide an assessment of affective to students.

II. Literature Review

Previously I will explain some of the theories and methods used in this study, where the explanation I write with reference to that I got through official sources such as journals, proceedings, conference and thesis or dissertation.

A. Elearning

E-learning is a learning process that is created by the interaction with digital content, network – based services and support guidance. E-learning is a catalyst for collaboration across boundaries – national and international boundaries as well as a facilitator for networks around the world in academic study [5]. Works net facilitate the development of high – quality programs and courses insuring money, relevance and catalogs a broad curriculum. Elearning is often called the use of network information and communication technologies in teaching and learning. A number of other terms are also used to describe this mode of learning. Including online learning, virtual learning, distributed learning, networking and web-based learnings [6]. Basically, all referring to the education process that utilizes information and communication technologies to mediate asynchronous and synchronous learning and teaching activities. On closer inspection pembelajarn will become clear if the educational process for a slightly different and thus can not be used synonymously with the term e-learning.

B. Instructional Design

Instructional design according to the terms can be defined a design process and systematically to create learning more effective and efficient, and to make learning activities more easily, which is based on what we know about the theory of learning, information technology, systematic analysis, research in the field of education and methods of management [7][3].

The term instructional system development (instructional system development) and instructional design (instructional design) are often considered to be the same, or at least not explicitly distinguished in its use, although he said there is a difference in meaning between "design" and "development". The word "design" means a pattern or a sketch or outline or plan introduction. Being "development" means to make grow regularly to make something bigger, better, more effective, and so on [3].

Also according to instructional design as a process is the development of teaching systematically used exclusively theory - learning theory to making sure the quality of learning [4] [8]. Implies that the preparation of the learning plan should be in accordance with the concept of education and learning that is adopted in the curriculum that is used.

There are several models in instructional design:

- Robert Gagne's step Model
- ADDIE Model
- ARCS Model
- Roger Schank's goal - based Scenarios
- Empathic Instructional Design

C. *Attitude*

The purpose of learning is essentially the formula qualification ability to be achieved by students after learning process. The formulation of the qualifications to be possessed abilities of students after participating in the learning of this lesson with the "behavior change" (change of behavior). The type of behavior change stretcher in outline covers the fields of knowledge (cognitive), attitudes (affective) and skills (pikomotor).

Goal-oriented cognitive thinking skills. It includes more modest intellectual abilities, such as remembering, to the high capability, such as the ability to solve problems that require students connecting and combining ideas, methods or procedures that have been studied to solve a problem. Affective objectives associated with feelings, emotions, value systems and attitudes that indicate the acceptance or rejection of something. Affective goals include the ability of the simplest level, such as attention to a fenemena, to the most complex level such as determining the attitude based on the conscience. Krathwohl, Bloom, and Masia. Psychomotor goal-oriented motor skills associated with members of the body, or actions that require coordination between nerves and muscles. Psychomotor behavior neuromascular emphasis on skills, namely skills concerned with the movement of muscles.

D. *Affective Learning*

Affective learning is demonstrated by behaviors Indicating attitudes of awareness, interest, attention, concern, and responsibility, ability to listen and respond in interactions with others, and ability to demonstrate Reviews those attitudinal characteristics or values of the which are Appropriate to the test situation and the field of study [1][9][2].

- **Receiving:** Refers to the student's willingness to attend to particular phenomena of stimuli (classroom activities, textbooks, music, etc.). Learning outcomes in this area range from simple awareness that a thing exists to selective attention on the part of the learner. Receiving represents the Lowest level of learning outcomes in the affective domain.
- **Responding:** Refers to active participation on the part of the student. At this level he or she not only Attends to a particular phenomenon but Also reacts to it in some way. Learning outcomes in this area may emphasize acquiescence in responding (reads assigned material), willingness to respond (voluntarily reads beyond assignment), or satisfaction in responding (reads for pleasure or enjoyment). The higher levels of this category include instructional Reviews those objectives that are commonly classified under "interest"; that is, Reviews those that stress the seeking out and enjoyment of particular activities.
- **Valuing:** is concerned with the worth or value a student attaches to a particular object, phenomenon, or behavior. This ranges in degree from the simpler acceptance of a value (desires to improve group skills) to the more complex level of commitment (assumes responsibility for the effective functioning of the group). Valuing is based on the internalization of a set of specified values, but clues to these values are expressed in the student's overt behavior. Learning outcomes in this area are concerned with behavior that is consistent and stable enough to make the value clearly identifiable. Instructional objectives that are commonly classified under “attitudes” and “appreciation” would fall into this category.
- **Organization:** is concerned with bringing together different values, resolving conflicts between them, and beginning the building of an internally consistent value system. Thus the emphasis is on comparing, relating, and synthesizing values. Learning outcomes may be concerned with the conceptualization of a value (recognizes the responsibility of each individual for improving human relations) or with the organization of a value system (develops a vocational plan that satisfies his or her need for both economic security and social service). Instructional objectives relating to the development of a philosophy of life would fall into this category.
- **Characterization:** The individual has a value system that has controlled his or her behavior for a sufficiently long time for him or her to develop a characteristic “life – style.” Thus the behavior is pervasive, consistent, and predictable. Learning outcomes at this level cover a broad range of activities, but the major emphasis is on the fact that the behavior is typical or characteristic of the student. Instructional objectives that are concerned with the student's general patterns of

adjustment (personal, social, emotional) would be appropriate here.

E. Reasons of choosing Affective and Elearning Media

The reason why choose affektive as assessment parameters and why to use elearning as a medium of research on instructional design is because at this time, the education system in Indonesia has begun authorizing and implementing a learning system elearning [9] [2]. The education system like this is very helpful for students who wish to continue their education but are hampered by distance apart with college or place of learning.

Besides of course has its advantages, it also has some drawbacks, namely:

- Still chaotic in terms of administration.
- Due to rely solely on electronic media and distance learning, so that a direct interaction between teachers and students is reduced, this can result in a reduction in emotional interactions between teachers and students.
- Reduction of emotional interaction can be resulted in the difficulty in assessing the attitude of students.

With some of these conditions, in this study will be an instructional design (instruktional design) on a distance learning system (e-learning) with more emphasis on attitude assessment (affektive). Assessment on affektive necessary, because with the good attitude of the students, there will be a conducive atmosphere during the learning process. By doing so, the provision of material would be easier, of course, would be beneficial to both parties. In addition, at this time very much found cases - criminal cases or violations of the law committed by students. And also found on officials - government officials or institutions - private institutions or mostly orng dimasyarkat. This of course caused due to a reduction in value - the value of their attitude. Obviously we do not want this kind of thing continues to happen. That's why in this study will attempt to perform a method that will be used to assess attitudes of students. And with the acquisition of some of the results of the assessment it will be known sberapa great attitude and expected value will be used as a reference for the increase in the value of the attitude of the students.

III. Methodology and Implementation

In this study, researchers will use one of the models of some of the existing models of instructional design, research models that researchers use is the ADDIE Model [3][8].

A. Analysis

Analysis is the first step that is used in the ADDIE model, the following are some things that will be analyzed in this study, before starting to make a design quizioner [4] [3].

- Object to be used is STIKOM Bali STMIK students who are following Elearning classes. Quizioner will be distributed to several classes with each class there are 40 students.
- Parameter measurement used is affektive learning with taxonomy as follows: receiving, responding, valuing, organization and characterization.
- Quizioner consists of several individual questions – each question has an elements measurement parameters predetermined.

B. Design

After analyzing the situation and problems in the field conditions, the next step is to create a design questionnaires that will be used as a medium to determine the output of the research results which will then take the results to be evaluated [8] [4] [3]. Design questionnaires that will be made are as follows:

Questions will be made using the system obyektive or multiple-choice questions, with two possible answers, and each answer given point so that later it can be seen the number of points that will be collected from student answers to some questions and knowable level affektivenya value. Sample Questions:

1) *Do you always obey the rules laid down in class?*
   a) Strongly Agree (5 points)
   b) Agree (4 points)
   c) Neutral (3 points)
   d) Disagree (2 points)
   e) Strongly Disagree (1 point)

Questions will be made as many as 20 pieces of questions, where each consisting of four pieces of the question for the type or types of questions based on the taxonomy of afektive learning [2]. From Table 1 it can be seen that the max value that may be achieved for each student is 100 points, and the minimum value to be obtained is 20 points, the results to be obtained will be given a range of values for the points obtained, the range of possible values is like: very good, good, fair and poor, as can be seen in Table 2. From table 2 it can be seen that later after getting the results of answers on quizioner will be determined the range of values obtained eachs student.

TABLE I. Design Question and Grade of Affective Learning

Taxonomi of Affective Learning	Question	Max Point	Min Point
Receiving	4	20	4
Responding	4	20	4
Valuing	4	20	4
Organization	4	20	4
Characterization	4	20	4
Total	20	100	20

TABLE II. DESIGN A RANGE OF VALUES BASED ON YHE TOTAL POINTS OF ANSWERS ON QUIZIONER

Value Range	Receiving	Responding	Valuing	Organization	Characterization	Total Point
Very good	15 < Value ≤ 20	15 < Value ≤ 20	15 < Value ≤ 20	15 < Value ≤ 20	15 < Value ≤ 20	75 < Value ≤ 100
Good	10 < Value ≤ 15	10 < Value ≤ 15	10 < Value ≤ 15	10 < Value ≤ 15	10 < Value ≤ 15	50 < Value ≤ 75
Enough	5 < Value ≤ 10	5 < Value ≤ 10	5 < Value ≤ 10	5 < Value ≤ 10	5 < Value ≤ 10	25 < Value ≤ 50
Less	0 ≤ Value ≤ 5	0 ≤ Value ≤ 5	0 ≤ Value ≤ 5	0 ≤ Value ≤ 5	0 ≤ Value ≤ 5	0 ≤ Value ≤ 25

C. Develop and Implementation

Develop is make or realize quizioner design that was created previously [10] [3] and the implementation is to implement or deploy questionnaires that have been made previously to the object of research or student. Here are the questions quizioner which will be given to students:

TABLE III. QUESTIONS IN QUIZIONER

No.	Type	Question	Answer option				
			VG	G	N	E	L
1	Receiving	Do you like to do the type of exercises in e-learning?					
2	Receiving	Are you happy to elaborate on their opinions in e-learning?					
3	Receiving	Are you happy and always read the material provided by the teacher?					
4	Receiving	Are you happy and always listen and watch the video given by the teacher?					
5	Responding	Do you always obey the rules laid down in class?					
6	Responding	Do you often respond to the opinion of the teacher or a friend in class in a discussion held in the classroom or e-learning?					
7	Responding	Do you always do all the tasks assigned by the teacher?					
8	Responding	If the misunderstanding that occurred as a result of mistakes you do, will you apologize for your mistake?					
9	Valuing	Are you happy to provide an assessment of a material that dberikan by the teacher?					
10	Valuing	Do you always appreciate or give an appreciation of the role played by the teachers during the learning process?					
11	Valuing	Are you happy to give attention or assessment of teachers or friends - classmates in the learning process?					
12	Valuing	Are you happy to give sympathetic to teachers or friends - classmates?					
13	Organization	If there is a problem if you are going to discuss these issues with your friends - to get a solution to these problems?					
14	Organization	Do you always disciplined, diligent and punctual in attending the discussion or video konference (class elearning) given by the teachers or held in the classroom?					
15	Organization	Are you happy to provide solutions to problems in the discussions held on the problems encountered?					
16	Organization	Are you happy to greet a friends or a teacher in the daily activities - day either directly or through the medium of the internet?					
17	Characterization	Are you going to explain the subject matter to your friends who do not understand about the material?					
18	Characterization	You will keep the spirit undergo the course even if you know little about these subjects?					
19	Characterization	You will not impersonate or cheating jobs friend even though you do not understand the lecture					
20	Characterization	You will always seek to understand how a lot of reading and ask the teacher or friend who understand					

After making as many as 20 questions on quizioner, the next step is the selection of the sample, the quizioner multiply and spread to some predetermined Students, Students will be given a few clues about the workmanship and given time to work. Quizioner deployment is done through two phases, the first one conducted are validity and reliability to determine whether quizioner who has made a valid and realible. Then, if it has been found that quizioner made is valid and realible, then spread quiozioner then performed again, to get results that can later be evaluated as to whether the value obtained from each student.

D. Evaluate

After implementation in the form of the spread questioner to some students and then the supervisor waited until the time expires, then all questionnaires were collected and subsequently be given the results of its implementation and assessment, kemdian of the results of the validation test will be conducted and the reliability [11][12].

Here are the results of the implementation and assessment of some of the students:

TABLE IV. RESULTS EVALUATION FOR TEST VALIDATION AND RELIABILITY

No.	Name	Recv				Resp				Valu				Orgn				Char				Total Point
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
1	ROMANDA SAGITA PUTRA	5	4	3	1	3	4	5	3	3	3	3	4	4	4	4	4	4	2	2	2	67
2	ALEXANDER PRANOTO SANJAYA	4	4	2	2	4	5	4	4	4	3	3	3	3	4	4	5	5	5	5	5	78
3	I PANDE GEDE HENDRA MAHARDIKA	3	4	4	4	2	4	3	3	3	3	3	5	5	5	5	5	5	2	2	3	73
4	M DZAKARIA ILHAMSYAH PUTRA	4	4	4	4	1	5	5	5	5	5	3	3	3	3	2	2	3	3	2	2	68
5	I WAYAN ANDRIANA	4	4	3	3	3	4	5	5	5	5	4	4	4	3	3	3	2	2	2	3	71
6	NI LUH ARYANI KUSUMA D	1	2	4	4	4	4	4	5	5	5	5	5	3	3	3	3	3	2	2	3	70
7	AZWAR ANAS	3	3	3	3	3	4	4	4	4	5	5	5	5	2	2	3	3	4	5	5	75
8	BAIQ DIAN ERI SAFITRI	3	3	3	4	4	4	4	3	3	3	3	2	2	3	4	5	5	5	5	5	73
9	VINGKY EKA SUSILAWATI	4	4	4	5	5	5	5	5	5	5	3	3	3	2	2	2	4	4	4	3	77
10	IDA BAGUS GEDE GIRI ASRAMA	5	5	5	3	3	3	2	2	3	3	4	4	4	5	5	4	3	3	3	4	73
11	M ARIK TRI SUTRISNO	5	5	3	3	3	4	4	4	4	4	4	3	3	3	3	4	4	4	5	5	77
12	I PUTU ANDRE IRAWAN	4	4	4	4	3	3	5	2	2	2	3	3	4	4	4	3	3	3	5	5	70
13	AHMAD FATHUR RIDHO	5	5	5	5	4	4	4	3	3	4	3	4	5	3	2	3	3	4	2	2	73
14	FIRMAN EVENDI	5	5	5	4	4	4	3	3	3	2	2	2	3	1	4	4	4	5	5	5	73
15	I PUTU AGUS WIDIANTARA PUTRA	4	4	4	2	3	4	5	5	5	5	4	4	4	3	3	3	2	2	3	3	72
16	HAIRIL MOHI	5	5	5	4	4	2	3	4	5	4	4	4	3	4	3	2	4	5	3	5	78
17	ESTI WULANSARI	4	5	5	5	4	4	4	5	5	5	3	4	3	4	5	2	3	3	2	2	77
18	NI MADE MARCELLINA DEWANTI	4	4	4	5	5	5	5	5	3	3	3	3	2	2	1	4	4	4	5	5	76
19	SLAMET RIYANTO	3	3	3	2	2	4	4	5	5	5	2	4	4	5	5	5	5	5	3	3	77
20	I NYOMAN GEDE ADIPRADNYA S.	2	2	2	3	3	3	4	5	5	5	5	4	4	4	3	3	3	2	2	3	67
21	NYOMAN TRIJATA ADI WIJOYO	5	5	5	5	5	5	4	4	4	3	2	2	3	3	4	4	3	3	3	5	77
22	ERICK HERYANTO PUTRA	4	4	4	4	3	3	2	2	1	2	4	4	3	3	3	4	5	5	4	5	69
23	ARIF NASRUDIN	3	3	3	3	4	4	4	4	5	5	5	5	2	2	1	4	4	3	3	3	70
24	NGAKAN GEDE EFANO YUDHA P.	1	4	4	3	2	5	2	5	5	5	2	3	4	4	4	4	5	4	5	5	76
25	ADI FEBRIANA RAMDANI	3	3	3	4	4	4	5	5	5	3	3	4	4	2	2	4	4	5	5	2	74
26	I GST AYU DIAH CANDRADEWI	3	4	4	5	5	5	5	5	4	4	4	4	4	3	5	4	3	3	2	1	77
27	LANNY JANNE CINTHIA GOSAL	4	4	4	5	5	2	2	3	3	3	2	2	3	3	4	5	3	3	4	5	69
28	RENANDA NUR RUMARA	3	3	3	2	2	2	3	3	5	5	5	5	4	4	5	4	4	5	4	2	73
29	I NYOMAN ARDIKA	3	3	4	4	4	3	3	3	3	4	4	3	2	2	5	5	5	5	2	1	68
30	NIKITA FITRIANI IMA BOE CHARI	4	4	4	2	2	3	3	1	4	4	5	5	5	5	4	3	4	4	4	2	72
31	KURNIA DITA SAPUTRA ASWAL	5	3	3	2	2	2	1	4	5	5	5	5	4	4	4	3	2	1	4	3	67
32	ANGGORO RAHMAN MUHAMAD	5	5	5	3	3	3	2	2	2	3	3	3	4	4	3	3	3	5	5	5	71
33	DIAN PRAMONO PUTRA	3	3	4	4	4	4	3	2	2	4	4	4	5	5	5	4	4	3	3	4	74
34	DEWA AYU KRISNA DEWI	5	4	4	4	3	5	5	5	4	4	3	3	4	4	3	3	4	3	4	1	75
35	FADIL AHMAD	4	5	3	4	5	4	4	4	3	3	2	2	4	4	5	5	5	4	3	2	75
36	CHAIRIL ANAM	4	4	3	2	4	4	3	5	4	4	5	5	4	4	3	3	2	1	2	4	70
37	NI LUH SINTA PURNILA DEWI	3	3	2	2	4	4	4	5	5	4	5	4	4	4	3	3	2	1	4	4	70
38	I GEDE SUDARMA YASA	5	3	4	4	4	4	3	3	3	2	3	4	5	5	4	4	5	4	2	1	72
39	I GEDE ASTAWA	3	2	4	4	5	2	5	2	5	5	5	2	2	2	3	4	4	4	2	3	68
40	I NYOMAN ARI SURYADI	3	4	4	4	5	5	5	5	5	5	5	4	4	4	3	5	4	4	2	1	81

Recv : Receiving   Orgn : Organization   Char : Characterization
Resp : Responding   Valu : Valuing

After getting the answer from quizioner deployment is done, the next step is to test the validity and reliability, validity testing performed to test whether quizioner made valid or not. Then the reliability testing conducted to determine the extent to which the measurement results remain consistent, if the measurement is done twice or more of the same symptoms using the same gauge as well.

1) Validity Test

The validity of the test can be done by calculating the value of r table and calculate the value of r count, if the product moment correlation coefficient or r count larger than r table then quizioner is declared valid [13] [11]. Some things need to be done to determine the validity of a quizioner are as follows:

a) Determining the value of a significant level, in this study it is determined that the value is a significant level α = 5%.

b) Determining the value of DF (Degree of Freedom), DF value can be determined by the formula DF = N - 2, where N is the number of respondents, in this study will be *determined that the number of respondents was N = 40 people. So it can be determined that the value of DF = 38.*

c) Determining the value of r table, table r value can be determined based on the value of the DF with reference to Table Product Moment, with a value of DF = 38 table shows that the value of r = 0.320.

TABLE V. PRODUCT MOMENT

DF	TF 5 %	DF	TF 5 %	DF	TF 5 %	DF	TF 5 %
1	-	11	0.602	21	0.433	31	0.355
2	-	12	0.576	22	0.423	32	0.349
3	0.997	13	0.553	23	0.413	33	0.344
4	0.950	14	0.532	24	0.404	34	0.339
5	0.878	15	0.514	25	0.396	35	0.334
6	0.811	16	0.497	26	0.388	36	0.329
7	0.754	17	0.482	27	0.381	37	0.325
8	0.707	18	0.468	28	0.374	38	0.320
9	0.666	19	0.456	29	0.367	39	0.316
10	0.632	20	0.444	30	0.361	40	0.312

d) Determining the value of count r, r count value may be carried out using SPSS, based on the results that have been obtained then the data can be entered into SPSS and then processed to obtain the following results:

TABLE VI. DESCRIPTIVE STATISTICS

	Mean	**Std. Deviation**	**N**
score question	3.90	.955	40
score question	4.25	.670	40
score question	3.38	1.254	40
score question	3.93	.764	40
score question	3.78	1.097	40
score question	3.88	.791	40
score question	3.75	.742	40
score question	3.88	.757	40
score question	3.88	.791	40
score question	3.78	.800	40
score question	3.80	.823	40
score question	3.88	.791	40
score question	3.80	.791	40
score question	3.83	.747	40
score question	3.78	.800	40
score question	3.83	.747	40
score question	3.90	.810	40
score question	3.78	.832	40
score question	3.75	.840	40
score question	3.93	.829	40
Total question	76.63	3.271	40

TABLE VII. CORRELATIONS

		SP 1	SP 2	SP 3	SP 4	SP 5	SP 6	SP 7	SP 8	SP 9	SP 10
Total Answers	Pearson Correlation	**.480****	**.340**	**.348***	**.355**	**.440****	**.348**	**.318**	**.319**	**.378***	**.361**
	Sig. (2-tailed)	.002	.900	.028	.112	.004	.067	.010	.011	.016	.004
	N	40	40	40	40	40	40	40	40	40	40

		SP 11	SP 12	SP 13	SP 14	SP 15	SP 16	SP 17	SP 18	SP 19	SP 20
Total Answers	Pearson Correlation	**.400**	**.370**	**.348**	**.340***	**.373**	**.317**	**.331**	**.334**	**.321**	**.346**
	Sig. (2-tailed)	.006	.095	.023	.032	.087	.017	.022	.034	.098	.078
	N	40	40	40	40	40	40	40	40	40	40

Based on the validity of the test results were performed using SPSS, the results obtained in the form of two tables as follows:

a) Descriptive Statistics table, in this can be seen that the number of respondents who became the sample is 10 people. And can also be seen the averages score of the answers obtained from each question there.

b) Correlations table can be analyzed that the validity of the test results data showed all the questions are valid, because the value of r count larger than r table = 0, 320.

2) Test Reliability

Reliability test was conducted to determine the extent to which the measurement results remain consistent [14], when the measurements were taken twice or more of the same symptoms using the same gauge as well.

Criteria for a research instrument is said to be reliable by using this technique, when the reliability coefficient (r11) > 0, 6. Here is the requirement to measure the measuring instrument according to Cronbach Alpha table [15].

TABLE VIII. ALPHA CRONBACH

Value	Description
r 11< 0.20	Very low
0.20 ≤ r 11 < 0.40	Low
0.40 ≤ r 11 < 0.70	Medium
0.70 ≤ r 11 < 0.90	Height
0.90 ≤ r 11 < 1.00	Very Height

Reliability testing can be performed with the SPSS, the following are the results of reliability testing using SPSS: Berdasarkan table Reliabilty Statistics yang telah didapatkan dari pengujian menggunakan SPSS, maka dapat diketahui bahwa instrument penelitian dinyatakan reliable, karena nilai r 11 = - 0. 650 > 0. 6.

TABLE IX. RELIABILITY STATISTICS

Cronbach's Alpha[a]	**N of Items**
.650	20

Based on the table reliabilty Statistics which have been obtained from test using SPSS, it is known that otherwise reliable research instrument, because the value r 11 = 0. 650 > 0. 6. Having done testing the validity and reliabitas against quizioner that have been made, then quizioner is declared to have valid and reliable, meaning quizioner this can be distributed to several respondents that later showed a value of answers pertanyanaan which have been answered by the respondents, the following are the results obtained from the spread quizioner has been done:

TABLE X. EVALUATION OF THE IMPLEMENTATION QUIZIONER

No.	**Name**	**Recv**				**Resp**				**Valu**				**Orgn**				**Char**				**Total Point**
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**	**20**	
1	ROMANDA SAGITA PUTRA	4	4	4	4	3	3	3	2	2	2	3	4	4	5	5	5	5	5	4	4	75
2	ALEXANDER PRANOTO SANJAYA	3	3	3	3	4	4	4	4	3	3	3	5	5	5	2	2	5	5	5	3	74
3	I PANDE GEDE HENDRA MAHARDIKA	4	4	4	3	3	3	2	4	4	4	5	5	5	3	3	4	4	4	5	2	75
4	M DZAKARIA ILHAMSYAH PUTRA	2	2	2	3	3	3	4	4	4	3	3	3	4	5	5	5	4	4	4	4	71
5	I WAYAN ANDRIANA	3	3	3	4	4	4	5	5	5	3	3	4	4	4	4	3	2	4	4	5	76
6	NI LUH ARYANI KUSUMA D	4	4	4	5	5	3	3	3	4	3	3	3	3	2	2	5	5	5	4	4	74
7	AZWAR ANAS	4	5	5	5	3	3	3	4	4	4	3	3	2	2	4	4	4	4	4	4	74
8	BAIQ DIAN ERI SAFITRI	4	1	4	2	3	4	3	4	4	5	4	4	3	3	3	4	4	5	5	5	74
9	VINGKY EKA SUSILAWATI	4	1	1	2	3	3	3	4	4	4	5	5	5	4	4	4	4	5	5	2	72
10	IDA BAGUS GEDE GIRI ASRAMA	4	4	4	5	5	5	3	3	4	4	2	2	3	3	4	4	5	5	5	2	76
11	M ARIK TRI SUTRISNO	5	5	5	5	3	4	4	4	5	5	5	4	4	4	3	3	2	4	1	1	76
12	I PUTU ANDRE IRAWAN	4	1	2	4	4	3	4	4	4	4	3	3	4	4	5	5	5	5	4	4	76
13	AHMAD FATHUR RIDHO	5	5	4	4	3	3	2	2	4	1	5	5	5	5	4	4	3	3	4	4	75
14	FIRMAN EVENDI	4	4	4	4	1	3	4	4	5	5	5	5	3	3	4	4	2	5	1	4	74
15	I PUTU AGUS WIDIANTARA PUTRA	5	4	1	2	3	5	5	4	4	5	5	5	4	4	2	2	2	3	4	4	73
16	HAIRIL MOHI	5	4	5	5	5	4	4	4	3	3	1	1	4	4	5	5	4	4	2	2	74
17	ESTI WULANSARI	4	4	4	5	5	4	4	2	2	2	1	2	3	3	4	4	4	5	5	5	72
18	NI MADE MARCELLINA DEWANTI	4	5	5	5	5	5	4	4	4	3	2	2	4	5	5	5	4	2	3	1	77
19	SLAMET RIYANTO	5	4	4	4	5	5	5	4	4	4	4	3	3	3	2	2	4	5	5	2	77
20	I NYOMAN GEDE ADIPRADNYA S.	5	4	3	4	4	4	3	3	4	4	4	5	5	5	5	5	5	4	3	3	82

No.	Name	Recv				Resp				Valu				Orgn				Char				Total Point
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	
21	NYOMAN TRIJATA ADI WIJOYO	4	4	4	4	4	3	3	3	4	4	5	5	5	5	3	3	4	4	2	1	74
22	ERICK HERYANTO PUTRA	4	5	2	3	3	3	3	4	4	4	4	4	5	5	5	5	5	4	3	1	76
23	ARIF NASRUDIN	1	5	2	3	3	4	4	4	4	4	5	5	5	5	4	4	5	5	3	1	76
24	NGAKAN GEDE EFANO YUDHA P.	1	5	2	2	3	3	4	4	2	4	5	5	5	5	5	5	4	4	4	4	76
25	ADI FEBRIANA RAMDANI	2	4	5	5	5	4	4	4	3	3	2	1	4	4	4	5	5	5	5	1	75
26	I GST AYU DIAH CANDRADEWI	2	4	1	2	3	3	4	4	4	4	4	5	5	5	5	4	4	4	4	4	75
27	LANNY JANNE CINTHIA GOSAL	3	4	4	4	4	4	5	5	5	5	5	4	4	4	3	3	2	2	4	1	75
28	RENANDA NUR RUMARA	3	3	1	2	2	3	3	4	4	4	5	5	5	5	5	4	4	4	4	4	74
29	I NYOMAN ARDIKA	5	3	4	4	4	5	5	5	5	4	4	4	3	3	2	2	4	4	4	2	76
30	NIKITA FITRIANI IMA BOE CHARI	5	2	2	4	4	3	5	5	5	5	4	4	2	2	4	4	2	4	4	4	74
31	KURNIA DITA SAPUTRA ASWAL	4	1	2	2	4	4	4	5	5	5	3	3	4	4	4	5	5	5	5	2	76
32	ANGGORO RAHMAN MUHAMAD	4	1	3	3	3	3	3	3	3	3	4	4	4	4	4	5	5	5	5	5	74
33	DIAN PRAMONO PUTRA	4	4	3	3	4	4	4	4	5	5	5	5	5	4	4	4	2	2	4	1	76
34	DEWA AYU KRISNA DEWI	5	5	5	5	4	4	4	3	3	2	2	1	4	4	4	5	5	5	4	1	75
35	FADIL AHMAD	5	5	5	5	2	4	4	3	3	4	4	4	4	4	3	2	4	4	3	3	75
36	CHAIRIL ANAM	5	5	5	5	4	4	4	2	2	4	4	5	5	5	4	4	4	4	2	1	78
37	NI LUH SINTA PURNILA DEWI	1	2	2	4	4	5	5	5	5	4	4	3	3	4	4	4	5	5	5	1	75
38	I GEDE SUDARMA YASA	5	5	5	4	4	4	3	3	3	2	2	5	5	5	5	3	4	4	4	1	76
39	I GEDE ASTAWA	4	2	4	4	4	4	3	5	3	3	4	3	3	5	5	5	5	3	4	4	77
40	I NYOMAN ARI SURYADI	1	2	3	5	4	4	5	5	3	4	3	2	4	4	1	4	5	5	5	5	74

Recv : Receiving Orgn : Organization Char : Characterization
Resp : Responding Valu : Valuing
Sangat Baik Cukup
Baik Kurang

TABLE XI. NUMBER AND PERCENTAGE VALUE OF AFFECTIVE LEARNING

Affective Learning	Very Good		Good		Enough		Less	
Receiving	19 People	47.50%	14 People	35.00%	7 People	17.50%	-	-
Responding	16 People	40.00%	**23 People**	57.50%	1 People	2.50%	-	-
Organization	19 People	47.50%	17 People	42.50%	4 People	10.00%	-	-
Valuing	22 People	55.00%	16 People	40.00%	2 People	5.00%	-	-
Characterization	20 People	50.00%	16 People	40.00%	4 People	10.00%	-	-

IV. CONCLUSIONS

Based on the evaluation of the questionnaires with as many as 20 questions that each contained four questions about each Affective Taxonomy has been created and distributed to 40 students. Where previously the questions that have been tested for validity and reliability. From the results of the evaluation showed that there is a maximum total value obtained by a student is 82 points (Very Good) obtained by one student and no students are getting less value. If viewed from each Affective Learning Affective Learning Taxonomy have maximum value is on Responding to the number of students as many as 23 people with a percentage of 57.50% with a range of values Good. Based on this study concluded that based on questions spread was found that the value of the attitude of the students already Very Good. Expected by this research will be used as a new knowledge or can be used as a reference so that later can be used further and can help in assessing the attitudes of students, especially in elearning learning system.

REFERENCE

[1] R. W. Picard, S. Papert, W. Bender, B. Blumberg, C. Breazeal, D. Cavallo, T. Machover, M. Resnick, D. Roy, and C. Strohecker, "Affective Learning — A Manifesto," BT Technol. J., vol. 22, no. 4, pp. 253–269, 2004.

[2] K. N. Allen and D. Ph, "Affective Learning: A Taxonomy for Teaching Social Work Values," J. Soc. Work Values Ethics, vol. 7, no. 2, 2010.

[3] B. M. Ngussa, "Application of ADDIE Model of Instruction in Teaching-Learning Transaction among Teachers of Mara Conference Adventist Secondary Schools , Tanzania," J. Educ. Pract., vol. 5, no. 25, pp. 1–11, 2014.

[4] B. S. Wang and H. Hsu, "Using the ADDIE Model to Design Second Life Activities for Online Learners," vol. 53, no. 6, 2009.

[5] H. Cetin and E. Ertekin, "The Relationship Between Eighth Grade Primary School Student's Proportional Reasoning Skills," Int. J. Instr., vol. 4, no. 1, 2011.

[6] H. Yueh, W. Lin, and W. L. Corporation, "Effect of Student Engagement on Multimedia - Assisted Instruction," Int. J. Knowl. Manag. Elearning, vol. 4, no. 3, pp. 346–358.

[7] R. N. Melo and L. B. Baruque, "Learning Theory and Instructional Design Using Learning Object," Int. J. Scholarsh. Teach. Learn., vol. 5, no. 1, 2011.

[8] K. T. Dewi, I. W. Suastra, N. M. Pujani, P. Pascasarjana, and U. P. Ganesha, "Pengaruh Model Pembelajaran Analyse ,Design ,Develop , Implement , Evaluate (ADDIE) Terhadap Keterampilan," E-Journal Progr. Pascasarj. Univ. Pendidik. Ganesha, vol. 3, no. 1, 2013.

[9] R. Schroeder, "Valuing Information Literacy: Affective Learning and the ACRL Standards," vol. 10, no. 2, pp. 127–146, 2010.

[10] P. Schaber, A. L. Whiteside, and D. C. Brooks, "Designing Learning Environments to Foster Affective Learning : Comparison of Classroom to Blended Learning Designing Learning Environments to Foster Affective Learning :," Int. J. Scholarsh. Teach. Learn., vol. 4, no. 2, 2010.

[11] Ö. Korkmaz, "A validity and reliability study of the Attitude Scale of Computer Programming Learning (ASCOPL) Halis Altun," Int. J. Educ., vol. 4, no. 1, pp. 30–43, 2014.

[12] A. Test, "Validity and Reliability Study of the Internet Addiction Test," Int. J. Educ., vol. 3, no. 4, pp. 207–222, 2013.

[13] S. Fan, "Developing a Valid and Reliable Instrument to Evaluate Users ' Perception of Web-Based Learning in an Australian University Context," J. Online Learn. Teach., vol. 7, no. 3, pp. 366–379, 2011.

[14] I. Kalender, "Reliability-Related Issues in the Context of Student Evaluations of Teaching in Higher Education," Int. J. High. Educ., vol. 4, no. 3, 2015.

[15] M. J. Warrens, "On Cronbach ' s Alpha as the Mean of All Possible - Split Alphas," Adv. Stat., vol. 2014, pp. 5–10, 2014.

[16] M. Tavakol and R. Dennick, "Making sense of Cronbach ' s alpha," Int. J. Med. Educ., pp. 53–55, 2011.

Enhancing Performance of GIS on Cloud Computing

Ahmed Teaima A.T
Institute of Statistical Studies and Researches.
Cairo Univ.
Cairo, Egypt

Hesham Ahmed Hefny. H. F. H
Institute of Statistical Studies and Researches.
Cairo Unive.
Cairo, Egypt

BahaaShabana B.S
Misr Higher Institute for Commerce & Computer, Mansoura Univ.
Mansoura, Egypt

***Abstract*—Cloud computing provides a way of determining dynamically scalable and virtualized resources as a service over the Internet. GIS is a technology, which could use Cloud Computing for distributed parallel processing of a large set of data, store and share the results with users around the world. GIS is beneficial and works well when it be available to everyone, everywhere, anytime and with downcast fee of minimal sized in terms of technology and outlay. Cloud Computing used to portray and help users to use GIS applications in an easy way. This paper will study some example of a data structure like a K-d tree and Quad trees of GIS application and compare between them when storing these data structures on Cloud computing, the paper also portrays the results of the study of data structure on cloud computing platforms to retrieve data from cloud computing. The paper provides an application for "finding neighborhood from existing data stored.**

Keywords—Cloud Computing; GIS; Kd-tree; Quadtree

I. Introduction

Geographical information system (GIS) is a group of Tools that analyzes, stores, manages, captures and presents visual data that are associated with geographical locations" this assumes a definition of the acronym of GIS [11]. GIS or geospatial information studies play a prominent role in many fields and widely adopted nowadays. In another view, it is any information system merging of statistical analysis, cartography hardware, software, and special types of DB (huge size-different shapes -…) and data to provide information and present the result of all these operations. GIS used in decision –making as in public health [12] which describe the relation between distribution diseases and concentration vary in different locations for making best possible decision by using spatial relations between it, visualizing the data to produce information and processing these data. In addition, in a pilot project designed to explore the potential for an information tool and educate sector engagement model to benefit the sector and its communities in the transport corridor to the north of Brisbane. By allowing participants, community, government and non-government organizations (NGO) to access information at a regional level to assist with decision making and the evaluation of shared cross-sector service provision and planning initiatives. [15]

Over a few decades, efforts made to upgrade applications of GIS in order to provide huge spectrum services to the users through the globe. For example, but not limited to, application of integration, GIS and hydrology, by monitoring of Surface water and Groundwater resources is dependent on dynamic and static parameters of these water systems as well as meteorological data sets. All this information is large in volume and spatial as well as temporally varying in Nature. [13] Another one of using GIS in watershed management. By studying the basic characteristic of watershed likes, drainage network and flow paths derived from readily available Digital Elevation Models (DEMs) and USGS's National Hydrography Dataset (NHD) program. [14]. Cloud computing has emerged as a paradigm to deliver on demand resources (e.g., infrastructure, platform, software, etc.) to customers similar to other utilities (e.g., water, electricity and gas). [16]

Cloud Computing can be used across the challenges in GIS applications. GIS is a complete System of Hardware, Software, and Spatial Data (topographic, demographic, graphic image, digitally...) performs processing and analysis operations on those data to produce reports, graphics, statistics, and controls geographic data processing workflows. [1] [8]

II. Preview on Different GIS Databases

The GIS has a special natural due to the large amount of data, the way of storing this data, and the experts who deal with it. The recent emergence of cloud computing brings new possibilities in service deployment. Services deployed in environments that made to scale up or down as required, with the service provider only charged for actual usage. Many types of database can be used for storing the spatial data like Quad tree, R tree and K-d tree I will look for those types as a background.

A *quad tree* is a tree data structure used to represent a picture successively in deeper level represented the best subdivisions of picture areas. Each node represents and links to the quadrant of its parent. A process of subdividing an image matrix into four quadrants parts recursively until every part has unique color fills up the tree. A Quad tree is trees whose nodes are either leaf (no children) or have four children Fig 1 show the shape of quad tree structure. The children are arranged one, two, three, and four. The region of the quad tree describes a piece of space in two dimensions (like X and Y) by dividing the region into four equal quadrants, sub quadrants, and so on with each leaf (which mean the node is the last one) node containing data corresponding to a specific sub region. Each node in the tree must have exactly four children, or have leaf node. [17]

The region quad tree is not strictly a 'tree' as the positions of subdivisions are independent of the data. They precisely called 'tries'. A region quad tree with a depth of n (number of levels) used to represent an image consisting of 2n × 2n pixels, where each pixel value is zero or one [2]. The advantage of the quad tree lies in reducing the complexity of the intersection process by enabling the pruning of certain objects or portions of objects from the query. The disadvantage of quad tree, it must planed scope beforehand. [18]

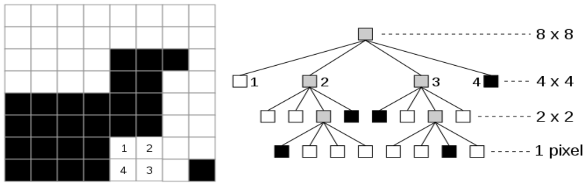


Fig. 1. Quad tree structure

Rtree is a real extension of B-trees (Comer 1979), which refers to binary search tree, in that a node can have more than two children, unlike self-balancing binary search trees, the B-tree optimize for systems that read and write large blocks of data. B-trees are a good example of a data structure for external memory. A B - tree commonly used in databases and file systems. B-tree does not need re-balancing as frequently as other self-balancing search trees, but may waste some space [19]. The *R tree* (**R**ectangle tree) is a data structure, use in multiple dimensions. Which is height-balanced tree [2]. It consists of two levels of storage, medium and last node (leaf node). The stats of last nodes and medium nodes stored the data objects built by gathering rectangles at the lower level. The collected data in rectangle shape as shown in Fig 2, which illustrate the structure of the R tree. The nodes can be covered, overlapping, or completely disjoint, no assumption about their properties. The Minimum Bounding Rectangles (MBRs) of the actual data objects assume stored in the last node of the tree. Each medium node is associated with some rectangle, which completely encloses all rectangles that correspond to lower level nodes. [3]

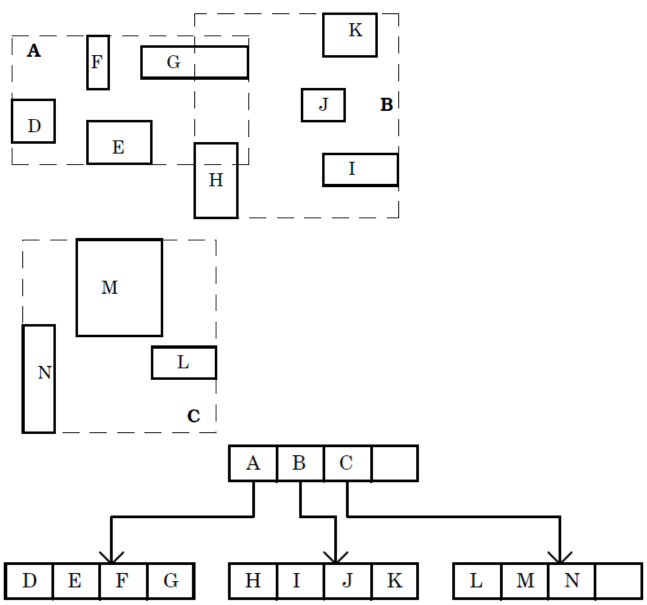


Fig. 2. R tree structure

k-d tree (short for k-dimensional tree) is a space-partitioning data structure for organizing points in a k-dimensional space. k-d trees are a useful data structure for several applications, such as searches involving a multidimensional search key (e.g. range searches and nearest neighbor searches). k-d trees are a special case of binary space partitioning trees [20].k-d tree also used in computer vision and machine learning algorithms consists of finding nearest neighbor matches to high dimensional vectors that represent the training data.[21]. However, a key problem of data driven tree structures is the capability of data update. Each point insertion or deletion requires the modification of large parts of the actual tree structure. [22]

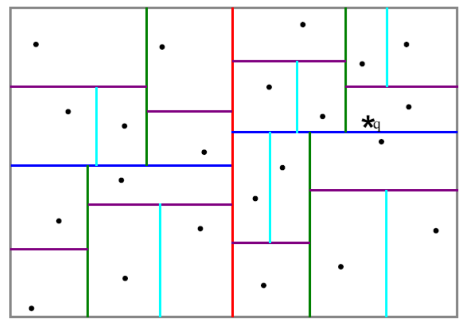


Fig. 3. k-d tree structure

III. PREVIOUS WORK

To implement spatial applications efficiently requires the use of spatial data structures, which used to store data objects that linked with location and an important class of data structures used in computer graphics, geographic information systems, and many other fields.

To improve performance of K-d tree and Quad tree have different shapes. For example, the building of the data structure which represent in the using of mathematical mean by using median of data [9]. Another shape of improving performance, in this paper the author use the same data structures K-d tree and Quad tree addition to Tile arrays and by ignoring the unnecessary objects, the time of retrieving data is decreased as in [10]

IV. METHODOLOGY

Fig 4 illustrates the GUI of The application .It builds to display the performance of different types of GIS data structures, quad tree, and K-d tree. The application has the ability to search, add, delete, and update the points. In addition to, it has to division the map to different scale from 1 to 9 and prints the name of site beside the location on the tree. The application contains the ability to find the nearest neighbor from any record exits in DB. All the process occurs on the cloud computing on the internet. The database structures storage on the “SOMEE.com.” The application can play online and off line.

The database of application uploaded to the cloud computing which represented by SOMEE.COM as a hosting DB. The application loaded the database in the first running of application Fig 5 shows the time of searching in database.

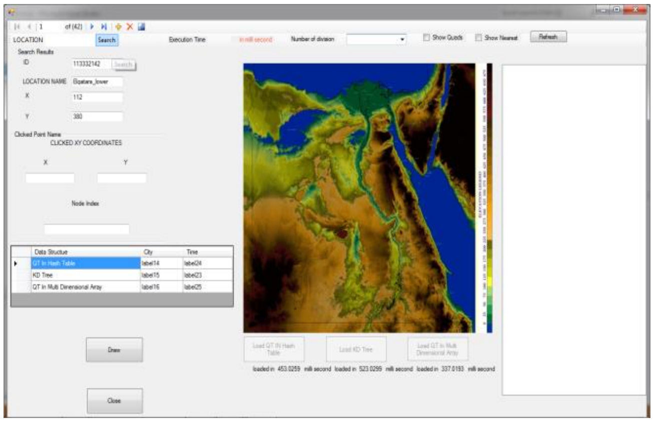


Fig. 4. GUI of the project

Hash table, and a Multidimensional array storing the data in data structure, the application illustrates the time response in Fig 5 represents the search time in DB.

Fig. 5. Search time in DB

Fig 6 illustrates the division in (4) degree. Moreover, the red point is the result of search for the point under search and its location in the tree.

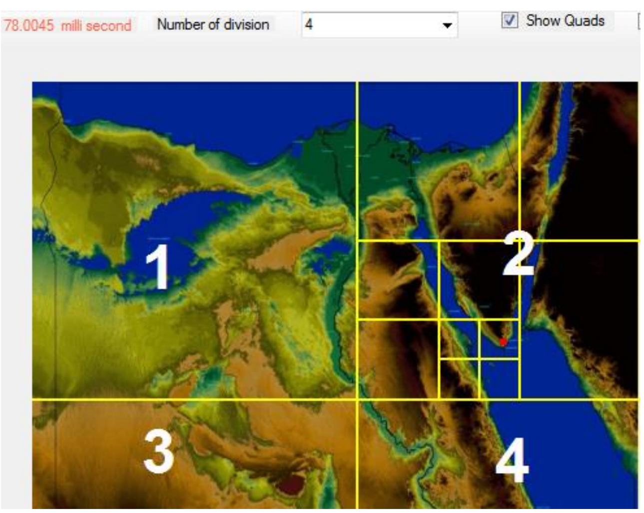


Fig. 6. Snap shoot for the result on map

We have one database but with different data structure.Actually, we use hash tables and multidimensional array.

A hash function is any function that used to map data of arbitrary size to data of fixed size. The values returned by a hash function called hash values, hash codes, hash sums, or simply hashes. One use is a data structure called a hash table, widely used in computer software for rapid data lookup. Hash functions accelerate table or database lookup by detecting duplicated records in a large file.

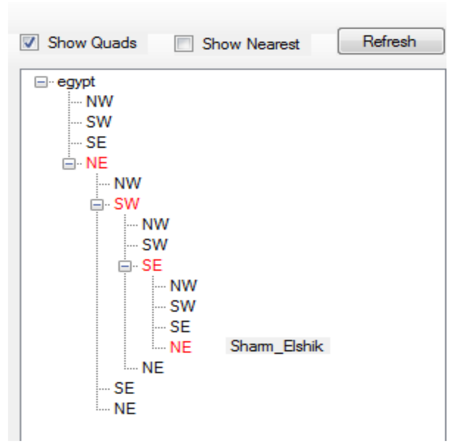


Fig. 7. locations in the tree

Hash functions used in hash tables, to find out a data record (a dictionary definition) given its search key (the headword). Specifically, the hash function used to plane the search key to an index. The index gives the address in the hash table where the suitable record should be stored. Hash tables, sequenced, used to implement associative arrays and dynamic sets. [23]

Typically, the domain of a hash function (the set of possible keys) is larger than its range (the number of different table, indexes), and so it will plan several different keys to the same index. Therefore, each slot of a hash table is associated with (implicitly or explicitly) a multi of records, rather than a single record. For this reason, each slot of a hash table often called a "bucket," and hash values called "bucket indices." The hash function only hints at the record's location — it tells where we can start looking for it. Still, in a half-full table, a good hash function will typically decrease the search down to only one or two entries. Hash table used in many applications like an approximate nearest neighbor.

Searching in large databases has become popular owing to its computational and memory efficiency. The famous hashing methods, e.g., *Locality Sensitive Hashing (LSH) and Spectral Hashing (SH)*, construct hash functions based on random or principal projections. [4]. The *complementary hashing approach*, is an approach used hash table, which is able to balance the precision and recall in a more effective way. The key idea is to employ multiple complementary hash tables, which are learned sequentially in a boosting manner, so that, given a query, its true nearest neighbors missed from the active bucket of one hash table are more likely to be found in the active bucket of the next hash table. [5]

Now we look for another type of data structure "*multidimensional array*." It is recognized in the past that, traditional database Management Systems (DBMSs) does not handle efficiently multi-dimensional data (which are geometrical shapes in our search) such as squares, Polygons, or even points in a multi-dimensional space. Multidimensional data arise in many applications, such as the most important fields:

1) Cartography, Maps could be stored and searched electronically answering efficiently geometric queries.

2) *Computer-aided Design (CAD).*
3) *Computer vision and robotics.*
4) *Rule indexing in expert database systems. [6].*

On some applications the focus on increase the memory storage and reduce the conflict of data access to use one dimensional array, but by use the automatic partitioning memory scheme for multidimensional arrays based on linear transformation to provide high data throughput of memory storage the experimental illustrate that saving in memory banks and digital signal processing (DSP). [7]

The application starts as in Fig 8.

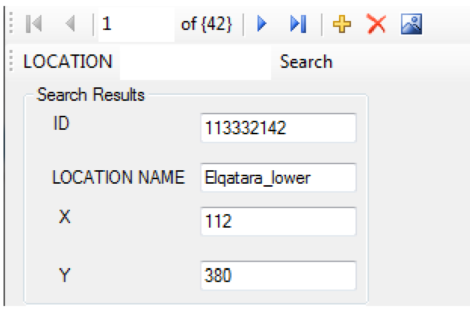


Fig. 8. Start of the project

We have 42 records which are the cities and famous place in Egypt, ID represent the index of location in DB Location name, x and y the coordinate of the location.

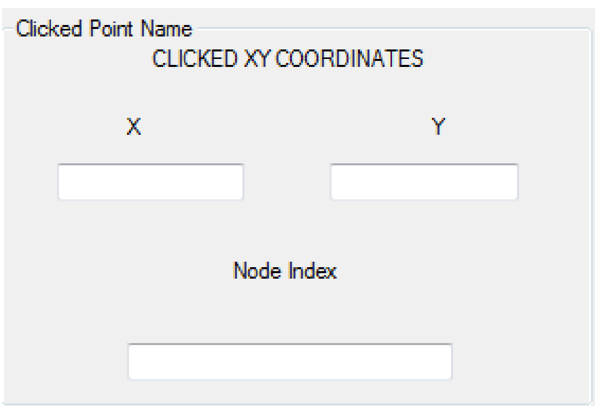


Fig. 9. coordinates of point form

Fig 8 gives the coordinate of the point when clicked on the map directly and also the index number of the location.

V. RESULTS

First, we select the name of city from location name and put it in the location in the first label and press "search" which bring the name of the location and display it on the map Fig 9 show the choosing process of "Kafr_Elshik" city from a database.

The name of a city under the search "Kafr_Elshik" Fig 11 and Fig 12 shown the result of choosing process.

Fig 13 shows the time response for the different type of database structures.

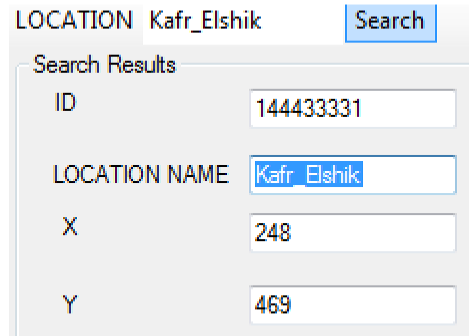


Fig. 10. information of "Kafr_Elshik" point

Fig. 11. The location on the map

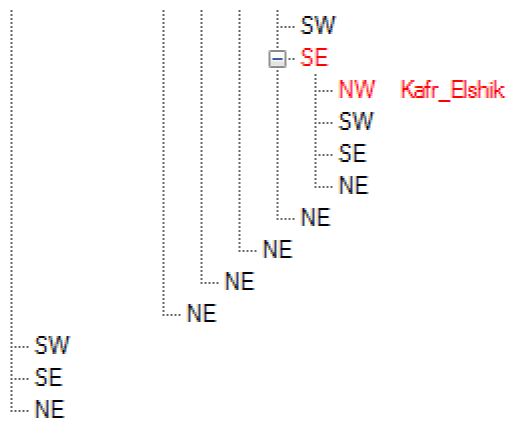


Fig. 12. The location in the tree

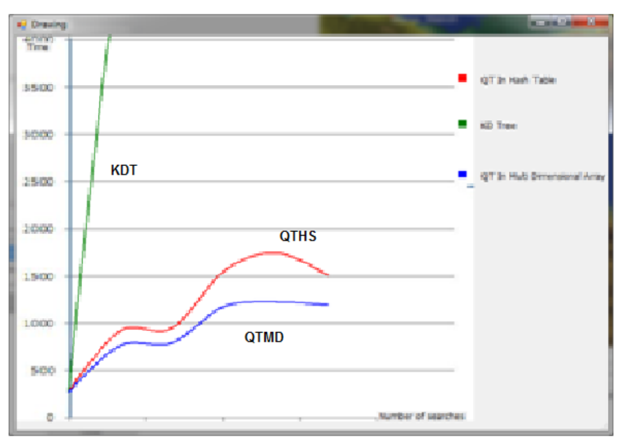


Fig. 13. time response for data structures

TABLE I. TIME RESPONSE

X	1	2	3	4	5
QT HT (mic.s)	979	1232	982	1334	1111
K-DT (mic.s)	999	4484	4481	880	4566
QTMD (mic.s)	724	897	724	976	1085

The curves illustrate the numbers of search (X-axis) with the time in (microsecond). The curves represent that the biggest time the system taken is in the data structure of K-d tree(K-DT)which green one and next it the Quad tree(QT HT) with data structure in hash table represented by red one and the smallest one is the Quad tree in multidimensional array(QTMD) in blue.

VI. CONCLUSION

Our goal is studying the performance of different kind of DB structure for GIS which storage in Cloud computing. The data type of GIS is huge and need to store in the data structure with a method to provide the performance goal:-

- Min. Mass storage.
- Min. Time search.

So, during store data map we use two data structures:-

- K-d tree to store points.
- Quad tree to store regions.

We find that quad tree is more useful and can store large regions with small data as shown in Fig 14.

Fig. 14. the map in quad tree division

The gauge of any query is the query time. We use 3 data structures to search in database Hash table (has a constant time "hash function") for regions, K-d tree (proportional to the length of the tree) query for XY points, and Multidimensional array (has constant time) to search for regions. As we say before we will treat with regions so we compare between hash table and multidimensional array. We find a multi dimensional array is the fastest one. The disadvantage of multidimensional array large memory size, but this memory is local on the computer not in the server. Maybe not as effective if the amount of divisions increase. The result illustrates that the data type of GIS should store in a "quad tree in multidimensional array" which give better performance than the two other types, K-d tree and Quad tree in the hash function.

REFERENCES

[1] Muzafar Ahmad Bhat,RazeefMohd Shah , Bashir Ahmad ."Cloud Computing: A solution to Geographical Information Systems (GIS)"ISSN: 0975-3397Vol. 3 No. 2 Feb 2011

[2] Lakshmi Balasubramanian, M. Sugumaran"A State-of-Art in R-Tree Variants for Spatial Indexing", International Journal of Computer Applications;Mar2012, Vol. 42, p35

[3] YannisTheodoridis, TimosSellis" Optimization Issues in R-tree Construction",http://link.springer.com/book/10.1007/3-540-58795-0http://link.springer.com/bookseries/558Volume 884, pp 270-273, 1994

[4] Jun Wang, Sanjiv Kumar "Semi-Supervised Hashing for Large Scale Search", Ieee Transactions On Pattern Analysis and Machine Intelligence Vol, 34, No 12, December 2012

[5] HaoXu , JingdongWang, Zhu Li Gang Zeng, Shipeng Li Nenghai Yu "Complementary Hashing for Approximate Nearest Neighbor Search",ICCV '11 Proceedings of the 2011 International Conference on Computer Vision Pages 1631-1638

[6] TimosSellis, Nick Roussopoulos, Christos Faloutsos"The R Tree: A Dynamic Index for Multi- Dimensional Objects", In The VLDB Journal (1987)

[7] Yuixnwang, Pengli,Peng Zhang,Jasf Con Cong " Memory partitioning for Multidimensional Arrays in High-level Synpaper". May 29 - June 07 2013, Austin, TX, USA. Copyright 2013 ACM

[8] P.S.V.S. Sridhar, T. N. Jowhar, AnantBhaskarGarg, U. Kedareswarudu"A Framework of Information Technology for Water Resources Management" International Journal of Computer Applications (0975 – 8887) Vol. 30– No.5, September. 2011

[9] Logamani. K, Punitha. S. C "Density Based Clustering using Enhanced K-d tree" Logamani. K, et al | International Journal of Computer Science, Engineering and Technology (IJCSET) | November 2014 | Vol 4, Issue 11,314-318

[10] TilmannZäschke, ChristophZimmerli, Moira C. Norrie "The PH-Tree – A Space-Efficient Storage Structure and Multi-Dimensional Index" SIGMOD'14, June 22–27, 2014, Snowbird, UT, USA. Copyright is held by the owner/author (s). Publication rights licensed to ACM. ACM 978-1-4503-2376-5/14/06

[11] "Geographic Information Systems as an Integrating Technology: Context, Concepts, and Definitions". Kenneth E. Foote and Margaret Lynch, The Geographer's Craft Project, Department of Geography, The University of Colorado at Boulder. Retrieved 21 Apr 2015.

[12] "GIS Applications in Public Health as a Decision Making Support System and It's Limitation in Iran" M.S. Mesgari, K. N. Toosi World Applied Sciences Journal 3 (Supple 1): 73-77, 2008

[13] "A Framework of Information Technology for Water Resources Management" P.S.V.S. Sridhar, T. N. Jowhar, Anant Bhaskar Garg, U. Kedareswarudu, International Journal of Computer Applications (0975 – 8887) Volume 30– No.5, September 2011

[14] "Application of GIS Technology in Watershed-based management and Decision Making" U. Sunday Tim, Sumant Mallavaram, AWRA Hydrology & Watershed Management Technical Committee Vol. 1, No. 5, July - August 2003

[15] "GIS as a Tool for Decision Making and Evaluation" Delforce, Wayne and Rich, Bruce and Savage, Bernadette and Adkins, Barbara and Hall, Carly and Caldeweyher, Daniel (2005) GIS as a Tool for Decision Making and Evaluation. In Bailey, C. and Barnett, K., Eds. Proceedings Social Change in the 21st Century, Brisbane, Australia.

[16] "A framework for ranking of cloud computing services" Saurabh Kumar Garg, Steve Versteeg, bRajkumar Buyyaa Cloud Computing and Distributed Systems Laboratory, Department of Computing and Information Systems, University of Melbourne, Australia, June 2012

[17] " Image Rotation Using Quad Tree" Journal of Engineering, Computers & Applied Sciences (JEC&AS) Volume 1, No.1, October 2012

[18] "Using a distributed quadtree index in peer-to-peer networks" Egemen Tanin • Aaron Harwood • Hanan Samet, Egemen Tanin • Aaron Harwood • Hanan SametPublished online: 4 April 2006 c Springer-Verlag 2005

[19] "AS B-tree: A Study of an Efficient B+-tree for SSDs", Hongchan Roh, Sungho Kim, Daewook Lee and Sanghyun Park, journal of information science and engineering 30, 85-106 ,2014

[20] "A Tool based Edge Server Selection Technique using Spatial Data Structure" Global Journal of Computer Science and Technology, Vol. 14 Issue 3 Ver. 1.0, 2014

[21] "Scalable Nearest Neighbor Algorithms for High Dimensional Data" Marius Muja, David G. Lowe, ieee transactions on pattern analysis and machine intelligence, vol. 36, no. 11, 2014

[22] "filtering of point clouds from photogrammetric surface reconstruction" K. Wenzel, M. Rothermel, D. Fritsch, N. Hale, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XL-5, 2014 ISPRS Technical Commission V Symposium, 23 – 25 June 2014, Riva del Garda, Italy

[23] https://books.google.com.eg/books?id=k6Yfuxaqp2QC&printsec=frontcover

A Survey of Quality Prediction of Product Reviews

H.Almagrabi
School of Computer Science
University of Manchester
Manchester, UK

A. Malibari
Dept. of Computer Science
King Abdulaziz University
Jeddah, Saudi Arabia

J. McNaught
National Centre for Text Mining
University of Manchester,
Manchester, UK

***Abstract*—With the help of Web-2.0, the Internet offers a vast amount of reviews on many topics and in different domains. This has led to an explosive growth of product reviews and customer feedback, which presents the problem of how to handle the abundant volume of data. It is an expensive and time-consuming task to analyze this huge content of opinions. Therefore, the need for automated sentiment analysis systems is vital. However, these systems encounter many challenges; assessing the content quality of the posted opinions is an important area of study that is related to sentiment analysis. Currently, review helpfulness is assessed manually; however the task of automatically assessing it has gained more attention in recent years. This paper provides a survey of approaches to the challenge of identifying the content quality of product reviews.**

Keywords—sentiment analysis; product reviews; content analysis; helpfulness detection

I. INTRODUCTION

The Internet has made it possible to discover opinions of others on a wide range of subjects, through social media websites, such as review sites and wikis, and through online social networks. According to a survey, 81% of Internet users have done research on a product at least once [1]. Studies have found that customers' reviews can form others' opinions and subsequently affect sales [2, 3, 4].

Understanding and analyzing public opinion is important for the prediction of future events. Consequently, this aids the process of making a decision that can involve improving services, handling political elections and calculating risk management. Organizations conduct consumer surveys to explore opinions about their products and/or services. However, the design and the supervision of these surveys are expensive, tedious and a time-consuming task [5]. It is easier for companies to utilize the freely available online consumer reviews. However, the explosive growth of opinion text on the Web makes it hard to manage. In addition, opinions posted on the Web in free-text style are less structured than those conducted from consumer surveys and focus groups, and require more effort to collect and analyze [6, 7, 8].

Sentiment analysis, also referred to as opinion mining, is a growing field in text mining technology, which is concerned with the analysis of people's opinions, attitudes, evaluations and emotions, expressed in free-text fashion towards different objects, such as organizations, product attributes, public events and even individuals [9]. Sentiment analysis is motivated by the fact that individuals and organizations are increasingly using the content of social media for decision-making. Since 2000, there has been much attention paid to sentiment analysis research, mainly because of the rise of machine learning (ML) technology in natural language processing (NLP). In addition, the public datasets available for training using ML algorithms have aided sentiment analysis research. There have been major breakthroughs and promising results in research into sentiment analysis, especially in opinion summarization, feature extraction and polarity identification. However, the extensive amount of uncontrolled user-generated reviews on the Internet has raised concerns about their quality and reliability.

There are many challenges in sentiment analysis research such as dealing with sarcasm and implicit opinions, domain dependency, subjectivity detection and entity identification. A significant challenge that has been studied is determining the quality, also called utility, helpfulness or usefulness, of product reviews [10, 11, 12, 13]. This survey will point out some of the significant research to handle the quality prediction problem.

The rest of this paper is organized as follows. In section 2 we have introduced the challenge of 'product reviews quality. Sections 3 and 4 give an overview of the methods used to predict the utility of product reviews. Section 5 discuses some of the research related to the quality prediction problem, while section 6 concludes the paper.

II. QUALITY OF REVIEWS

The topic of quality of reviews is related to opinion spam detection, which makes it an important area of research. However, according to Liu [9], it is different from spam detection, as spam reviews may not be of low quality. Fake reviews may be of high quality, especially if they are well written, which makes them hard to identify. Determining good quality reviews saves readers time and effort by discarding noisy and low quality content. It is useful to have a mechanism to automatically assess a review's helpfulness as soon as it is written.

Some aggregation and hosting websites rank reviews according to their perceived helpfulness by readers, such as Amazon.com, Epinions, IMDB CitySearch, etc. Users manually assess reviews by responding to a question, such as "Was this review helpful to you?". Readers can respond with "YES or NO" and the feedback results are calculated and displayed next to each review (e.g., "12 of 20 people found the following review helpful"). Although this helpfulness evaluation method has been used by many websites, it is still a meaningful task to automatically determine the quality of each review for the following reasons:

1) Many reviews have little or no helpfulness evaluation, especially in low-traffic items. According to [11], some *reviews lack a helpfulness evaluation: 38% of 20,000 Amazon MP3-player reviews only received three or less votes in three months [11]. In addition, consumers are not obligated to respond to the feedback question to determine a review's perceived helpfulness, even if they found it helpful or not.*

2) Human generated helpfulness evaluation may be fake, which makes the helpfulness voting score untrustworthy [1]. Spammers can click on the helpfulness voting buttons (Yes or No) to increase or decrease the helpfulness of a review. Therefore, depending on the helpfulness feedback to identify helpful reviews can be problematic.

3) Biases can be found in the manual helpfulness evaluations [10]. Reviews with high helpfulness score are prominently displayed, which would have a disproportionate influence on readers and consequently on the helpfulness voting score itself. This type of bias is referred to as "winner circle" bias in [10]. In addition, an in-depth analysis of Amazon's highly-voted reviews, lead to discover that some of the reviews are not of as good quality as the helpfulness voting score indicates. Readers tend to value others' reviews positively, which makes the distribution of helpfulness evaluation skewed towards the helpful vote, known as the "imbalance vote bias". The third type of bias identified is called "early bird bias" [10]. The helpfulness voting score may take a long time to accumulate, particularly in newly posted reviews. Earlier posted reviews are displayed to readers for a longer time than newly posted reviews.

4) The use of robust review quality prediction systems will facilitate ranking reviews according to their utility, and thus users can easily and quickly access them. Furthermore, applications such as sentiment extraction and opinion summarization will benefit from such systems by operating on high quality content rather than spammed and misleading reviews. For example, in the process of opinion summarization it is useful to only use good quality reviews and discard low quality ones, including reviews with high helpfulness voting score, which are subject to the previously mentioned biases. Therefore, automatically classifying reviews according to their quality would aid and speed up the quality of opinion summarization [10].

III. Quality as Regression Problem

Generally, the problem of determining the quality of reviews is seen as a regression problem. The method uses machine learning models to assign a score to each review. These scores can be used in recommendation and ranking systems [9]. Researchers have used different types of features to train and test models on datasets from different domains. A Support Vector Machine (SVM) regression model to rank reviews according to their helpfulness was employed by [11]. They used structural features (e.g., review length, number of sentences), lexical features (e.g., unigrams and bigrams), syntactic features (e.g., nouns, verbs, etc.), semantic features (e.g. sentiment words) and meta-data features (e.g., number of stars) [9]. The most useful features used were the length of the review, its unigram and its product rating.

Zhang and Varadarajan [13] use a similar feature set to that proposed in [11]. However, they did not include any meta-data information. Their study assumes that a good quality review should discuss many aspects of the product. Thus, a comparison between the review and the product specifications was considered. However, the results show that this feature did not improve system performance [1]. Furthermore, the study includes review similarity to editorial reviews, which did not improve the system performance either. They found that the perceived helpfulness in product reviews depends greatly on its linguistic style (e.g., word count, comparatives and superlatives, proper nouns, etc). In contrast with the results of [11], there is a weak correlation between review length and utility score [1, 13]. According to Pang and Lee [1], the difference in domain choice affected the results in the two studies. Electronic product reviews used in [11] do not include as sophisticated language as found in book and movie reviews, which were used in [13].

A different approach was proposed by Ghose and Ipeirotis [14], who studied the relationship between the subjectivity of a review and its helpfulness. A classifier determines the subjectivity of a sentence, and then the standard deviation of the subjectivity score of the sentences in a given review is computed. The results indicate that the standard deviation score and a readability score have a strong effect on utility evaluation. Building on their previous research, Ghose and Ipeirotis [15] expanded their work by examining multiple product categories and by adding textual features, such as history information about the author, readability metrics and spelling errors, etc. They found that reviews including a mixture of subjective and objective information influence sales and the perceived usefulness. In addition, readability and informativeness features were found to correlate positively with sales and the perceived usefulness. An important finding of this research is that the type of product affects the perceived helpfulness of a review. For feature-based products (e.g., electronics), reviews that include objective more than subjective information increase the usefulness of the review. However, in experience products (e.g., movies), it was found that subjectivity matters the most, as users prefer to read personalized and highly sentimental comments that describe the reviewer experience and provide more information about the product.

Looking at the problem from a different perspective than the above approaches, the work in [16] introduced three main factors affecting the helpfulness of a review: reviewer's expertise, review timeline and review style based on part-of-speech tags. A nonlinear regression model was used to integrate the proposed factors. Extensive experiments on movie reviews (IMDB data-set) show the efficiency of the proposed model. They argue that their model is general enough to be employed in other domains, by replacing the genres of movies with the categories of products and by modeling the timelines and the writing style using their proposed algorithm.

Previous research efforts focus only on the meta-data and on the review text itself to analyze various properties of product reviews in order to predict quality. Other studies tried to tackle the quality problem from different perspectives.

For example, the study by [17] incorporates another information resource: the author's behaviour on e-commerce sites, such that information derived from their online transactions helped to assess the quality of reviews and to identify spammed ones. Three features were used to assess the quality of reviews: personal reputation, seller degree and expertise degree. The correlation between each feature and the helpfulness votes was examined using a linear regression analysis.

Lu, Tsaparas, Ntoulas, and Polanyi [18], investigate if the social context of reviews can enhance the performance of quality prediction. In their view, important information can be obtained from the social context about the quality of reviewers, which affects the quality of their reviews. In order to incorporate social context in predicting review quality, regularization constraints, based on a set of experimentally-validated hypotheses, were employed. An example is the “author constraint hypothesis”, which assumes that reviews from the same author are similar in quality [18]. The results show accuracy improvement in predicting the quality of reviews using a text-based classifier (linear regression model). They argue that the proposed regularization technique can be applicable and generalized for quality evaluation of other user generated content. However, this method cannot be employed to review sites that do not have a trusted social network [9].

Previous studies have proven that online reviews affect the sales of products, however many studies fail to consider the quality of reviews. Another group of interested researchers [19] proposed a regression model that incorporates the quality factor for predicting sales performance of products being reviewed. Results indicate the positive correlation between review quality and prediction accuracy of sales performance.

IV. Classification and Other Methods

In addition to ranking reviews according to their quality and utility, researchers also used classification methods to determine the quality of reviews. In most previous studies, helpfulness votes were used as the ground-truth data for training and testing regression models [11, 14]. However, in a different approach, these approaches were unreliable because of the previously mentioned types of biases discovered from their extensive analysis [10]. Thus, they did not use user-helpfulness feedback (helpfulness votes) as the ground-truth in training and testing their model. Their work focused on improving the quality of opinion summarization by detecting and discarding noisy and low quality reviews using a classification based approach. The proposed approach explores three features of product reviews: readability, informativeness and subjectivity. A set of specifications was proposed for judging the quality of reviews, and four categories defined: “best reviews”, “good reviews”, “fair reviews” and “bad reviews”. A SVM was used to perform binary classification, with the “bad review” category as the low quality class and the remaining categories as the high quality class. After the classification step, only high quality reviews were used in generating opinion summarization.

O'Mahony and Smyth [20] proposed a classification-based recommender system to recommend the most helpful reviews to the end user. Many features were used to train a classifier to distinguish between helpful and non-helpful reviews: reputation, content, social and sentiment features. The reputation and sentiment features achieved better classification performance than content and social features. A significant finding was that the classification performance remained high, even in the absence of the reputation features, which are not always available.

Chien and Tseng [21] treated the quality problem as a classification problem by employing a multiclass SVM model to classify product reviews. In order to derive informative review features, an information quality "IQ" framework adopted from [22] was employed. Table 1 illustrates the information quality categories along with their dimensions. The authors defined five classes of quality: high, medium, low, duplicated and spam, and the specifications of review quality were adopted from [10]. Furthermore, factors that shape high quality reviews were analyzed and the findings show that helpful reviews need to be subjective and provide detailed comments on a number of product aspects.

TABLE I. Wang and Strong's IQ Framework [22]

IQ Category	IQ Dimensions
Intrinsic IQ	Believability, accuracy, objectivity, reputation
Contextual IQ	Value-added, relevancy, timeliness, completeness, appropriate amount of information
Representational IQ	Interpretability, ease of understanding, representational consistency, concise representation
Accessibility IQ	Accessibility, access security

In a recent study, Bayesian inference was used to measure the probabilities of the reviews belonging to certain classes [72]. In addition, an extended fuzzy associative classifier was developed to train a review helpfulness classification model. The model incorporated features from previous studies [10, 21, 32, 52, 54, 69], for example, subjectivity features, emotion features and stylistic features.

In a different approach from the previous supervised methods, Tsur and Rappoport [23] introduced a fully unsupervised method to rank book reviews according to review helpfulness. First, the proposed REVRANK algorithm identifies the dominant terms in a set of review documents. These important terms represent a “virtual optimal” or a core review representation. Subsequently, reviews are mapped to this optimal representation and a ranking score is given to each review according to distance between the review and the virtual review. All reviews of a given book were explored to generate a lexicon of the dominant concepts.

This is relevant to keyphrase extraction proposed in the TextRank and the CollabRank systems [24, 25]. Both systems employ a graph-based unsupervised ranking algorithm which ranks keyphrases, using the co-occurrence links between words in the TextRank system and the collaborative knowledge from multiple documents in the CollabRank system.

In recent work, a new problem of personalized review quality prediction was addressed to recommend helpful reviews [26]. The authors argue that the quality of reviews may not be the same for different readers, while all the previous studies assume that it is. They found that there are some latent features that affect the user's evaluation of the quality of the review. Based on this assumption, a series of probabilistic graphical models, based on matrix factorization and tensor factorization, were proposed. The experiment was conducted on a real-life dataset from Eopinion.com, and the results show that the proposed technique outperformed the existing state-of-the-art approaches, at that time, using textual and social features.

All the above studies did not consider that highly ranked reviews may include highly redundant information [9]. Another method was proposed to solve this problem by selecting a small comprehensive set of high quality reviews [27]. These reviews cover many different aspects and viewpoints of the reviewed product. The authors of [27] extended existing algorithms for maximizing coverage to handle this problem. Their work is different in that they selected a set of comprehensive reviews rather than scoring each review. Furthermore, the proposed approach is different from opinion summarization because it aims to identify a subset of reviews that cover the different aspects of a product rather than summarizing the opinions on the extracted features of a product. The most related work to [27] is the work of Lappas and Gunopulos [28]. However, whereas the goal of the former authors is to cover the product aspect from a fixed size set with both negative and positive opinions, that of the latter is to cover all product aspects while preserving opinion distribution.

Miao, Li, and Dai [29], introduced a sentiment mining and retrieval system which is concerned with mining useful information from customers' reviews. They employed both data mining and information retrieval techniques to build a novel temporal opinion quality and relevance ranking system, which mines customers' preferences.

Wu, Greene, and Cunningham [30] compared two aggregation methods for combining sets of features in order to identify untruthful opinions about hotels. Their solution was to build a useful suspicious-review ranking system. The results show that the best features to identify suspicious reviews are: proportion of positive singleton reviews, truncated rating, and reactive positive singleton reviews. Furthermore, it was found that singular value decomposition outperforms the unsupervised hedge algorithm for combining features to identify suspicious reviews about hotels. Although this work falls under spam detection, it is related to identifying qualitative reviews. Determining criteria for identifying suspicious reviews would improve the identification of reliable and trustworthy reviews.

Lau, Zhang, Xia, and Song [31] proposed a method to detect non-informative online opinionated expressions. The proposed multi-facet quality metric utilizes both the intrinsic properties of opinionated expressions and association with other opinionated expressions posted on the Internet. Furthermore, to avoid the biases mentioned in [10], helpfulness votes were not used as the ground-truth for quality assessment.

A novel approach to assess the quality of product reviews was proposed by Min and Park [32]. The proposed metric employs linguistic clues to capture time expressions related to the use of the product and product aspects during different purchase times. They found that tense and time expressions are the most useful linguistic clues to assess the customer's previous purchase experiences. This approach is similar to the work done by the group in [18], however there is a difference, because this work uses the reviewer's social context information based on the social network-based review website, such as the PageRank score of the author. Features used in this work improve system performance, however, they have limitations similar to the “helpfulness votes”. In contrast, the method proposed by Min and Park extracts the reviewer's characteristics directly from the textual content of a review by utilizing his/her experience.

The authors in [33] proposed a method to evaluate the helpfulness of online reviews based on the domain user's perspective, such as manufacturing engineers and product designers. They conducted an exploratory study to understand what makes reviews useful to designers. Four categories of features were proposed, based on their experiment, to identify helpful reviews: product features, linguistic features, features using information theory and features based on information quality. Machine learning algorithms were employed using both classification and regression to evaluate the proposed method. The results show a strong correlation between the designer’s rating and the proposed method.

Another classification approach to modeling the helpfulness problem was proposed by Zeng and Wu [34]. A three-class classification framework was introduced to find the helpful positive reviews, the helpful negative reviews and to filter out the unhelpful reviews. Table 2 lists the features used in the classification approach. Some of the features were adopted from [11] and other features were added based on the findings of [35]. The study uses the list of common ideas related to helpfulness and unhelpfulness proposed in [35]. The performance of the three-class problem is quite high and the results show that helpful reviews (positive and negative) can be identified with high precision from unhelpful ones [35].

Recently, the work of Krishnamoorthy [69] has developed a new method for extracting linguistic features based on the linguistic category model (LCM), proposed by Semin and Fiedler [70]. A binary classifier was built and evaluated using the LCM with a combination of other features, namely: metadata features (e.g., review extremity), readability features, and subjectivity features (i.e., the total number of subjective words normalized by review length). The experiment on two real-life review datasets shows that linguistic category features are better predictors for the helpfulness of product reviews. Table 3 presents the linguistic category features used in the LCM.

TABLE II. CLASSIFICATION FEATURES OF ZENG AND WU [34]

Features	Description
Unigram (Product Description)	The number of unigrams used between the review and the corresponding product description
Bigram (Product Description)	The number of bigram used between the review and the corresponding product description
Trigram (Product Description)	The number of trigrams used between the review and the corresponding product description
Length	The length of a review
Comparisons	The review uses the string "compare to" or "ADJ + er than"
Degree of detail	Defined by formula
Use of Ratings	The "Star" ratings of the review
Pros and Cons	The review contains exactly the strings "Pros" and "Cons"

TABLE III. LINGUISTIC CATEGORIES AND THEIR DESCRIPTION [69]

Category	Description
Adjectives (ADJ)	Qualifies a noun; highly subjective and abstract
SV (State verbs)	Refers to mental or emotional state
SAV (State Action Verbs)	Describes the emotional consequences of an action; high positive or negative connotation
IAV (Interpretive Action Verbs	Multitude of actions that have the same meaning; have a positive or negative connotation
DAV (Descriptive Action Verbs)	Objective description of a specific action; no positive/negative connotation

V. RELATED TASKS

Researchers from many fields have shown great enthusiasm in studying and analyzing the quality of online reviews [10, 11, 13, 14, 16, 18, 36]. Pang and Lee [37] carried out the first study related to this problem. They studied the prediction of product rating, which correlates with the perceived helpfulness of reviews [11]. Automatically scoring essays is another related study which has been used to rate the quality of reviews [38]. The work group of [11] built a regression model to rank reviews according to their quality, employing the same set of features used in essay scoring by [38].

By using statistical methods, Jindal and Liu [39] discovered that the extracted features from reviews, such as rating and title length, improved the identification of spammed and duplicated reviews. The task of evaluating the quality of Web posts and the quality of answers in question and answering systems is also related to predicting the helpfulness of reviews [40, 41, 42, 43, 44, 45]. A study by Hoe, Li, and Zou [46] was to determine whether a review would achieve helpfulness votes, or not, through examining the posting time and textual features of a review. The work in [44] integrates user and community features with the review textual features to assess the quality of questions and answers using classification methods. The model utilizes community features which were driven from the answer text, for example the length of the answer and number of points received, in addition to the user's features, such as number of answers given. The authors of [45] introduce a co-training method to model the quality of both the review and the reviewer in an attempt to use community information for extracting features.

The identification of high quality reviews influences feedback and reputation systems. It was found that seller and buyer behaviour changes in response to the change in a seller's feedback profile [47]. In another similar study [48], it was found that personal reputation positively influences buyers to purchase a product, thus a small amount of negative feedback will not affect sales. However, most research on the quality of reviews analyzes the personal reputation impact on the review itself, rather than on the seller's/buyer's transactions [17]. In addition, the work in [49] has also examined the impact of high quality reviews on purchase decisions and thus on sales. The results show that reviews with a high proportion of helpful votes are more important in making purchase decisions than the aggregate star rating of a product. Another important finding is that the reviewer's reputation has no impact on the consumer purchase decision, which contradicts the findings of [17].

The work in [19] proposed a regression model for predicting sales performance using product reviews. This was one of the first attempts to examine the economic impacts of review quality. The authors of [50] studied the problem of product review search. The retrieved reviews are ranked according to their quality based on a given query. Other studies tackle the problem of finding good answers from question-answering systems such as Yahoo! Answers[1]. A method of determining the quality of the answer was proposed by [51], which considered both textual features from the answers and social features from the answerer.

Some studies are interested in determining the fundamental characteristics of helpful reviews from a theoretical and practical point of view [52, 53, 54, 55, 71]. For example, Mudambi and Schuff examined factors contributing to review helpfulness, by employing a linear regression model. The study shows that product type, review extremity and review depth affect review helpfulness [52]. They analyzed the different impact of product type (search or experience goods) on review helpfulness in the multistage consumer decision process. Search goods are defined as products about which consumers can easily obtain information about their quality before purchasing. Experience goods are products that require purchasing in order to evaluate their quality, for example, books and movies. Thus, the type of product affects a consumer's textual evaluation of a product and consequently affects the perceived helpfulness of that review. The study in [54] examines what factors affect the number of helpfulness votes reviews receive. It is argued that knowing the most important factors to attract helpfulness votes will help website designers to improve their helpfulness voting mechanism. For example, the findings show that semantic features have the most impact on the number of helpfulness votes. The conclusion is that websites should provide more ranking options to rank reviews rather than ranking the most recent review. Siering and Muntermann [55] confirm that product type influences the perceived helpfulness as was discovered in [52]. Furthermore, they find that perceived helpfulness is affected by other textual aspects related to review sentiment, product quality and review uncertainty. Lee and Choeh [71]

[1] http://answer.yahoo.com/

employed a multilayer perceptron neural network model to improve helpfulness prediction accuracy. The product review metadata and the textual characteristics of both the review and the reviewer were used as features in the proposed model. The study found that the list price and the sales rank of the product are important for helpfulness prediction. In addition, the results show that the neural network model outperforms the conventional regression models in helpfulness prediction.

Other studies have discovered other factors that impact review helpfulness, such as specific emotions [56, 57] and review readability [15, 53]. From a theoretical perspective, the authors of [53] proposed a model to investigate the relationship between the textual content of a review and its perceived helpfulness votes. Four readability measures were employed to validate the proposed model. Figure 1 demonstrates the four readability measures. The model was based on three specific aspects: conformity, understandability and extensiveness. The results show that helpfulness is affected by review readability more than its length and that reviews with extreme votes receive a higher score than the less helpful ones. Some studies assume that focusing on product quality in online reviews will provide diagnostic value and thus impact the review helpfulness [58, 59].

Racherla and Friske [60] explore which factors contribute the most to consumer perception of the utility of online reviews. Furthermore, they investigate if the impact of these factors varies according to the type of service offered to the consumer. An important finding of this research is that reviews provided by reviewers with high expertise and good reputation are significantly helpful reviews. In addition, reviews with a great amount of information are not particularly considered more helpful than reviews containing less information.

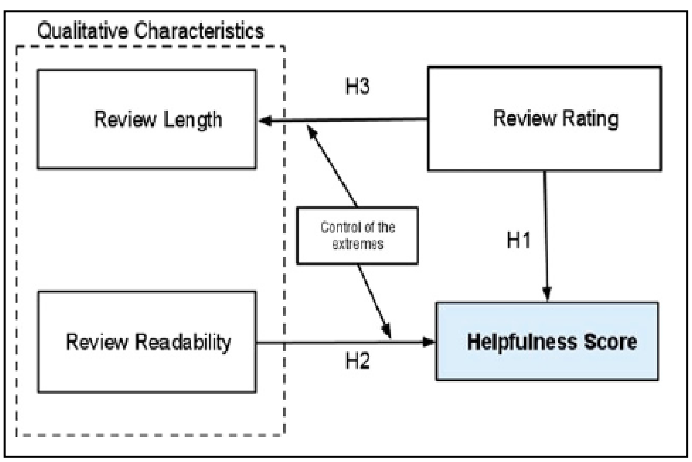


Fig. 1. Theoretical model [53]

The work of Otterbacher [61] has focused on understanding helpfulness rather than just predicting it. A well-established framework for quality assessment by Wang and Strong [22] is adopted. The framework includes four categories of data quality developed from the end user's perspective: Intrinsic quality, Contextual quality, Representational quality and Accessibility. Each category contains several dimensions. Only the first three categories are used because accessibility is not an issue as reviews are accessible in the virtual community environment [22]. A simple linear regression analysis shows that the helpfulness score correlates with several data quality dimensions.

In another study to identify the specification of a helpful review, Connors, Mudambi, and Schuff [35] conducted an open-ended analysis. A group of 40 participants were asked to rate reviews and describe what factors they found in helpful and unhelpful reviews. Table 4 shows the results of the analysis. For example, the “product usage information” was used 30 times by all participants. Categorizing these textual descriptions highlighted the influence of three criteria for helpfulness: the credibility of the reviewer, the similarity between the author and the reader, and the use of positive and negative comments about the product.

In the next stage of their study, a controlled experiment was conducted to study the effect of the three criteria on the helpfulness rating. The findings show that both the characteristics of the reviewer and of the review influence the reader’s understanding of review helpfulness. For example, the study recommends that reviewers should disclose their self-described expertise as part of their review. Furthermore, review sites must favour content over balance, a review expressing both positive and negative statements towards a product is not perceived as more helpful as a one-sided review. This idea was also proposed by Schlosser [62], who argues that reviews including two-sided arguments (pros and cons), are not necessarily more helpful, credible and persuasive than a one-sided argument review. The study proves that reviews with two-(versus one-) sided arguments receive higher helpfulness votes only if the reviewer rating is fairly favourable, while reviews written by extremely favourable reviewers are considered more helpful even if they include only one-sided argument.

While preceding studies have examined review characteristics, Ngo-Ye and Sinha [73] incorporate review related engagement features in their proposed hybrid text-regression models to predict review helpfulness. Furthermore, this study uses a bag-of-words representation as part of the textual features. The proposed hybrid model including the textual features and the reviewer engagement characteristics enhanced helpfulness prediction. This work offers new factors that contribute to the prediction of helpfulness. In another attempt to examine the factors affecting the helpfulness of reviews, a recent study by Liu and Park [74] suggested that the combination of the messenger and message features correlates positively with the helpfulness prediction of reviews.

Features such as the reviewer identity and the length of the review were used to build a textual regression model to predict helpfulness. Valence (positive or negative) consistency, a new aspect to the problem of utility prediction, was recently investigated in [75]. That study examined the influence of other nearby reviews on the perceived helpfulness of the review itself. The results show that, whether or not the reviews are being positive or negative, consistent reviews are more helpful than inconsistent ones, which was not the case in prior studies [76, 77]. For example, Scholz and Dorner [78] found that positive reviews achieve better helpfulness scores than negative ones.

TABLE IV. IDEAS RELATED TO HELPFULNESS AND UNHELPFULNESS [35]

Helpfulness	Times Mentioned
Pros and Cons	36
Product Usage	30
Information Detail	24
Good Writing Style	13
Background Knowledge of product	12
Personal Information About Reviewer	12
Comparisons	10
Lay-Man's Terms	9
Conciseness	8
Lengthy	7
Use of Ratings	7
Authenticity	5
Honesty	5
Miscellaneous	4
Unbiased	4
Accuracy	3
Relevancy	3
Thoroughness	3
Unhelpfulness	**Times Mentioned**
Overly Emotional/Biased	24
Lack of Information	17
Irrelevant Comments	9
Not Enough Detail	6
Poor Writing Style	6
Using Technical Language	6
Low Credibility	5
Problems With Quantitative Rating	5
Too Much Detail	5
Too Short	4

Martin and Pu [79] suggested that emotional words are powerful parameters in helpfulness prediction. They propose a framework to extract the emotionality from the textual content of reviews. GALC, a general lexicon of emotional words, was employed to represent a model of 20 categories using supervised classification methods. The results show that the emotion-based method outperforms the previous structural-based methods by 9%. The work of Mertz, Korfiatis, and Zicari [80] examined the helpfulness prediction problem by evaluating the performance of dependency bigrams and discourse connectives. A binary classifier was introduced using the previously mentioned novel text-based features. This study shows that various types of discourse relations are useful set features for predicting review helpfulness. Moreover, there is a strong correlation between high star ratings and helpful reviews. Another study has investigated how misalignment between the star rating and the textual content of the review can lower the overall helpfulness of the review [81]. It found that misalignment between star rating and review text often occurs in reviews of experience goods and in reviews with high star ratings. This theoretical analysis suggests that highly rated experience goods reviews are perceived more helpful than other reviews.

Recently, a study of factors contributing to online review helpfulness was carried out [82]. Specifically, the goal of the study is to examine the joint effect of the message length of a review (word count) together with reviewer characteristics and the patterns of the review on the utility of the review. While prior studies have suggested that there is a positive correlation between word count and the perceived helpfulness of a review [52, 77], the results of this study point out that the association between word count and helpfulness is valid only in reviews with 144 or less words [82]. The hypothesis results of this study are listed in table 5.

TABLE V. RESULTS OF HYPOTHESIS TESTING [82]

H1a: For reviews written by all reviewers, word count is a significant predictor of review helpfulness when the review is shorter than average	Supported
H1b: For the reviews written by top reviewers, word count is a significant predictor of review helpfulness	Not supported
H2: For top reviewers, reviewer experience is a significant predictor of review helpfulness	Not supported
H3: For top reviewers, reviewer impact is a significant predictor of review helpfulness	Not supported
H4: For top reviewers, reviewer cumulative helpfulness is a significant predictor of review helpfulness	Supported
H5: For the top reviewers, product rating is a significant predictor of review helpfulness	Supported

In an approach proposed recently by Tang, Gao, Hu, and Liu [63], the prediction of review helpfulness for each user was investigated. A context-aware helpfulness prediction framework utilizes both the social and the content context of a review. The review content affects the perceived helpfulness by other users. Furthermore, information about the author, rater and their relationship can provide the social context of reviews. For example, it is more likely that raters find reviews from their connected authors helpful. However, this framework has some limitations that need to be addressed. The social context is likely to change over time, such as user preference.

VI. CONCLUSION

With e-commerce websites growing rapidly, reviews about products are becoming an important source of information for making informed purchase decisions. Sentiment analysis research is concerned with extracting and summarizing the massive content of product reviews. However, the explosive growth of product reviews raises concerns about their reliability and quality. Although review helpfulness is currently assessed manually by many retailer websites, it is important to automatically assess the quality of reviews at least for two reasons. The first reason is to provide helpfulness evaluation when human evaluations are lacking. The second reason is to correct skews in human helpfulness evaluations mentioned in [10].

Regression and classification methods have been examined to rank and classify reviews according to their helpfulness. Major challenges in quality prediction include feature weighting, which affects the performance of classifiers enormously.

Another challenge is to consider the quality of the reviewer. Some trust metrics from other studies may be of use to determine the helpfulness of product reviews, for example, research into peer-to-peer and reputation networks [64, 65, 66, 67, 68].

Investigating what factors determine review helpfulness could improve review systems. Therefore, retailers' websites could introduce automatic helpfulness scoring systems to reduce customers' search cost. This would affect customers' satisfaction and purchase behaviour [49, 78]. Furthermore, there should be a method to increase the efficiency of reviews to support online purchase decisions. For example, introducing a mechanism to allow an easier comparison between reviews would affect the process of making a purchase decision [83].

References

[1] B. Pang and L. J. Lee, Opinion Mining and Sentiment Analysis. Hanover, MA: Now Publishers, 2008.

[2] P. Chatterjee, "Online Reviews: Do Consumers Use Them?," Social Science Research Network Working Paper Series, May/09/ 2006.

[3] P.-Y. Chen, S.-Y. Wu, and J. Yoon, "The impact of online recommendations and consumer feedback on sales," in ICIS 2004 Proceedings, p. 58, 2004.

[4] C. Dellarocas, N. Awad, and X. Zhang, "Exploring the value of online reviews to organizations: Implications for revenue forecasting and planning," in ICIS 2004 Proceedings, p. 30, 2004.

[5] M. Gamon, A. Aue, S. Corston-Oliver, and E. Ringger, "Pulse: mining customer opinions from free text," in Advances Intelligent Data Analysis Madrid, Spain, 2005.

[6] Y. Chen and J. Xie, "Online Consumer Review: Word-of-Mouth as a New Element of Marketing Communication Mix," Manage. Sci., vol. 54, pp. 477-491, 2008.

[7] J. A. Chevalier and D. Mayzlin, "The effect of word of mouth on sales: Online book reviews," Journal of Marketing Research, vol. 43, pp. 345-354, 2006/08// 2006.

[8] Y. Liu, "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," Journal of Marketing, vol. 70, pp. 74-89, 2006.

[9] B. Liu, Sentiment analysis and opinion mining. San Rafael, Calif.: Morgan & Claypool, 2012.

[10] J. Liu, Y. Cao, C.-Y. Lin, Y. Huang, and M. Zhou, "Low-Quality Product Review Detection in Opinion Summarization," in Proceedings of (EMNLP-CoNLL), 2007, pp. 334-342.

[11] S.-M. Kim, P. Pantel, T. Chklovski, and M. Pennacchiotti, "Automatically assessing review helpfulness," in the Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, Sydney, Australia, 2006.

[12] A. Ghose and P. G. Ipeirotis, "Designing novel review ranking systems: predicting the usefulness and impact of reviews," in the Proceedings of the ninth international conference on Electronic commerce, Minneapolis, MN, USA, 2007.

[13] Z. Zhang and B. Varadarajan, "Utility scoring of product reviews," in the Proceedings of the 15th ACM international conference on Information and knowledge management, Arlington, Virginia, USA, 2006.

[14] A. Ghose and P. G. Ipeirotis, "Designing novel review ranking systems: predicting the usefulness and impact of reviews," in the Proceedings of the ninth international conference on Electronic commerce, Minneapolis, MN, USA, 2007.

[15] A. Ghose and P. G. Ipeirotis, "Estimating the helpfulness and economic impact of product reviews: Mining text and reviewer characteristics," Knowledge and Data Engineering, IEEE Transactions on, vol. 23, pp. 1498-1512, 2011.

[16] Y. Liu, X. Huang, A. An, and X. Yu, "Modeling and Predicting the Helpfulness of Online Reviews," in the Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, 2008.

[17] S. Huang, D. Shen, W. Feng, Y. Zhang, and C. Baudin, "Discovering clues for review quality from author's behaviors on e-commerce sites," in the Proceedings of the 11th International Conference on Electronic Commerce, Taipei, Taiwan, 2009.

[18] Y. Lu, P. Tsaparas, A. Ntoulas, and L. Polanyi, "Exploiting social context for review quality prediction," in the Proceedings of the 19th international conference on World Wide Web, Raleigh, North Carolina, USA, 2010.

[19] X. Yu, Y. Liu, X. Huang, and A. An, "A quality-aware model for sales prediction using reviews," in the Proceedings of the 19th international conference on World wide web, Raleigh, North Carolina, USA, 2010.

[20] M. P. O'Mahony and B. Smyth, "Learning to recommend helpful hotel reviews," in the Proceedings of the third ACM conference on Recommender systems, New York, New York, USA, 2009.

[21] C. C. Chen and Y.-D. Tseng, "Quality evaluation of product reviews using an information quality framework," Decision Support Systems, vol. 50, pp. 755-768, 2011.

[22] R. Y. Wang and D. M. Strong, "Beyond accuracy: what data quality means to data consumers," J. Manage. Inf. Syst., vol. 12, pp. 5-33, 1996.

[23] O. Tsur and A. Rappoport, "RevRank: A Fully Unsupervised Algorithm for Selecting the Most Helpful Book Reviews," in Proceedings of the International AAAI Conference on Weblogs and Social Media 2009.

[24] R. Mihalcea and P. Tarau, "TextRank: Bringing Order into Texts," in Empirical Methods in Natural Language Processing Conference 2004, pp. 404-411.

[25] X. Wan and J. Xiao, "CollabRank: towards a collaborative approach to single-document keyphrase extraction," in the Proceedings of the 22nd International Conference on Computational Linguistics - Volume 1, Manchester, United Kingdom, 2008.

[26] S. Moghaddam, M. Jamali, and M. Ester, "ETF: extended tensor factorization model for personalizing prediction of review helpfulness," in the Proceedings of the fifth ACM international conference on Web search and data mining, Seattle, Washington, USA, 2012.

[27] P. Tsaparas, A. Ntoulas, and E. Terzi, "Selecting a comprehensive set of reviews," in the Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining, San Diego, California, USA, 2011.

[28] T. Lappas and D. Gunopulos, "Efficient confident search in large review corpora," in the Proceedings of the 2010 European conference on Machine learning and knowledge discovery in databases: Part II, Barcelona, Spain, 2010.

[29] Q. Miao, Q. Li, and R. Dai, "AMAZING: A sentiment mining and retrieval system," Expert Systems with Applications, vol. 36, pp. 7192-7198, 4// 2009.

[30] G. Wu, D. Greene, and P. Cunningham, "Merging multiple criteria to identify suspicious reviews," in Proceedings of the fourth ACM conference on Recommender systems, 2010, pp. 241-244.

[31] R. Y. Lau, W. Zhang, Y. Xia, and D. Song, "Multi-facets quality assessment of online opinionated expressions," in Web Information Systems Engineering–WISE 2010 Workshops, 2011, pp. 212-225.

[32] H.-J. Min and J. C. Park, "Identifying helpful reviews based on customer's mentions about experiences," Expert Systems with Applications, vol. 39, pp. 11830-11838, 11/1/ 2012.

[33] Y. Liu, J. Jin, P. Ji, J. A. Harding, and R. Y. K. Fung, "Identifying helpful online reviews: A product designer's perspective," Computer-Aided Design, vol. 45, pp. 180-194, 2// 2013.

[34] Y.-C. Zeng and S.-H. Wu, "Modeling the Helpful Opinion Mining of Online Consumer Reviews as a Classification Problem," in Sixth International Joint Conference on Natural Language Processing, 2014, p. 29.

[35] L. Connors, S. M. Mudambi, and D. Schuff, "Is It the Review or the Reviewer? a Multi-Method Approach to Determine the Antecedents of Online Review Helpfulness," in System Sciences (HICSS), 2011 44th Hawaii International Conference on, 2011, pp. 1-10.

[36] C. Danescu-Niculescu-Mizil, G. Kossinets, J. Kleinberg, and L. Lee, "How opinions are received by online communities: a case study on amazon.com helpfulness votes," in the Proceedings of the 18th international conference on World wide web, Madrid, Spain, 2009.

[37] B. Pang and L. Lee, “Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales,” in the Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, Ann Arbor, Michigan, 2005.

[38] Y. Attali and J. Burstein, “Automated essay scoring with e-rater v.2,” Journal of Technology, Learning, and Assessment, vol. 4, 2006/02// 2006.

[39] J. Nitin and B. Liu, “Product Review Analysis,” ed: Technical Report, UIC, 2007.

[40] M. Weimer, I. Gurevych, and M. Mühlhäuser, “Automatically assessing the post quality in online discussions on software,” in Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions, 2007, pp. 125-128..

[41] M. Weimer and I. Gurevych, “Predicting the Perceived Quality of Web Forum Posts,” in the Proceeding of Recent Advances in Natural Language Processing Conference (RANLP), 2007.

[42] J. Jeon, W. B. Croft, J. H. Lee, and S. Park, “A framework to predict the quality of answers with non-textual features,” in the Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, USA, 2006.

[43] L. Hoang, J.-T. Lee, Y.-I. Song, and H.-C. Rim, “A model for evaluating the quality of user-created documents,” in the Proceedings of the 4th Asia information retrieval conference on Information retrieval technology, Harbin, China, 2008.

[44] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, “Finding high-quality content in social media,” in the Proceedings of the 2008 International Conference on Web Search and Data Mining, Palo Alto, California, USA, 2008.

[45] J. Bian, Y. Liu, D. Zhou, E. Agichtein, and H. Zha, “Learning to recognize reliable users and content in social media with coupled mutual reinforcement,” in the Proceedings of the 18th international conference on World wide web, Madrid, Spain, 2009.

[46] Y. Y. Hao, Y. J. Li, and P. Zou, “Why Some Online Product Reviews Have No Usefulness Rating,” in the Proceeding of the Pacific Asia Conference of Information Systems 2009.

[47] T. Khopkar, X. Li, and P. Resnick, “Self-selection, slipping, salvaging, slacking, and stoning: the impacts of negative feedback at eBay,” in the Proceedings of the 6th ACM conference on Electronic commerce, Vancouver, BC, Canada, 2005.

[48] P. Resnick, R. Zeckhauser, J. Swanson, and K. Lockwood, “The value of reputation on eBay: A controlled experiment,” Experimental Economics, vol. 9, pp. 79-101, 2006/06/01 2006.

[49] P.-Y. Chen, S. Dhanasobhon, and M. Smith, “All Reviews are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.Com,” Social Science Research Network Working Paper Series, July/20/ 2006.

[50] S. Huang, D. Shen, W. Feng, C. Baudin, and Y. Zhang, “Improving product review search experiences on general search engines,” in the Proceedings of the 11th International Conference on Electronic Commerce, Taipei, Taiwan, 2009.

[51] M. A. Suryanto, E. P. Lim, A. Sun, and R. H. L. Chiang, “Quality-aware collaborative question answering: methods and evaluation,” in the Proceedings of the Second ACM International Conference on Web Search and Data Mining, Barcelona, Spain, 2009.

[52] S. M. Mudambi, “What makes a helpful online review? A study of customer reviews on amazon.com,” MIS Quart Manage Inf Syst MIS Quarterly: Management Information Systems, vol. 34, pp. 185-200, 2010.

[53] N. Korfiatis, E. García-Bariocanal, and S. Sánchez-Alonso, “Evaluating content quality and helpfulness of online product reviews: The interplay of review helpfulness vs. review content,” Electronic Commerce Research and Applications, vol. 11, pp. 205-217, 5// 2012.

[54] Q. Cao, W. Duan, and Q. Gan, “Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach,” Decis. Support Syst., vol. 50, pp. 511-521, 2011.

[55] M. Siering and J. Muntermann, “What Drives the Helpfulness of Online Product Reviews? From Stars to Facts and Emotions,” in Wirtschaftsinformatik, 2013, p. 7.

[56] Y. Dezhi, B. Samuel, and Z. Han, “Dreading and Ranting: The Distinct Effects of Anxiety and Anger in Online Seller Reviews, ” in the Poceeding In International Conference on Information Systems, 2011.

[57] P. F. Wu, H. V. d. Heijden, and N. Korfiatis, “The Influences of Negativity and Review Quality on the Helpfulness of Online Reviews,” in the International Conference on Information Systems, 2011.

[58] Z. Yang and X. Fang, “Online service quality dimensions and their relationships with satisfaction: A content analysis of customer reviews of securities brokerage services,” International Journal of Service Industry Management, vol. 15, pp. 302-326, 2004.

[59] R. M. Schindler and B. Bickart, “Perceived helpfulness of online consumer reviews: The role of message content and style,” Journal of Consumer Behaviour, vol. 11, pp. 234-243, 2012.

[60] P. Racherla and W. Friske, “Perceived 'usefulness' of online consumer reviews: An exploratory investigation across three services categories,” Electron. Commer. Rec. Appl., vol. 11, pp. 548-559, 2012.

[61] J. Otterbacher, “'Helpfulness' in online communities: a measure of message quality,” in the Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Boston, MA, USA, 2009.

[62] A. E. Schlosser, “Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments,” Journal of Consumer Psychology, vol. 21, pp. 226-239, 7// 2011.

[63] J. Tang, H. Gao, X. Hu, and H. Liu, “Context-aware review helpfulness rating prediction,” in the Proceedings of the 7th ACM conference on Recommender systems, Hong Kong, China, 2013.

[64] C. Dellarocas, “The digitization of word of mouth: Promise and challenges of online feedback mechanisms, ” Management science, vol. 49, pp. 1407-1424, 2003.

[65] R. Guha, R. Kumar, P. Raghavan, and A. Tomkins, “Propagation of trust and distrust, ” in Proceedings of the 13th international conference on World Wide Web, 2004, pp. 403-412.

[66] S. D. Kamvar, M. T. Schlosser, and H. Garcia-Molina, “The Eigentrust algorithm for reputation management in P2P networks, ” in the Proceedings of the 12th international conference on World Wide Web, Budapest, Hungary, 200

[67] S. Kurohashi, K. Inui, and Y. Kato, eds., Workshop on Information Credibility on the Web, 2007.

[68] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman, “Reputation systems,” (CACM), vol. 43, pp. 45–48, (ISSN 0001-0782), 2000.

[69] S. Krishnamoorthy, “Linguistic Features for Review Helpfulness Prediction,” Expert Systems with Applications, 2015.

[70] G. R. Semin and K. Fiedler, “The linguistic category model, its bases, applications and range,” European review of social psychology, vol. 2, pp. 1-30, 1991.

[71] S. Lee and J. Y. Choeh, “Predicting the helpfulness of online reviews using multilayer perceptron neural networks,” Expert Systems with Applications, vol. 41, pp. 3041-3046, 2014.

[72] Z. Zhang, Y. Ma, G. Chen, and Q. Wei, “Extending associative classifier to detect helpful online reviews with uncertain classes, ” 2015.

[73] T. L. Ngo-Ye and A. P. Sinha, “The influence of reviewer engagement characteristics on online review helpfulness: A text regression model, ” Decision Support Systems, vol. 61, pp. 47-58, 2014.

[74] Z. Liu and S. Park, “What makes a useful online review? Implication for travel product websites, ” Tourism Management, vol. 47, pp. 140-151, 2015.

[75] S. Quaschning, M. Pandelaere, and I. Vermeir, “When Consistency Matters: The Effect of Valence Consistency on Review Helpfulness, ” Journal of Computer‐Mediated Communication, vol. 20, pp. 136-152, 2015.

[76] K. Carlson and A. Guha, “The ratings paradox: Why we prefer reading negative reviews, but then subsequently rate these reviews as less useful,” Advances in Consumer Research, vol. 37, 2010.

[77] Y. Pan and J. Q. Zhang, “Born unequal: a study of the helpfulness of user-generated product reviews,” Journal of Retailing, vol. 87, pp. 598-612, 2011.

[78] M. Scholz and V. Dorner, "The Recipe for the Perfect Review?," Business & Information Systems Engineering, vol. 5, pp. 141-151, 2013/06/01 2013.

[79] L. Martin and P. Pu, "Prediction of helpful reviews using emotions extraction," in Twenty-Eighth AAAI Conference on Artificial Intelligence, 2014.

[80] M. Mertz, N. Korfiatis, and R. V. Zicari, "Using Dependency Bigrams and Discourse Connectives for Predicting the Helpfulness of Online Reviews," in E-Commerce and Web Technologies, ed: Springer, 2014, pp. 146-152.

[81] S. M. Mudambi, D. Schuff, and Z. Zhang, "Why Aren't the Stars Aligned? An Analysis of Online Review Content and Star Ratings," in System Sciences (HICSS), 2014 47th Hawaii International Conference on, 2014, pp. 3139-3147.

[82] A. H. Huang, K. Chen, D. C. Yen, and T. P. Tran, "A study of factors that contribute to online review helpfulness," Computers in Human Behavior, vol. 48, pp. 17-27, 2015.

[83] G. Lackermair, D. Kailer, and K. Kanmaz, "Importance of Online Product Reviews from a Consumer's Perspective," 2013.

Medical Image De-Noising Schemes Using Different Wavelet Threshold Techniques

Nadir Mustafa
School of Computer Science & Technology, UESTC
Chengdu, 611731, China

Jiang Ping Li
School of Computer Science & Technology, UESTC
Chengdu, 611731, China

Saeed Ahmed Khan
School of Microelectronics & Solid Sate Electronics, UESTC
Chengdu, 611731, China

Mohamed Tag Elsir
School of Electronic Engineering, UESTC
Chengdu, 611731, China

***Abstract*—In recent years most of researcher's has done tremendous work in the field of medical image applications such as Magnetic Resonance Imaging (MRI), Ultra Sound, CT scan but still there are many research and experiments in medical imaging field and diagnosing of human health by Health Care Institutes. There is a growing interest for medical imaging de-noising as a hot area of research and also imaging equipment as a device. It is used for better image processing and highlighting the important features. These images are affected with random noise during acquisition, analyzing and Transmission process. This results in blurry image visible in low contrast. Wavelet transforms have effective method to separate the noise from the original medical image by using threshold techniques without affecting the important data of an image. Wavelet transform enables us to use the forward wavelet transform to represent sub-band of the original image in decomposition process then reconstructing this sub band coefficients to original image using inverse wavelet transform. In this work, the quality of medical image has been evaluated using filter assessment parameters like Variance, standard deviation, the squared difference error between original medical image & de-noised image (MSE) and the ratio between original image & noisy image. From numerical results, we can see that the algorithm is efficient de-noising of noisy medical image. When, investigating with Baye's threshold techniques it achieved the Best value of peak signal to noise ratio (PSNR). For best medical image de-noising, the wavelet based de-noising algorithm has been investigated and results of Baye's techniques and hard & soft threshold methods have been compared.**

Keywords—Baye's Wavelet threshold; Discrete Wavelet; Medical Image De-noising; Magnetic Resonance Imaging (MRI)

I. Introduction

Most of medical diagnostic equipment has applications such as magnetic resonance imaging (MRI), criminal identification systems (CIS), agricultural and biological research (ABR) uses the concept of digital image processing. The term image de-noising is the best tool used in these applications, where it effectively captures the noise from corrupted medical image and preserving with the valuable data and important features of the medical image [1][2].

The motivation of using Medical Resonance Imaging (MRI) as a hot area because of it's related with human health. The medical Resonance imaging (MRI) very useful and low cost in diagnosis the human health and mapping the diagnosis output of the medical image in real and refine it as image quality [3][4].

During image acquisition and transmission, it has been usually observed that random noise always occurs at another end. In previous work, many researchers achieved good results in PSNR but not in MSE or visa-versa. Our work gives Good results in both PSNR and MSE [6][7].

Most of medical images are vulnerable to noise. This noise causes problems such as a blurred vision of images. Therefore, it is not easy for the medical doctors to examine the abnormalities in human in the invisible image. Most of medical imaging applications have been affected with random noises during Acquisition and transmission process that required improve and recover hidden data and details coefficients from noisy medical image.

Baye's threshold techniques provide good results when compared to Soft and Hard thresholds in terms of MSE and PSNR values as shown in the simulation results.

II. Wavelet Transform

The main difference between wavelet transform and windowed Fourier transform lies in signal analysis, the wavelet transforms using decomposition process to localize the signal in real time domain and frequency domain. In contrast windowed Fourier analysis has ability to localize the signal in Fourier space domain. Both Fourier and wavelet analysis represent the signal in different version such as sine wave and shifted and scale version which is kind of mother wavelet. But they are similar in windowing scales. The sine wave Fourier transforms have unlimited duration compared with waveform of wavelet transform and the wavelet forward to be irregular waveform. The wavelet Mathematical formula can be written as follows [5]. The equation (1) has been expressing a mathematic syntax of Fourier analysis and transform:

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-j\omega t} dt, \tag{1}$$

Which mathematically is the sum over all time of the signal f(t) multiplied by Exponential formula. (In mathematic the Exponential formula can be expressed as real and imaginary of sinusoidal vectors).

The Fourier coefficients F (ω) and sinusoid components of original image are results of applying Fourier transform in image processing, when multiplied by sinusoid of frequency. Similarly, the equation (2) expressing mathematic syntax of continuous wavelet transforms which is integration over whole time of the signal multiplied by shift and scale version of the wavelet function.

$$C(scale, position) = \int_{-\infty}^{\infty} f(t)\psi(scale, position, t)dt \qquad (2)$$

Many wavelet coefficients C are generated as function of scale and position, while Appling CWT in image processing by multiplied each coefficient with CWT function which consists of scale and shift as wavelet parameter [10].

In 1976 Croiser, Esteban, and Galand proposed discrete wavelet transform (DWT) as an effective technique to decompose the discrete time signals. On the other hand, Crochiere, Weber, Flanagan proposes the coding of speech signals technique which achieved as same time. The discrete wavelet transform DWT is sample of wavelet series and CWT version, which support sufficient and frequent information for wavelet decomposition and reconstruction of the original image, DWT requires sufficient amount of computations and resources. Furthermore, the DWT have ability to computerize the continuous wavelet transform as discretized version of signals.

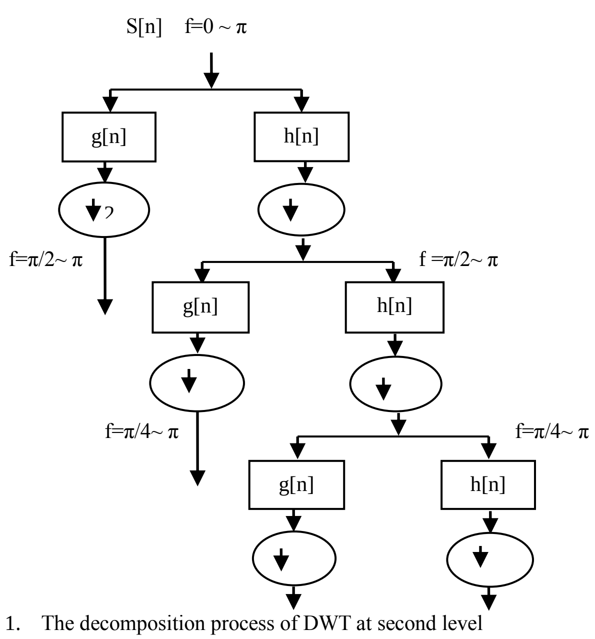


Fig. 1. The decomposition process of DWT at second level

Fig. (1) Depicts the decomposition of the signals by using high pass and low pass filters respectively to analyze the low and high frequencies and then measuring the amount of detailed information in the signal by using up sampling and down sampling operation. Both up sampling and down sampling terms respectively defined as reduction or remove the sample of signal, on other hand it means increase or adding new sample to the signal using Quadratic Mirror Filters which consist of two filters high pass and low pass. The DWT decompose the signal S[n] simultaneously by passing through to the high pass filter h[n] and low pass filter g[n], which generate the details coefficients and approximation coefficients that respectively named [HH, HL, LH] and [LL]. While using the Quadratic Mirror Filters the half of frequencies signal are removed [9][11]. The following figure is describing the decomposition of DWT at second level.

III. THRESHOLDING

The threshold techniques is an effective and necessary tools in wavelet transform which used to calculate the wavelet coefficients using 1-D or 2-D or 3-D dimensional wavelet array A [i,j] sized with i,j=1 to M that is define the element numbers of array. The Tb is a threshold parameter which remains the wavelet coefficient and wavelet power in thresholding process. The scientists Donoho and Johnstone gave the best choice of threshold in (1994). They showed that the threshold discards the smaller wavelet coefficients and preserve the larger coefficients than threshold level [12] [13].

A. Hard threshold

The hard threshold techniques deal with wavelet coefficients that less than threshold level after computed the wavelet transform and inverse wavelet transform by sitting all the coefficients to zero. Mathematically, the hard threshold formula is represented as follow in equation (3).

$$w_t = \begin{cases} w_t & |w_t| > T_b \\ 0 & |w_t| \le T_b \end{cases} \qquad (3)$$

B. Soft threshold

The soft threshold work similar to hard threshold discarded all the coefficients less than threshold level to zero. It’s also minimizing the magnitude of preserve wavelet coefficients to be equal with largest discarded coefficient. The soft threshold formula is represented as follow in equation (4)

$$w_t = \begin{cases} \operatorname{sgn}(w_t) \quad (|w_t| - T_b) & |w_t| > T_b \\ 0 & |w_t| \le T_b \end{cases} \qquad (4)$$

IV. PROPOSED DE-NOISING ALGORITHM

The Baye’s threshold technique has a ability to forming the threshold wavelet sub band coefficients from medical image in Bayesian Frame Work (BFW) during the decomposition process for wavelet coefficients, In Fig. 2, De-nosing algorithm model using discreet wavelet transform (DWT) at 2nd- level of decomposition have been proposed. It has been assumed in Generalized Gaussian distribution (GGD) and carefully finds the threshold Tb to reduce the thresholding risk, which affect to the important data in wavelet details and approximation coefficients. Finally, the performance of this thresholding is better for de-noising medical image compared with other threshold techniques [14][16].

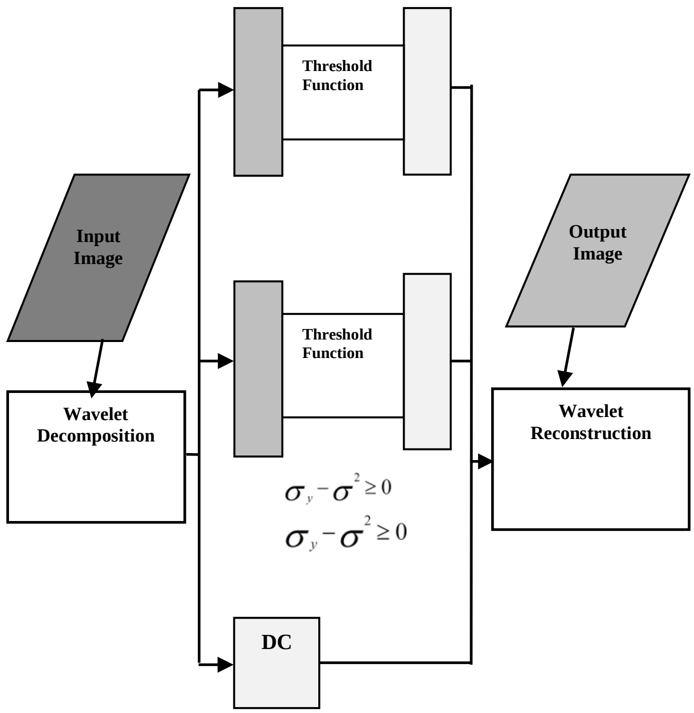


Fig. 2. Proposed De-nosing Algorithm Model using discreet wavelet transform (DWT) at 2nd- level of decomposition

A. *Estimation Parameters for Proposed Algorithm*

This section defines the adaptive parameters to proposed de-noising algorithm for medical image decomposed sub band coefficients which determine the threshold level (T_b) in equation (5) for different details and approximation sub band coefficients that depend on sub branch equation, which compare the max value of difference variances compared to zero σ_y - σ2 ≥ 0, by equation (3).

$$T_b = \sigma^2 / \sigma_x \qquad (5)$$

$$T_b = \max(abs(X)) \qquad (6)$$

Apply the wavelet transforms to medical image you have to generate details coefficient and approximation coefficients. From sub-band HH the noise variance measurement σ^2 is computed by the median estimator shown in equation (7).

$$\sigma^2 = \left[\frac{X - median(|Y_{ij}|)}{0.6745}\right]^2, Y_{ij} \in subband \quad \text{HH} \quad (7)$$

The details coefficients and sub band [LH, HL and HH] are generated while using wavelet decomposition process, the signal variance measurement σ_x estimated by equation (8) using standard MATLAB command.

$$\sigma_x = \sqrt{(\max(\sigma_y - \sigma^2, 0))} \qquad (8)$$

And σ x is the signal variance without noise defined as the squired max value of differences variance of (noise image and additive white noise image) which compared to zero, performed using standard MATLAB command.

$$\sigma_y = \frac{\sum (X)^2}{l_k} \qquad (9)$$

The standard deviation is defined as the average amount by which individual data from medical image sub band confidents differ from the arithmetic mean of all the data in the set. Lk is the length of the sub band at kth scale.

V. Image De-Noising Algorithm

The Image De-noising Algorithm is used to extract the noise from during acquisition and transmission and achieve the best image quality in medical image processing, after performed the discrete wavelet transform for original image, and then implementing the bayes wavelet threshold to remove the noise, which affected the visibility of the medical image. And finally the inverse wavelet transform performed to recover the clarity of decomposed medical image. As following steps:

A. *Perform the decomposition on the two of original medical images X, Y corrupted by Additive Gaussian Noise using wavelet transform*

B. *Chose the bayes threshold from equation (5) or (6) by estimate the measurements*

a) *Estimate the noise variance σ2 using equation (7).*

b) *Estimate the additive white noise variance σ y using equation (9).*

c) *Estimate the signal variance σ x using equation (8).*

C. *Perform the Reconstruction on the two of de-noised medical image X^, Y^ using inverse wavelet transforms to fine the clear vision of the image.*

VI. Experiments Analysis and Outputs

The laboratory work implemented on the two of medical images with same size (256×256) named (gland pituitary & Prostate). These images are defined as Xs, r & Ys, r matrix, which taken s, r = 1 to N parameter as image size, and it is a gray scale image. The problems were occurred while transmission and acquisition of original image. The image was affected by noise which could not be diagnosed by doctors and health institutes due to invisibility in image while investigating patient. To overcome on this issue, the forward wavelet transforms, bayes wavelet threshold and inverse wavelet transform techniques are effective tools to separate the noise from original images using wavelet decomposition & reconstructions process. After decomposition, bayes threshold have been applied for wavelet details coefficients and to assess the performance, the de-noising threshold algorithm have to add the variance noise measurement σ 2 = 0.04 of white Gaussian noise to original image for getting high peak signal to noise ratio (PSNR) measurement an d low mean square error (MSE) to assess the quality of reconstructed image compared with wavelet soft and hard threshold that measured by the equation (10)

$$PSNR = 10 \log_{10} \frac{255^2}{mse} \qquad (10)$$

Where the signal to noise ratio measurement is defined as the ratio of signal power to the de noise power of the de-noised medical image, often expressed in decibels

$$MSE = \frac{s}{s \times r} \sum_{s=1}^{N} \sum_{r=1}^{N} \left[X(s,r) - X^{\wedge}(s,r) \right]^2 \qquad (11)$$

The mean square error (MSE) of medical image in equation (11) is an estimator, which measures the average of the square "errors", that is, the difference error between original medical images (X) and synthesis image (X^). The Experimental work have been carried out by using bayes threshold to estimate the medical image quality using PSNR & MSE measurements with various wavelet threshold methods and various wavelet packages which have been presented in table I, II. From the tables, it can be understand that the lower value of MSE and higher value of PSNR in bayes threshold is vice versa in soft threshold and hard threshold. Which reveals that, bayes threshold in medical image de-noising is effective tool based on the experimental results. The bayes threshold presents the visual quality while compared with soft and hard. Moreover, Bayes threshold has high efficiency and good performance than soft and hard threshold. The MSE ratio of Soft threshold, hard threshold and Bayes threshold of gland pituitary and prostate image have been shown in figure (3&4). While figure (5&6) depicts the PSNR ratio of Soft threshold, hard threshold and Bayes threshold of gland pituitary and prostate image.

TABLE I. MSE RESULT OF GLAND PITUITARY & PROSTATE IMAGES TEST WITH VARIENCE 0.04

Medical Images	Wavelet Package	Soft Threshold	Hard Threshold	Bayes Threshold
gland pituitary Image	haar	0.0587	0.0570	0.0548
	sym4	0.0570	0.0556	0.0537
	db4	0.0559	0.0548	0.0530
	bior3.7	0.0541	0.0534	0.0524
Prostate Image	haar	0.0652	0.0620	0.0601
	sym4	0.0645	0.0616	0.0593
	db4	0.0626	0.0594	0.0575
	bior3.7	0.0595	0.0577	0.0566

TABLE II. PSNR RESULT OF GLAND PITUITARY & PROSTATE IMAGES TEST WITH VARIENCE 0.04

Medical Image	Wavelet Package	Soft Threshold	Hard Threshold	Bayes Threshold
gland pituitary Image	haar	32.1694	32.2001	32.2284
	sym4	32.1575	32.1703	32.2034
	db4	32.1404	32.1689	32.1921
	bior3.7	32.1204	32.1404	32.1704
Prostate Image	haar	31.1594	31.1671	31.1904
	sym4	31.1575	31.1703	31.1834
	db4	31.1454	31.1589	31.1661
	bior3.7	31.1342	31.1404	31.1584

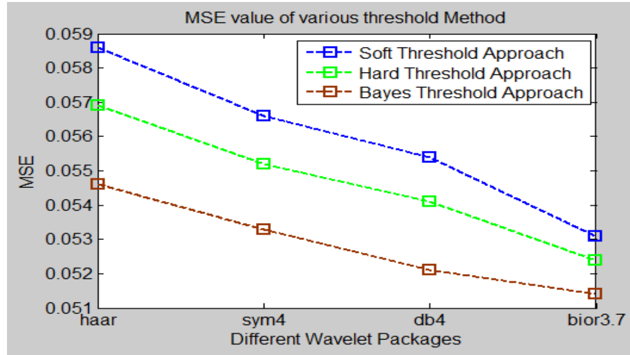


Fig. 3. Comparison of MSE of gland pituitary image de- noising; using different threshold techniques

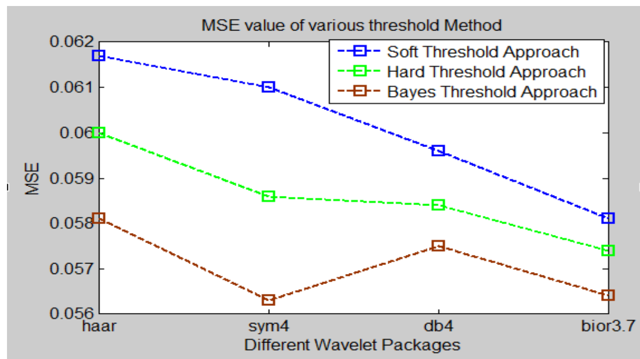


Fig. 4. Comparison of MSE of Prostate image de- noising; using different threshold techniques

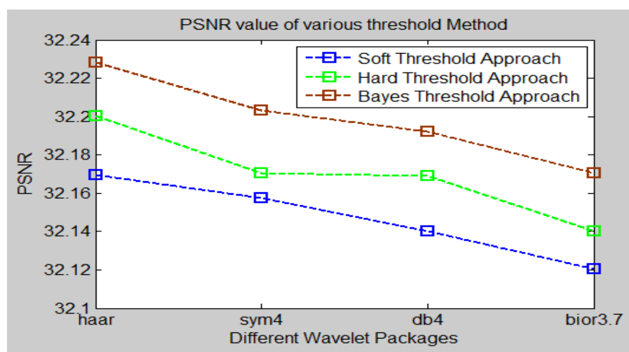


Fig. 5. Comparison of PSNR of Gland Pituitary image de- noising; using different threshold techniques

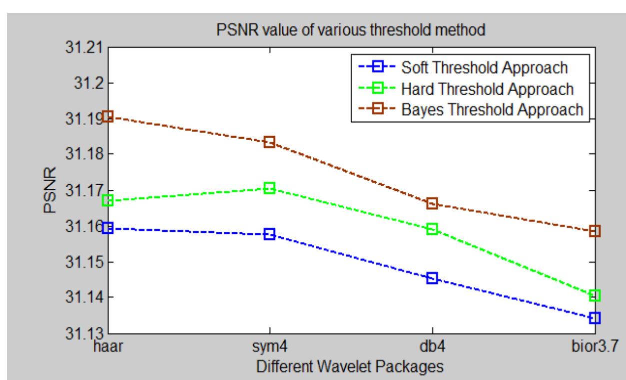


Fig. 6. Comparison of MSE of Prostate image de- noising; using different threshold techniques

The Comparison analysis of soft, hard and bayes threshold can be visualized from figure 7 & 8. It can be observed that, the soft and hard threshold in gland pituitary and prostate images are blurred than the bayes threshold. It reveals that, according to the experimental results, bayes threshold technique has much worth over soft and hard threshold for medical image de-nosing.

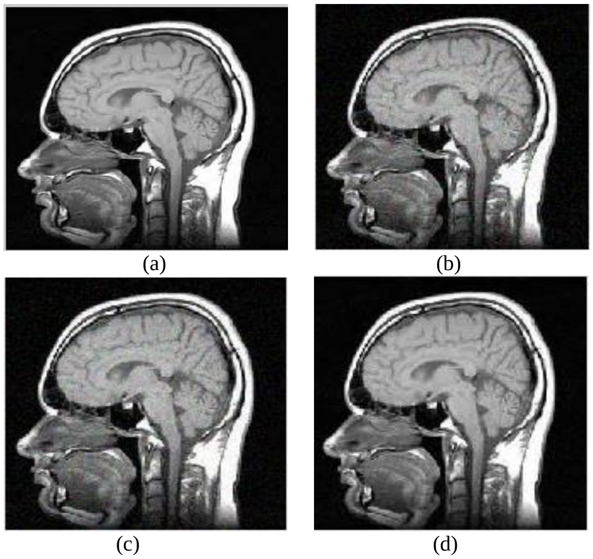


Fig. 7. Comparing the performance of gland pituitary image de- nosing (a) Original image (b) Soft threshold (c) Hard threshold (d) Bayes threshold

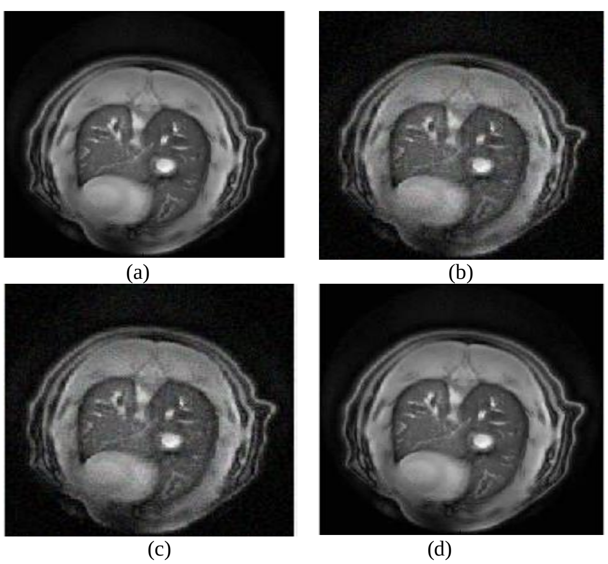


Fig. 8. Comparing the performance of prostate image de- nosing (a) Original Image (b) Soft threshold (c) Hard threshold (d) Bayes threshold

VII. Conclusion

Nowadays, Magnetic Resonance Imaging (MRI) plays an important role in medical image processing applications. In this work Bayes wavelet methods provide powerful tools for data de-noising problems. As we have applied various method on brain & prostate image using different wavelet threshold techniques (Soft, Hard and Baye's) and different wavelet packages (Haar, Sym4, db4 and bior3.7) and among these threshold techniques, the Baye's threshold has an optimal performance compared to the soft & hard threshold. In this method PSNR has maximum value & MSE lower value and the vision of medical image is much clear as compared to the soft and hard thresholds. Bayes threshold technique removed the noise significantly and remains the important features of data and medical image.

For future work, I suggest using different kinds of images. Like CIS, Agricultural, biological and geographical images for mapping and navigation.

Acknowledgment

This paper was supported by National Natural Science Foundation of China (grant NO: 61370073), the national high technology research and development program of china (Grant No: 2007AA01z423)

References

[1] Maarten Jansen. Noise Reduction by Wavelet Thresholding, volume 161. Springer Verlag, United States of America, 1 edition, 2001.

[2] Martin Vetterli S Grace Chang, Bin Yu. Adaptive wavelet Thresholding for image de-noising and compression. IEEE Transactions on Image Processing, 9(9):1532–1546, Sep 2000.

[3] Iain M.Johnstone David L Donoho. Adapting to smoothness via wavelet shrinkage. Journal of the Statistical Association, 90(432):1200–1224, Dec 1995.

[4] David L Donoho. Ideal spatial adaptation by wavelet shrinkage. Biometrika,

[5] David L Donoho. De-noising by soft Thresholding. IEEE

[6] S. G. Chang, B. Yu, M. Vetterli: IEEE Trans. Image Processing, (2000) 9 p. 1532

[7] S. G. Chang, B. Yu, M. Vetterli: IEEE Trans. Image

[8] D.L. Donoho, De-Noising by Soft Thresholding, IEEE Trans. Info. Theory 43, pp. 933-936, 1993.

[9] Javier Portilla, Vasily Strela, Martin J. Wainwright, Eero P. Simoncelli, Adaptive Wiener Denoising using a Gaussian Scale Mixture Model in the wavelet Domain, Proceedings of the 8th International Conference of Image Processing Thessaloniki, Greece. October 2001.

[10] Michel Misiti, Georges Oppenheim, and Jean-Michel Poggi, Wavlet Toolbox User's Guide,□□COPYRIGHT 1996 - 1997 by The Math Works, Inc. All Rights Reserved.

[11] Robi Ploikar, Index to series of tutorials to wavelet tranfrom ,The Wavelet Tutorial was originally developed and hosted (1994-2000) .

[12] S. Grace Chang, Bin Yu and M. Vattereli, Adaptive Wavelet Thresholding for Image Denoising and Compression, IEEE Trans. Image Processing, vol. 9, pp. 1532-1546, Sept. 2000.

[13] D.L. Donoho and I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, Journal of American Statistical Assoc., Vol. 90, no. 432, pp 1200-1224, Dec. 1995.

[14] S. Grace Chang, Bin Yu and M. Vattereli, Wavelet Thresholding for Multiple Noisy Image Copies, IEEE Trans. Image Processing, vol. 9, pp.1631- 1635, Sept.2000.

[15] S. Grace Chang, Bin Yu and M. Vattereli, Spatially Adaptive Wavelet Thresholding with Context Modeling for Image Denoising,, IEEE Trans. Image Processing, vol. 9, pp. 1522-1530, Sept. 2000.

[16] M. Vattereli and J. Kovacevic, Wavelets and Subband Coding. Englewood Cliffs, NJ, Prentice Hall, 1995

[17] Pooja Kaushik and Yuvraj Sharma,Comparison Of Different Image Enhancement Techniques Based Upon Psnr & Mse,International Journal of Applied Engineering Research, ISSN 0973-4562 Vol.7 No.11 (2012)

[18] Lakhwinder Kaur , Savita Gupta and R.C. Chauhan, Image Denoising using Wavelet Thresholding,

Handwriting Word Recognition Based on SVM Classifier

Mustafa S. Kadhm
Computer Science Department, University of Technology/ Baghdad

Asst. Prof. Dr. Alia Karim Abdul Hassan
Computer Science Department, University of Technology/ Baghdad

***Abstract*—this paper proposed a new architecture for handwriting word recognition system Based on Support Vector Machine SVM Classifier. The proposed work depends on the handwriting word level, and it does not need for character segmentation stage. An Arabic handwriting dataset AHDB, dataset used for train and test the proposed system. Besides, the system achieved the best recognition accuracy 96.317% based on several feature extraction methods and SVM classifier. Experimental results show that the polynomial kernel of SVM is convergent and more accurate for recognition than other SVM kernels.**

Keywords—Arabic Text; Preprocessing; Feature Extraction; SVM

I. Introduction

This Handwriting recognition is the process of converting the handwriting text images into a text file that understandable by the computer and used for many purposes. There are a lot of applications that depends on handwriting that are postal address reading for mail sorting purposes, cheque recognition and word spotting on a handwritten text page, etc. Naturally, handwriting is cursive and more difficult than printed recognition due to several factors that are the writer's style, quality of the paper and geometric factors controlled by the writing condition its very unsteady in shape and quality of tracing. Several steps taking place in handwriting recognition system, starting with preprocessing, feature extraction and classification.

To develop pattern recognition methods and systems, a lot of sample patterns is essential. In the same way as off-line character pattern databases such as IESK_arDB [1], AHDB [2], IFN/ENIT [3], KHATT [4] and so on, have been playing significant roles for off-line handwriting recognition. The purpose of this database is to enable the community to challenge the problem of object classification and recognition. Therefore, in this paper AHDB dataset has been selected for our proposed system. The dataset has the most popular Arabic words that written by many writers.

Furthermore, preprocessing is the first step in handwriting recognition systems it is helpful to reduce the variability of handwriting by correct these factors, and it will help to enhance the accuracy of segmentation and recognition methods. The second step in recognition system is the features extraction that extract a helpful information from the image text word to distinguish it from the other words. The last step of the recognitions is the classification that make the decision to sign the text word into the correct class label. [5]

The main contribution of the work presented in this paper is the integration of using multi scale features with a SVM for handwriting word recognition.

In order to show that features for Arabic script can be learned with the HOG descriptor, we evaluate our method on the AHDB dataset. The remainder of this paper is organized as follows:

Section II gives an overview of previous publications covering part-based classifiers for handwritten word recognition. Afterwards, our work is discussed in Section III and evaluated in Section IV. A final conclusion is given in Section V.

II. Related Work

The number of work for Arabic handwriting word recognition is quite limited comparing to the Latin script. The first works was given by Farah. [6] who introduced a system based on the combination of three Multi-Layer Perceptron for the recognition of Arabic words with recognition accuracy 94%. El-Hajj. [7] have used Neural Networks to combine three homogeneous HMM-based classifiers, that have different features as input , and they achieved recognition accuracy 94,44%. Alma'adeed [8] combined a rule based recognizer with a set of HMMs to recognize words in bank check of 47 words. The system tested on 4700 words and achieved 60% recognition accuracy.

Souici-Meslati [9] present a hybrid approach to the recognition of literal amounts on bank-checks. Three classifiers ran in parallel: neural networks, K-nearest neighbor, and Fuzzy K-nearest neighbor. The outputs were combined by word level score summation. 1200 words by 100 writers used for training and 3600 words for testing .The recognition accuracy was 96%. Burrow [10] applied on K-nearest Neighbors classification to each sub-word. The author scores at 74% for sub-word on correctly represented classes.

III. Proposed Handwriting Recognition System

The proposed method for handwriting word recognition has several major steps. Each of the recognition step affect the accuracy and the performance of the recognition. First of all the input images converted into grayscale it pass through several process as shown in figure1.

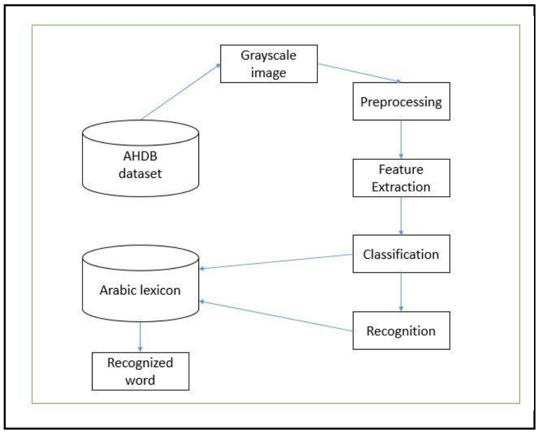


Fig. 1. Flowchart of the proposed system

The proposed system involves several steps which are; preprocessing, feature extraction, classification and recognition. Besides that, each step has it benefits for the recognition process. Here the proposed method steps described in details:

A. Preprocessing

Preprocessing is an essential step in the handwriting word recognition due to the effectiveness of this process on the recognition rate. Several steps has been taken place in the preprocessing phase that make the proposed method obtain a high accuracy.

The input to the handwriting word recognition system is a grayscale text image which has the Arabic word. During preprocessing the image converted to binary by thresholding method. The benefit of the thresholding is reducing the image diamantine to make it easy to process.

In the proposed system Fuzzy C-Means clustering (FCM) in [11] has been used to for thresholding purpose. After that, some noise appear due to the thresholding. 3X3 median used to remove undesired information from the binary image as shown in figure2.

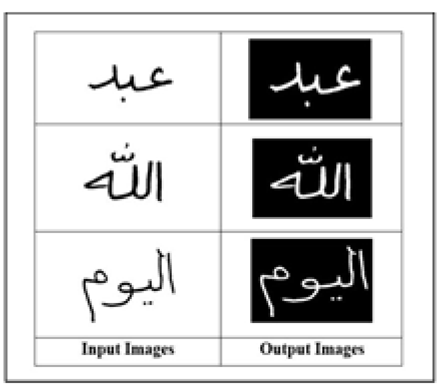


Fig. 2. Image Thresholding

Black space around the written word in an image does not help in any recognition process. So this unwanted black space around the word was eliminated. To eliminate this black space, bounding boxes were used. From each side of the binary image, the first pixel of the written word was located. This produced four points which formed the boundaries of the bounding box. The black area around this box could then be eliminated using these four values. The elimination of black space in an image is shown in Fig. 3, below:

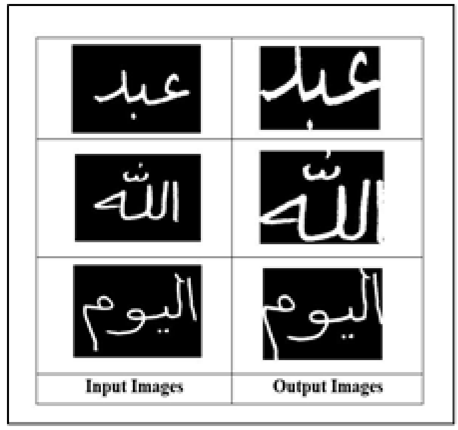


Fig. 3. Black space elimination

Moreover, image thinning is used which is the process of reducing image size by remove the redundant pixels without losing the representation of the original image. 3*3 mask used to scan the whole image and find the 4 connected pixels. After that the unaffected pixels are eliminate from the image this process must save the geometry and the connections between the words and the location of original word , based on border pixels removing recursively taking into account saving the geometry, location and connections. Image thinning method in [12] has been used.

The last step in preprocessing is the image normalization. AHDAB dataset has various image sizes. It important to make all the image in the dataset in the same size and make the recognition process fast. After testing several sizes (32x32, 64x64 and 128x128) the 128*128 size gave best recognition rate. Therefore, all the dataset images normalize into size 128*128.

B. Feature Extraction

The most important process in handwriting word recognition system is the feature extraction step. The best recognition depends on a successful feature extractions methods. A lot of feature extractions methods has been proposed for recognition purpose. However, there are three main types of features that can obtained from the character images.

1) Structural Features: Structural features describe the geometrical and topological characteristics of a pattern by describing its global and local properties. The structural features depend on the kind of pattern to be classified [13].

For Arabic words, the features consist of zigzag, dots, loops, end points, intersection points and strokes in many directions.

2) Statistical Features: Statistical features are numerical measures computed over images or regions of images. They include, but are not limited to, histograms of chain code directions, pixel densities, moments, and Fourier descriptors [14] .Statistical features are easy to compute and text independent. In the proposed system two types of statistical feature has been used which are:

- Connected Components Feature: The Arabic words has different shapes. Several shapes has different numbers of connected components pixels (segments). The idea behind of the connected component is to scan the whole image from left to right to find the groups of connected pixels (8 – connected neighbors). After that, each group of the connected pixels will get a label number. Therefore, the feature that obtained from this method is the number of connected components. This method is useful in Arabic words, since there are several words has different number of connected components.

- Zoning Features: In zoning features the image divided into number of zones and a particular features extracted from each zone. Several features extracted in this method which increased the recognition accuracy.

First the image divided into four zones figure5 then for each zone summation of the diagonal pixels has been calculated as a feature for that zone.

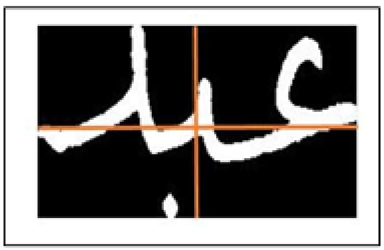


Fig. 4. Divide image into 4 zones

Second, the image divided into sixteen (16) vertical and horizontal blocks figure (14) then the summation of each block pixels will be the feature of that block.

3) Global Transformation: The transformation schemes convert the pixels transformation of the pattern to a more compact form which reduces the dimensionality of features [15].

- The Discrete Cosine Transform Features (DCT): The DCT converts the pixel values of an image in the spatial domain into its elementary frequency components in the frequency domain. Given an image f (i, j), its 2D DCT transform is defined as follows:

$$F(u,v)=\frac{2}{N}C(u)C(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1}f(x,y)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (1)$$

The inverse transform is defined by:

$$f(i,j)=\frac{2}{N}\sum_{u=0}^{N-1}\sum_{v=0}^{N-1}C(u)C(v)F(u,v)\cos\left[\frac{(2x+1)u\pi}{2N}\right]\cos\left[\frac{(2y+1)v\pi}{2N}\right] \quad (2)$$

Where

$$\alpha(u),\alpha(v)=\begin{cases}\sqrt{\frac{1}{N}} & \text{For } u,v=0\\ \sqrt{\frac{2}{N}} & \text{Otherwise}\end{cases} \quad (3)$$

Due to its strong capability to compress energy, the DCT is a useful tool for pattern recognition applications. The DCT can contribute to a successful pattern recognition system with classification techniques such as Support Vector Machine and Neural Network [16]. In the proposed system the DCT applied for the whole image that produced from the previous phase. The output of the DCT is an array of DCT coefficients. The features are extracted in a vector sequence by arranging the DCT coefficient in zigzag order, so that most of the DCT coefficients away from the beginning are small or zero. After testing the coefficients it found that the best number of DCT coefficients to represent the handwriting word as feature vector is the first 50 coefficients.

The DCT features extracted by the following steps:

```
Algorithm  DCT_FEXT
Step1: Read input image
Step2: Compute DCT for the input image(binary image)
Step3: Convert the DCT image into 1D array by zigzag order
Step4: Choose the first 50 DCT coefficients as a features
Step 5: Save the result features in 1D array
End
```

- Histogram of Oriented Gradient (HOG): Histogram of Oriented Gradient (HOG) was first proposed by Dalal and Triggs [17] for human body detection but it is now one of the successful and popular used descriptors in computer vision and pattern recognition. HOG counts occurrences of gradient orientation in part of an image hence it is an appearance descriptor. Before applying the HOG, the binary images converted to grayscale then filtered by using proposed edge detection mask filter in figure6 which gave better recognition rate than Sobel and Roberts filters.

$$\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix} \qquad \begin{bmatrix}-1 & 0\\ 0 & 1\end{bmatrix}$$

Fig. 5. proposed edge detection filter

HOG divides the input image into small square cells (here 6×6 has been used with bin=9 directions) and then computes the histogram of gradient directions or edge directions based on the central differences.

For improve accuracy, the local histograms have been normalized based on the contrast and this is the reason that HOG is stable on illumination variation. By applying this step, the total size of the feature set in the feature vector will be (6x6x9) = 324 .It is a fast descriptor in compare to the SIFT and LBP due to the simple computations, it has been also shown that HOG features are successful descriptor for detection.

C. Features Normalization

An important step to make the mathematical computing simple and fast a feature normalization (scaling) has been used to make the features ranges [0 1] by applying the following formula:

$$A' = \frac{A - \text{Min}(A)}{\text{Max}(A) - \text{Min}(A)} \tag{4}$$

Where A is an original value, A′ is the normalized value.

D. Classification and Recognition

After the feature extraction, the major task is the make decision to classify the word to which class it belongs. There are various classifiers that can applied in word recognition. The most important and more effective classifier is Support Vector Machine (SVM).

1) SVM Classifier: Vapnik and Cortes developed SVMs [18, 19] as a statistical learning machine in the late 1990s. Within a short time, they became one of the most popular classification systems in data mining and pattern recognition applications, due to their high classification rates. Researchers successfully applied SVMs in many modern learning applications such as Optical Character Recognition (OCR), bioinformatics, document analysis, and image classification.

SVM commonly used with linear, polynomial, RBF and sigmoid kernels. A multiclass SVM classification (libsvm) has been used in the proposed system [20] with different kernels of 1) linear, 2) polynomial, 3) RBF, 4) sigmoid and it achieves a very high recognition accuracy.

The final step is the recognition which is matching the selected class by the SVM with the character ASCII and find the desired word in the Arabic lexicon.

IV. Experimental Results and Discussions

The proposed method is implemented using Matlab R2015a version, under windows7 64-bit Operating System, with RAM 6GB, CPU 2.50GHz core i5 and it achieved fast and effective results.

The proposed dataset has 2913 handwriting word images. Each word has 105 images written in different style. In the handwriting word recognition system 70% of the dataset used for training purpose (2044) and 30% for testing (896) and it achieved 96.317% recognition accuracy. When we increase the number of the training data, the recognition accuracy is also increase.

However, we applied the system on our proposed dataset that has 1675 word images. Besides, the dataset has many images with noise. The proposed system also achieves height recognition accuracy 96.3% with the proposed dataset.Figure6 shows the samples of the proposed dataset.

Furthermore, in the proposed system SVM classification work with different kernels and each kernel achieved different accuracy. Besides that, there are an important parameters which make the SVM work perfectly.

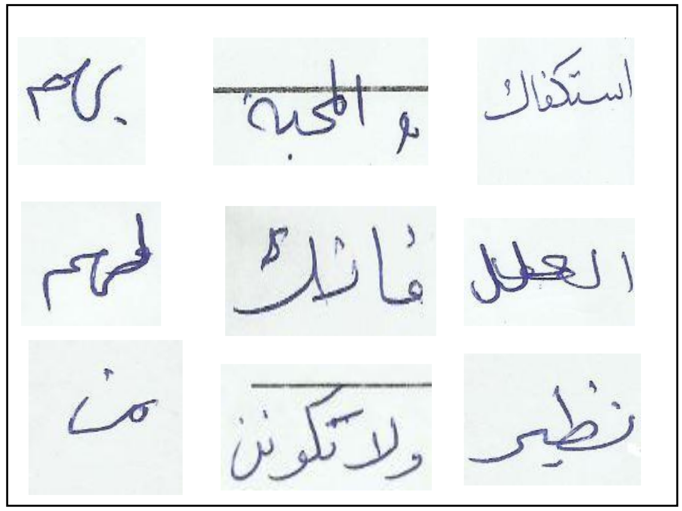


Fig. 6. proposed handwriting dataset

The most important parameters in SVM are: cost(c) and gamma (γ). After many testing of the system the best values of the parameters was $c = 8$ and $\gamma = 0.0625$.

Furthermore, different SVM kernels has been tested and the best achievement was by using SVM polynomial kernel.

In Table1 it noted that the highest accuracy obtained by our proposed system this is due to the use of efficient features and SVM classifier. Our system achieves good accuracy compared with [6, 7, 8, 9, 10] systems which prove the performance of Polynomial kernel of SVM classifier.

TABLE I. Comparison Results

	Classifier	**Accuracy**
El-Hajj	3HMMs	94.44%
Farah	3MLP	94%
Burrow	Several KNN	74%
Souici-Meslati	NN+KNN+Fuzzy	96%
Alma'adeed	HMMS	60%
Proposed System	SVM	96.317%

V. Conclusion

In this a paper, a proposed a high accurate handwriting word recognition system. The system use 70% of the dataset for training and 30% for testing and obtained high accuracy with SVM polynomial kernel. The high accuracy achieved by several factors starting from the efficient preprocessing stage with the use of FCM the with efficient feature extraction methods and finally with more accurate recognition classifier .Experiments, our proposed system gave best recognition accuracy than the existing systems.

TABLE II. COMPARISON BETWEEN DIFFERENT KERNELS OF SVM

SVM Kernels	Linear	Polynomial	RBF
Recognition Accuracy	92.63%	**96.317%**	91.5%

In addition, we are planning to apply our proposed work in future for recognition the text in documents not just word recognition. The future work need efficient segmentation method that can segment each text word without any overlapping or missing parts. Neural networks can be used for the segmentation and classification which may improve our results.

REFERENCES

[1] M. Elzobi, A. Al-Hamadi, Z. Al Aghbari, and L. Dings, "IESK-ArDB: a database for handwritten Arabic and an optimized topological segmentation approach," International Journal on Document Analysis and Recognition, vol. 16, no. 3, pp. 295–308, 2013.

[2] S. Al-Ma'adeed, D. Elliman and C. Higgins, "A Data Base for Arabic Handwritten Text Recognition Research," Proceedings of the 8th International Workshop on Frontiers in Handwriting Recognition (IWFHR 2002), pp. 485-489, August 2002.

[3] S. Mozaffari, H. E. Abed, V. Maergner, K. Faez and A. Amirshahi, "IfN/Farsi-Database: a Database of Farsi Handwritten City Names," Proceedings of the 11th International Conference on Frontiers in Handwriting Recognition (ICFHR 2008), pp. 397-402, August 2008.

[4] S. A. Mahmoud, I. Ahmad, M. Alshayeb, W. G. Al-Khatib, M. T. Parvez, G. A. Fink, V. Margner and H. E. Abed, "KHATT: Arabic Offline Handwritten Text Database," Proceedings of the 2012 International Conference on Frontiers in Handwriting Recognition (ICFHR 2012), pp. 449-454, September 2012.

[5] Lorigo, L.M. and Govindaraju, V. (2006). Off-line Arabic Handwriting Recognition: A Survey. IEEE TRANSACTIONSON PATTERN ANALYSIS AND MACHINE INTELLIGENCE. , 712- 724.

[6] Aziz N., Farah N., Sellami N., and Ennaji A., " Using Diversity in Classifier set Selection for Arsbic handwrittien recognition" in proceedings of multi classifiers system , lecture note in CS, Berlin,pp.235-244,2010.

[7] El-Hajj R., Mokbel C., and Likforman-Sulem L., "Combination of Arabic Handwritten words". In proceedings of the 19th International Conferemnce on Document Analysis and Recognition,Parana,pp. 959-963,2007.

[8] S.Alma'adeed,C.Higgens,and D.Elliman,"offline recognition of handwritten Arabic words using multiple hidden markov models",knowledge based systems ,vol.17,pp. 75-79,2004.

[9] L. Souici-Meslati, M. Sellami, " A Hybrid NeuroSymbolic Approach for Arabic Handwritten Word Recognition ", JACIII, Journal of Advanced Computational Intelligence and Intelligent Informatics, FujiPress, Vol. 10, No. 1, pp. 17-25, 2006

[10] Peter Burrow. Arabic handwriting recognition. Technical report, School of Informatics, University of Edinburgh,, 2004

[11] Elzobi, Moftah, Ayoub Al-Hamadi, Zaher Al Aghbari, and Dinges, Laslo (2012) viewed 5 March 2015, at. http://www.iesk-ardb.ovgu.de/

[12] Niall, Gunter, Innsbruck: Optical Character Recognition. Informatics Research Institute (IRIS) at University of Salford(2011)

[13] Vinciarelli, A. B. (2004). Offline recognition of unconstrained handwritten texts using HMMs and statistical language models. IEEE Transactions on Pattern Analysis and Machine Intelligence , 26 (6), 709–720.

[14] Issam Bazzi ,Richard Schwartz and John Makhoul. An "Omnifont Open-Vocabulary OCR System for English and Arabic". IEE Trans. On pattern analysis and machine intteligance. Vol 21 , no. 6 , pp 495-504,1999

[15] F.zaki, S. Elkonyaly, A. A. Elfattah, and Y. Enab. "A new technique for Arabic handwriting recognition". In processing of the 11th international conference for statistics and computer science ,(cairo , Egupt), pp. 171-180,1986

[16] JIANG, J., WENG, Y. & LI, P. (2006) Dominant colour extraction in DCT domain. Image and Vision Computing, 24, 1269-1277.

[17] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In CVPR, 2005

[18] Cortes C, Vapnik V. Support-vector networks. Mach Learn. 1995;20(3):273–297.

[19] Vapnik V. The nature of statistical learning theory. Springer; 1999.

[20] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. ACM T. Intell. Syst. Technol., 2(3):1–27, 2011

Creating a Knowledge Database for Lectures of Faculty Members, Proposed E-Module for Isra University

Dr. Amaal Al-Amawi
Management Information Systems Department
Isra University
Amman, Jordan

Dr. Salwa Alsmarai
Management Information Systems Department
Isra University
Amman, Jordan

Dr. Manar Maraqa
Management Information Systems Department, Amman College for Financial and Administrative Sciences, Al-Balqa' Applied University, Amman, Jordan

***Abstract*—Higher education in Jordan is currently expanding as new universities open and compete for offering the best learning experience. Many universities face accreditation challenges, hence, they attend to recruit lecturers who may not have a solid teaching experience. Experienced lecturers tend to have high turnover rate, which cause knowledge loss. To prevent such loss, this research presents a knowledge repository framework. This framework will serve as a reference and a vessel of knowledge that builds and develops the educational and teaching capacities of professors/lecturers. It can also be seen as part of the electronic learning system, which brings benefits to students and enables them to retrieve any lectures they need. The main question we aim to answer is whether a knowledge memory can be designed and created to contribute in supporting the educational and teaching capacities of university lecturers. In order to answer this question, this research creates an electronic knowledge database to store explicit knowledge taken from lectures (written, audio and visual). These lectures are prepared and circulated or presented by university professors/lecturers throughout all university colleges and departments. This knowledge database resembles a cognitive memory that grows and develops with time.**

Keywords—Knowledge; knowledge database; electronic knowledge database; Knowledge sharing

I. Introduction

Information and communication technology is an essential tool considering that it has two key abilities. The first is the ability to code knowledge and make it virtual and available. The second is the ability to create networks, with which knowledge can flow and be transferred within organizations (universities). By creating electronic knowledge databases, data can be provided to build the capacities of new and existing lecturers. Through the database, they can access the lecturing methods used by the university's qualified and distinguished professors. This would contribute in building an outstanding teaching staff that is able to provide the university with a competitive feature that favors it over others. This is especially important with the large number of universities currently present and the difficulty to find exceptional staff. Adding to this is the high turnover among lecturers in most universities caused by job opportunities and advantages provided by other universities to attract the best teaching staff available. Therefore, building an electronic knowledge database for all university lecturers and specializations is necessary.

A. Study Question

The research question can be determined by asking a number of questions. Some of these questions will be answered in the theoretical part of the research. The remaining questions will be tackled in the practical part of the research. The questions are:

1) What is knowledge and how important is it considering that it is part of the intellectual capital?

2) Does the university under study have a knowledge database?

3) Can a knowledge memory (knowledge database) be designed and created to contribute in supporting and building the educational and teaching capacities of university lecturers?

This research shows the importance of adopting tools that are best able to improve the performance of the university. Such tools are tasked with storing knowledge electronically, maintaining it, and ensuring its flow and exchange among human resources who are hungry for this knowledge. Consequently, this would feed into developing and expanding their thoughts and ideas through adopting an electronic medium suitable for this purpose.

We decided to address this issue especially because it became a point of interest to many universities who wish to keep and store knowledge in an electronic manner and share it among staff. This directly feeds into promoting excellence as well as encouraging development among staff members.

B. Methods Used to Collect Data and Information

To address the study question and achieve its objectives, Arabic and foreign scientific sources, such as relevant periodicals, references, internet sites, were consulted to build the theoretical framework.

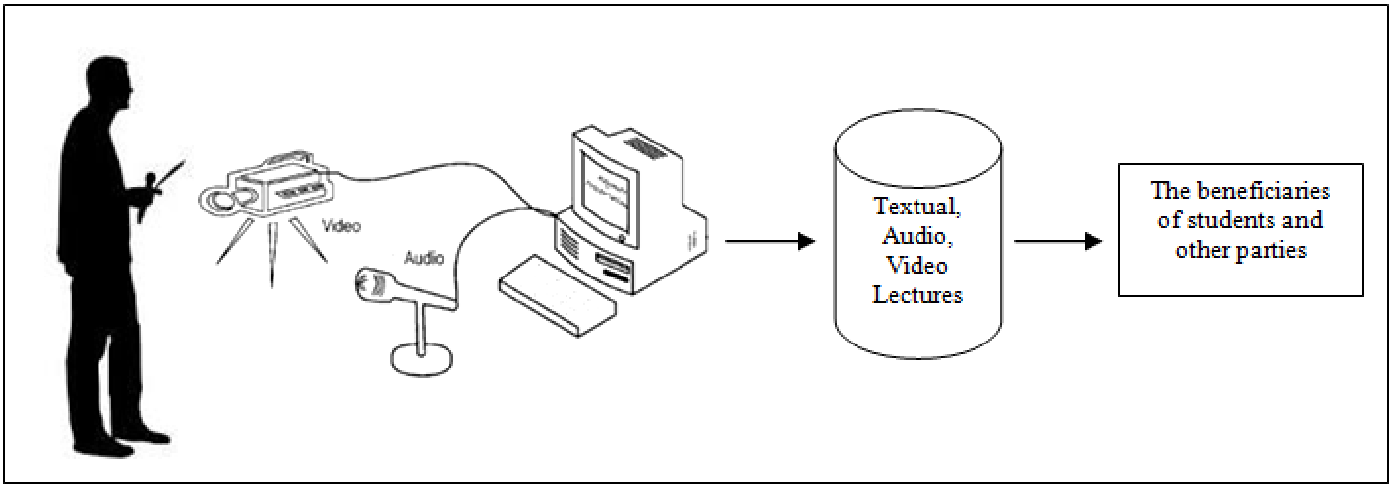


Fig. 1. The proposed system

Concerning the practical framework, data was collected from educational electronic websites or experienced lecturers at the university as shown in Fig. 1. This supports the creation of a computerized module to serve the university concerned with this study.

The study is divided into five parts starting with the introduction as part one. Part two presents the study's theoretical background and part three illustrates the fundamental principles of the intellectual capital. Part four shows practical steps of designing a database for a faculty, while part five presents the conclusions and the implementation ideology.

II. Part Two: Study's Theoretical Framework

A. *Definition of Knowledge*

The concept of knowledge is very old and was the topic of ancient Greek, Roman and Muslim philosophers. The verses of the Holy Quran and the teachings of Prophet Muhammad (peace and blessings be upon him) stressed the importance of knowledge. As a definition, the meaning of knowledge is explained in dictionaries and lexicons as a clear and affirmed understanding of things, facts and behaviors [1]. According to Oxford dictionary, it is the process or act of acquiring information using ones' senses and intuition. Derver psychology dictionary says that it is a term that covers all different forms of learning ranging from understanding, imagination, remembrance, belief, contemplating, judgment and reasoning [2].

The importance of knowledge is found in the fact that it shows the level of people's awareness of what goes around them, and their ability to understand and deal with circumstances. That is because the operations of the brain play a critical role in human behavior. The human brain is the organ that stimulates people to make sound comparisons and decisions that are suitable for finding new problem-solving techniques [3].

The major development of ancient and modern civilizations is only a proof of their ability to learn and share knowledge. This idea can be seen as an integrative accumulative process that happens and forms on the course of relatively long periods of time so that it becomes usable and applicable for solving problems and specific circumstances [4].

Daft mentioned that knowledge consists of two parts: an apparent part that can be directly handled, moved and transferred into paper for easy reference, and a second part that is implicit difficult to document or transfer. This part includes skills, experiences and wisdom of individuals. According to this concept of knowledge, what an organization holds, whether internal system information, skills, or mental capacities of staff, form in their entirety a resource for the organization if they were used and managed in a scientific and reasonable way. This would certainly reflect on the performance of the organization by giving it an advantage over other competitors [5].

B. *Types of Knowledge*

There are various types of knowledge and those who develop it should be familiar of its differences. In fact, during the process of acquiring knowledge, it is important to know how to select it.

Bolisani & Scarso [6], Lundvall & Johnson [7] and Drew [8] identified four types of knowledge: Know-what, Know-why, Know-how and Know-who. Whereas, Kingston & Macintosh [9] added two extra types of knowledge: Know-when and Know-where. All will be described briefly, as follows [10]:

1) Know-what: includes knowledge of facts which is very similar in meaning to the term information; therefore, it is transferrable. Kingston & Macintosh describe it as the explicit knowledge of things; it includes concepts, as well as material topics and situations. Drew points out that it is used by experts and specialists.

2) Know-why: refers to the scientific knowledge of principles, laws, rationals, reasons, debates, applied research, justification of why things happened and how do they work.

3) Know-how: it's the knowledge of required actions for specific incidents and situations. According to Bolisani &

Scarso, it consists of applied capabilities to execute certain activities and Drew specifically assigns it to creative processes. The know-how relies on lengthy experiments that are the prerequisites of building expert systems. However, the problem with scientific experiment is that they are seldom documented. In short, the know-how is not the knowledge of books but rather a practical experiment.

4) Know-who: comprises of information on who knows what, and who knows how. Drew on this type indicates that it refers to the importance of individuals and social and political relations.

5) Know-when: is the knowledge needed to control actions and incidents [that happened or should happen] at the right timing.

6) Know-where: includes areas where knowledge is needed and asks questions to find where do communications come from? and where are the inputs and outputs of knowledge?

Lundvall & Johnson point out that types 1 and 2 can be easily codified and are included in explicit knowledge, while types 3 and 4 are difficult to be encoded and eventually fall under tacit knowledge.

Another more widely accepted taxonomy on knowledge, first introduced by Polanyi [11] and popularized by Nonaka and Takeuchi [12], is their categorization of explicit and tacit knowledge. According to these two, explicit knowledge is that which can be expressed in words and numbers and can be easily communicated and shared in the form of hard data, scientific formulas, codified procedures or universal principles whereas tacit knowledge is highly personal and hard to formalize. Personal insights, intuitions and hunches fall into this category of knowledge.

C. The SECI Model and Knowledge Conversion

The model was based on Polanyi's definition of tacit and explicit knowledge. In fact, Nonaka mentioned it in his book entitled "The Knowledge Creating Company" [13]. He then further explained it in his joint book with Takeuchi under the same title [12]. The main idea here is that there is a spiral interactive movement between explicit and tacit knowledge. This movement is accompanied by four conversion processes of knowledge that lead to the formation of the new organizational knowledge. The model consists of four blocks each representing a fundamental process of conversion (See Fig. 2). The mechanism of knowledge conversion in the SECI Model is explained as follows [14]:

1) Socialization: transfers tacit knowledge owned by an individual to a group of individuals working at the organization. This is possible at inter-organizational level through social interaction between employees, , and between employees and customers/suppliers or any recipient of the organization's services.

2) Externalization: knowledge is transferred from tacit to explicit on the individual to group level. Externalization is the second process of knowledge conversion. Moreover, Nonaka & Takuech indicated that it is an essential process as it converts knowledge from one type into a totally new type [tacit to explicit] [12].

3) Combination: here, explicit written knowledge [whether codified or digital] is converted to encoded bundles or configurations of explicit knowledge, which is also written. The conversion into explicit knowledge requires previous cognitive and mental processes.

4) Internalization: converts explicit knowledge to tacit knowledge from a group to individual's level. Nonaka &Takeuchi stress that continuous learning and collecting of the working individual's expertise through on-the-job-learning support the internalization process [12]. Conversion is achieved by redoing tasks and following relevant instruction. Furthermore, conversion takes the form of intellectual models that are known as the technical method.

D. Knowledge storage (Knowledge memory)

It includes processes of retention, sustainability, accessibility, retrieval and location. An institution may put in a lot of effort to acquire knowledge. Nonetheless, knowledge might be vulnerable to being lost either by failure to recall/remember it or retrieve it when needed [15]. Consequently, storing knowledge and the ability to retrieve it is a critical element in Knowledge Management. Often, this element is referred to as the "Organizational Memory", which is a term that refers to storing and recording intellectual capital. It comprises of the information which individuals use to work with, in addition to the knowledge in the systems and structures of that specific organization [16].

By the same token, to ensure continuity of the organization's knowledge base and achieve maximum benefit, this knowledge is ought to be stored, saved and retrieved at all times. Therefore, Knowledge Management should establish Organizational Memory [17].

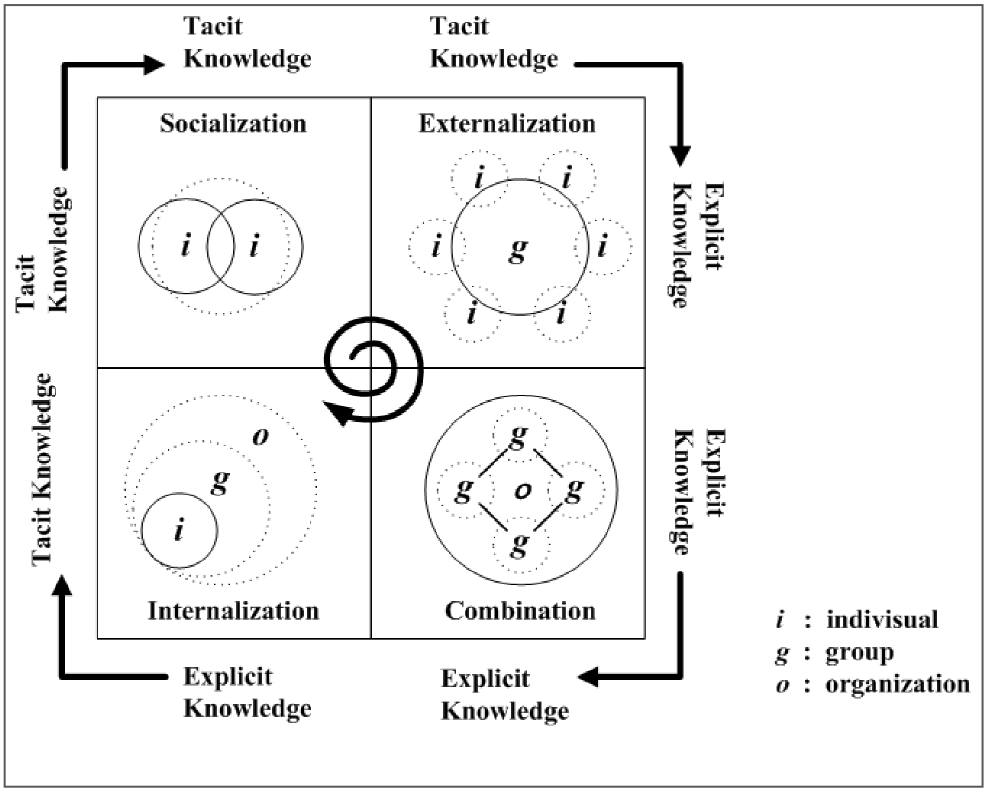


Fig. 2. The SECI Model and Knowledge Conversion Source: Nonaka and Konno [13]

III. Fundamental Principles of Intellectual Capital

Identifying the organization's assets, namely its intellectual assets such as knowledge depicted in an equation, a trade secret, an invention, a program or a process etc, is crucial to the organization's vision, its strategic plan, as well as its endeavors to gain competitiveness advantage.

Hence, the future of knowledge-dependant organizations relies heavily on the ideas they possess. In this age of information, greater attention is given to intellectual assets that are not only measured but might also outweigh its material wealth [18].

A. Firstly: The definition of Intellectual Capital (IC):

According to the Organization for Economic Co-operation and Development (OECD), Intellectual Capital is a subset of "intellectual assets". Malone considers IC as intellectual materials such as knowledge, information, intellectual property and experience, which can be used to create wealth [20].

Furthermore, IC can also be identified as "the knowledge that could be converted into value" [21]. Conversely, Hansen, Nohria and Tierney [22], demonstrated that "intellectual capital is the competitive measure that operates the creative and strategic development of an organization which depends on creativity and innovation, which are considered the key elements that lead to the survival of an organization in the quickly changing work environment."

Furthermore, knowledge is considered a new type of intellectual capital. Since knowledge is the information one has and benefits from, it is dissimilar to data that can be converted to information. It is rather a combination of categorized experiences derived from one's vision or that of a company. The aim is to reuse the available information resources - of a company per say - in a positive manner. Knowledge generates new knowledge; thus, the intellectual capital never depletes, it is rather increased by accumulation [18].

B. Secondly: Components of Intellectual Capital:

Writers and researches did not agree on a unified classification of IC and ultimately identify its components. Davenport & Pursak [21] came up with a model of three main components: Human Resources, Intellectual Assets and Intellectual Property. Edvinsson & Sullivan [23], however, provided a model for companies' capitals, which is divided into two main components: the first is the material capital, while the second is dedicated to intellectual capital that includes both human capital and customers' capital. As to Stewart [24], he noted that IC breaks down to three components. The first is ***workers***: who provide knowledge, innovations or solutions to customers' problems and eventually achieve profits for the company. Secondly, the ***working system***: meaning the organizational structure or hierarchy that's being adopted by the organization. It is also the set of rules and principles the company abides by for its internal procedures, whether inter-organizational or with costumers in external environments. Lastly, the ***customers:*** who are an important source of information that is employed to further develop the company's production [24].

The human intellectual capital (education, capabilities and skills) is converted into a structural capital, i.e. a systematic documented knowledge and rules. In other words, it converts tacit knowledge into explicit knowledge to avoid losing valuable knowledge owned by human competencies should they leave the place of work [25].

C. Thirdly: The Importance of Intellectual Capital

Intellectual capital is a good long-term profitable investment. Nonetheless, to accomplish the rewarding results, sacrifice should be made; it can be in the form of costs that are put upfront to obtain it. Yet, intellectual capital is the single resource that does not fall under the law of diminishing returns. Indeed, it never depletes; quite the contrary; it is an ever-growing accumulative resource that is used in generating and developing new ideas. Therefore, intellectual capital is regarded now as the actual capital on which companies' failures and successes depend [18]. Thus, modern administrations have based their selection processes on individuals with knowledge capabilities and are keen on providing more knowledge-stimulating environments. Given that, institutions and also countries have shifted their focus to intellectual capital since it is a primary indicator of a country's wealth.

Despite the downsizing of human resources by modern technology, as they have replaced many staff members, the remaining human resources are thought to be the ones with better qualities and competencies. Their stay is determined by their ability to grasp and apply modern technologies efficiently. The importance of intellectual capital is on the rise due to the domination of information technologies in this day and age [26].

Researchers see that knowledge has become an essential source for competitiveness. Therefore, organizations should equally own knowledge and be well-experienced on how to manage it. Those two skills together form what is known as "intellectual capital".

IV. PART THREE/ PRACTICAL ASPECT-STEPS OF DESIGNING A DATABASE FOR FACULTY LECTURES

This topic will be discussing the main steps -suggested by the researchers as shown in Fig. 3- on designing an electronic database for Isra University (IU), as follows:

A. Facts about Isra University Knowledge database

According to the researchers at Isra University, there is a number of knowledge sources that could be adopted to build a reliable knowledge database, such as:

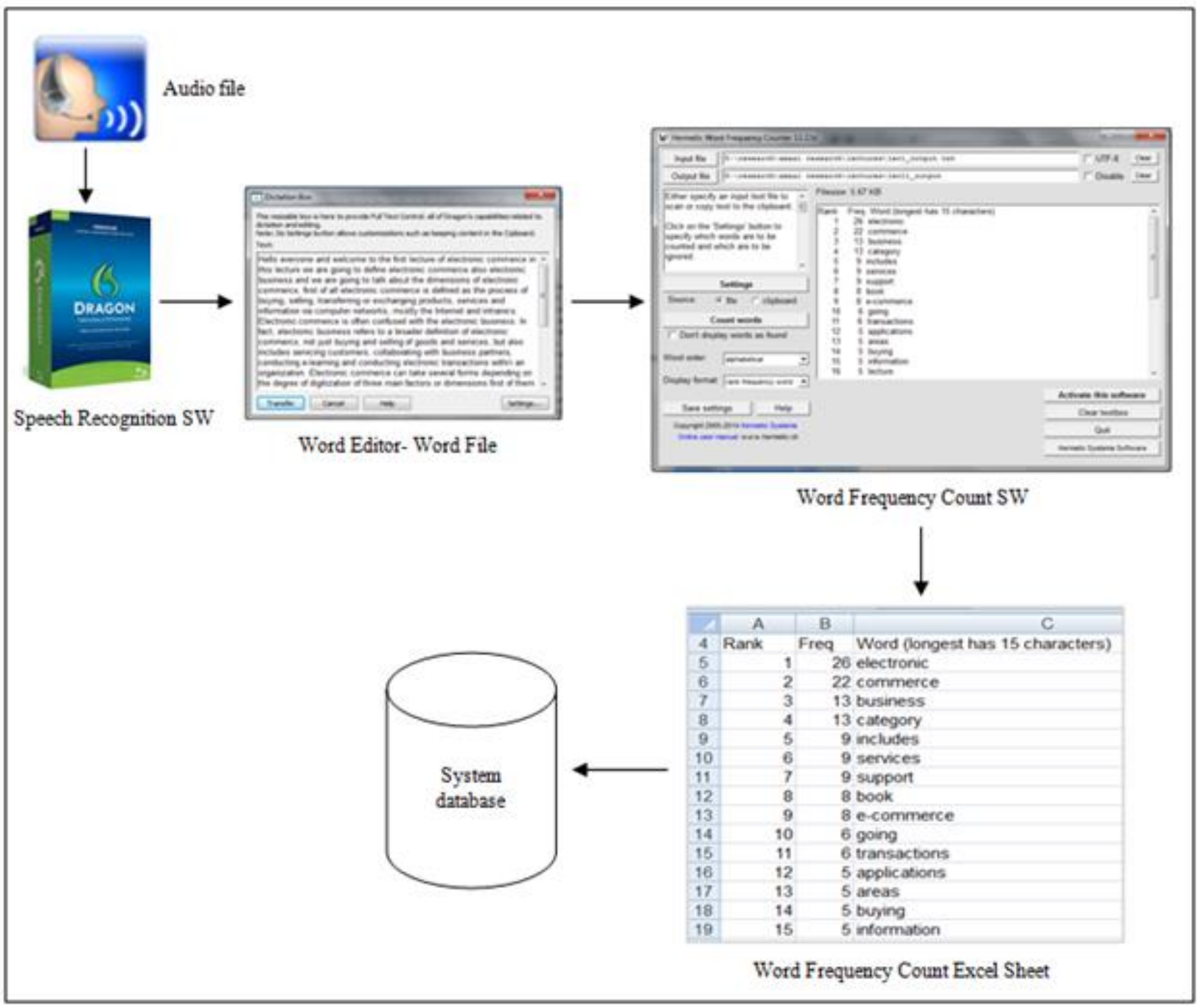


Fig. 3. System Scheme

1) E-learning websites that is used by IU departments like the colleges of Science and Information Technology, Pharmacology, Nursing and Engineering. A page is dedicated for every lecturer. Lectures are being uploaded to the professor's designated page either as texts or as a Power Point presentation.

2) Since some departments do not have direct access to such websites, and in order to communicate with the students, researchers had no choice but to obtain the materials separately from the faculty and then upload the lectures in an electronic format.

3) The third and key source is the audio and visual lectures, which is thought to be the hardest to obtain as they not only require the approval of the faculty member providing them, but also the permission of the presidency of the university (IU). Moreover, cameras should be installed at the lecture halls in order to acquire the audio/visual materials. But since the halls are used by multiple academics, instructors that are not listed on the website [with little or no experience at all] can turn off the camera during their class. At the same token, they themselves can benefit from the service and see how lectures are being given professionally.

4) Another way to obtain recorded lectures is to ask the instructors and professors themselves to record some lectures in their offices during their free office hours. Doing this is very important even before the department decides to start recording lectures. The reason is that in order to convert an audio file into a written document (word file), we need to use speech recognition programs which are usually user-dependent. This means that the user has to train the program to recognize his/her voice and the way he/she talks before using it.

One more thing users need to know is that speech recognition programs work best when users dictate what they want to say- users should not talk fast- and they should keep the microphone close to them; otherwise the number of errors in the recognition process will increase.

B. Practicality

This topic focuses on the materials suggested by the researchers pertaining to designing a data of knowledge for universities, as follows:

1) System Analysis

Fig. 3 illustrates a scheme suggested by researchers for the process of lecture storage. It was translated electronically through a system specifically designed for universities and colleges as per specialization.

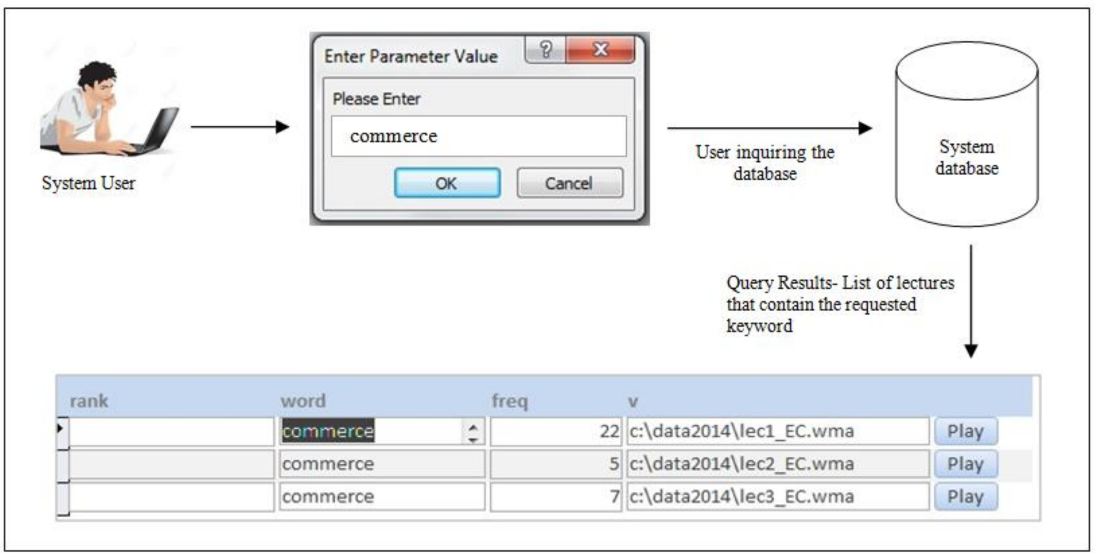


Fig. 4. User inquiring the system

The proposed system as shown in Fig. 3 captures the recorded video and audio files from the instructor. The audio file is then passed to a speech recognition program that will convert it to a word file. Afterwards, a word count program is used to count the frequency of each word in that file so that when a user wants to search for some keywords he/she will be able to do so.

The speech recognition software that was used in our proposed system is Dragon NatuarallySpeaking 12.5. It was used to convert the audio lecture to a word document. The word file was then passed to Hermetic word Frequency Counter 13.223 software to create the word frequency excel file. The database of the system was created using Microsoft Access. Now when a user (student, instructor, etc.) wants to retrieve a lecture that has the keyword 'commerce' for instance (see Fig. 4), it will retrieve all lectures that contain this keyword according to its number of occurrences (frequency) and then the user decides which lecture he/she wants to watch or listen to.

2) System design

Several tables have been designed to store data according to their type, (for instance digital, analog, history, etc..) Each table has a unique primary key, and is interlinked with the rest of the tables to facilitate the inquiry and information retrieval processes later on.

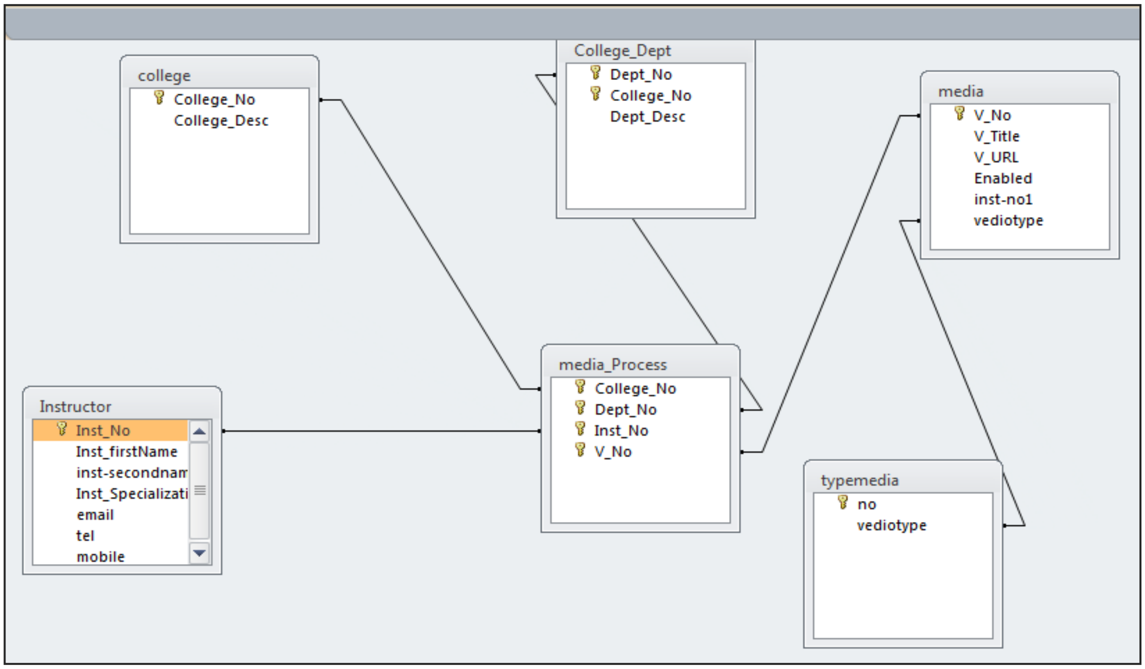


Fig. 5. The tables of the system using Microsoft Access

3) Establish relationships between tables

Relationships between tables are created upon providing primary and foreign keys. In this fashion, the main table and its sub tables will be identified easily and more efficiently done quickly on one hand, and it will keep the information under control on the other hand. Fig. 5 illustrates the relationships between tables.

4) System's interface and windows

Below is a brief description of the interface used in the system:

a) Main interface of the system:

It includes the list of lecturers/instructors, departments, colleges, and courses of each department, with learning materials according to its type; audio, video, and text-based materials (see Fig. 6).

a) Faculty interface:

The interface allows adding, omitting and amending the list of faculty members and provides a bio on every instructor (as shown in Fig. 7).

C. Advantages and limitations of using the proposed system

In this section we will be discussing the advantages and limitations that were identified by the researchers during the development of this system. The advantages are:

- The programming model is composed of a database to store various information and knowledge as per university department.
- Streamlined informative and user-friendly interface to offer users fast and easy access of to the knowledge they are looking for.
- The module saves time, effort and costs.

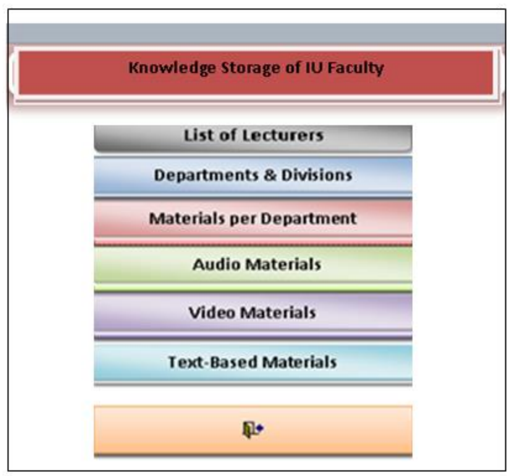


Fig. 6. The main interface

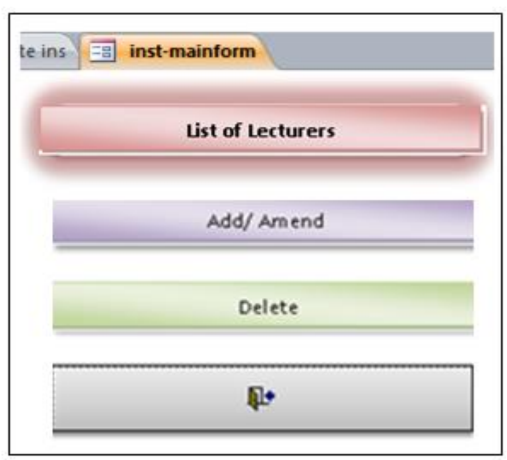


Fig. 7. Lecturers interface

- It offers quick and continuous updates. More so, password changes could be done from time to time.
- The model is adjustable, as future amendments can be easily applied.
- Elements can be added and/or deleted and authorities can be revoked and/or changed.

Whereas limitations are:

- Personal resistance from instructors and professors towards recording their lectures.
- The accuracy rate of speech recognition software plays a major role in the success of this database.
- Speech recognition software is user dependent; it means that each user of this software has to train the program on the way he/she talks and pronounce the words.
- Speech recognition software recognize only one language at a time (i.e. English in our case) while many of the lectures held in Isra' university are mixed (both languages Arabic and English are used at the same time) which makes it hard for the speech recognition program to accurately recognizes every word said by the instructor at that lecture. In this case, instructors have to go back to the word file generated by the speech recognition program and correct the words that were wrongly recognized.

V. Conclusions and Future Work

Transferring and sharing knowledge is one of the most important stages of knowledge management in any institution or organization. The main objective of the study was to provide a proposed solution to what our Arab universities suffer from in terms of exchanging and sharing information between instructors and students and between instructors themselves. This study was able to achieve this objective through:

1) Presenting a theoretical framework about the definition of knowledge, its types and importance considering that it is part of the intellectual property of organizations.

2) Analyzing the situation on the field and study how knowledge flows currently in Isra University. Study the ability to computerize and store this knowledge in a database to build a knowledge memory approved by the university.

3) Building a proposed knowledge database that holds the lecturers' notes in all its formats whether texts, presentations, videos or audio recordings of lectures.

The proposed system was tested by the three researchers including recording lectures, storing them into the system and retrieving them. The system fulfilled all the users' requirements and the user-interface was easy to interact with. The MIS department at Isra University will start using the system on Oct. 2015 so that researchers are able to test the system over a whole academic year. After that, the system will be evaluated by the three researchers at the end of the 2015/2016 academic year and later on results will be published.

References

[1] S. E. Stumpf and J. Fieser, " Philosophy: History and ," 8th ed., McGraw-Hill Humanities/Social Sciences/Languages, 2011.

[2] R. D. Hanna, "The use of information and communication technology in the flow of knowledge applications: A software model for a number of Mosul University magazines," unpublished MA thesis, Administration and Economy College, Mousil University, Iraq, 2011.

[3] I. Litvaj and D. Stancekova, "Decision - Making, and Their Relation to The Knowledge Management , Use of Knowledge Management in Decision - Making," Procedia Economics and Finance, Vol.23, pp. 467 – 472, 2015.

[4] M. Ibrahim, A. O. Yusuf and A. R. Dahlan, "Human Component of Knowledge Management for Innovation in Institutions of Higher Learning: A Literature Review," International Journal of Computer Science and Information Technology Research, Vol. 3, No. 2, pp. 562-568, 2015.

[5] R. L. Daft, "Organization Theory & Design," 9th ed., South Western College Publishing, USA, 2006.

[6] E. Bolisani and E. Scarso, "Information technology management: a knowledge-based perspective," Technovation, Vol. 19, No. 4, pp. 209-217, 1999.

[7] B.A. Lundvall and B. Johnson, "The learning economy," Journal of Industry Studies, Vol. 1, No. 2, pp. 23-42, December 1994.

[8] R. L. Drew and C. K. Thompson, "Model-based semantic treatment for naming deficits in aphasia," Journal of Speech, Language, and Hearing Research, Vol. 42, pp. 972-989, 1999.

[9] J. Kingston and A. Macintosh, "Knowledge Management Through Multi-Perspective Modelling: Representing and Distributing Organisational Memory," In: B.R. Gaines et al.(eds.), 12th Workshop on Knowledge AcquisitionModeling and Management (KAW'99), Knowledge Science Institute, University of Calgary, 1999.

[10] I. M. Ajam, "Information technology and knowledge management impacts on the strategic choice," unpublished PhD thesis, Administration and Economy College, Al-Mustansiriya University, Iraq, 2007.

[11] M. Polanyi, "The Tacit Dimension, London," University of Chicago Press, 1966.

[12] I. Nonaka and H. Takeuchi, "The Knowledge Creating Company :How Japanese Companies Create the Dynamics of Innovation," Oxford University Press, New York, 1995.

[13] I. Nonaka and N. Konno, "The Concept of 'Ba': Building a foundation for knowledge creation," California Management Review, Vol. 40, No. 3, pp. 43, 1998.

[14] A. M. Uthman, "Attitudes of public secondary schools' principals towards the application of knowledge management in the northern provinces in Palestine," unpublished MA thesis, Postgraduate Studies College, An-Najah National University, Nablus, Palestine, 2010.

[15] S. Al-Kubaysi and S. Al-Mihyawi, "Knowledge Management," Arab Administrative Development Organization, 1st ed., Cairo, Egypt, 2005.

[16] A. Al-Ali, G. Al-Omari and A. Qandelji, "Introduction to Knowledge Management," Al-Masira Publishing and Distribution House, Amman, Jordan, 2006.

[17] I. B. Fernandez and R. Sabherwal, "Knowledge management systems and processes," M.E. Sharpe, Inc. , 2010.

[18] S. Al-Sha'bani, "The Impact of the Development of Intellectual Capital on Technological Proficiency and its Reflections on the Cost Reduction: Application to the National Company for Furnishing Industries in Nineveh," AL-Anbar University journal of Economic and Administration Sciences, Vol. 4, No. 7, pp. 378-402, 2011.

[19] H. H. Al-Khaffaf, "Accounting treatments to measurement and disclosure problems of intellectual capital (case study)," unpublished MA thesis in accounting, Administration and Economy College, Mousil University, Iraq, 2006.

[20] M. Malone, "New metrics for a new age," Forbes Magazine, pp. 40-41, April 1997.

[21] T. H. Davenport and L. Prusak, "Working Knowledge: How Organization Manage What the Know," Harvard Business School Press, Boston, MA, USA, 1997.

[22] M. Hansen, N. Nohria and T. Tierney, "What's your Strategy for Managing Knowledge," Harvard business Review, Vol. 77, No. 2, pp. 106-116, 1999.

[23] L. Edvinsson and P. Sullivan, "Developing Model for Managing Intellectual Capital," European Management Journal, Vol. 14, No. 4, pp. 356–364, 1996.

[24] T. A. Stewart, "The Wealth of Knowledge: Intellectual Capital and the Twenty-first Century Organization," Nicholas Brealey Publishing, 2002.

[25] A. A. Al-Daqqaf, "A preface to cost management in industrial plants in order to support their competitiveness," unpublished PhD thesis in accounting philosophy, College of Commerce in Al-Isma'iliyah, Sues Canal University, Egypt, 2004.

[26] G. Al-Omari, "The role of information technology and knowledge in building the organizational memory to improve problem-solving," a research presented to the 2nd Scientific Conference, Israa University, Business and Finance Sciences College, Jordan, 2007.

A Comparative Study of Relational and Non-Relational Database Models in a Web- Based Application

Cornelia Győrödi
Department of Computer Science and Information Technology, University of Oradea
Oradea, Romania

Robert Győrödi
Department of Computer Science and Information Technology, University of Oradea
Oradea, Romania

Roxana Sotoc
Department of Computer Science and Information Technology, University of Oradea
Oradea, Romania

***Abstract*—The purpose of this paper is to present a comparative study between relational and non-relational database models in a web-based application, by executing various operations on both relational and on non-relational databases thus highlighting the results obtained during performance comparison tests. The study was based on the implementation of a web-based application for population records. For the non-relational database, we used MongoDB and for the relational database, we used MSSQL 2014. We will also present the advantages of using a non-relational database compared to a relational database integrated in a web-based application, which needs to manipulate a big amount of data.**

***Keywords*—MongoDB; MSSQL; NoSQL; non-relational database**

I. Introduction

As technology nowadays is tireless and evolves more and more every day, the amount of data is increasing and an application to handle a huge volume of data efficiently it is important to choose the right model of the database. Relational database model has a quite rigid schema that means that a schema must be designed in advance before data had been loaded and all attributes of the schema are uniform for all elements, in the case of missing values null values are used instead [5]. It is difficult to change the schema of databases, especially when, it is a partitioned relational database that spreads across multiple servers. If our data capture and management requirements are constantly evolving, a rigid schema can quickly become an obstacle to change [6].

Generally a web application must support millions of users simultaneously and to handle a huge volume of data a relational database model is still widely in use today, even though it has serious limitations when to handle a huge volume of data.

Google, Amazon, Facebook and LinkedIn have been among the first companies that discovered those limitations of the relational database model as far as the demands of new applications. These limitations have led to the development of non-relational databases, also commonly known as NoSQL (Not Only SQL) [7].

Non-relational databases do not use the RDBMS principles (Relational Data Base Management System) and do not store data in tables, and have schema-less approach to data management. Non-relational databases do not require schema definition before inserting data nor changing the schema when data collection and management need evolve [6] [10]. Instead, they use identification keys and data can be found based on the keys assigned [8].

NoSQL could be categorized in 4 types [2]:

1) Key-Value databases – which are the simplest NoSQL data stores to use, from an API perspective. The most popular ones are Redis, Riak, etc.

2) Document databases – which store and retrieve documents as XML, JSON, BSON and so on. The most popular document database is MongoDB, which provides a rich query language.

3) Column family stores – these databases store data in column families as rows that have many columns associated with a row key. One of the most popular is Cassandra.

4) Graph Databases – allows you to store entities and relationships between these entities. There are many graph databases, but between this type of database we can mention OrientDB, FlockDB. etc. [2]

According to the article was written by Matt Asay, “NoSQL databases eat into the relational database market” [1] the NoSQL databases, especially MongoDB, occupy more and more space on the market, but with all these Oracle and SQL Server are still constant. In Fig 1. we can see the popularity of MongoDB and its evolution from 2014 to 2015.

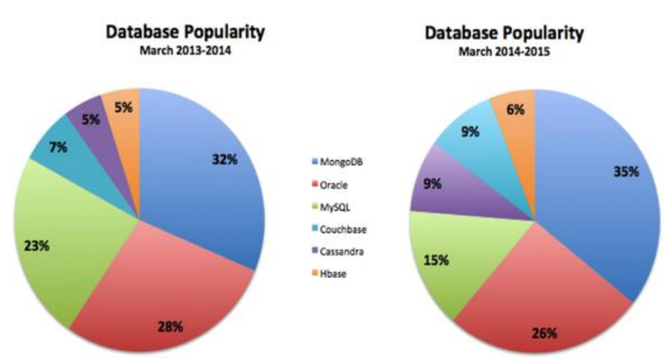


Fig. 1. Database Popularity [1]

In Fig. 1. we can see a growth of 3% for MongoDB from 2014 until 2015 and a decrease of 2% for Oracle and 8% for MySQL. This could mean that the companies are leaning toward NoSQL databases, so they can manipulate more data at a lower price [1].

A big advantage of non-relational databases is that they are more scalable and provide superior performance and their data model addresses several issues that the relational model is not designed to address like large volumes of structured, semi-structured and unstructured data, agile sprints, quick iteration, frequent code pushes, object-oriented programming, efficiency, monolithic architecture and so on [3] [9].

In this paper, we focus on one of the non-relational databases, namely MongoDB, and we make a comparison with one of the relational databases, namely MSSQL to highlight advantages and disadvantages of the two models. The study was based on the implementation of a web-based application for population records, which needs to manipulate a big amount of data.

II. Application Development Using MongoDB Vs. MSSQL

We created a comparative study between relational databases, namely MSSQL, and non-relational databases, namely MongoDB. The study has based on the implementation of a website for population records, which needs to manipulate a big amount of data. To highlight the advantages of using the non-relational database MongoDB compared to the relational database MSSQL, various operations were performed on the two databases. These operations are the four basic operations that can be performed on any database, namely: SELECT, INSERT, UPDATE and DELETE. These operations were made on 1, 100, 500, 1.000, 5.000, 10.000, 25.000 and 50.000 records. The application has been developed, in ASP.NET MVC 4 with C# programming language and we implemented the same methods for both MSSQL and MongoDB databases. For non-relational databases, we used the MongoDB C# Driver that is the officially supported C# driver for MongoDB. The version of the driver is 2.0 and the version of MongoDB is 3.0.

To be able to connect to MongoDB we used a MongoClient. In code, we used two namespaces that are specific to MongoDB: MongoDB.Driver and MongoDB.Bson. Finally we added the connection string in web.config file, like this:

```
<connectionStrings>
<add name = "MongoDB" connectionString = "
mongodb://localhost:27017/"/>
</connectionStrings>
```

We used this code connection string through this code:

```
    public static string connString =
System.Configuration.ConfigurationManager.Conn
ectionStrings["MongoDB"].ConnectionString;
```

After retrieving the correct connection string we used it in methods for calling the collections that we need like the one bellow:

```
public static IMongoCollection<BsonDocument>
ConnectToServer()
{
var client = new MongoClient(connString);
var db = client.GetDatabase("Dizertatie");
   IMongoCollection<BsonDocument> collection =
   db.GetCollection<BsonDocument>("People");
return collection;
}
```

In the method above, we call the database called “Dizertatie” and we get the collection called “People” and the results that will be returned will be a IMongoCollection of BsonDocument type which is a specific format for MongoDB.

Next, we created asynchronous methods to work with the data from MongoDB. For example, for deleting all the registered people from the collection we created a method like this:

```
public static async Task<DeleteResult>
DeleteAllConsumers()
{
       var collection = ConnectToServer();
       var filter = new BsonDocument("_id",
       new BsonDocument("$exists", true));

       var result = await
       collection.DeleteManyAsync(filter);
       return result;
}
```

The MongoDB method that we call for deleting the registered people, DeleteManyAsync, will delete multiple documents inside the collection that we are connected to. The number of the deleted documents depends on the filter that we need to create and provide when we want to call the method. In our case, we will delete all registers from this collection.

In the following section we will mainly focus on the performance results for both databases, MongoDB and MSSSQL, that we obtained after we executed operations of SELECT, INSERT, UPDATE and DELETE.

III. Comparative Study: MongoDB Vs MSSQL

Generally, depending on the scope of the application that we want to develop, when the project is in the planning stage, each company will establish the resources and the limits for the project. We also need to choose the database for the application that will be developed. Here, we should consider the amount of data that will be manipulated, the rapidity that the project needs and the budget.

Considering these factors, for an application that will need to store a large amount of data, if the application have to handle a huge volume of data, then we should think how we could achieve this in an efficient way or how many resources should be allocated for this scope. For example, in the application that we created, we took in consideration the fact that we need a huge space for storing the data and the queries that will be made every day like adding new data, deleting, updating and so on. All these queries are expensive and we should also, think about the rapidity at which they are processed. Another important fact that we need to consider is the number of users that will access the application, like the employees from all the country, plus the usual users that will need different information.

Considering all these facts, it is important to analyse the performance of the application in terms of insertion, update, deletion and selection operations.

The performance of the database that we have chosen can be a very important fact because of the storage space, all the hardware and other components that we need.

We began testing with the creation of databases both MongoDB and MSSQL after that we executed operations of INSERT, UPDATE, DELETE and SELECT on both databases. All these operations have been made on 1, 100, 500, 1.000, 5.000, 10.000, 25.000 and 50.000 records.

In MongoDB, for inserting a list of people into the database we write the following code:

```
public static async Task<string>
InsertPeople(List<BsonDocument> pplList)
{
   var collection = ConnectToServer();
   var result = await
collection.InsertManyAsync(pplList)
.ContinueWith(x => x.ToJson());
   return result;
}
```

And in MSSQL we write the following code:

```
using (SqlConnection sqlConn = new
SqlConnection(connStr))
{
 sqlConn.Open();
 using (SqlCommand sqlCmd = new SqlCommand())
{
sqlCmd.Connection = sqlConn;
sqlCmd.CommandType =
CommandType.StoredProcedure;
sqlCmd.CommandText =[dbo].[Set_people]";
sqlCmd.Parameters.AddWithValue("@peopleTbl",
data); sqlCmd.Parameters.Add("@isInsert",
SqlDbType.Bit).Value = isInsert;
int result = sqlCmd.ExecuteNonQuery();}
sqlConn.Close();}
}
```

In MongoDB the following method will insert one record:

```
public static async Task<string> AddUser(User
model) {
       var collection = ConnectToServer();
       var result = await
       collection.InsertOneAsync(model)
       .ContinueWith(x => x.ToJson());
       return result;
       }
```

After we executed these operations, we obtained the following results that are shown in Table 1.

The graphical representation of the results from Table 1 is shown in Figure 2. We notice that until 1.000 records, we obtained a maximum difference of 200 milliseconds, but we can see that the difference is more significant after 1.000 records and for 50.000 records we obtained a difference of 7 seconds and half.

TABLE I. THE RESULTS OF THE INSERT OPERATION

Insert	MongoDB - sec	SQL - sec
1 user	00:00:00:003	00:00:00:402
100 users	00:00:00:005	00:00:00:096
500 users	00:00:00:018	00:00:00:183
1.000 users	00:00:00:033	00:00:00:387
5.000 users	00:00:00:162	00:00:00:736
10.000 users	00:00:00:521	00:00:01:085
25.000 users	00:00:00:816	00:00:03:378
50.000 users	00:00:01:835	00:00:08:306

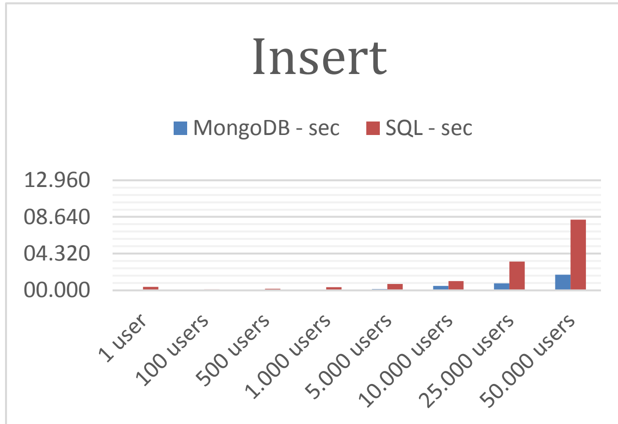


Fig. 2. MSSQL vs MongoDB insert

In MongoDB to select records we write the following code:

```
List<People> pplList = new List<People>();
var collection = ConnectToServer();
var filter = new BsonDocument("_id", new
BsonDocument("$exists", true));
using (var cursor = await
collection.FindAsync(filter)) {
while (await cursor.MoveNextAsync()){
var batch = cursor.Current;
foreach (var document in batch) {
People ppl = new People();
ppl.ID =
document["_id"].AsObjectId.ToString();
ppl.CNP = document["Cnp"].AsString;
ppl.Nume = document["Nume"].AsString;
ppl.Prenume = document["Prenume"].AsString;
ppl.Varsta =
document["Varsta"].AsInt32.ToString();
ppl.Sex = document["Sex"].AsString;
ppl.Adresa = document["Adresa"].AsString;
ppl.Ocupatie =
document["Ocupatie"].AsString;
ppl.AdresaOcupatie =
document["AdresaOcupatie"].AsString;
ppl.Telefon = document["Telefon"].AsString;
ppl.StareCivila =
document["StareCivila"].AsString;
ppl.OrasLocalitate =
```

```
document["OrasLocalitate"].AsString;
ppl.Judet = document["Judet"].AsString;
pplList.Add(ppl); } }}
```

In MSSQL to select records we write the following code:

```
DataTable dt = new DataTable();
try{
using (SqlConnection sqlConn = new
SqlConnection(connStr)) {
sqlConn.Open();
using (SqlCommand sqlCmd = new
SqlCommand()){
sqlCmd.Connection = sqlConn;
sqlCmd.CommandType =
CommandType.StoredProcedure;
sqlCmd.CommandText = "[dbo].[Get_People]";
SqlDataAdapter da = new
SqlDataAdapter(sqlCmd);
da.Fill(dt); }
sqlConn.Close();}}
catch (Exception) { }
return dt;
```

After we executed these operations for select we obtained the following results that are shown in Table 2:

TABLE II. THE RESULTS OF THE SELECT OPERATION

Select	MongoDB - sec	SQL - sec
1 user	00:00:00:003	00:00:00:083
100 users	00:00:00:004	00:00:00:002
500 users	00:00:00:017	00:00:00:005
1.000 users	00:00:00:031	00:00:00:006
5.000 users	00:00:00:206	00:00:00:028
10.000 users	00:00:00:291	00:00:00:052
25.000 users	00:00:00:830	00:00:00:190
50.000 users	00:00:01:616	00:00:00:327

The graphical representation of the results from Table 2 are shown in Figure 3.

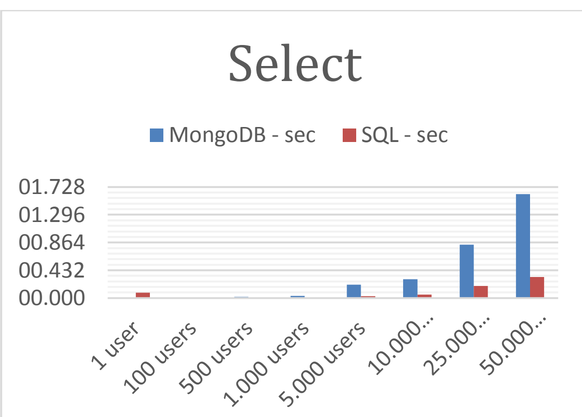


Fig. 3. MSSQL vs MongoDB select

From figure Fig.3 we notice that the select operation is more fast and efficient in MSSQL than in MongoDB. We also see that at 1.000 records, it is a difference 26 milliseconds and this grows up to 1.3 seconds when we select 50.000 records.

To update the records in MongoDB we write the following code:

```
var collection = ConnectToServer();
var filter = new BsonDocument("_id", new
BsonDocument("$exists", true));
var update = Builders<BsonDocument>.Update
    .Set("Nume", "Updated")
    .Set("Adresa", "Updated");
var result = await
collection.UpdateManyAsync(filter, update);
```

The methods that we created for MSSQL to insert a new record or a list of records will be used for UPDATE operation too, but the parameter *@isInsert* will be equal with false in that case. The results of the update operation are shown in Table 3.

TABLE III. THE RESULTS OF THE UPDATE OPERATION

Update	MongoDB - sec	SQL - sec
1 user	00:00:00:005	00:00:00:039
100 users	00:00:00:007	00:00:00:048
500 users	00:00:00:059	00:00:00:059
1.000 users	00:00:00:042	00:00:00:159
5.000 users	00:00:00:245	00:00:02:219
10.000 users	00:00:00:463	00:00:04:634
25.000 users	00:00:01:294	00:00:19:946
50.000 users	00:00:02:224	00:00:31:205

The graphical representation of the results from Table 3 is shown in Figure 4.

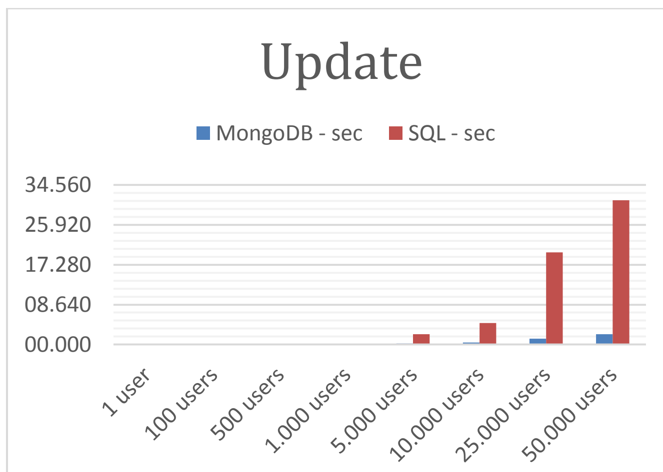


Fig. 4. MSSQL vs MongoDB update

For the update operations we can see a bigger difference from 5.000 records, which is approximatively 2 seconds and until 50.000 records it grows up to 29 seconds and gives MongoDB an advantage. We notice that MongoDB spends less time than MSSQL, for performing update operation as shown in figure Fig. 4.

To delete the records in MongoDB we write the following code:

```
var collection = ConnectToServer();
```

```
var filter = new BsonDocument("_id", new
BsonDocument("$exists", true));
var result = await
collection.DeleteManyAsync(filter);
return result;
```

And in MSSQL to delete records we write the following code:

```
using(SqlConnection sqlConn = new
SqlConnection(connStr)) {
sqlConn.Open();
using(SqlCommand sqlCmd = new SqlCommand()){
sqlCmd.Connection = sqlConn;
sqlCmd.CommandType =
CommandType.StoredProcedure;
sqlCmd.CommandText =
"[dbo].[Del_All_Consumers]";
int result = sqlCmd.ExecuteNonQuery();
success =
!string.IsNullOrEmpty(result.ToString()) ?
true : false; }
sqlConn.Close(); }
```

The results of the delete operation are shown in Table 4.

TABLE IV. THE RESULTS OF THE DELETE OPERATION

Delete	MongoDB - sec	SQL - sec
1 user	00:00:00:004	00:00:00:081
100 users	00:00:00:003	00:00:00:019
500 users	00:00:00:007	00:00:00:063
1.000 users	00:00:00:017	00:00:00:082
5.000 users	00:00:00:053	00:00:00:143
10.000 users	00:00:00:106	00:00:00:200
25.000 users	00:00:00:317	00:00:00:350
50.000 users	00:00:01:508	00:00:01:787

The graphical representation of the results from Table 4 is shown in Figure 5.

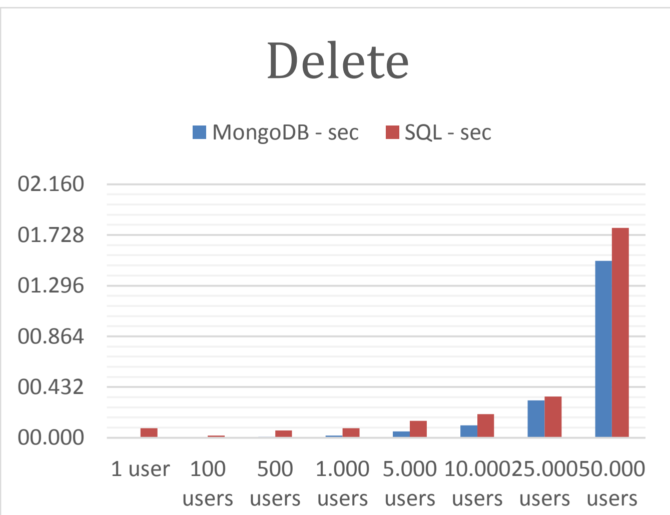


Fig. 5. MSSQL vs MongoDB delete

From figure Fig. 5 we notice that MongoDB provided lower execution times than MySQL in delete operations, especially when the number of records increases, which is essential when an application should provide support to thousands of users simultaneously.

IV. CONCLUSIONS

In this paper, we showed the results of different operations that had been applied for MongoDB and MSSQL databases. MongoDB provided lower execution times than MSSQL in INSERT, UPDATE and DELETE operations, which is essential when an application should provide support to thousands of users simultaneously. The only time when the MSSQL obtained an advantage was with the SELECT operations, the other ones gave advantages to MongoDB.

We can also notice that the difference between the results of each database was not noticeable until around 1.000 records. Thus, we can say that relational databases, namely MSSQL is suitable for small and medium applications. Relational databases are widely used in most of the applications and they have good performance when they handle a limited amount of data.

We need to be carefull when we want to choose a model of the database for the application that we will want to create. We should take into consideration main factors as the amount of data, the flexibility of schema, the budget, the amount of transactions that would be made and how frequent they are called. These days, companies, depending on the application that they want to develop, have the possibility to choose the most suitable database from a wide range of databases. Generally, for smaller and medium applications, a relational database would be chosen and for big applications, that use and manipulate large quantities of data, a non-relational database will be chosen. Of course, these are not the only criteria for choosing a database, but it depends on each company and the purpose of the application that would need to be developed.

Considering that, in our days, the request for storing a big volume of data at a low price is bigger every day, we tend to choose a non-relational database. Also, MSSQL being commercial at a pretty big price and MongoDB being an open source solution, it is a big disadvantage for MSSQL.

In the end, for choosing the correct database that can satisfy all the needs that an application demands in order to have been developed, all the things presented above should be taken into consideration before the start of development of the project. Depending on what each application needs, we can choose the most suitable database, a non-relational or a relational database.

REFERENCES

[1] Matt Asay, "NoSQL databases eat into the relational database market", Published March 4, 2015, Available: http://www.techrepublic.com/article/nosql-databases-eat-into-the-relational-database-market/ , accessed july 2015.

[2] Pramod Sadalage,"NoSQL Databases: An Overview", Published Octomber 1, 2014, Available: http://www.thoughtworks.com/insights/blog/nosql-databases-overview, accessed july 2015

[3] S. Hoberman, "Data Modeling for MongoDB", Publisher by Technics Publications, LLC 2 Lindsley Road Basking Ridge, NJ 07920, USA, ISBN 978-1-935504-70-2, 2014.

[4] Michael Kennedy, "MongoDB vs. SQL Server 2008 – Performance Shutdown", Published April 29, 2010, Available: http://blog.michaelckennedy.net/2010/04/29/mongodb-vs-sql-server-2008-performance-showdown/, accessed july 2015

[5] R. D. Bulos, J. Bonsol, R. Diaz, A. Lazaro, V. Serra,"Comparative analysis of relational and non-relational database models for simple queries in a web-based application", Research Congress 2013, de la Salle University Manila, march 7-9, 2013.

[6] C. Bazar, C. Iosif, "The transition from RDBMS to NoSQL. A comparative analysis of three popular non-relational solutions: Cassandra, MongoDB and Couchbase", Database Systems Journal, vol. V no 2/2014, pp.49-59.

[7] N. Jatana, S. Puri, M. Ahuja, I. Kathuria, D. Gosain, "A survey and comparison of relational and non-relational databases", International Journal of Engineering Research &Technology (IJERT) ISSN: 2278-0181, vol 1 Issue 6, august 2012, pp. 1-5.

[8] Cornelia Győrödi, Robert Győrödi, George Pecherle, Andrada Olah, "A comparative study: MongoDB vs. MySQL", IEEE - 13th International Conference on Engineering of Modern Electric Systems (EMES), 2015, Oradea, Romania, 11-12 June 2015, ISBN 978-1-4799-7649-2, pag. 1-6.

[9] MongoDB, NoSQL Database Explained, Available: https://www.mongodb.com/nosql-explained , accessed july 2015.

[10] R. P Padhy, M. R. Patra, S. C. Satapathy, "RDBMS to NoSQL: Reviewing Some Next-Generation Non-Relational Database's", International Journal of Advance Engineering Sciences and Technologies, Vol. 11, Issue No. 1, 015-030, 2011.

Role of Secondary Attributes to Boost the Prediction Accuracy of Students' Employability Via Data Mining

Pooja Thakar, Research Scholar
Department of Computer Science
Banasthali University
Jaipur, Rajasthan, India

Prof. Dr. Anil Mehta, Professor
Department of Management
University of Rajasthan
Jaipur, Rajasthan, India

Dr. Manisha, Associate Professor
Department of Computer Science
Banasthali University
Jaipur, Rajasthan, India

***Abstract*—Data Mining is best-known for its analytical and prediction capabilities. It is used in several areas such as fraud detection, predicting client behavior, money market behavior, bankruptcy prediction. It can also help in establishing an educational ecosystem, which discovers useful knowledge, and assist educators to take proactive decisions to boost student performance and employability.**

This paper presents an empirical study that compares varied classification algorithms on two datasets of MCA (Masters in Computer Applications) students collected from various affiliated colleges of a reputed state university in India. One dataset includes only primary attributes, whereas other dataset is feeded with secondary psychometric attributes in it. The results showcase that solely primary academic attributes don't lead to smart prediction accuracy of students' employability, once they square measure within the initial year of their education. The study analyzes and stresses the role of secondary psychometric attributes for better prediction accuracy and analysis of students' performance. Timely prediction and analysis of students' performance can help Management, Teachers and Students to work on their gray areas for better results and employment opportunities.

Keywords—Data Mining; Education; Prediction; Psychometric; Educational Data Mining

I. INTRODUCTION

Every year sizable amount of graduates and postgraduates from numerous professional institutes competes in the job market for good employment opportunities. Conversely, the world economy isn't generating enough employment for this young unemployed populace. To boost the possibility of obtaining the right job that matches the qualification and training of these students, institutes not only ought to add to their academic qualification but should equip them with essential employability skills.

Employability skills are necessary across all areas and kinds of jobs. These skills don't seem to be only important to employees, but employers also seek them in candidates before recruitment. Thus, education system should promote a course of study that emphasizes and nourishes the development of employability skills along with fundamental scholastic skills.

An educational institute contains a large number of student records that remains unutilized. This data can be utilized for the betterment of students if analyzed well. The data can be mined and pruned to guide the students at the right time. Customized guidance to every student in the very first year of their professional college education can give them a rational chance to improve themselves and fetch better employment.

In the higher education system, MCA (Masters in Computer Applications) is a professional degree course that provides theoretical and practical knowledge of Computers and its Applications and makes students ready for IT (Information Technology) and ITeS (Information Technology Enabled Services) Companies. In the last year of the course, every student aspire and compete to obtain a decent job before they pass out. The prediction of students' employability factor and timely steerage by educators can greatly facilitate in rising students efforts in right direction. At the same time better placed and high performer students will bring laurels to the institute reciprocally.

A number of researches have been done to predict students performance, but most of them include only primary academic attributes for prediction purposes. Whereas, many researches have evidently shown positive association between non-academic factors and employability factors. The present study analyzes the role of various factors in improving the prediction accuracy of employability. We applied various data mining classification techniques on student datasets; first with only primary attributes and then once more by adding secondary psychometric attributes to it. Comparative analysis is done by applying classification algorithms in Weka Tool to point out the impact of secondary psychometric attributes on prediction accuracy.

Further, this paper is structured as follows: Section II presents literature review in educational data mining for employability factors prediction and analysis. Section III describes data mining classification techniques used in the study Section IV presents the prediction process of data mining Section V demonstrates and analyzes the results in the form of comparative table and charts. Section VI concludes with an outline and a view on future work.

II. LITERATURE REVIEW

Data mining has spread its wings in the sector of education very well and lots of work have been done to explore the correlation among attributes, predicting academic performance, finding best mining technique for performance monitoring.

In year 2014, Emerald Group Publishing Limited published a paper stating that emotional intelligence, self management, work and life experiences are necessary factors for Employability Development Profile [1]. Another paper published within the same publishing house and described that employability is joined with competences and tendencies [2]. Cairns, Gueni, Fhima, David and Khelifa analyzed employees' profiles and found positive correlation in employees' jobs, assignments and history [3]. Potgieter & Coetzee, revealed a number of important relationships between personality and employability [4]. David, Hamilton, Riley and Mark disclosed that highest weight is given to soft skills by employers [5]. Denise Jackson and Elaine Chapman in 2012 steered prominent skill gap between professional institutes and corporate [6]. V. K. Gokuladas in his first paper reflected that graduates ought to possess special skills beyond basic academic education [7]. In his next paper he showcased that GPA and proficiency in English language as important factors for employability [8]. Bangsuk Jantawan and Cheng-Fa Tsai, designed a model for prediction of the employees' performance using data mining techniques [9].

Bhardwaj and Pal applied Bayesian classification and found that factors like living location, medium of teaching, mother's qualification, family income and status are highly correlated with academic performance of students [10]. Tongshan Chang, & Ed. D experimented and provided evidences that data mining is an effective technology for college recruitment [11]. Hijazi and Naqvi conjointly found that the factors like mother's education and family income are highly correlated with the student's academic performance [12]. Khan implemented clustering and found that girls with high socio-economic background perform better in science stream and boys with low socio-economic background are usually better academic achiever [13]. Z. J. Kovacic applied CHAID and CART techniques on students enrolment data and presented two decision trees, which classified successful and unsuccessful students. The accuracy obtained with these techniques was only 59.4 % and 60.5 % respectively [14]. Al-Radaideh, et al predicted the final grade of students using ID3, C4.5, and the Naïve Bayes algorithms and found that Decision Tree provide better prediction than any other model [15]. Sudheep Elayidom , Sumam Mary Idikkula & Joseph Alexander applied data mining to assist students in selecting an appropriate branch as per personal skill set for better placement later [16].

These studies reveal great potential of data mining in education sector. More work is needed to establish it as a customized guiding tool for students. With continued research, it'll be ready to support student community very well in near future.

III. Data Mining and Classification

Data mining is often divided into predictive and descriptive techniques. Predictive data mining analyses the data and help in building models, which tries to predict the behaviour of novice instance. The technique of classification belongs to this group and is widely used for predictive modeling. Classification is a supervised learning approach in which data is classified into known classes. Classification rules are identified from training data and are tested for the remainder of data.

The classification accuracy of an algorithm is majorly dependent on the type of dataset rather than chosen algorithm itself. The most important characteristics of a dataset are its predictors, number of classes and number of instances [17]. The used datasets of MCA students do not have much impact of classes, because it has only two output classes namely "employed" and "unemployed". It's binary in nature, therefore works well for most of the algorithms. The number of instances have very little role on the accuracy of classification, the quality of instances matter more [17]. The quality of attributes will offer more information and is a vital factor [17].

Information gain can offer a very good knowledge about the quality of attributes. To find out the quality of attributes in present data sets, information gain is calculated in WEKA Tool for all the attributes. Thereafter, they are stratified from best (Rank 1) to worst (Rank 29) as per the information gain received. The ranks of all attributes present in two datasets are depicted in Fig. 1.

Fig. 1 clearly shows that secondary attributes rank much superior as compared to primary attributes. The attributes like Permanent Address, Attention to Detail, Logical Ability and Score in English are more significant than percentages earned at senior secondary and graduation level. It signifies that only primary academic attributes might not be enough to achieve higher classification accuracy.

To investigate further, two datasets with the same set of instances (214 instances) are taken for comparative analysis between primary and secondary attributes. One dataset contains only primary academic attributes (12) such as percentages earned at secondary, senior secondary and graduation levels, whereas, second dataset includes both primary and secondary psychometric attributes (12 primary and 17 secondary). They're described in Table I.

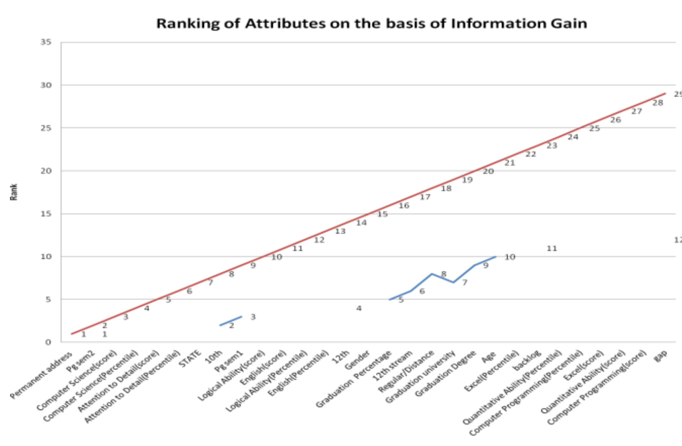


Fig. 1. Graph of attributes ranked as per Information Gain

TABLE I. LIST OF PRIMARY AND SECONDARY ATTRIBUTES

Attribute	Values
Primary Attributes	
Age	{<25,>=25}
Secondary Percentage	{A,B,C,D,O}
Senior Secondary Percentage	{A,B,C,D,O}
Stream in Senior Secondary	{Commerce,Science,'Vocational…}
Graduation Degree	{CS,'NON CS'}
Graduation %	{A,B,C,D,O}
Type (Regular/Distance)	{Regular,Distance}
Graduating University	{State,Central,Deemed,Private}
Post Graduate Sem 1 %	{A,B,O,D,C}
Post Graduate Sem 2 %	{A,B,O,D,C}
No. of Supplimentaries in 1st Year	{Numeric}
Gap in years between Graduation and PG	{Numeric}
Secondary Attributes	
Gender	{MALE,FEMALE}
Permanent Address	{'EAST DELHI','OUTSIDE DELHI',..}
State	{DELHI,RAJASTHAN,MP,BIHAR,…}
English	{Numeric}
Quantitative Ability	{Numeric}
Logical Ability	{Numeric}
Attention to Detail	{Numeric}
Computer Programming Skills	{Numeric}
Computer Science Knowledge	{Numeric}
Psychometric Score	{Numeric}
English(P) P→Percentile	{Numeric}
Quantitative Ability(P)	{Numeric}
Logical Ability(P)	{Numeric}
Attention to Detail(P)	{Numeric}
Computer Programming(P)	{Numeric}
Computer Science(P)	{Numeric}
Psychometric (P)	{Numeric}

Thereafter, various types of classifiers are applied on these two datasets to check which sort of attribute set will herald higher classification accuracy. WEKA Tool is used to implement these classifier algorithms. Ten most widely used base classification algorithms used in educational data mining are chosen. They're Naïve Bayes from Bayes Category, RBF Network and Multilayer Perceptron from Functions Category, IB1 and IBk from Lazy Category, PART and DTNB from Rules Category, lastly J48, Random Tree and Random Forest from Trees Category.

The Classifies are described as below:

Naïve Bayes: This classifier is predicated on Bayes' theorem and believes in individual possibilities of every attribute pair. It is simple to build and easy to understand, withal provides excellent classification results.

RBFNetwork: Radial Basis Function (RBF) networks have proven to be valuable neural network. These are feed-forward networks, which are trained with supervised training algorithm. The algorithm generally trains in no time and is less susceptible to issues with non-stationary inputs.

Multilayer Perceptron: MLP algorithm is also widely used neural network algorithm. The input layer is of attributes, classes make output layers, hidden layers are interconnected through several neurons. The back propagation algorithm is applied to optimize the weights. This algorithm suits well for approximating a classification.

IB1: The algorithm uses distance measure to find the training instance closest to the given test instance. Thus predicts class that is same as training instance. If multiple instances are same, the first one found is used.

IBk: This is K-nearest neighbours, an instance based classifier. This can select appropriate value of K based on cross-validation and can also do distance weighting.

PART: This algorithm builds a partial C4.5 decision tree with every iteration and makes the "best" leaf into a rule. It uses separate-and-conquer methodology.

DTNB: This is decision table/naive bayes hybrid classifier. The algorithm uses forward selection search. At every step, selected attributes are modeled by naive Bayes. The rest are modeled by decision table. The attribute may be dropped entirely.

J48: J48 decision tree is an implementation of C4.5 algorithm. The tree is structured by training instances and is compatible for dataset with few samples. It doesn't overfit on given dataset.

Random Tree: In this algorithm, K randomly chosen attributes are taken to construct a tree. It doesn't perform any pruning.

Random Forest: This algorithm is used for constructing a forest of random trees.

Robustness of the classifier is usually calculated by applying cross validation on the classifier. During this study, 10-fold cross validation is employed, that split the data set randomly into 10 subsets of equal size. Nine subsets are used as training set and one subset is used as test set. This procedure is performed 10 times to incorporate every subset for test once.

IV. PREDICTION WITH DATA MINING

Students' employability prediction can facilitate students, academician and management to take proactive actions. This can improve the success percentage of students to get employed in excellent companies. The present study shows that employability can be predicted well, if secondary psychometric parameters are also taken into consideration. Prediction models that embody personal, social, psychological and other environmental variables show better results as compared to models considering only academic parameters.

A. Data Collection

The data set used in this study is obtained from the affiliated colleges of a reputed State University in Delhi, India. Colleges offering three years MCA Degree Course were contacted and complete details of 214 students for the session 2012-2015 were collected.

B. Data Selection and Transformation

The obtained data set is then divided into two data sets, one with only primary academic attributes (12 attributes) and another with primary and secondary attributes (29 attributes) with the same number of instances (214 instances). Before proceeding for mining, the irrelevant attributes such as name, phone number were removed. Some derived variables like age was added. The data within attributes were also made meaningful by converting them into categories such as marks converted to grades, addresses converted to regions and states. All the predictors are described in Table1.

C. Implementation of Classification Algorithms

WEKA is an open source mining tool that implements a large collection of machine leaning algorithms. The algorithms used for classification purpose for the present study are Naïve Bayes, RBF Network, Multilayer Perceptron, IB1, IBk, PART, DTNB, J48, Random Tree and Random Forest. The 10-fold cross-validation is chosen as an estimation approach to obtain a reasonable idea of accuracy, since there's no separate test data set. This technique divide training set into 10 equal parts, 9 are applied as training set for making machine algorithm learn and 1 part is used as test set. This approach is enforced 10 times on same dataset, where every training set act as test set once.

V. Results

The performance of ten classification algorithms for predicting students' employability on two datasets (one with only primary attributes and second with both primary and secondary attributes) were experimented upon and results were calculated.

The percentage of correctly classified instances is commonly known as accuracy or sample accuracy of a model. The accuracy percentage of all the classifiers were calculated for both datasets and results are shown in Table II.

The table is plotted as a graph and is delineated as Fig. 2. It depicts the performance of classifiers with two datasets, initial dataset with only primary attributes and second dataset with both primary and secondary attributes. The graph clearly shows the improvement in performance of most of the classifiers, when second dataset is chosen. This further proves that the information gain shown by the secondary attributes earlier (Fig. 1), additionally facilitate in better performance of classifiers. Thus, we can say that secondary attributes play very important role in improving prediction accuracy of classifiers with respect to employability. Along with the results of accuracy, the training and simulation errors with the help of Kappa Statistic, Mean Absolute Error (MAE) Root Mean Squared Error (RMSE), Relative Absolute Error (RAE) and Root Relative Squared Error (RRSE) were calculated. The results of simulation are shown in Table III and Table IV. The percentage differences of RAE% and RRSE% are further plotted in a graph and are depicted in Fig. 3 and Fig. 4. The graphs clearly show that Error Percentages (RAE and RRSE) reduces significantly, when we include secondary psychometric parameters in dataset.

TABLE II. Prediction Accuracy of Classifiers

Classifier Name	Accuracy % (Only Primary Attributes)	Accuracy % (Primary and Secondary Attributes)
NaïveBayes	77.10%	84.50%
RBFNetwork	83.10%	85.04%
Multilayer Perceptron	73.80%	82.71%
IB1	74.20%	83.17%
IBk	80.30%	83.17%
PART	81.30%	85.90%
DTNB	78.90%	86.40%
J48	84.50%	84.50%
Random Tree	79.90%	80.37%
RandomForest	79.40%	85.90%

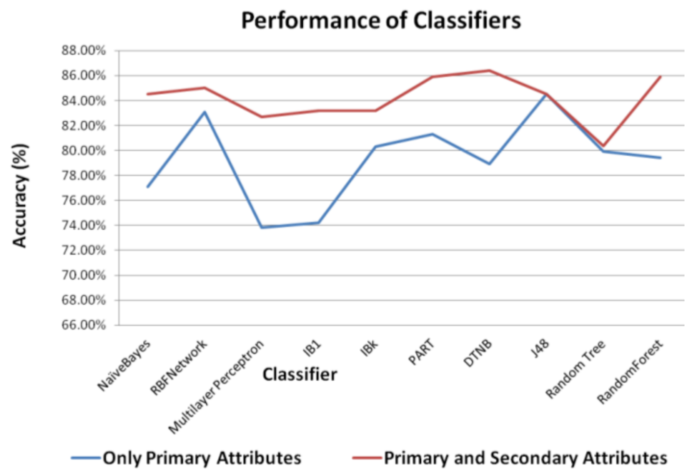


Fig. 2. Graph depicting Classifiers Performance

TABLE III. Training and Simulation Error Percentage

Classifier	RAE (Only Primary Attributes)	RAE (Primary and Secondary Attributes)	RRSE (Only Primary Attributes)	RRSE (Primary and Secondary Attributes)
NaïveBayes	100.41%	66.40%	111.04%	99.07%
RBFNetwork	95.60%	76.30%	103.20%	93.30%
Multilayer Perceptron	96.60%	67.60%	126.96%	102.90%
IB1	97.50%	63.80%	140.30%	113.50%
IBk	92.09%	65.10%	119.80%	112.90%
PART	98.21%	84.28%	101.47%	92.74%
DTNB	130.50%	102.20%	111.90%	98.07%
J48	99.06%	94.94%	99.99%	99.10%
Random Tree	92.60%	86.04%	121.60%	114.90%
RandomForest	99.60%	90.20%	109.60%	95.10%

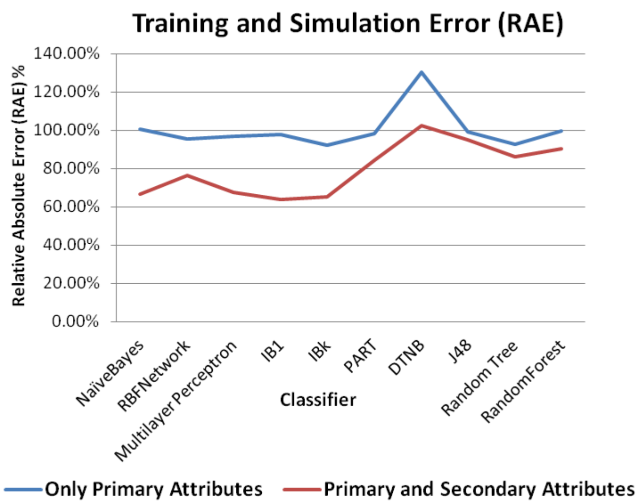


Fig. 3. Graph of Relative Absolute Error Percentages for classifiers

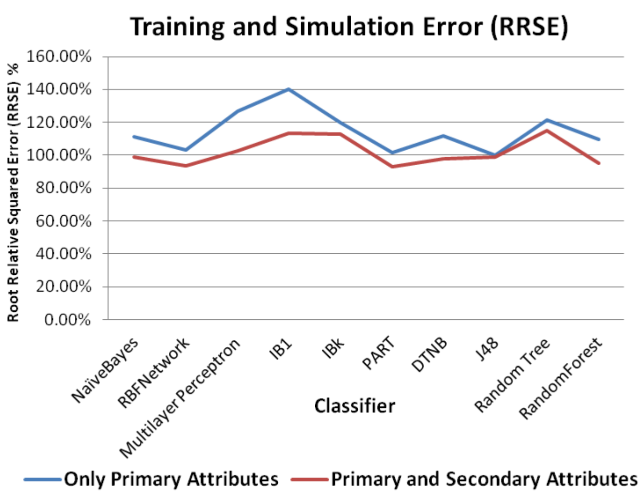


Fig. 4. Graph of Root Relative Squared Error Percentages for classifiers

Kappa Statistics: Kappa is a normalized value of agreement for chance. It can be described as

K= (P (A)- P(E))/(1-P (E))

Where,

P (A) is percentage agreement and P (E) is chance agreement.

If K =1 than agreement is ideal between the classifier and ground truth.

If K=0, it indicates there's a chance of agreement.

Table IV represents the Kappa Statistics calculated for every classifier used in the current study, once for initial dataset with primary academic attributes only, thenceforth with second dataset, which includes both primary and secondary attributes. Each classifier produces K value greater than 0 i.e. each classifier is doing better than the chance for training set [18], once second dataset is chosen. Fig. 5 additionally depicts the values in graph form.

TABLE IV. KAPPA STATISTIC OF CLASSIFIERS

Classifier	Kappa Statistic (Only Primary Attributes)	Kappa Statistic (Primary and Secondary Attributes)
NaïveBayes	0.0389	0.386
RBFNetwork	-0.0264	0.1544
Multilayer Perceptron	0.0211	0.3116
IB1	-0.0503	0.0126
IBk	-0.0002	0.0126
PART	0	0.2612
DTNB	-0.0543	0.2978
J48	0	0.0773
Random Tree	0.0818	0.1872
RandomForest	-0.0835	0.2072

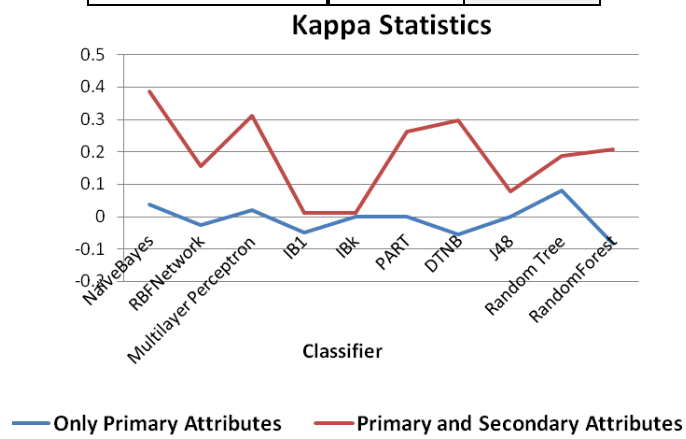


Fig. 5. Graph of Kappa Statistics Percentages for classifiers

Once prognostic model is formed, it is necessary to ascertain its accuracy. It is generally calculated based on the precision, recall values of classification matrix.

Precision is a fraction of retrieved instances that are relevant. It is calculated as

PRECISION= (TP)/(TP+FP)

Where,

TP is total number of true positives.

FP is total number of false positives.

Recall is a fraction of relevant instances that are retrieved. It is usually expressed in percentages and is calculated as

RECALL= (TP)/(TP+FN)

Where,

TP is total number of true positives.

FN is total number of false negatives.

These methods are not very apposite, if dataset is imbalanced [19]. The datasets utilized in the current study is imbalanced with only few instances for “employed” class and large number of instances for “unemployed” class.

Receiver Operating Characteristic (ROC) Curve/Area is suggested to be a better choice of evaluation with such dataset [19].

ROC curves can represent the family of best decision boundaries for relative costs of True Positive (TP) and False Positive (FP).

In ROC curve the X-axis represents

% FP=(False Positive)/(True Negative+False Positive)

and the Y-axis represents

% TP=(True Positive)/(True Positive+False Negative)

The ideal point on the ROC curve is (0,100) that is when all positive examples are classified correctly and no negative examples are misclassified as positive.

Area under the ROC Curve (AUC) is a very useful metric for judging classifier performance. It is independent of the decision criterion selected and prior probabilities. The AUC comparison can ascertain a dominance relationship between classifiers.

Comparison of evaluation measure ROC Area for minority class “employed” is presented in Table V. It is also further depicted as graph in Fig. 6.

The graph (Fig. 6) clearly illustrates the increase in ROC Area values for almost all the classifiers towards 1, when second dataset is chosen as compared to the first dataset with only primary attributes.

This also implies and proves that the performances of learning techniques are highly dependent on the nature of the dataset used.

TABLE V. PERFORMANCE OF CLASSIFIERS W.R.T ROC AREA

Classifier	ROC Area (Only Primary Attributes)	ROC Area (Primary and Secondary Attributes)
Naïve Bayes	0.619	0.868
RBFNetwork	0.596	0.784
Multilayer Perceptron	0.595	0.765
IB1	0.476	0.504
IBk	0.566	0.545
PART	0.501	0.719
DTNB	0.602	0.701
J48	0.466	0.635
Random Tree	0.579	0.697

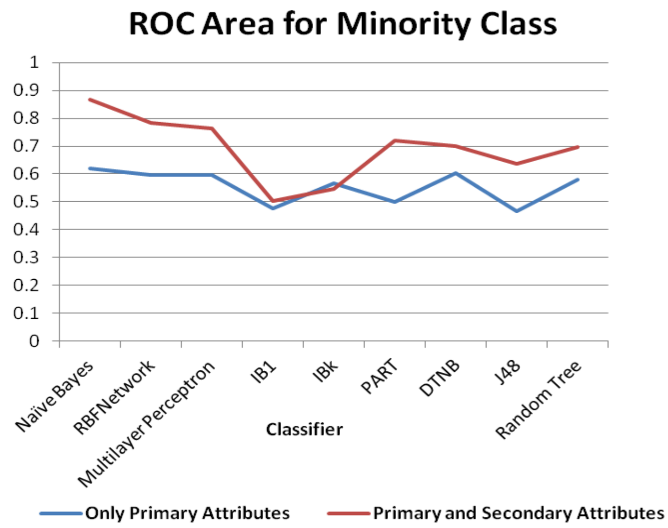


Fig. 6. Graph of ROC Area for minority class of classifiers

VI. CONCLUSION AND FUTURE SCOPE

The results prove that prediction accuracy for students’ employability can be enhanced with the inclusion of secondary attributes such as personal, social, psychological and other environmental variables in the dataset. Accuracy percentage shows incredible improvement with all types of classifiers. Error Percentage also reduces remarkably. Kappa Statistics and ROC Area shows great signs of improvement. Hence, proves that secondary psychometric attributes play the essential role in boosting the prediction accuracy of students’ employability.

Due to imbalanced datasets, classifiers could not attain high percentage accuracy with low error percentage. The maximum accuracy percentage attained in the study is by DTNB, which is 86.4% with very high error percentages, that crosses the minimal limit. Thus may not be helpful enough to be converted into prediction rules. Some technique is required to handle the problem of the imbalanced dataset.

In future, the dataset can be improved by adding more significant attributes to enhance the accuracy percentage of classifiers with low error percentage. Moreover, some automated technique is also required at the preprocessing stage to identify the best attributes for finest performance of classifiers; which also handles the problem of an imbalanced dataset.

REFERENCES

[1] Pool, Lorraine Dacre, Pamela Qualter, and Peter J. Sewell. "Exploring the factor structure of the CareerEDGE employability development profile." Education+ Training 56.4 (2014): 303-313.

[2] Vanhercke, Dorien, Nele De Cuyper, Ellen Peeters, and Hans De Witte. "Defining perceived employability: a psychological approach." Personnel Review 43, no. 4 (2014): 592-605.

[3] Hicheur Cairns, Awatef, et al. "Towards Custom-Designed Professional Training Contents and Curriculums through Educational Process Mining." IMMM 2014, The Fourth International Conference on Advances in Information Mining and Management. 2014.

[4] Potgieter, Ingrid, and Melinde Coetzee. "Employability attributes and personality preferences of postgraduate business management students." SA Journal of Industrial Psychology 39.1 (2013): 01-10.

[5] Finch, David J., Leah K. Hamilton, Riley Baldwin, and Mark Zehner. "An exploratory study of factors affecting undergraduate employability." Education+ Training 55, no. 7 (2013): 681-704.

[6] Jackson, Denise, and Elaine Chapman. "Non-technical skill gaps in Australian business graduates." Education+ Training 54.2/3 (2012): 95-113.

[7] Gokuladas, V. K. "Technical and non-technical education and the employability of engineering graduates: an Indian case study." International Journal of Training and Development 14.2 (2010): 130-143.

[8] Gokuladas, V. K. "Predictors of Employability of Engineering Graduates in Campus Recruitment Drives of Indian Software Services Companies." International Journal of Selection and Assessment 19.3 (2011): 313-319.

[9] Jantawan, Bangsuk, and Cheng-Fa Tsai. "The Application of Data Mining to Build Classification Model for Predicting Graduate Employment." International Journal Of Computer Science And Information Security (2013).

[10] B.K. Bharadwaj and S. Pal., Data Mining: A prediction for performance improvement using classification", International Journal of Computer Science and Information Security (IJCSIS), Vol. 9, No. 4, pp. 136-140, 2011.

[11] Tongshan Chang, Ed.D, Data Mining: A Magic Technology for College Recruitment's, Paper of Overseas Chinese Association for Institutional Research (www.ocair.org), 2008.

[12] S. T. Hijazi, and R. S. M. M. Naqvi, "Factors affecting student's performance: A Case of Private Colleges", Bangladesh e-Journal of Sociology, Vol. 3, No. 1, 2006.

[13] Z. N. Khan, "Scholastic achievement of higher secondary students in science stream", Journal of Social Sciences, Vol. 1, No. 2, pp. 84-87, 2005.

[14] Z. J. Kovacic, "Early prediction of student success: Mining student enrollment data", Proceedings of Informing Science & IT Education Conference, 2010.

[15] Q. A. AI-Radaideh, E. W. AI-Shawakfa, and M. I. AI-Najjar, "Mining student data using decision trees", International Arab Conference on Information Technology (ACIT'2006), Yarmouk University, Jordan, 2006

[16] Sudheep Elayidom, Sumam Mary Idikkula & Joseph Alexander "A Generalized Data mining Framework for Placement Chance Prediction Problems" International Journal of Computer Applications (0975– 8887) Volume 31– No.3, October 2011.

[17] Tanwani, Ajay Kumar, et al. "Guidelines to select machine learning scheme for classification of biomedical datasets." Evolutionary Computation, Machine Learning and Data Mining in Bioinformatics. Springer Berlin Heidelberg, 2009. 128-139.

[18] Sugandhi, C., P. Ysodha, and M. Kannan. "Analysis of a Population of Cataract Patients Databases in Weka Tool." International Journal of Scientific and Engineering Research 2.10 (2011).

[19] Chawla, Nitesh V. "Data mining for imbalanced datasets: An overview." Data mining and knowledge discovery handbook. Springer US, 2005. 853-86

Image Processing Based Customized Image Editor and Gesture Controlled Embedded Robot Coupled with Voice Control Features

Somnath Kar[1]
Department of Electrical Engineering
Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal, India

Ankit Jana[2]
Department of Electrical Engineering
Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal, India

Debarshi Chatterjee[3]
Department of Electrical Engineering
Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal, India

Dipayan Mitra[4]
Department of Electrical Engineering
Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal, India

Soumit Banerjee[5]
Department of Electrical Engineering
Maulana Abul Kalam Azad University of Technology
Kolkata, West Bengal, India

Debasish Kundu[6]
Gobindapur Saphali Memorial Polytechnic Gobindapur,
Guskara, Burdwan,
West Bengal, India

Sudipta Ghosh[7]
Bengal Institute of Technology and Management,
Santiniketan, Gopalnagar, Birbhum,
West Bengal, India

Sauvik Das Gupta[8]
Electrical and Computer Engineering
Oklahoma State University
Stillwater, OK, USA

***Abstract*—In modern sciences and technologies, images gain much broader scopes due to the ever growing importance of scientific visualization (of often large-scale complex scientific/experimental data) like microarray data in genetic research, or real-time multi-asset portfolio trading in finance etc. In this paper, a proposal has been presented to implement a Graphical User Interface (GUI) consisting of various MATLAB functions related to image processing and using the same to create a basic image processing editor having different features like, viewing the red, green and blue components of a color image separately, color detection and various other features like noise addition and removal, edge detection, cropping, resizing, rotation, histogram adjust, brightness control that is used in a basic image editor along with object detection and tracking. This has been further extended to provide reliable and a more natural technique for the user to navigate a robot in the natural environment using gestures based on color tracking. Additionally, Voice control technique has been employed to navigate the robot in various directions in the Cartesian plane employing normal Speech recognition techniques available in Microsoft Visual Basic.**

***Keywords*—Image Processing; Image Editor; Gesture Control; Embedded Robot; Voice Control**

I. Introduction

In imaging science, image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image.

MATLAB based image processing [1] is a very convenient platform and suitable for programming. An image is a matrix of pixel values. MATLAB considers every input as a matrix. For this reason, MATLAB provides an easy tool for image processing as a user can easily access each and every pixel value from the image matrices and edit it. Moreover, there is an 'image processing tool box' built in MATLAB for this purpose. Guerrero, J. [2] has demonstrated the use of Matlab and GUI for image processing and implemented in a deep vein thrombosis screening system.

Object tracking is a mature discipline aiming to define techniques and systems for processing videos from cameras placed in a specific environment. The need for high power computers, the availability of high quality and inexpensive video cameras, and the increasing need for automated video analysis has generated a great deal of interest in object tracking algorithms. Tracking an object in video has a variety of real world applications; these include autonomous aerial reconnaissance, remote surveillance, and advanced real time collision avoidance systems. There are three key steps in video analysis viz. detection of interesting moving objects, tracking of such objects from frame to frame, and analysis of frames to recognize their behavior. In its simplest form, tracking can be defined as the problem of estimating trajectory of an object in the image plane as it moves around a scene. The main aim is to track the real-time moving objects in different video frames with the help of a proposed algorithm. P.R.V. Chowdary [3] showed us an example to implement image processing algorithm for gesture recognition. The work of T. Mahalingam [4] in vision based moving object tracking through enhanced color image segmentation using Haar classifiers is very helpful to understand the underlying principle of color tracking which is also this project's backbone.

On the other hand, hand gestures can be interpreted as described by Luigi Lamberti1 and Francesco Camastra in their paper [5] where they have modeled a color classifier performed by Learning Vector Quantization. J.S. Kim [6] et al. developed a pattern recognizing algorithm to study the features of the hand. Some gesture recognition systems involve adaptive color segmentation [7], hand finding and labeling with blocking, morphological filtering, and then gestures are found by template matching. These processes do not provide dynamicity for the gesture inputs.

Several other gesture controlled robotic system uses various ways to recognize the gesture commands such as hand finding and labelling with blocking, showing specific number of fingers [8] for specific command, hand position and orientation which are measured by ultrasonic for gesture recognition. M. Mahalakshmi [9] used CAMSHIFT algorithm for real time vision based object tracking. T. Said [10] proposed a different way of controlling multi robots through multi-object color tracking.

Some other technologies use Microsoft Xbox 360 Kinect(C) [11] for gesture recognition. Kinect gathers the color information using an RGB and depth information and Infra-Red camera respectively. This system though is not very cost effective.

Keeping all previous works in mind, here a system has been developed where the main objective is to provide reliable and a more natural technique for the user to navigate a robot in the environment using gestures. Certainly, building a less costly and robust system was the motive behind this work. The primary focus on building this gesture controlled robot is on the type of gestures. The gesture, this work mostly concentrated on was color tracking. A system is proposed where the robot will track the movement of the particular color and will move along that direction. In absence of any command the robot will stop.

The system implementation involves the design of an advanced User Interface that controls the robot's movement with the help of either a specific color tracking through image processing or specific voice commands. This process has been tested for two separate colors. The color detection and tracking algorithm has been evaluated on a self-developed embedded prototype built on an open source AVR microcontroller-based platform *(ARDUINO)*. To understand the underlying principles of image processing, at first an image editor has been developed with the help of Matlab Graphical User Interface (GUI). This image editor demonstrates the functions that has been used in the main color detecting and tracking algorithm. The next section describes the hardware components required, followed by section III, describing the methodology to design the system. Then comes the experimental evaluation section followed by section V with conclusion and section VI with future works.

II. Hardware Platform

The hardware part mainly consists of a digital computer, an Arduino Uno board, Dual H-Bridge Motor Driver, Integrated laptop webcam, DC Geared Motors, Metal chassis, two wheels and a free wheel which are being discussed along with their specific functions.

A. Arduino Uno with Arduino Cable

An Arduino Uno is an Atmel 8-bit, 16-bit or 32-bit AVR microcontroller based board, with complementary components which helps in circuit incorporation. The Arduino Uno can be programmed with the Arduino Software. The ATmega328 on the Arduino Uno comes preprogrammed with a bootloader that allows users to upload new code to it without the use of an external hardware programmer. This board has a 5V linear regulator and a 16 MHz crystal oscillator.

B. Dual H-Bridge Motor Driver Circuit

The dual H-bridge motor driver circuit used here consists L293D IC as main driver. The Device is a monolithic integrated high voltage, high current four channel driver designed to accept standard DTL or TTL logic levels and drive inductive loads (such as relays, solenoids, DC and stepper motors) and switching power transistors. It uses two bridges; each pair of channels is equipped with an enable input. A separate supply input is provided for the logic, allowing operation at a lower voltage and internal clamp diodes are included. It takes digital signal as an input from the Arduino and gives digital output to the DC motors of the robot. It also amplifies voltage (from 5V to 12V) and current (from 40-50 mA to 250mA) per pin. So, overall power is amplified. Though the current amplification could be done by transistors in Darlington pair connection, which is less costly, the main purposes of using this IC are

- Unlike transistors it accepts bidirectional current and
- It can control two motors simultaneously.

There's a PWM input per driver so one can control motor speed. It runs at 5V logic. This holds good for motor voltages from 4.5V up to 36V. This works well for the 12V motors which have been used here.

C. Integrated laptop Webcam

The Webcam used here is a basic webcam associated with the HP 430 notebook. This webcam provided necessary color tracking facility for this project.

D. *DC Geared Motors*

The DC geared motors serve the main function of this project. Two DC motors with metallic gear head were used to run the two back wheels of the robot. The motors used were of 12 Volt, 100 rpm rating. It generates 1.5 kg-cm torque, which is enough to drive the wheels of the robot. Its no-load current is 60 mA (maximum) and full-load current is 300 mA (maximum). By adjusting motors rotations, the robot was being able to move forward, backward, right or left.

E. *DC adapter with header*

A DC adapter of 12V output is used to give power to the Dual H-Bridge Motor Driver Circuit.

F. *Metal Chassis*

The metal chassis worked as the main building block of the motor. It holds the circuit components on one back and the motors on the other side.

G. *Two Wheels and One Free Wheel*

The two motors drive two wheels differently at the same time. A free wheel is introduced for proper balancing of the motor.

III. METHODOLOGY

Gesture is a movement of a body part especially the hand to express an idea. Here, in this project this gesture phenomenon is for giving command.

The whole system implementation could be divided into several steps. The possible functions of image processing were realized and tested in a Matlab GUI based image editor.

Image processing operations can be roughly divided into three major categories, Image Compression, Image Enhancement and Restoration, and Measurement Extraction. Image defects which could be caused by the digitization process or by faults in the imaging set-up can be corrected using Image Enhancement techniques. Once the image is in good condition, the Measurement Extraction operations can be used to obtain useful information from the image like color tracking. The use of MATLAB as a Digital Image Processing Tool has made the development of many applications which incorporate different Image Enhancement function very easy. Unlike coding programs the user of a GUI need not to understand the details of how the tasks are performed. GUI components can include menus, toolbars, push buttons, radio buttons, list boxes, and sliders, just to name a few. GUIs created using MATLAB tools can also perform any type of computation, read and write data files, communicate with other GUIs, and display data as tables or as plots. The fig. 1 illustrates the work of the GUI created for image enhancement. Here auto histogram adjustment process shows the resultant enhanced image.

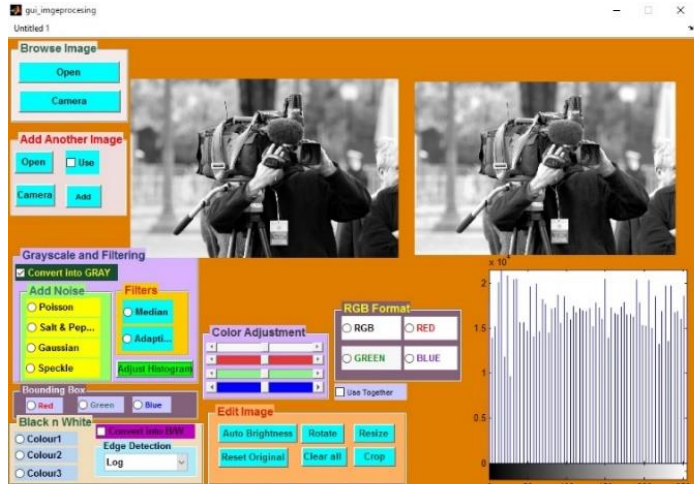


Fig. 1. GUI of Image Processing

This proposed system consists of two main hardware components: the computer which runs the Matlab, Visual Basic, and the Arduino Uno microcontroller board which is flexible, inexpensive, offers a variety of digital and analog inputs, serial interface and digital outputs. Arduino Uno controls the robot by controlling the action of motor driver circuit. Arduino Uno also enables the user to control the motor through voice commands. The computer communicates with the Arduino Uno microcontroller board through USB data transfer cable.

A. *The Color Tracking Through Image Processing*

Movement of Image frame with particular color is taken as an input and processed using Image Processing. This image processing is the fundamental part of tracking. The tracking system uses the Matlab coding. Matlab tracks the command direction and send the direction to the Arduino software which directly gives command to the Arduino Uno board. With the help of motor driver circuit, the Arduino controls the movement of the robot in the required direction as given by the command.

The Matlab program first opens the video capturing frame. This frame captures video through laptop associated webcam in RGB format for infinite time until user commands to stop. From this captured video continuous snapshots are taken out. This RGB snapshots are the main tools for the process done by Matlab.

The RGB images are flipped in both rows and columns to correct the effect of flipped image taken by webcam. Then the particular color (viz. Red, Blue, Green) upon which the Matlab color tracking program is based on, is extracted from it; the color is extracted from the RGB picture. Generally, this image contains dusty noise. So a median filter is used to filter out the noise from this image. This monochromatic image is then converted to black and white. From this black and white image, the area, co-ordinates of the centroid and the bounding box containing the color used for giving command can be easily found out. The centroid contains the *x* and *y* co-ordinates, which are printed over the bounding box. The changes of *x* co-ordinate represent the movement of hand along *x*-axis i.e. in right or left direction and similarly the changes in *y* co-ordinate represent the movement of hand in upward or downward direction.

The tracking program functions this way: the program tracks the centroid and so the bounding box and according to which direction the centroid moves, the program sends commands to the robot to move in that direction via Arduino software. Our hand movement is never absolute in one direction. For example, if we move our hand in the right, some upward or downward movement occurs along with. But here the maximum change occurs in *x* co-ordinate and slight change take place in *y* co-ordinate. So a random threshold value of *y* co-ordinate (say 20) has been taken to omit the changes of *y* co-ordinate. So the Matlab will send command (exact direction to move horizontally along the right only) to Arduino. Similar pattern is used for other directions.

B. *Working of Arduino*

The Arduino software programing takes the directions, which came from Matlab, as input. According to the case it runs the required function and sends accurate instruction to the motor.

With the Arduino software the Arduino Uno sends necessary information to the motor driver control circuit, which then controls the motions of the motors and thus the movement of the robot.

C. *Voice control*

At the end of the work, a special feature of giving command to the motor has been introduced. This is voice command. The robot can accept voice commands through Microsoft Visual Basic 2008 Express Edition. Microsoft Visual Basic accepts voice commands and sends instructions to the Arduino software. This Arduino software then recognizes the input and sends the required instruction to the Arduino Uno board, which with the help of motor control circuit runs the motor in the direction accordingly.

IV. Experimental Evaluation

The whole system that has been developed is given in fig. 2. It shows us different parts and connections of the circuit. The input pins of dual H-bridge motor driver circuit are connected to Arduino's digital pins whereas the output pins are fed to the two motors. This can be seen in the bottom part of the system which is given in the fig 3.

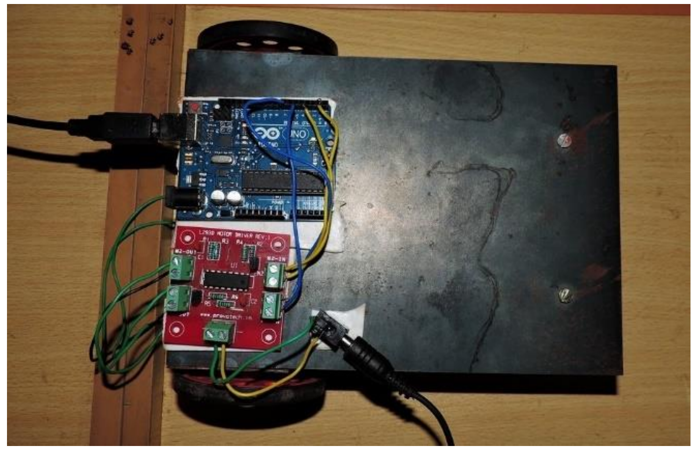


Fig. 2. Top view of the whole robot

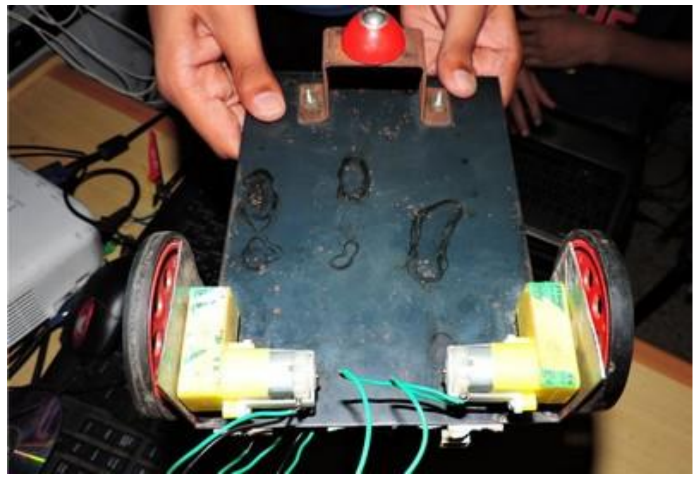


Fig. 3. The view of the bottom of the robot

Fig. 4 shows the input given to the Matlab for tracking of red color. The rectangular box in the picture is the bounding box and in it its area and the co-ordinates of the centroid is printed.

In the additional part of controlling the robot with voice command, the Microsoft Visual Basic has been used and it is shown in fig. 5, where the robot was given command to move forward.

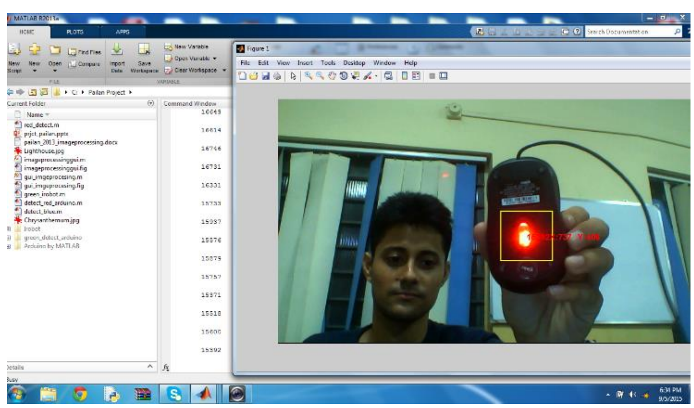


Fig. 4. Color tracking using Matlab

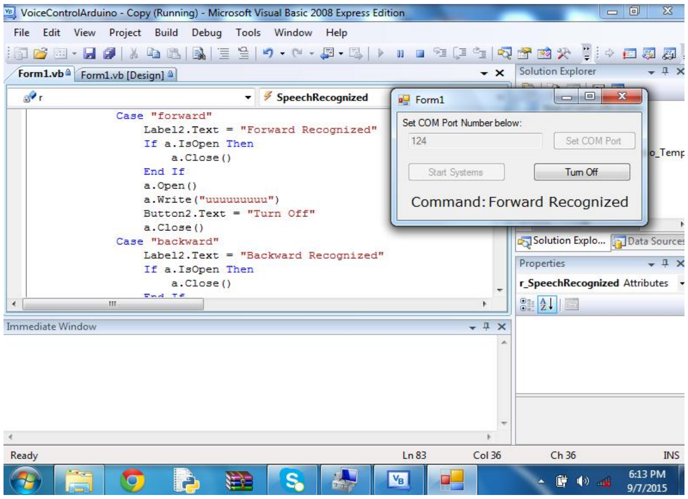


Fig. 5. Voice command using Microsoft Visual Basic

V. Discussion and Conclusion

A low cost computer vision system that can be executed with low power in-built laptop web cam was one of the main objectives of this work, which has been implemented successfully. The system has been experimented with around 6 colors and the results achieved higher average precision.

From the reults obtained, it can be concluded that the tracking algorithm is quite efficient. The main advantage of this algorithm is that even tilted colored objects can easily be recognized and analyzed. The colored object used to give the command for movement direction, if it is multicolored, only one specific color is detected. This overrules the possibility of interfering of other moving objects nearby, provided it is of different color.

However, there are several factors which limit the efficiency of the detection rate. Some of them are discussed below.

- Ambient lighting intensity vastly effects the outcome.
- the colored objects should be properly placed in front of the webcam so that the entire region is captured.
- Gesture made in this method involves only one color at a time and this reduces the number of gestures that can be sent within a definite time interval.
- Monocromatic color dependency affects the subsequent detection rates.

Whereas, the voice command recognition based implementation is much more simple and direct. It is independent of ambient conditions. Unless it gets the exact moving commands it does not send signals to the robot. So, the ambient noise leaves no interference.

This system has many potential application in many fields, from home applications to big industrial applications. Being at one particular position, many robotic applications at a time can be controlled to carry out many jobs. The type of jobs are somewhat specific like lifting things up, moving and putting them down, pushing things, stacking things, positioning things and vehicles. In industries, in harsh environments, instead of sending people, the robots can be sent to carry out jobs like pulling the hot items and put them outside to be cooled down.

With a commercial hardware package and wireless facility, this system could be used in homes and in some small industries to control robots to move appliances.

VI. Future Works and Applications

Controlling a robot, in real time, through the gestures is a novel approach and its applications are myriad. The use of service robot to domestic users and industries in the upcoming years would need such methods extensively. The approach has huge potential once it gets further optimized, as its time complexity is higher, with the help of hardware having better specifications.

There are several possibilities for future renovation. The tracking algorithm could be implemented through GUI. This would make it more user-friendly and simultaneously, a robust interface for interaction. The speech recognition system can be used in the voice command algorithm to make it user dependent for more security. Implementing this concept on an i-Robot create would be more efficient and versatile.

Coupling this sytem with a wheelchair can be very effective for physically challenged people, whose legs are not working but the limbs, to become independent for moving from one place to other.

On the other hand, with camera facilities and more intelligent algorithm to work independently, this system can be very useful in big industries for carrying out some jobs and to carry things and place it at specific place. This would require the wireless communication technology to free the robots from direct wired connection with the control system.

Use of more efficient wireless communication technique and a camera on the robot unit would improve the performance of the system to a great extent and can be incorporated in future applications like investigating lives of wild animals, investigating through narrow tunnels and recovering necessary items from a place under devastating fire etc. where human beings can not be present.

As a whole it can be concluded that the system has a huge scope of further research and application which can prove to be effective in various fields.

Acknowledgment

We would like to thank e-school learning (ESL), Kolkata for assisting us with all necessary technical details and working platform to build this project successfully.

Authors' Contributions

Somnath, Ankit, Debarshi contributed in designing the image editor in MATLAB, implementing it in the gesture controlled robot, in testing and writing the report. Soumit and Dipayan assisted in assembling the hardware parts of the robot and in data collection. Debasish and Sudipta designed the Voice control part in Visual Basic. Sauvik conceptualized the problem, designed it and helped in making the final coding in MATLAB.

REFERENCES

[1] Justyna Inglot, 'Advanced Image Processing with Matlab', in Bachelor's Thesis Information Technology, May 2012, Date of the bachelor's thesis 07.05.2010, Mikkeli University of Applied Sciences.

[2] J. Guerrero, "Image Processing and Analysis with Matlab", International Conference on Electrical, Communications, and Computers, 2009. CONIELECOMP 2009, Cholula, Puebla, pp. xvi, 26-28 Feb. 2009.

[3] P.R.V. Chowdary, M.N. Babu, T. V. Subbareddy, B.M. Reddy, V. Elamaran, "Image processing algorithms for gesture recognition using Matlab", 2014 International conference on Advanced Communication Control and computing Technologies (ICACCCT), Ramanathapuram, pp. 1511-1514, 8-10 May 2014.

[4] T. Mahalingam, M. Mahalakshmi, "Vision based moving object tracking through enhanced color image segmentation using Haar classifiers", Conference on Trendz in Information Sciences & Computing (TISC), 2010, Chennai, pp. 253-260, 17-19 Dec 2010.

[5] Luigi Lamberti1 and Francesco Camastra, "Real-Time Hand Gesture Recognition using a Color Glove," *Department of Applied Science, University of Naples Parthenope*, 2010, pp.321-344.

[6] J.S. Kim, C.S. Lee, K.J. Song, B. Min, Z. Bien, "Real-time hand gesture recognition for avatar motion control," *Proceedings of HCI'97*, pp. 96-101, February 1997.

[7] Chao Hy Xiang Wang, Mrinal K. Mandal, Max Meng, and Donglin Li, "Efficient Face and Gesture Recognition Techniques for Robot Control‖," CCECE, 1757-1762, 2003.

[8] Jagdish Lal Raheja, Radhey Shyam, Umesh Kumar and P Bhanu Prasad, "Real-Time Robotic Hand Control using Hand Gestures", Second International Conference on Machine Learning and Computing, 2010.

[9] M. Mahalakshmi, "Real time vision based object tracking using CAMSHIFT algorithm with enhanced color image segmentation", 2010 International Conference on Computing Communication and Networking Technologies (ICCCNT), Karur, pp. 1-8, 29-31 July 2010.

[10] T. Said, S. Ghoniemy, G. Ismail, I. A. EL-Ghafar, 13th International Conference on Aerospace Sciences & Aviation Technology, ASAT- 13, May 26 – 28, 2009.

[11] Gesture controlled robot using Kinect http://www.e-yantra.org/home/projects-wiki/item/180-gesture-controlled-robot-using-firebirdv-and-kinect

Gesture Recognition Based on Human Grasping Activities Using PCA-BMU

Nazrul H. ADNAN
Kolej Kemahiran Tinggi Mara (KKTM),
Mukim Serom 4 & 5, 84410 Ledang, Johor
Darul Takzim, MALAYSIA

Mahzan T.
Kolej Kemahiran Tinggi Mara (KKTM),
Mukim Serom 4 & 5, 84410 Ledang, Johor
Darul Takzim, MALAYSIA

***Abstract*—This research study presents the recognition of fingers grasps for various grasping styles of daily living. In general, the posture of the human hand determines the fingers that are used to create contact between an object at the same time while developing the touching contact. Human grasping can detect by studying the movement of fingers while bending during object holding. Ten right-handed subjects are participated in the experiment; each subject was fitted with a right-handed *GloveMAP*, which recorded all movement of the thumb, index, and middle of human fingers while grasping selected objects. *GloveMAP* is constructed using flexible bend sensors placed back of a glove. Based on the grasp human taxonomy by Cutkosky, the object grasping is distinguished by two dominant prehensile postures; that is, the power grip and the precision grip. The dataset signal is extracted using *GloveMAP*, and all the signals are filtered using Gaussian filtering method. The method is capable to improving the amplitude transmission characteristic with the minimal combination of time and amplitude response. The result was no overshoot in order to smoothen the grasping signal from unneeded signal (noise) that occurs on the input / original grasping data. Principal Component Analysis – Best Matching Unit (PCA-BMU) is a process of justifying the human grasping data involves several grasping groups and forming a component identified as nodes or neuron.**

Keywords—recognition; grasp; grasp taxonomy; human finger; dimensionality reduction

I. INTRODUCTION

There are too many applications in this era that related to the human gesture that is include parts of the human body such as hands, face, body and many more. The hand gesture is one of the famous gestures used in daily life. People use the hand gesture to enhance the communication with others to deliver the information of thoughts effectively. This hand gesture will give a lot of important information of fingers or hand movement that can implement in the industrial applications such as video games industry, biomedical instrument, sports science, surveillance systems and many more. A device known as a Dataglove is presented as a medium to measure the information gain from hand or fingers activities. Dataglove is known as cyber glove or wired glove, and it is a device that can be donned by human. Any of physical data can capture by this technology using various sensors such as bending sensor or mostly known as flexible bend sensor, a force sensor or force resistive sensor, a tactile sensor and other types of sensor. Dataglove is also known as parts of “Haptic Science”, which is give meaning as science of applying tactile sensation to human interaction through computer.

GloveMAP can recognize and classify human grasping of some selected objects (Ball, Cylinder, Pen, Key, Disc, Scissors, Pins and Paper). The principle of *GloveMAP* is based on the concept of human grasping activities which is considered basic in daily living and was studied for a decade [1][2]. *GloveMAP* exploits the advantage of flexible bend sensors attached on *GloveMAP* hand surface to capture finger movement information when the user performs grasping activities. *GloveMAP* outputs are in a form of voltage over time that varies when the surface of flexible bending sensor is bending. The bending of flexible bend sensor will alter the resistivity of the sensor. The grasping activities will produce a signal that represents the characteristic of the grasping objects. The output signal must be filtered to reduce the noise produced by the thermal motion of electrons. The output signal is term as human grasping data. In finalizing the grasping signal, Principal Component Analysis (PCA) is used to reduce the data redundancy. PCA generally functions as to reduce the dimensionality of dataset in which there are a large number of interrelated variables, while maintaining as much as possible in dataset changes [3][4].

This research paper is structured as follows: Section 2 addresses the literature review of the related researches to the several approaches, applications and problems of recognizing the fingers grasping force signal. Section 3 describes the methodologies of the system. Section 4 describes the experiment of the research. Section 5 will present the results and discussion. Finally, section 6 described the conclusions and proposing some possible future work.

II. LITERATURE REVIEW

Nowadays, computer and related technology has become so distinctive, Human Computer Interaction (HCI) is one of the associated technology. Since HCI has such a common phenomenon, there was a need to make it as seamless as possible so that it was close to the natural human-to-human interaction. However, one big hurdle that must be overcome in order to achieve this objective was the lack of human grasping perception in today’s computers. If computers can recognize human grasping activities, not only the interaction becomes more natural and easy through improved HCI, but many other useful applications could be developed. For example, a rehabilitation system device can be developed to assist the elderly and highly disabled people. Intelligent tutoring systems

Kolej Kemahiran Tinggi Mara Ledang (KKTM), Majlis Amanah Rakyat (MARA)

can take into account by transforming the system into education site. Regarding to this, the system is capable to be a smart tutorial for those who wanted a system with smart application using virtually hand motion or fingers movement. Human hand can be fitted with intelligent devices that are able to transform the users' grasping motion into many applications. In the last decade, there were numerous literatures on grasping force analysis, grasping force optimization and grasping force stability. The key problems in the last twenty years were force analysis [5] and grip strength [6], usually problem occurs when multi-fingered grasping takes place. Generally speaking, advances in the recognition of human activities such as motion control [7], hand grasping [8][9] and robot grasping [10] are progressing. They are demonstrated using popular methods such as EMG [11], Dataglove [12][13][14][15], and humanoid hand [10].

According to [16], the direct kinematics of fingertips is used to grasp the objects. They also proposed the position and orientation as the best methodologies for the study. Meanwhile for the data reduction and classifier method for finger grasping data, [17] stated that PCA is the best reduction method especially for the motionless position synergy angle configuration of the physical posture and contour of human hand / fingers whilst grasping the object.

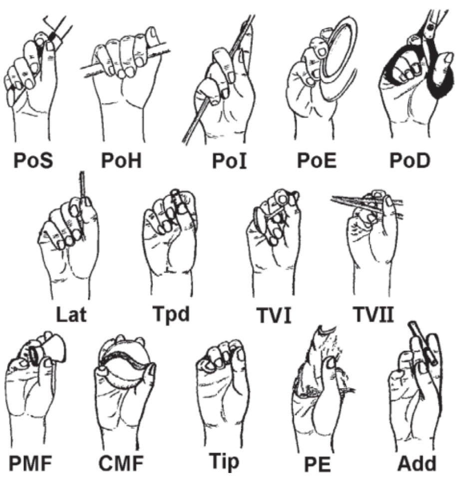


Fig. 1. Kamakura Grasp Taxonomy [2]

Fig. 2. Resistive interface glove [18]

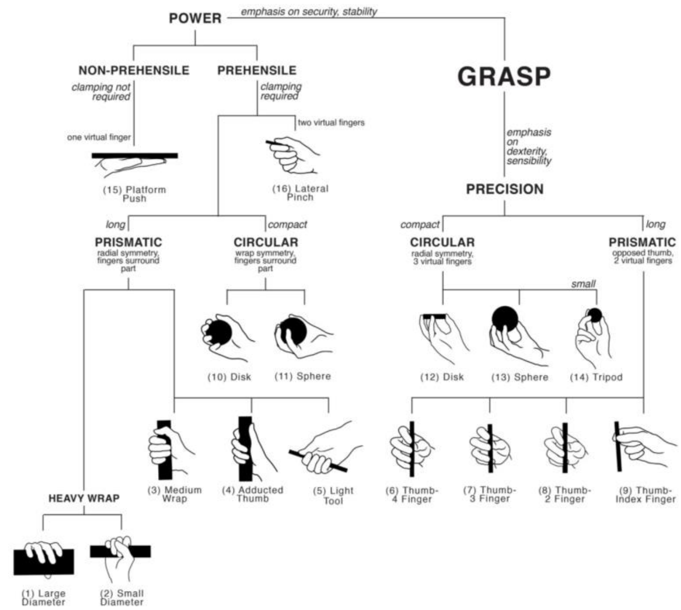


Fig. 3. Cutkosky Grasp Taxonomy [1]

Most of the research efforts on grasping taxonomy for the analysis of human grasping activities include the selection of experiment objects. According to Miller et al., to solve the grasping analysis in research by using grasp pre-shape which is defined as the finger configuration when a hand begins the fingers bending activities [19]. Napier, classified human hand shapes into a precision grasp and a power grasp [10]. In the power grasp, the object is grasped by the whole hand, including the fingers, thumb, and palm. In contrast, in the precision grasp, the object is grasped only by the fingertips. Kamakura et al., proposed an occupational therapist grasp taxonomy which consists of 14 hand shapes used in daily life [2]; this taxonomy is shown in Fig. 1. Fig. 2 shows the resistive interface glove which is used in conducting the experiment of human grasping activities. Meanwhile Fig. 3 illustrates the Cutkosky grasp taxonomy. Cutkosky, proposed the improved taxonomy on grasping activities using 16 hand shapes used by humans working with tools and metal parts [1]. Both taxonomies are identified as grasp types. Dataset of grasp pre-shapes for a human hand is produced from real human grasp activities.

III. METHODOLOGY

This section presents method used in characterizing fingers grasp capabilities for various grasping styles of daily living. In general, the posture of the human hand determines the fingers that are used to create contact between object at the same time develops the touching relation. The relationship between human grasp and selection of grasping object is based on the philosophy of grasping which is known as Cutkosky Grasp Taxonomy. The taxonomy distinguishes two dominant prehensile postures which are the power grip and the precision grip. According to Cutkosky (1989), all subjects should confine to single-handed operations and there should have been a better appreciation of how task requirements and object geometry combine to justify the grasp choice for better result of human grasp [1]. The next process flow is to eliminate or minimize the unwanted signal and noise by using Gaussian Filter. Gaussian Filtering makes grasping signal become smoother and lessens the abrupt changes in signal frequency. Then the grasping signals are analyzed using PCA. Since PCA functions as data reduction, PCA becomes the first choice method in reducing the redundancy in grasping signal. PCA is capable to generate an *"Eigenfinger"* for thumb, index and middle fingers of grasping data. Fig. 4 shows the example of fingers grasp testier and Fig. 5 shows a sample of human grasp object.

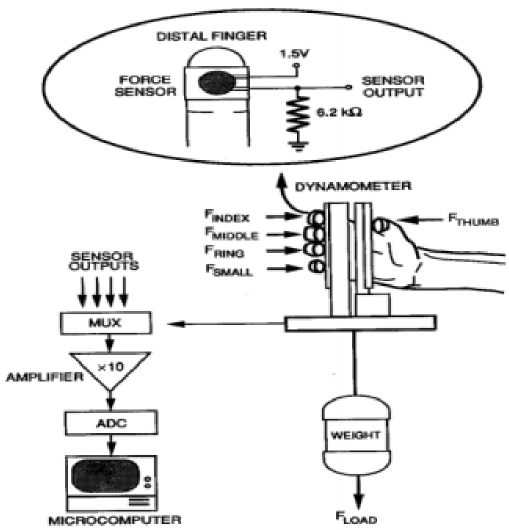


Fig. 4. Experiment of fingers grasps [20]

Fig. 5. Object grasping

A. Gaussian Filtering Techniques

GloveMAP signal is prepared with Gaussian filtering method in order to remove noise produced by random thermal motion of charge inside the electrical conductor. Noise within signal could affect the performance of objects' feature and classification. Resistors used in *GloveMAP* also would produce noise as heat inside resistors buildup. Each data collection from 8 objects will be filtered using Gaussian Filtering. Fig. 6(a) and Fig. 6(b) show unfiltered and filtered voltage produced from human grasping. Both figures demonstrate the result of Gaussian Filtering into raw voltage to reduce noises and overshoot. Gaussian has an advantage of reducing noises and overshoot of the input grasping signal.

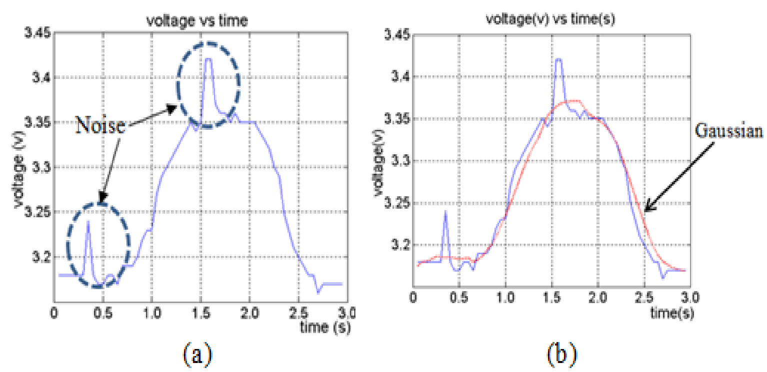


Fig. 6. a) Unfiltered voltage output with noise b) Voltage with Gaussian Filter

B. Kinematical Finger Structure

For the kinematical finger structure, finger joints consist of two main joints namely the Proximal Interphalangeal (PIP) and Interphalangeal (IP) joint of human finger. The other joint is the Metacarpophalangeal (MCP) joint. Fig. 7 illustrates a DIP flexion of human finger. Although the hand has so many flexions, the finger movement is highly constrained, below is the list of two motion finger constrains so that it cannot make arbitrary gestures:

*1) **Intrafinger constraints***: This is the constraint between the joints of the same finger and the movement of joints can be approximated by the following equation:

$$\theta_{DIP} = 0.67\theta_{PIP} \quad (1)$$

Where θ_{DIP} known as the DIP bending angle meanwhile θ_{PIP} is the bending angle of the PIP joint.

*2) **Angle range constraints:*** This type of constraint refers to the limits of the ranges of finger motions as a result of hand anatomy. Fig. 8 shows the PIP flexion of human hand and it is usually within the following ranges.

$$0^0 \leq \theta_{PIP\_Flexion} \leq (90^0 \sim 100^0) \quad (2)$$

and, meanwhile Fig. 9 shows the MCP flexion of human hand.

$$0^0 \leq \theta_{MCP\_Flexion} \leq 90^0 \quad (3)$$

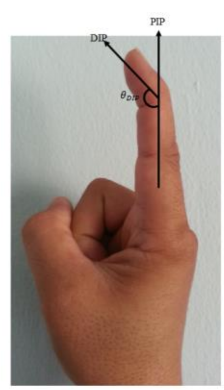


Fig. 7. DIP flexion

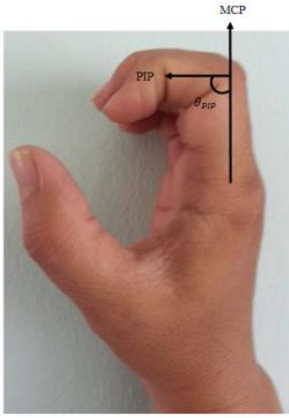


Fig. 8. PIP flexion

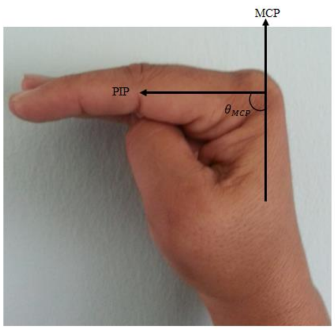


Fig. 9. MCP flexion

Fig. 10. Finger Abduction

Another kinematical finger movement is the finger abduction; Fig. 10 shows how fingers move in abduction action. Abduction is functioning by 2 conditions moving the index finger away from the middle finger or bringing the index finger close to the middle finger. Abduction also can be Thumb-Index, Thumb-Middle, or Index-Middle finger movement as shown in Fig. 11.

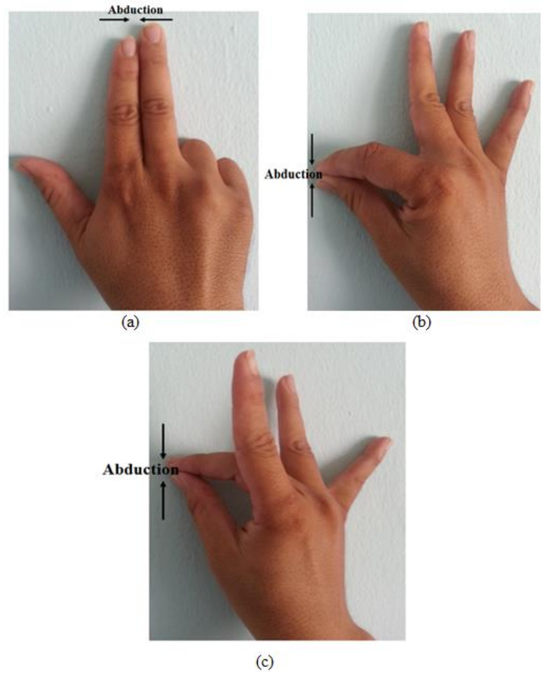


Fig. 11. Abduction (a) Index-Middle (b) Thumb-Index (c) Thumb-Middle

C. Principal Components Analysis (PCA)

PCA is found useful in many applications, but the basic application is the dimensionality reduction method. The grasping data signal could be calculated by converting the coordinate of the finger bending. It is defined as the space of *Eigenfingers* (feature spaces). For the example let the dataset, consisting of p observation variables and q observations for each variable stack into a matrix $X \epsilon R^{q \times p}$, with q is column and p is row of data and it is expressed in Equation (4):

$$A = \begin{bmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \\ \vdots & \vdots & \vdots \\ x_{q1} & x_{q2} & x_{q3} \end{bmatrix} \quad (4)$$

Where A is a covariance matrix for 3 dimensional dataset, using usual dimension x_{11}= x dimension, x_{12} = y dimension and x_{13} = z dimension.

The principal component transform is defined as A equal to the *Eigenfingers* matrix and defined as the roots of:

$$determinant(A - \lambda I) = |A - \lambda I| = 0 \quad (5)$$

Where;
I = (q x p) identity matrix
λ = eigenvalue of A

Otherwise, $\lambda_1 > \lambda_2 ... > \lambda_n$ could be identified as the eigenvalues of the covariance (some other researchers termed this as the diagonal covariance matrix) of A. The analysis of PCA could be used by both *Eigenfingers* and Eigenvalues are requisite. Whereas Eigenvalues can be simplified as ***Eigenvalues = Eigenfingers*original data***. Some analysis of the real numbers is dependent to both concepts (vectors and linear transformations). *Eigenfingers J* of *A* and Eigenvalues λ can be determined as:-

$$A_J = \lambda_J \tag{6}$$

and simplified as:

$$(A - \lambda I)X = 0 \tag{7}$$

The concept of Jacobi method [21] is applied where λ and A were calculated and I is known as the identity matrix. Lastly, it is simply to find the *Eigenfingers* determinant as shown in Equation (8).

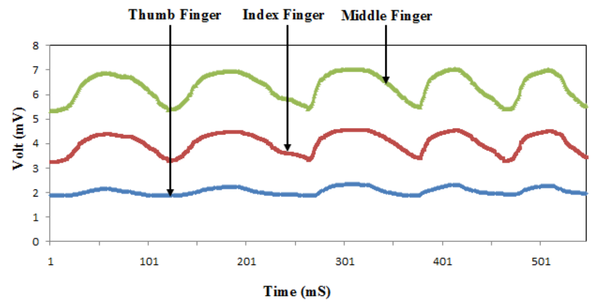
$$\det(A - \lambda I) = 0 \tag{8}$$

IV. RESULT AND DISCUSSION

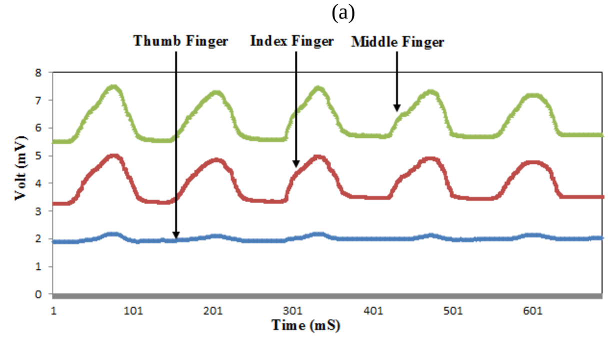
In this section, the analyses of overall step results are started accordingly from data acquisition, data analysis, features usage, and finally classifier recognition result.

A. Human Grasping Data

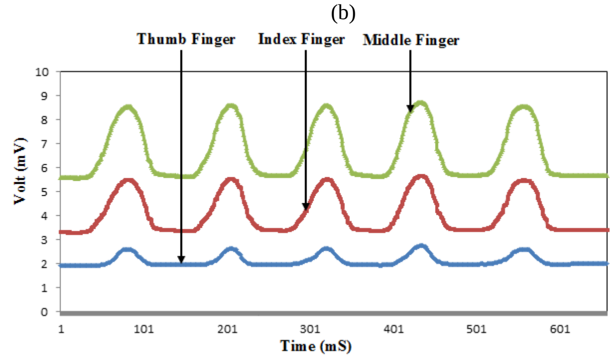
Fig. 12 shows the sample of 3 out of 8 objects grasping data. The figures show three main fingers results (thumb, index and middle) involved in the experiment. All figures show that middle and index finger were given more bending compared to the thumb. Basically index and middle fingers could be defined as the two strongest fingers meanwhile the thumb functioned as the main supportive finger whilst grasping the objects. Based on Fig. 12, it is proven that the signal for both fingers (index and middle finger) was more functioning compared to thumb fingers. Naturally, the thumb cannot bend more compared to the index and middle finger, however the thumb at the same time is moderately flexible (when the hand was spread, the thumb was easily standing a fair distance from the rest of the fingers).



(a)



(b)



(c)

Fig. 12. Grasping signals of the object (a) "Ball" (b) "CD" (c) "Cylinder"

B. PCA-Best Matching Unit (PCA-BMU)

The process of justifying the human grasping data involves several grasping groups and forming a component identified as nodes or neuron. The group of neurons basically has one main neuron located at the center of group of neurons, which is the winning neuron or centroid. Based on the explanation in the previous chapter, centroid or the winning neuron is formed by the competition of each neuron for representation of the group of data. The process of competing occurs until a next competing between other neuron except the centroid or winning neuron had been finalized.

To determine the PCA-Best Matching Unit (PCA-BMU) of the grasping data, the concept of neighborhood between neuron was applied. So, the next step was calculating all neurons nearby the centroid or winning neuron. One method to calculate the neighborhood between nodes and centroid or winning neuron is the Euclidean distance.

The steps to determine the nearest node to the winning neuron or centroid are stated below.

1) All nodes were justified using Euclidean Distance to winning neuron or centroid.

2) The equivalent or nearest node matching with any of the centroids were justified.

3) The nearest nodes to the winning neuron or centroid will form a group of node identified as "Cluster".

From a group of neuron, set of five nearest distances with a weight vector closest to the input vector of the centroid or winning neuron was tagged as the PCA-BMU. Lastly, to justify the features, the total sums of distance for all five nearest points were calculated and the result is known as PCA-BMU features. Fig. 13 to Fig. 15 show the PCA-BMU features with 5 nearest points for all objects. Meanwhile Table 1 to Table 3 shows the total sums of five nearest distance of object "Ball", "CD" and "Cylinder".

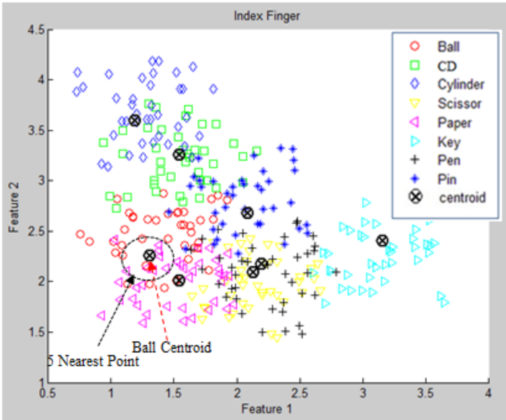


Fig. 13. 5 nearest points for PCA-BMU feature of the object "Ball"

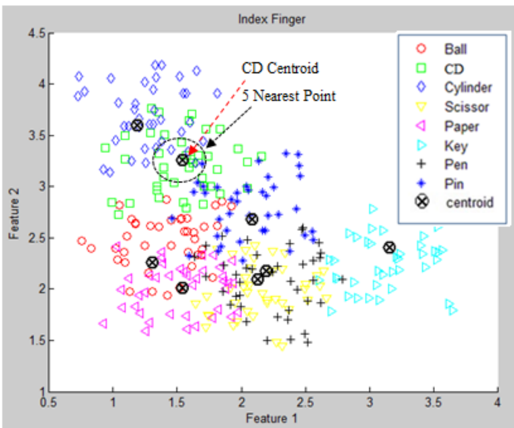


Fig. 14. 5 nearest points for PCA-BMU feature of the object "CD"

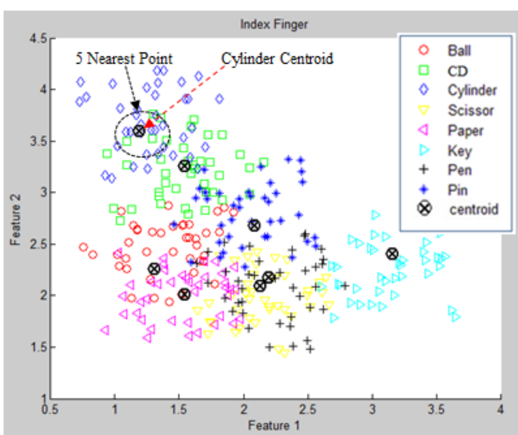


Fig. 15. 5 nearest points for PCA-BMU feature of the object "Cylinder"

TABLE I. SAMPLE OF 10 REPETITIONS GRASPING PERFORMANCES OF PCA-BEST MATCHING UNIT (PCA-BMU) FEATURES FOR OBJECT "BALL"

Ball Object		**5 nearest distance point**					**PCA-BMU Features**
		1	**2**	**3**	**4**	**5**	
Repetition 1	Thumb	0.0482	0.0901	0.1111	0.1202	0.1213	**0.4910**
	Index	0.0815	0.0824	0.0823	0.0945	0.0956	**0.4361**
	Middle	0.2746	0.2887	0.2938	0.3108	0.3215	**1.4894**
Repetition 2	Thumb	0.1315	0.1382	0.1940	0.1968	0.3005	**0.9610**
	Index	0.0673	0.0951	0.1166	0.1226	0.1245	**0.5262**
	Middle	0.0840	0.0891	0.0912	0.0938	0.0954	**0.4534**
Repetition 3	Thumb	0.0392	0.0733	0.0785	0.0847	0.885	**0.3641**
	Index	0.1079	0.1200	0.1406	0.1582	0.2014	**0.7281**
	Middle	0.0382	0.0608	0.0835	0.0840	0.0937	**0.3602**
Repetition 4	Thumb	0.1080	0.1152	0.1154	0.1163	0.1165	**0.5714**
	Index	0.0967	0.1072	0.1251	0.1594	0.1996	**0.6879**
	Middle	0.0840	0.0938	0.0954	0.1001	0.1044	**0.4776**
Repetition 5	Thumb	0.0197	0.0353	0.0429	0.0430	0.0440	**0.1849**
	Index	0.0908	0.1266	0.1739	0.2977	0.3022	**0.9912**
	Middle	0.1031	0.1061	0.1283	0.1498	0.1792	**0.6666**
Repetition 6	Thumb	0.0810	0.0909	0.1817	0.1846	0.1928	**0.731**
	Index	0.0434	0.0459	0.0486	0.0611	0.0688	**0.2678**
	Middle	0.3403	0.3484	0.3555	0.3709	0.3815	**1.7966**
Repetition 7	Thumb	0.0864	0.1008	0.1021	0.1193	0.1303	**0.5389**
	Index	0.1404	0.1894	0.1929	0.1999	0.2011	**0.9237**
	Middle	0.0811	0.0811	0.1097	0.1097	0.1201	**0.5017**
Repetition 8	Thumb	0.0864	0.1008	0.1021	0.1193	0.1303	**0.5389**
	Index	0.1448	0.1489	0.1737	0.2080	0.2100	**0.8854**
	Middle	0.0520	0.0892	0.1079	0.1095	0.1095	**0.4681**
Repetition 9	Thumb	0.0520	0.0534	0.0565	0.0593	0.0687	**0.2899**
	Index	0.0315	0.1041	0.1182	0.1515	0.2593	**0.6646**
	Middle	0.1380	0.1711	0.2279	0.2454	0.2683	**1.0507**
Repetition 10	Thumb	0.0864	0.1008	0.1021	0.1193	0.1209	**0.5295**
	Index	0.0971	0.1076	0.1250	0.1597	0.1992	**0.6886**
	Middle	0.1090	0.1165	0.1297	0.1378	0.1502	**0.6432**

TABLE II. SAMPLE OF 10 REPETITIONS GRASPING PERFORMANCES OF PCA-BEST MATCHING UNIT (PCA-BMU) FEATURES FOR OBJECT "CD"

CD Object		**5 nearest distance point**					**PCA-BMU Features**
		1	**2**	**3**	**4**	**5**	
Repetition 1	Thumb	0.0201	0.0240	0.0434	0.0518	0.0641	**0.2034**
	Index	0.1177	0.1773	0.1819	0.1983	0.1996	**0.8749**
	Middle	0.0431	0.0714	0.2177	0.2274	0.2428	**0.8024**
Repetition 2	Thumb	0.0775	0.0873	0.0957	0.0962	0.1003	**0.4571**
	Index	0.1581	0.1715	0.1726	0.2260	0.2538	**0.9820**
	Middle	0.1190	0.1637	0.1865	0.2217	0.2634	**0.9543**
Repetition 3	Thumb	0.0957	0.0962	0.1003	0.1003	0.1024	**0.4949**
	Index	0.1345	0.1426	0.1665	0.2176	0.2356	**0.8967**
	Middle	0.0915	0.2628	0.2793	0.2856	0.2938	**1.2130**
Repetition 4	Thumb	0.0599	0.0703	0.0839	0.0919	0.0957	**0.4016**
	Index	0.2548	0.2581	0.2745	0.2828	0.2928	**1.3629**
	Middle	0.1140	0.1534	0.1538	0.1678	0.2781	**0.8672**
Repetition 5	Thumb	0.0957	0.0962	0.1003	0.1003	0.1024	**0.4949**
	Index	0.1859	0.2417	0.3167	0.3460	0.3534	**1.4437**
	Middle	0.0398	0.1496	0.1590	0.1793	0.1985	**0.7262**
Repetition 6	Thumb	0.0323	0.0443	0.0698	0.0788	0.0920	**0.3172**
	Index	0.1153	0.1748	0.1854	0.1947	0.1997	**0.8699**
	Middle	0.0424	0.0700	0.1117	0.1131	0.1156	**0.4528**
Repetition 7	Thumb	0.0323	0.0331	0.0443	0.0698	0.0788	**0.2583**
	Index	0.1382	0.1419	0.1743	0.2027	0.2339	**0.891**
	Middle	0.1117	0.1131	0.1156	0.1166	0.1213	**0.5783**
Repetition 8	Thumb	0.0989	0.1012	0.1026	0.1043	0.1043	**0.5113**
	Index	0.1304	0.1830	0.2015	0.3182	0.3353	**1.1684**
	Middle	0.1090	0.1117	0.1131	0.1149	0.1156	**0.5643**
Repetition 9	Thumb	0.0989	0.1012	0.1026	0.1043	0.1043	**0.5113**
	Index	0.2092	0.2174	0.2244	0.2344	0.2456	**1.131**
	Middle	0.1059	0.1166	0.1213	0.1253	0.1328	**0.6019**
Repetition 10	Thumb	0.0902	0.0989	0.1012	0.1026	0.1043	**0.4972**
	Index	0.1991	0.2086	0.3261	0.3572	0.3608	**1.4518**
	Middle	0.0390	0.1467	0.1523	0.1551	0.1556	**0.6487**

TABLE III. SAMPLE OF 10 REPETITIONS GRASPING PERFORMANCES OF PCA-BEST MATCHING UNIT (PCA-BMU) FEATURES FOR OBJECT "CYLINDER"

Cylinder Object		5 nearest distance point					PCA-BMU Features
		1	2	3	4	5	
Repetition 1	Thumb	0.0774	0.0819	0.0940	0.0957	0.0962	**0.4451**
	Index	0.2201	0.2541	0.3065	0.3624	0.3660	**1.5090**
	Middle	0.0342	0.0634	0.1072	0.1237	0.1244	**0.4529**
Repetition 2	Thumb	0.0228	0.0421	0.0643	0.0742	0.0934	**0.2968**
	Index	0.1540	0.1896	0.2205	0.2792	0.3114	**1.1546**
	Middle	0.0506	0.0525	0.0970	0.1693	0.1863	**0.5557**
Repetition 3	Thumb	0.0957	0.0962	0.0980	0.1003	0.1003	**0.4905**
	Index	0.0453	0.1442	0.1976	0.1985	0.2476	**0.8332**
	Middle	0.0426	0.0591	0.1002	0.1227	0.1772	**0.5018**
Repetition 4	Thumb	0.0558	0.0593	0.0597	0.0637	0.0703	**0.3088**
	Index	0.0840	0.1022	0.1481	0.1972	0.2568	**0.7883**
	Middle	0.0418	0.1043	0.1175	0.1273	0.2167	**0.6077**
Repetition 5	Thumb	0.1024	0.1082	0.1082	0.1096	0.1096	**0.5379**
	Index	0.0683	0.0881	0.1011	0.1076	0.1633	**0.5285**
	Middle	0.0478	0.0479	0.0501	0.0525	0.0576	**0.2560**
Repetition 6	Thumb	0.0931	0.0948	0.0970	0.0989	0.1012	**0.485**
	Index	0.1626	0.2519	0.2697	0.3068	0.3410	**1.332**
	Middle	0.0390	0.0482	0.0706	0.1071	0.1453	**0.4102**
Repetition 7	Thumb	0.0055	0.0133	0.0556	0.0675	0.0717	**0.2136**
	Index	0.1515	0.1861	0.2201	0.2798	0.3076	**1.1451**
	Middle	0.0495	0.0521	0.0953	0.1589	0.1605	**0.5163**
Repetition 8	Thumb	0.0455	0.0948	0.0970	0.0989	0.1012	**0.4374**
	Index	0.0423	0.0587	0.0986	0.1208	0.1746	**0.495**
	Middle	0.0863	0.1175	0.1334	0.1461	0.1551	**0.6384**
Repetition 9	Thumb	0.0578	0.0948	0.0970	0.0989	0.0997	**0.4482**
	Index	0.0519	0.0776	0.1101	0.1224	0.1886	**0.5506**
	Middle	0.0654	0.0723	0.0815	0.0942	0.0946	**0.408**
Repetition 10	Thumb	0.0436	0.0521	0.0528	0.0649	0.0928	**0.3062**
	Index	0.0524	0.0702	0.0958	0.1537	0.2783	**0.6504**
	Middle	0.0373	0.0556	0.0889	0.0913	0.0925	**0.3656**

V. CONCLUSION AND FUTURE WORK

In this paper, we proposed the method to classify human grasping signal for several selected objects which based on PCA-BMU techniques. The chosen of both PCA and Best Matching Unit (BMU) for this research capable to generating the best method to smoothen up the grasp signal of features. Experimental results show that the both method works well in defining grasp signal with only a usage of few principal components and also capable to identifying the grasp type of an input motion data. For next plan works, the results are by adding the signal processing technique in the research. This signal processing employ in solving the problem of analyzing more sophisticated signal pattern especially on the signal produced during transition gesture and continuing gesture

ACKNOWLEDGMENT

Special thanks to all members of KKTM Ledang and MARA for providing the research equipment's and internal foundations. This work was supported by the Majlis Amanah Rakyat (MARA).

REFERENCES

[1] Cutkosky, Mark R. "On grasp choice, grasp models, and the design of hands for manufacturing tasks." Robotics and Automation, IEEE Transactions on 5, no. 3 (1989): 269-279.

[2] Kamakura, N., Matsuo, M., Ishii, H., Mitsuboshi, F., & Miura, Y. "Patterns of static prehension in normal hands", American Journal of Occupational Therapy, 34(7) (1980), 437-445.

[3] Cobos, S., Ferre, M., Angel Sanchez-Uran, M., Ortego, J., & Aracil, R. "Human hand descriptions and gesture recognition for object manipulation".Computer methods in biomechanics and biomedical engineering, 13(3) (2010), 305-317.

[4] Turk, M., & Pentland, A. "Eigenfaces for recognition". Journal of cognitive neuroscience, 3(1) (1991), 71-86.

[5] Xu, Jijie, Michael Y. Wang, Hong Wang, and Zexiang Li. "Force analysis of whole hand grasp by multifingered robotic hand." In Robotics and Automation, 2007 IEEE International Conference on, pp. 211-216. IEEE, 2007.

[6] McDowell, Thomas W., Bryan M. Wimer, Daniel E. Welcome, Christopher Warren, and Ren G. Dong. "Effects of handle size and shape on measured grip strength." International Journal of Industrial Ergonomics 42, no. 2 (2012): 199-205.

[7] Capisani, Luca Massimiliano, and Antonella Ferrara. "Trajectory planning and second-order sliding mode motion/interaction control for robot manipulators in unknown environments." Industrial Electronics, IEEE Transactions on 59, no. 8 (2012): 3189-3198.

[8] Kerr, Jeffrey, and Bernard Roth. "Analysis of multifingered hands." The International Journal of Robotics Research 4, no. 4 (1986): 3-17.

[9] Song, D., Huebner, K., Kyrki, V., & Kragic, D. "Learning task constraints for robot grasping using graphical models". In Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on (pp. 1579-1585). IEEE.

[10] Tsoli, Aggeliki, and Odest Chadwicke Jenkins. "Neighborhood denoising for learning high-dimensional grasping manifolds." In Intelligent Robots and Systems, 2008. IROS 2008. IEEE/RSJ International Conference on, pp. 3680-3685. IEEE, 2008.

[11] Cipriani, Christian, Franco Zaccone, Silvestro Micera, and M. Chiara Carrozza. "On the shared control of an EMG-controlled prosthetic hand: analysis of user–prosthesis interaction." Robotics, IEEE Transactions on 24, no. 1 (2008): 170-184.

[12] Park, Jongwoo, and Joono Cheong. "Analysis of collective behavior and grasp motion in human hand." In Control Automation and Systems (ICCAS), 2010 International Conference on, pp. 2514-2518. IEEE, 2010.

[13] Adnan, Nazrul H., Khairunizam Wan, A. B. Shahriman, S. K. Zaaba, Shafriza nisha Basah, Zuradzman M. Razlan, D. Hazry, M. Nasir Ayob, M. Nor Rudzuan, and Azri A. Aziz. "Measurement of the flexible bending force of the index and middle fingers for virtual interaction". Procedia Engineering 41 (2012): 388-394.

[14] Häger-Ross, Charlotte, and Marc H. Schieber. "Quantifying the independence of human finger movements: comparisons of digits, hands, and movement frequencies." The Journal of neuroscience 20, no. 22 (2000): 8542-8550.

[15] Todorov, Emanuel, and Zoubin Ghahramani. "Analysis of the synergies underlying complex hand manipulation." In Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE, vol. 2, pp. 4637-4640. IEEE, 2004.

[10] Napier, John R. "The prehensile movements of the human hand." Journal of bone and joint surgery 38, no. 4 (1956): 902-913.

[11] Iberall, Thea. "Human prehension and dexterous robot hands." The International Journal of Robotics Research 16, no. 3 (1997): 285-299.

[12] Tortora, Gerard J., and Sandra Reynolds Grabowski. "Principles of anatomy and physiology." (1993).

[13] Chao, Edmund Y. Biomechanics of the hand: a basic research study. World Scientific, 1989.

[14] Ratnasingam, Sivalogeswaran, and T. M. McGinnity. "Object recognition based on tactile form perception." In Robotic Intelligence In Informationally Structured Space (RiiSS), 2011 IEEE Workshop on, pp. 26-31. IEEE, 2011.

[15] Burke, David., S. C. Gandevia, and G. Macefield. "Responses to passive movement of receptors in joint, skin and muscle of the human hand."The Journal of physiology 402, no. 1 (1988): 347-361.

[16] Cobos, S., Ferre, M., Angel Sanchez-Uran, M., Ortego, J., & Aracil, R.. "Human hand descriptions and gesture recognition for object manipulation". Computer methods in biomechanics and biomedical engineering, 13(3) (2010), 305-317.

[17] Jerde, T. E., Soechting, J. F., & Flanders, M. "Biological constraints simplify the recognition of hand shapes". Biomedical Engineering, IEEE Transactions on, 50(2) (2003), 265-269.

[18] Reference from Fishpond Ltd., 2015

[19] Miller, A. T., Knoop, S., Christensen, H. I., & Allen, P. K. "Automatic grasp planning using shape primitives". In Robotics and Automation,

2003. Proceedings. ICRA'03. IEEE International Conference on (Vol. 2, pp. 1824-1829). IEEE.

[20] Radwin, Robert G., Seoungyeon Oh, Todd R. Jensen, and John G. Webster. "External finger forces in submaximal five-finger static pinch prehension."Ergonomics 35, no. 3 (1992): 275-288.

[21] Jacobi, J., Fraser, G. L., Coursin, D. B., Riker, R. R., Fontaine, D., Wittbrodt, E. T., ... & Lumb, P. D. “Clinical practice guidelines for the sustained use of sedatives and analgesics in the critically ill adult”. Critical care medicine, 30(1) (2002), 119-141.

Robust Convolutional Neural Networks for Image Recognition

Hayder M. Albeahdili
Dep. of Electrical and computer Engineering
University of Missouri, Columbia
Columbia, Missouri, 65211, USA

Haider A. Alwzwazy
Dep. of Electrical and computer Engineering
University of Missouri, Columbia
Columbia, Missouri, 65211, USA

Naz E. Islam
Dep. of Electrical and computer Engineering
University of Missouri, Columbia
Columbia, Missouri, 65211, USA

***Abstract*—Recently image recognition becomes vital task using several methods. One of the most interesting used methods is using Convolutional Neural Network (CNN). It is widely used for this purpose. However, since there are some tasks that have small features that are considered an essential part of a task, then classification using CNN is not efficient because most of those features diminish before reaching the final stage of classification. In this work, analyzing and exploring essential parameters that can influence model performance. Furthermore different elegant prior contemporary models are recruited to introduce new leveraging model. Finally, a new CNN architecture is proposed which achieves state-of-the-art classification results on the different challenge benchmarks. The experimented are conducted on MNIST, CIFAR-10, and CIFAR-100 datasets. Experimental results showed that the results outperform and achieve superior results comparing to the most contemporary approaches.**

Keywords—Convolutional Neural Network; Image recognition; Multiscale input images

I. Introduction

Convolutional Neural Network (CNN) has been widely used in many real world applications, including face recognition [1, 2], image classification and recognition [3-6] and object detection [7] because it is one of the most efficient methods for extracting critical features for non-trivial tasks. CNN consists of a pipeline of alternative several different layers. Unlike neural network, CNN has three different types of layers which are considered a constituent element of CNN. Usually, Convolutional layer, subsampling layers, and fully connected layer are the main components of CNN. Also, there are some intermediate layers between those main layers that will be shown later. Then for a given task, images are passed into CNN to be processed. Passing images through several squish functions incorporated within CNN layers can lead to not leveraging some critical information used for recognition and some of the small features disappear after few layers. The reason for that is because the CNN architecture that implies like those restrictions. Specifically, both convolutional layers and max-pooling layers impose diminishing small features. To implement a robust model, small features must survive for long stages of CNN. To alleviate weaknesses inherited from former CNN models, in this work, different parameters that can influence features surviving for longer distance are explored. Deeper analysis for convolutional and max-pooling layers are presented, and then we introduce a model that has more chance for small features to survive until the final stage of CNN; specifically directly before fully connected layer.

The rest of the paper will be into five sections. In section II, prior works are presented. The most recent contemporary works are obtained. Then in section III, motivation and contribution of this work are introduced. The answer for questions, what have proposed and why it is proposed are presented in this part. Then in section IV, deploying different CNN architectures are presented. Different CNN structures are obtained in this section. Finally, experimental setup and conclusion are presented.

II. Related Work

The most dominant recent works achieved using CNN is a challenge work introduced by Alex Krizhevsky et al. [8] used CNN for challenge classification ImageNet. Various other techniques are proposed later to enhance CNN performance as demonstrated in [9, 10, 11, 12]. Recently vast works have been proposed to improve image recognition accuracy results using different methods. Thus several proposed methods are proposed for variety of applications such as image recognition [13, 14, 15, 16, 8], object detection [17, 18, 19], scene labeling [20], segmentation [21, 22], and variety of other tasks [23, 24, 25]. In addition, image recognition can be accomplished using different other approaches such as Pedro F. Felzenszwalb et al. [26] proposed a method for image recognition using Deformable Part Models (DPM). Further works are devoted using different strategies of using DPM as demonstrated in [27, 28, 29]. Varity of other methods are used for image classification such as SVM [30, 31, 32, 33], boosting [34], spatial pyramid matching [35] and different other works described in [36-39].

III. Motivation and Contribution

The state-of-the-art of image recognition specifically achieved on CIFAR-10, CIFAR-100, and MNIST is achieved using different technique as proposed in [40, 41, 42]. This work has some common procedures with prior works which can be described as follows:

- The first step is applying the pre-processing to the input images such as local contrast normalization. There are different pre-processing steps that can be applied before input images passed into deep model. In this work, pre-processing steps demonstrated by Goodfellow et al. [6] is followed.

- After pre-processing, images are fed into CNN to be trained. Multi-stage CNN is used to train and extract critical features from the input patterns used later to final scoring results. In this work different CNN architectures are used for training and extracting features. Specifically very contemporary works are recruited and incorporated for introducing new unified model which achieves the state-of-the-art image classification on the datasets used in this work. Furthermore, a robust CNN is proposed at the end of this work which accomplishes superior results comparing with recent works.
- Finally, the final outputs of CNN are evaluated for final scoring results. There are different methods to score final results of CNN either using SVM or using CNN itself by using soft-max layer build on the top of CNN. Thus soft-max layer are used in this work to evaluate and score the final recognition results.

One of the most contemporary work used CNN for image classification called Network In Network (NIN) introduced in [3] achieves superior results over several prior existing models that use deep neural network for image recognition because NIN uses different connection technique between convolutional layers than what is in conventional CNN. However, there are several factors that can influence and impact model performance leading to degrade model accuracy. Thus, in this work NIN will be recruited after diminishing its shortcomings. Weaknesses of general CNN used for image classification are various such as CNN's depth, width of the network, filter sizes, and network topology. All these are vital factors that can highly impact recognition accuracy. Consequently to diminish lethargy inherited from CNN architecture; this work endeavors to alleviate shortcomings of former networks by eliminating most limitations described earlier. Therefore in this work the most recent and very efficient methods are ensemble to be used for not trivial object recognition tasks. Variety of techniques is delved to enhance image recognition. Starting from leveraging models proposed in [3, 17] both models have several deterministically advantages over prior models as elucidating later. Both concrete models are adapted in this work for image recognition. In addition, extensive work is deliberated for exploring the impact of different parameters that can drastically influence model performance. Virtuous model is mainly instantiated to overcome drawbacks of prior deep neural network architecture used for image recognition. Finally a robust paradigm of CNN architecture is proposed at the end of this work. It achieves superior results comparing with all existing models.

Elegant CNN architectures are adapted to be used for image recognition are originally proposed for image classification [3] and object detection [17]. They are considered the robust deep neural networks models. It is worth mentioning that SPPnet proposed in [17] recruited in this work to provide multi-scale input to the image recognition model. Consequently, to best of our knowledge that image classification such as CIFAR-10, CIFAR0-100, and MNIST are trained with this like method. Providing multi-resolution input images to CNN enhances CNN accuracy drastically as it will be shown later. Furthermore, digging deeper for investigating and exploring most influential parameters is also devoted. Carefully exploring influential parameters can be best suited for mole recognition. Different model architectures are extensively analyze and investigated.. After obtaining best suited parameters, a robust model is proposed to enhance recognition performance. Proposed CNN architecture achieves best results and outperforms over most existing models. The proposed model is compared to the prior efficient works specifically compared to the prior deep neural network models. In addition, the experiments are conducted on different benchmarks for evaluation purpose. The experiments are mainly conducted on CIFAR-10, CIFAR-100, and MNIST datasets.

IV. Deploy Different CNN Architectures for Image Recognition

As illustrated earlier, this work principally is recruited two different deep neural network models named NIN and SPPnet explored in [3] and [17] respectively and implemented new unified model. Next sections start exploring in depth the influence and leveraging of incorporating both models on network architectures and how they can influence classification performance. Then the unified proposed model is an elegant model because it shortens some weaknesses inherited from former models. Thus exploring both architectures is accomplished next sections to show model's robustness on image classification.

A. Pipeline Steps of image classification

The basis CNN architecture is depicted in fig. 1. It fundamentally consists of series of stages. Part (a) presents images with multiscale to the network. Providing multi-resolution input is an essential step to gain higher accuracy. Part (b) trains the network with fed images. After choosing different scales for input images, they will feed to the CNN to extract features from different resolutions which increase the chance for small features to be enlarged using this technique. It is worth mentioning that using multi-scale input images is a method showed in [17] to increase object detection accuracy. However, we utilize it to be recruited in image recognition task. Then, finally part (c) classifies and scores input pattern. To look deeper for operations accomplished by CNN, the following steps are applied:

1) Input images are pre-processed using Goodfellow et al. [6] to be prepared for the next step.

2) After pre-processing, input images are fed to CNN. In this work a new architecture is proposed as shown in fig. 1. In addition, a robust and an efficient code are used for this purpose called Caffe [43]. It is very fast implementation which can process huge amount of data efficiently. In addition, it is very flexible to be easily adapted to different CNN architecture. The final layer of CNN has n-dimensional feature vector which is used for final classification results, where n is the number of classes for a given dataset.

3) Soft max layer is used for final scoring output. However, the length of final feature vector is anticipated to be n to match the number of classes.

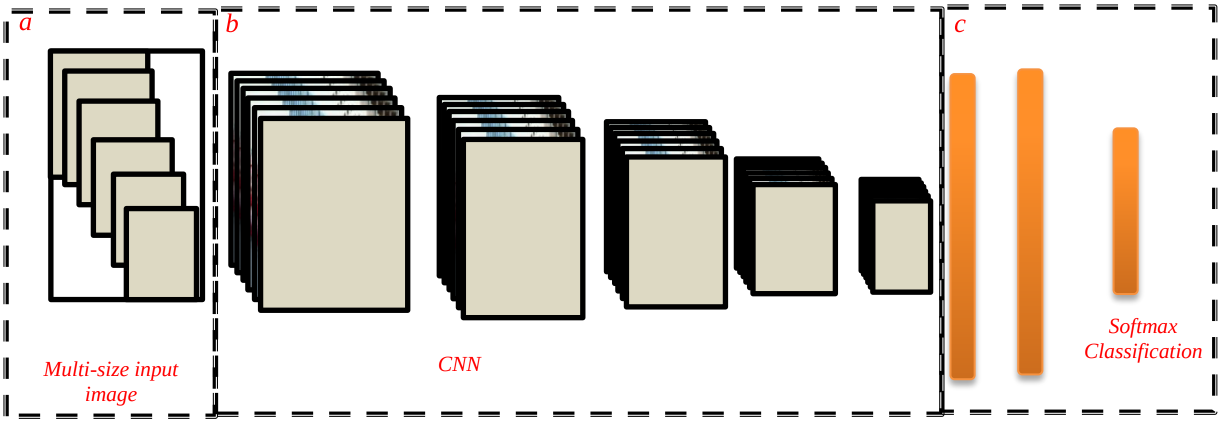


Fig. 1. simple CNN used for image classification

Fig. 1 describes CNN which has the architecture defined as 96C-96S-256C-256S-96C-96S-90F-120F-x-softmax, where C stands for Convolution layer, S is for subsampling layer, and F is for full conned layers.

It is worth mention that a dropout technique demonstrated in [12] is used in this work also to increase model performance by enhancing internal parameters and introducing more solid model. The accuracy achieved using CNN depicted in fig.1 is 0.9953, 0.83, and 0.528 on MNIST, CIFAR-10, and CIFAR-100 respectively. It is obvious that this model achieves competitive results to the most recent works. Next section provides deeper analysis and investigation for exploring and proposing more robust model.

B. Exploring different CNN architectures

It is obvious that the proposed network in fig. 1 achieves competitive results comparing to prior works. In addition, it accomplishes results which outperform accomplished work in [44] specifically it dominants over deep neural network approaches. Moreover, it achieves competitive results to many other approaches. The stimulating results are supportive to dig deeper and to investigate influential parameters and explore more robust model. In this part, recruited models will be used for further investigation and more effort will be put to explore more appropriate architecture for image classification. Leveraging CNN architecture is proposed in this section used for image recognition. It achieves state-of-the-art results on given benchmarks. Consequently, more parameters that can influence model performance are discussed next.

This work proposes a new topology for CNN architecture. Fig. 2 depicts the proposed model and it has drastically changes comparing with one implemented and explored in fig. 1. The proposed model inherits some leverage points from NIN. Instead of using conventional connection between convolutional layers as describe in [12, 9, 10, 11]. the robust connection proposed in NIN is incorporated in this work to increase and gain more accuracy on image classification. The size of CNN is kept the same as depicted in fig.1. The merit of this CNN architecture combines more than one elegant method such as multi-scale input images and nonlinear transformation between convolutional layers as demonstrated in [3] as shown in fig. 2.

To look deeper inside CNN and investigate the most critical parameters that can influence model performance. Fig.3 shows both convolutional and sub-sampling layers of CNN.

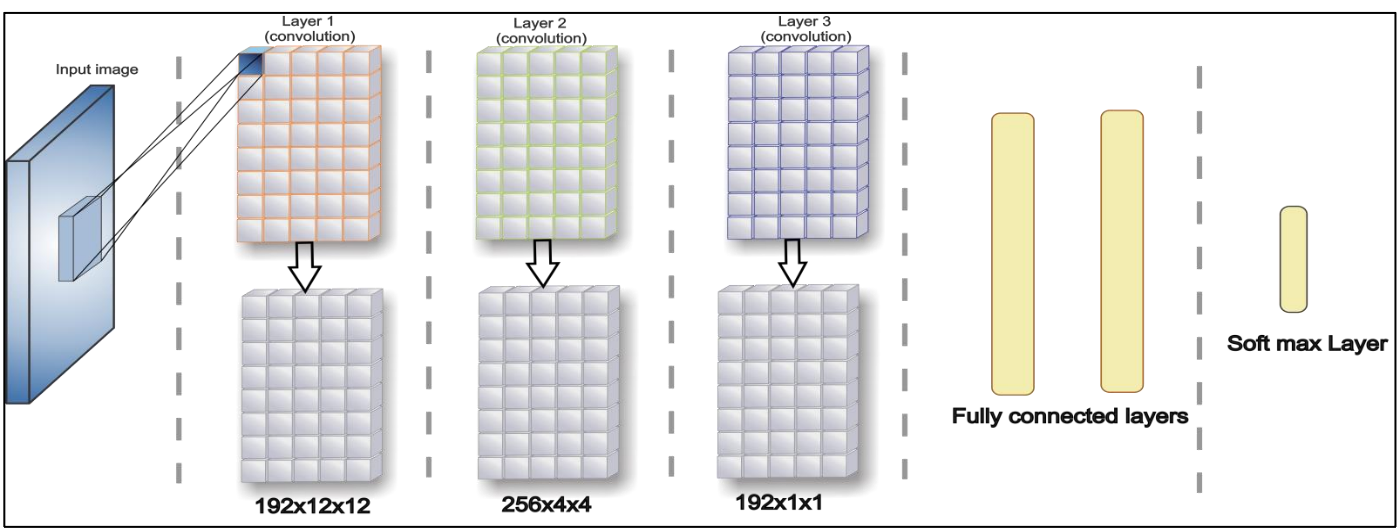


Fig. 2. CNN incorporated with two roboust models

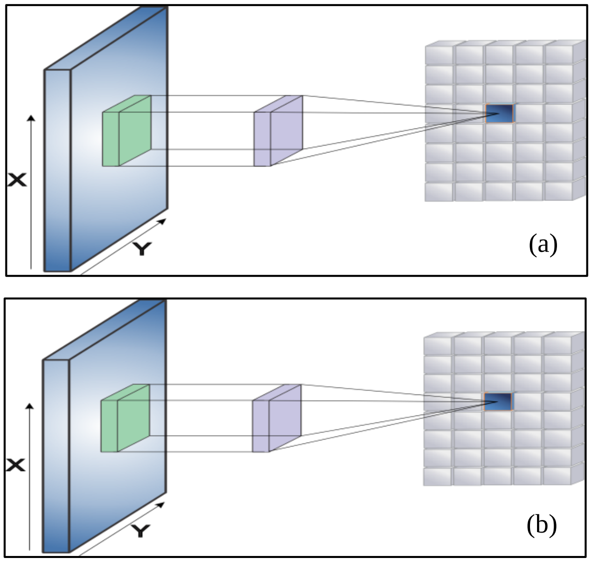


Fig. 3. CNN's layers (a) convolutional (b) max-pooling

It is clear the subsequent of alternative between these kinds of layers; it quickly diminishes the input images after few stages of CNN leading to losing vital information useful for final stage of classification. Specifically this work is dealing with small image sizes as will be obtained later. All the benchmarks used in this work have image sizes of 32x32 pixels. Consequently the small features will be not available after few stages. Therefore an elegant model of CNN architecture is proposed in this work as shown in fig. 4. It is clear that new model propose different connection than standard connection of conventional CNN. Some layers are received their connections not only directly from the layer below but also from two and three layers below. The reason for this kind of connections because small features within the input images can survive longer and will be part of the final scoring detection results. Furthermore, the first layers of CNN extract global features of input objects but as the images advance toward final fully connect layers, more accurate features are extracted.

C. Exploring Different CNN Sizes

In order to precisely analyze the influence of different CNN architectures, a new CNN architecture is proposed and carefully selected their parameter because same CNN architecture might work sufficiently for some tasks and inadequately for other tasks. Hence, in this part different deep model architectures is investigated that can fit for image recognition. Accordingly, CNN architectures are explored to be best suited for image classification. There are two model architectures are used in our experiments. They are shown in table I. In addition to the structure obtained in table 1, each network has more additional two fully connected layers build on the top of the final max-pooling layer. Then finally, soft-max layer is built on the top of final fully connect layer used for final scoring results. It is clear that there are two CNN architectures detailed in table1 called Network1 and Network2. It is obvious that network1 is smaller than the network2. Where, network1 consists of three convolutional layer and three max-pooling layers.

TABLE I. TWO CNN ARCHITECTURES. THE ABBREVIATION CON REFERS TO CONVOLUTION. XXYXY: X REPRESENTS NUMBER OF FEATURE MAPS AND Y IS THE KERNEL SIZE. LRN AND RELU ARE ABBREVIATION FOR LOCAL RESPONSE NORMALIZATION AND RECTIFIED LINEAR UNIT RESPECTIVELY

Model name	Input size	Con1/pool1	Con2/pool2	Con3/pool3	Con4/pool4	Con5/pool5
Network1	32x32	192x5x5,str:1, ReLU	256x5x5,str:1, ReLU	192x3x3,str:1, ReLU	-	-
		2x2, LRN	2x2, LRN	2x2, LRN	–	-
Network2	32x32	192x5x5,str:1, ReLU	256x1x1,str:1, ReLU	384x1x1,str:1, ReLU	256x1x1, str:1, ReLU	192x3x3, str:1, ReLU
		3x2	3x2	–	–	Spp layer

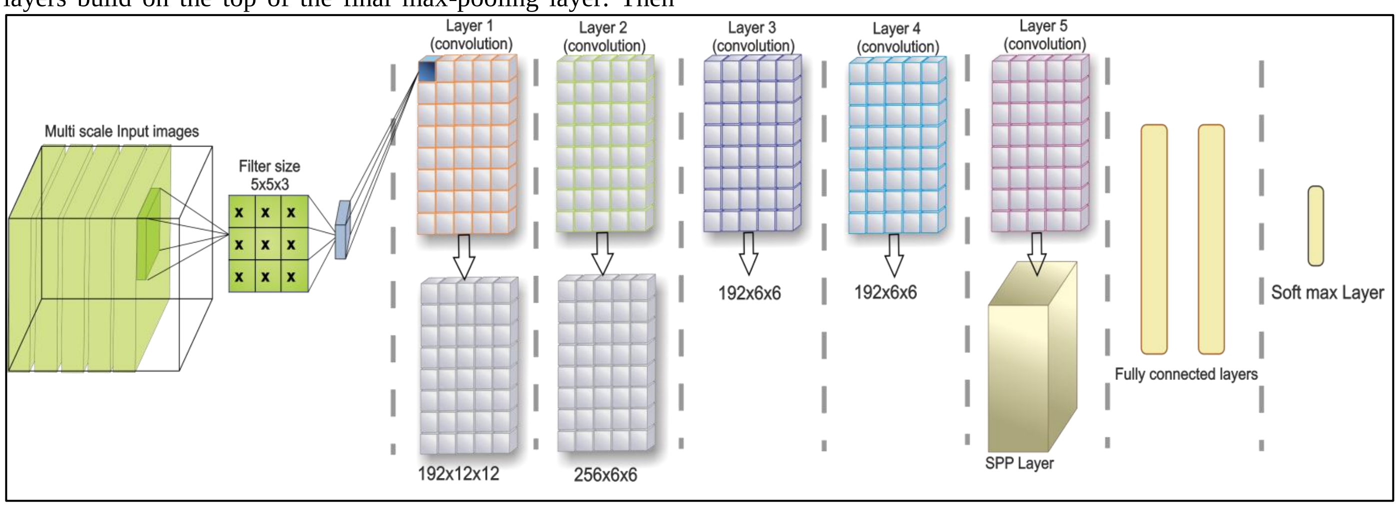


Fig. 4. CNN with five convoutional layers

V. Experiments Setup

In order to evaluate the proposed architecture models, extensive experiments are conducted on different challenge datasets. The most popular datasets are used for evaluation. MNIST, CIFAR-10, and CIFAR-100 are the benchmarks used in this work. To obtain the challenge accompanied with those datasets, next parts explain the related details for the datasets such as size number of image samples. It is worth mentioning that data augmentation is not used in these experiments.

A. MNIST dataset

MNIST [18] is a hand written digits 0-9. The dataset consists of 60000 samples. 50000 samples are used for training and the rest used for testing. All samples have the same size which is 28x28 pixels. The pixels are scaled to be between [0, 1] before the training. There is no preprocessing or data augmentation used in this work. The first CNN, which is named network1, structure is 192C-192S-256C-256S-192C-192C-200F-128F-10-soft-max, where C stands for Convolution layer, S is for subsampling layer, and F is for full conned layer. In this dataset, the size of mini-batches is 128 images. Test accuracy is 0.9961 % for MNIST dataset. This result is superior comparing with results [44]. A summary of the best published results on MNIST dataset is shown in Table II. Network2 which has a structured described as 192C-192S-256C-256S-384C-256C-192C-192S-400F-128F-10-soft-max achieves lower results than the prior model because MNIST might not require large network. The result achieved using network2 is 0.9958 on MNIST dataset. Comparing with other results, Table II shows the final results on MNIST. From Table II, it is obvious that both introduced morels achiever better results than what have been accomplished by Hayder et al. in [44] using their method hybrid training algorithm called Hybrid PSO-SGD which represent training algorithm using Particle Swarm Optimization and Scholastic Gradient Descent.

TABLE II. Results on MNIST Dataset

Method	Ref. #	Test Accuracy
Unsupervised Learning	[21]	0.64
What is the Best Multi-Stage	[22]	0.53
2-Layer CNN + 2-Layer NN	[23]	0.53
Stochastic Pooling	[23]	0.47
NIN + Dropout	[23]	0.47
Conv. maxout + Dropout	[24]	0.45
Hybrid PSO-SGD	[44]	0.43
Network1	**Ours**	**0.39**
Network2	**Ours**	**0.42**

B. CIFAR-10 Dataset

CIFAR-10 dataset consists of 10 classes of natural 32x32 RGB images with 50,000 samples for training and 10,000 samples for testing [19]. The same structure of network1 is used first for evaluation. The same steps are followed as in MNIST for CNN training. The performance achieved on this dataset is 86.73%.

On the other hand, the test accuracy on CIFAR-10 using network2 is 88.13% which is higher than network1 because CIFAR-10 is more challenge dataset than MNIST. Thus it requires more complicated structure. From table III, it is evident that the proposed method surpasses the other state-of-the-art works.

TABLE III. Test Set Accuracy Rates on CIFAR-10 Dataset

Method	Reference #	Accuracy
Tiled CNN	[25]	73.10
Improved LCC	[26]	74.50
KDES-A	[27]	76.00
PCANet-2 (combined)	[28]	78.67
PCANet-2	[28]	77.14
K-means (Triangle, 4000 features)	[29]	79.60
Cuda-convnet2	[30]	82.00
Hybrid PSO-SGD	[44]	82.41
Network1	**[ours]**	**86.73**
Network2	**[ours]**	**88.13**

C. CIFAR-100

CIFAR-100 is one of the most challenge dataset and it has 100 classes. Images are similar to CIFAR-10 even with size. However, the main difference is that number of image samples per class are very few comparing with CIFAR-10. The total number of images is 50,000 training examples. Thus each class has 500 samples only. Testing samples has 10,000 samples. Like CIFAR-10, the pixels are scaled to be between [0, 1] before the training. Since CIFAR-100 is similar to CIFAR-10 are similar, the same setting of CNN was used for both networks.

Table IV shows the final results achieved using the proposed two models. The first network achieves 53.52% test accuracy on CIFAR-100 while network2 achieves higher accuracy which is 59.85%.

TABLE IV. Test Set Accuracy Rates on CIFAR-100 Dataset

Method	Reference #	Accuracy
CONV. NET + PROBOUT	[45]	61.86%
Baseline + learned tree	[46]	63.15%
NOMP encoder	[47]	60.8%
Stochastic Pooling	[23]	57.49%
NIN	[3]	64.32%
Smooth Pooling Regions	[48]	56.29%
Beyond Spatial Pyramids	[49]	54.23%
Maxout Networks	[5]	61.43%
Network1	ours	53.52%
Network2	ours	59.85%

VI. Conclusion

In this work, image recognition using the deep neural network is introduced. Different model architectures are proposed by incorporating different prior elegant CNNs. Specifically both NIN and SPPnet are incorporated in a single unified model that achieves superior results comparing to former results. Then a new model is presented and outperforms prior work and accomplishes state-of-the-art results on the datasets. Also, different model architectures are introduced, and extensive parameters are discussed that can influence model performance. Deeper exploring different parameters that can be suited for CNN recognition model are presented as well. For evaluation, the experiments are conducted on challenge datasets. MNIST, CIFAR-10, and CIFAR-100 are the datasets used in this work.

FUTURE WORK

In feature work, more effort will be devoted in exploring more powerful network to handle more challenge tasks. More enhancements can be achieved by utilizing more technique to be recruited together and implemented the final model. Future works could also include more details such as reporting time consumption for each method and whether it is suitable for real-time applications or not. Also, those implemented models can be re-adapted to be used in object detection tasks.

REFERENCES

[1] S. L. Phung and A. Bouzerdoum, "A pyramidal neural network for visual pattern recognition," IEEE Transactions on Neural Networks, vol. 27, no. 1, pp. 329343, 2007

[2] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, Lior Wolf , "DeepFace: Closing the Gap to Human-Level

[3] Min Lin, Qiang Chen, and Shuicheng Yan "Network In Network" arXiv 1312.4400v3,4 Mar 2014

[4] Dan Cires¸an, Ueli Meier and Jurgen Schmidhuber, "Multi-column Deep Neural Networks for Image Classification" CVPR 2012

[5] Julien Mairal, Piotr Koniusz, Zaid Harchaoui, and Cordelia Schmid, "Convolutional Kernel Networks" arXiv 14 Nov 2014 .

[6] Ian J. Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, Yoshua Bengio "Maxout Networks" ICML 2013

[7] Dumitru Erhan, Christian Szegedy, Alexander Toshev, and Dragomir Anguelov, "Scalable Object Detection using Deep Neural Networks" CVPR, 2014.

[8] Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey. ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25 (NIPS'2012). 2012.

[9] Marc'Aurelio Ranzato, Fu-Jie Huang, Y-Lan Boureau, Yann LeCun, "Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition" CVPR, 2007

[10] Matthew D Zeiler and Rob Fergus. Stochastic pooling for regularization of deep convolutional neural networks. arXiv preprint arXiv:1301.3557, 2013.

[11] Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. arXiv preprint arXiv:1302.4389, 2013.

[12] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov" Dropout: A Simple Way to Prevent Neural Networks from Overfitting" Journal of Machine Learning Research 15 (2014) 1929-1958.

[13] Fabien Lauer, Ching Y. Suen, and G´erard Bloch "A trainable feature extractor for handwritten digit recognition" Journal Pattern Recognition 2007.

[14] Chen-Yu Lee, Saining Xie, Patrick Gallagher, Zhengyou Zhang, Zhuowen Tu, " Deeply-Supervised Nets "NIPS 2014

[15] M. Fischler and R. Elschlager, "The representation and matching of pictorial structures," IEEE Transactions on Computer, vol. 22, no. 1, 1973

[16] Kevin Jarrett, Koray Kavukcuoglu, Marc'Aurelio Ranzato and Yann LeCun " What is the Best Multi-Stage Architecture for Object Recognition?" ICCV'09, IEEE, 2009.

[17] Kaiming, He and Xiangyu, Zhang and Shaoqing, Ren and Jian Sun "Spatial pyramid pooling in deep convolutional networks for visual recognition" European Conference on Computer Vision, 2014

[18] Ross Girshick, "Fast R-CNN "arXiv preprint arXiv:1504.08083, 2015

[19] X. Wang, M. Yang, S. Zhu, and Y. Lin. Regionlets for generic object detection. In ICCV, 2013. 8

[20] Karen Simonyan and Andrew Zisserman "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION" arXiv:1409.1556v5 [cs.CV] 23 Dec 2014.

[21] C. Couprie, C. Farabet, L. Najman, and Y. LeCun. Indoor semantic segmentation using depth information. Internatinal Conference on Learning Representation, 2013. 2.

[22] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. CoRR, abs/1311.2524, 2013. 4

[23] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. FeiFei. Imagenet: A large-scale hierarchical image database. In CVPR, 2009. 2.

[24] L. N. Clement Farabet, Camille Couprie and Y. LeCun. Learning hierarchical features for scene labeling. PAMI, 35(8), 2013. 1, 2

[25] Honglak Lee, Roger Grosse, Rajesh Ranganath, and Andrew Y. Ng "Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations".

[26] Pedro F. Felzenszwalb, Ross B. Girshick, David McAllester and Deva Ramanan ," Object Detection with Discriminatively Trained Part Based Models"

[27] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in IEEE Conference on Computer Vision and Pattern Recognition, 2003

[28] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun. OverFeat: Integrated recognition, localization and detection using convolutional networks. arXiv preprint arXiv:1312.6229, 2013. 1, 2.

[29] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," International Journal of Computer Vision, vol. 77, no. 1, pp. 259–289, 2008.

[30] S. Lazebnik, C. Schmid, and J. Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In CVPR, 2006. 1, 2, 5, 6, 7

[31] G. Griffin, A. Holub, and P. Perona. Caltech-256 object category dataset. Technical report, CalTech, 2007.

[32] A. S. Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. CNN features off-the-shelf: an astounding baseline for recognition. arXiv preprint arXiv:1403.6382, 2014. 1, 2.

[33] Y. Lin, T. Liu, and C. Fuh. Local ensemble kernel learning for object category recognition. In CVPR, 2007

[34] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer. Weak hypotheses and boosting for generic object detection and recognition. In ECCV, 2004.

[35] K. Grauman and T. Darrell. The pyramid match kernel: Discriminative classification with sets of image features. In ICCV, 2005.

[36] P. Sermanet, S. Chintala, and Y. LeCun. Convolutional neural networks applied to house numbers digit classification. In ICPR, 2012

[37] Q. V. Le, J. Ngiam, Z. Chen, D. Chia, P. W. Koh, and A. Y. Ng, "Tiled convolutional neural networks," in NIPS, 2010.

[38] Tsung-Han Chan, Kui Jia, Shenghua Gao, Jiwen Lu, Zinan Zeng, and Yi Ma "PCANet: A Simple Deep Learning Baseline for Image Classification? "arXiv:1404.3606v2 [cs.CV] 28 Aug 2014.

[39] Adam Coates, Honglak Lee, and Andrew Y. Ng "An Analysis of Single-Layer Networks in Unsupervised Feature Learning" In AISTATS 14, 2011.

[40] Marc'Aurelio Ranzato, Fu-Jie Huang, Y-Lan Boureau, Yann LeCun, "Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition" CVPR, 2007

[41] Matthew D Zeiler and Rob Fergus. Stochastic pooling for regularization of deep convolutional neural networks. arXiv preprint arXiv:1301.3557, 2013.

[42] Ian J Goodfellow, David Warde-Farley, Mehdi Mirza, Aaron Courville, and Yoshua Bengio. Maxout networks. arXiv preprint arXiv:1302.4389, 2013.

[43] Jia, Yangqing and Shelhamer, Evan and Donahue, Jeff and Karayev, Sergey and Long, Jonathan and Girshick, Ross and Guadarrama, Sergio and Darrell, Trevor "Caffe: Convolutional Architecture for Fast Feature Embedding" arXiv preprint arXiv:1408.5093, 2014.

[44] Hayder M. Albehadili and Naz Islam "Hybrid Algorithm For the Optimization of Training Convolutional Neural Network". Volume 6 IJACSA,No. 10, October 2015

[45] Jost T. S. and Marth R, "improving deep neural networks with probabilistic Maxout units ", ICLR 2014

[46] Nitish Srivastava and Ruslan SalaKhudinov "Discriminative transfer learning with tree-based priors, NIPS,2013

[47] Tsun-Han Lin and H. T. Kung “stable and efficient representation learning with nonnegativity constraints”, ICML 2014.

[48] M. Malinowski and M. Fritz “learning smooth pooling regions for visual recognition”, BMVC 2013

[49] Y. Jia, C. Huang, and T. Darrell, “Beyond spatial pyramids: Receptive field learning for pooled image features”, CVPR 2012

L-Bit to M-Bit Code Mapping

To Avoid Long Consecutive Zeros in NRZ with Synchronization

Ruixing Li
Department of Electrical Engineering,
University of Nevada, Las Vegas,
Las Vegas, U.S.

Yun Lun
Department of Electrical Engineering,
University of Nevada, Las Vegas,
Las Vegas, U.S.

Shahram Latifi
Department of Electrical Engineering,
University of Nevada, Las Vegas,
Las Vegas, U.S.

Ming Lun
Department of Electrical Engineering,
University of Nevada, Las Vegas,
Las Vegas, U.S.

***Abstract*—we investigate codes that map L bits to m bits to achieve a set of codewords which contain no consecutive n "0"s. Such codes are desirable in the design of line codes which, in the absence of clock information in data, provide reasonable clock recovery due to sufficient state changes. Two problems are tackled- (i) we derive n_{min} for a fixed L and m and (ii) determine m_{min} for a fixed L and n. Results benefit telecommunication applications where clock synchronization of received data needs to be done with minimum overhead.**

Keywords—overhead; mapping; synchronization; consecutive "0"

I. Introduction and Background

In serial communications, data is transferred on a medium that carries a signal varying with time. For digital signal, each bit is represented as a high or low voltage for a fixed amount of time. We call this time period a clock cycle. The clock of the communication line is very important as it tells the transmitter when to transmit a new bit, and it tells the receiver when to read. For short distance transfer, such as communication within a digital system, we can have a clock signal between the transmitter and receiver to synchronize the clock; for example, the Serial Peripheral Interface (SPI) uses a clock signal for clock synchronization. However when it comes to long distance communication, adding signaling only for clock synchronization consumes part of the bandwidth. It is impossible to exactly match the clock speed for the transmitter and receiver. On the other hand, employing codes that contain explicit clock information (ex. Manchester coding) will waste half of the available bandwidth [1]. In practice, the clock information is embedded within the data so that, at the receiver end, the clock can be extracted and used to clock in the received data (using devices such as Phased Locked Loop or PLL). Nonetheless, having a long period of flat signal (which may correspond to consecutive "0"s) may cause the synchronization to be lost. For that purpose, the signal that carries the data must also have sufficient transitions or state changes to allow a PLL to lock onto the incoming data. In the event that a long sequence of "0"s is encountered, there will be a risk of losing synchronization.

As one of the scrambling techniques for data encoding, transmitter should provide sufficient amount of signal transitions for the receiver to maintain clock synchronization [2]. Line coding is applied on data before transmission especially in High Speed Serial Links to ensure a maximum Run Length (RL) to guarantee frequent transitions for Clock and Data Recovery (CDR) in asynchronous links [3], for example, B8ZS and HDB3, which substitute a long sequence with a code violation of the encoding rule. These types of techniques either require increase of signal rate for the same data rate, or require more than 2 signal levels to represent binary data. For Manchester and Differential Manchester, the signal rate is twice the data rate (50% overhead). For B8ZS and HDB3, having 3 signal levels to represent a single binary bit creates a 33% overhead. Though line codes can generate adequate timing information for clock recovery and error detection [5] [6], it usually comes at the cost of additional bits. In this paper, we will discuss how to minimize the overhead with the same clock recovery performance.

Another technique is to eliminate long sequence of zeros by encoding the data so that the transmitted data does not contain long sequences of *"0"*s. The 8b/10b encoding [4] which is widely used, adds 2 bits for every 8 bits resulting in 2/8 = 25% overhead while ensuring a maximum RL of 5. One other example would be mapping 4-bit data to 5-bit codes such that a sequence of *3 "0"*s is avoided ($L = 4, m = 5, n = 3$). There are total $2^4 = 16$ possible codes in 4-bit data.

In 5-bit code space, we have *24 (32 – 8)* codes without *"000"* sequence available. So mapping 4-bit all data with 5-bit codes is possible. Table I shows the 5-bit codes with the absent *"000"* patterns.

TABLE I. 5-Bit Codes with the Absent "000" Patterns

00100	00101	00110	00111	01001	01010
01011	01100	01101	01110	01111	10010
10011	10100	10101	10110	10111	11001
11010	11011	11100	11101	11110	11111

The overhead of this technique is *1/5 = 20%*[1], which is

[1] Calculated by (Total transmitted data size – Actual represented data size)/(Total transmitted data size)

lower than the *50%* of forced transition techniques and the *33%* of substitution techniques. We can try to reduce the overhead of code mapping techniques by mapping larger size data. Questions of interest might be - Can we do 9-bit to 10-bit mapping? If not, how about 9-bit to 11-bit, 61-bit to 64-bit, etc.

This paper proposes an empirical method of calculating the minimum overhead to avoid a given number of consecutive "0"s. The rest of this paper is organized as follows: Section 2 discusses the basic theory of avoid long sequence of "0"s. Section 3 introduces the methodology to achieve two empirical formulas for our concerns. The results and conclusion are given in Section 4 and 5 separately.

II. Theory

The research question here is- Given a specific size of a code, what is the smallest overhead to avoid a given number of consecutive *"0"*s.?

Example 1. We are given a 9-bit code and we want to avoid the sequence *"000".* First we will check if the 10-bit code has enough space to hold the 9-bit code and also avoid the sequence *"000".* If 10-bit is not possible, we will consider the 11-bit code and continue checking until we find the smallest size of code that can hold the 9-bit code and avoid the sequence *"000".*

To check if 10-bit code is enough, first we will enumerate the codes in the 10-bit code that has *"000"* sequence in. The calculations are depicted in Table II. Note that in the patterns given in the table, X can be 0 or 1 and each line must exclude the cases that had been counted in the previous lines.

TABLE II. Enumeration of 10-Bit Codes that Contain "000"

Pattern	The number of occurrences
XXXXX XX000	2^7=128
XXXXX X0001	2^6=64
XXXXX 0001X	$2^5 \times 2^1$=64
XXXX0 001XX	$2^4 \times 2^2$=64
XXX00 01XXX	$2^3 \times (2^3-1)$=56
XX000 1XXXX	$2^2 \times (2^4-3)$=52
X0001 XXXXX	$2^1 \times (2^5-8)$=48
0001X XXXXX	$2^0 \times (2^6-20)=44$
Total	520

Note that for X's of length 3 or more on the right side, we have to exclude any codes that have the "000" sequence, because they were already covered in previous lines.

Subtracting the total number of *"000"* patterns- 520 from the total code space $2^{10}=1024$, we get only 504 codes which is not enough for mapping all 9-bit codes to 10-bit codes.

Generalization

To answer the general question of if it is possible to map all *L*-bit codes to *m*-bit codes that avoid sequence of n consecutive zeros, we first have to find the number of codes without *n*-zero sequence by subtracting the number of codes with *n*-zero sequence from the total *m*-bit code space 2^m.

To answer another question of finding the smallest number of consecutive *"0"*s, we can start with 2 zeros and work upward. Say if we cannot avoid 2 zeros, test if we can avoid 3, 4, 5, etc. Repeat until we can find that smallest number of consecutives *"0"*s we can avoid through L-bit to m-bit mapping.

To find the number of codes with sequence(s) of n *"0"*s in m-bit space, use a similar step from the 9B-10B mapping example to obtain the solution (Table III).

TABLE III. Enumeration of M-Bit Codes that Contain N Consecutive "0"S

Pattern	The number of occurrences	Distribute the multiplication	k
XXXXXXXXX…X00 …0	2^{m-n}	2^{m-n}	k=0
XXXXXXXX…X00 …01	$2^{m-n-1} \times [2^0 - f(0,n)]$	$2^{m-n-1} - 2^{m-n-1} \times f(0,n)$	k = 1
XXXXXXX…X00… 01X	$2^{m-n-2} \times [2^1 - f(1,n)]$	$2^{m-n-1} - 2^{m-n-2} \times f(1,n)$	k = 2
XXXXXX…X00…0 1XX	$2^{m-n-3} \times [2^2 - f(2,n)]$	$2^{m-n-1} - 2^{m-n-3} \times f(2,n)$	k = 3
XXXXX…X00…01 XXX	$2^{m-n-4} \times [2^3 - f(3,n)]$	$2^{m-n-1} - 2^{m-n-4} \times f(3,n)$	k = 4
⋮			
X00…01XX…XXX XXX	$2^1 \times [2^{m-n-2} - f(m-n-2,n)]$	$2^{m-n-1} - 2^1 \times f(m-n-2,n)$	k= m-n-1
00…01XX…XXXX XXX	$2^0 \times [2^{m-n-1} - f(m-n-1,n)]$	$2^{m-n-1} - 2^0 \times f(m-n-1,n)$	k = m – n

We can see that calculating $f(m,n)$ requires recursive calculation of the number of codes with n-zero sequence in the code lengths less than m. We must define the basis for the recursive function, otherwise we will go to endless loop of calculations. We know that there cannot exist n-zero sequence in the code if the code length m is shorter than n bits. So $f(m,n)=0$, for $m<n$.

For $m \geq n$, continue our generalization. Adding the terms and simplifying gives

$$2^{m-n} + (m-n) \times 2^{m-n-1} - \sum_{k=1}^{m-n} 2^{m-n-k} \times f(k-1,n)$$

Now that we obtain the piecewise recursive function $f(m,n)$:

$$f(m,n) = \begin{cases} 0, & m < n \\ 2^{m-n} + (m-n) \times 2^{m-n-1} - \sum_{k=1}^{m-n} 2^{m-n-k} \times f(k-1,n), & m \geq n \end{cases}$$

The calculation of the function $f(m,n)$ seems very complicated with the summation. We can make the calculation easier and more efficient by observing the following 2 special cases of m and n.

Case (i) if $m=n$, we know that there is only one code that has the n-zero sequence; that is the n-zero sequence itself.

Case (ii) if $n < m \leq 2n$, the last addition to the sum is $2^0 \times f(m-n-1,n)$. Let's look at the last largest $m=2n$, we substitute m with 2n and obtain $2^0 \times f(n-1,n)$, the recursive call to the function returns *0* because $n-1<n$. Similarly, all m value between n and *2n* make recursive call to function *f* with first argument less than second argument n, which will ultimately give *0* as the result. So for $n<m\le 2n$, the summation evaluates to *0.*

By separating the domain of the function, we produce a new formula with 4 pieces but is easier to calculate or more efficient to compute digitally.

$$f(m,n)=\begin{cases} 0, & m<n \\ 1, & m=n \\ 2^{m-n}+(m-n)\times 2^{m-n-1}, & n<m\le 2n \\ 2^{m-n}+(m-n)\times 2^{m-n-1} \\ -\sum_{k=1}^{m-n} 2^{m-n-k}\times f(k-1,n), & m>2n \end{cases}$$

We can use the code in Appendix to calculate the function *f.* Let $m=10, n=3$, the function returns *520* which match our previous calculation for 9-bit to 10-bit mapping example. We can also check our result by counting the number of codes with "000" pattern by using a brute force checking program, created by Edgar Solorio (See Appendix) Our result matches the number counted by this checking program ($alpha=520$). Similarly when $m=11, n=3$, the function returns *1121*, that means there are *927* codes available for mapping. While this is not enough to map 10-bit, it is sufficient for 9-bit codes, giving *18.2%* overhead.

Generally speaking, for given m and n,

$$L_{max}=floor\{log_2[2^m-f(m,n)]\} \quad (1)$$

Floor is the greatest integer function, mapping a real number to the largest previous integer.

We define the minimum overhead bits

$$h=m-L_{max} \quad (2)$$

If we try *19*-bit to *22*-bit mapping, which is possible, there is only *13.6%* (*3* bits) overhead. Similarly *64*-bit codes has about $2^{56.4}$ codes without *"000"*, while we cannot map *61*-bit codes to *64*-bit codes, there is enough to map 56-bit codes. Also 8-bit to 9-bit mapping is possible, 9-bit space has *238* codes with *"000",* leaving *274* codes available to map *8*-bit (*256* possible) codes. The overhead of for 8B9B is *11.1%.* Since 9B-10B is impossible, code mapping with *1*-bit overhead stops at with *8*-bit to *9*-bit mapping.

III. Methodology

The result from the 9B-10B example in the Work section shows that mapping from *9*-bit codes to *10*-bit codes cannot avoid all codes with 3 consecutive zeros. If we want to avoid *3* consecutive zeros, the minimum overhead to map *9*-bit code is *2* bits. The following 2 questions are our main concerns about avoiding consecutive zero level signal transmitted in regard to maintain synchronization.

A. *For a given pattern length L and mapping from L bits to $m=L+h$ bits, what is the minimum number of consecutive "0"s we can avoid? (Fixed L and m, find n_{min})*

The code in Appendix shows how to solve this question. Table IV shows the minimum avoidable zeros with *1* to *9* bits overhead for mapping data of lengths from $L=2$ to *24* bits. The jumps in n_{min} are highlighted and bold faced.

TABLE IV. Minimum Avoidable Zeros with 1 to 9 Bits Overhead for Mapping Data of Lengths from $L=2$ To **24** Bits

L	h=1	h=2	h=3	h=4	h=5	h=6	h=7	h=8	h=9
2	2								
3	2								
4	**3**	2							
5	3	2							
6	3	**3**	2						
7	3	3	2						
8	3	3	3	2					
9	**4**	3	3	2					
10	4	3	3	**3**	2				
11	4	3	3	3	2				
12	4	3	3	3	2				
13	4	3	3	3	**3**	2			
14	4	3	3	3	3	2	2		
15	4	3	3	3	3	**3**	2		
16	4	3	3	3	3	3	2		
17	4	**4**	3	3	3	3	**3**	2	
18	4	4	3	3	3	3	3	2	
19	4	4	3	3	3	3	3	**3**	2
20	4	4	3	3	3	3	3	3	2
21	**5**	4	3	3	3	3	3	3	2
22	5	4	3	3	3	3	3	3	**3**
23	5	4	3	3	3	3	3	3	3
24	5	4	**4**	3	3	3	3	3	3

If we rearrange and extend the data, we can get the following table:

TABLE V. Minimum L for Specific N_{MIN} and H

h	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9
1	2	4	9	21	43	88	177	355
2	4	6	17	38	82	171	348	702
3	6	8	24	56	122	254	519	
4	8	10	31	74	161	337	690	
5	10	13	38	92	201	420	861	
6	13	15	46	110	240	503		
7	15	17	53	127	279	586		
8	17	19	60	145	319	669		
9	19	22	68	163	358	753		
10	22	24	75	181	397	836		
11	24	26	82	199	437	919		
12	26	28	89	216				
13	28	31	97	234				
14	31	33	104	252				
15	33	35	111	270				
16	35	38	118	288				
17	38	40	126	305				
18	40	42	133	323				
19	42	44	140					
20	44	47	148					

We plot the data both in horizontal (Fig. 1) and vertical direction (Fig. 2),

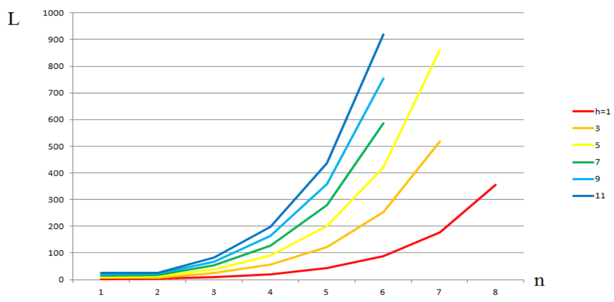


Fig. 1. Plotting using data from Table V (horizontal)

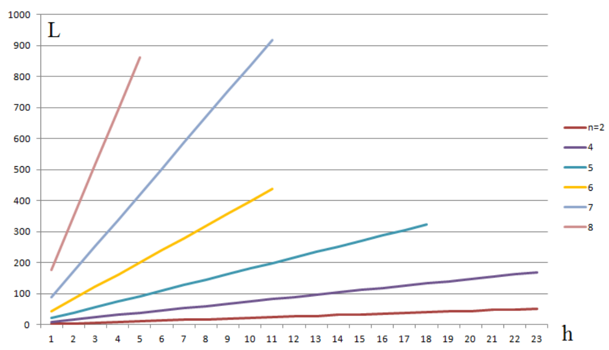


Fig. 2. Plotting using data from Table V (vertical)

Obviously, we can observe that L is approximating linear with h and is relatively exponential with n. This means for a fixed n, the overhead bits h should be proportional with L. We can assume that

$$L = \left(e^{(n-c)/a} - b\right) * h \qquad (3)$$

a, b, c in this formula is the coefficient to be determined.

We can choose any three points in Table V to determine the coefficient. For example, we substitute *(3, 20, 47)*, *(4, 20, 148)* and *(5, 11, 199)* [in *(n, h, L)* order] in (3), we can get $a = 1.333, b = 2.171, c = 0.988$. Therefore,

$$\left(\mathrm{e}^{\frac{\mathrm{n}-0.988}{1.333}} - 2.171\right) * \mathrm{h}$$

Thus,

$$n = 1.333 * \ln\left(\frac{L}{h} + 2.171\right) + 0.988 \qquad (4)$$

When L, m is given, we can determine n_{min} via (4) and $n_{min} = floor(n)$.

B. *For a given number of consecutive "0"s to avoid, how to minimize the overhead? Solve the problem for a specific case of avoiding two "0"s first. Under what conditions can we map* $"L"$ *to* $"L+1"$*? If not, what about* $"L"$ *to* $"L+2"$ *or* $"L"$ *to* $"L+3"$*? (Fixed* L, n*, find* h_{min}*)*

The code in Appendix shows how to solve question 2. Table VI shows the minimum overheads to avoid n zeros with $n = 2$ to 10 for mapping data of lengths from $L = 4$ to 24 bits. The jumps in overheads are highlighted and bold faced.

TABLE VI. MINIMUM OVERHEADS TO AVOID N CONSECUTIVE ZEROS

L	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9	n=10
4	2	1	1	0	0	0	0	0	0
5	2	1	1	**1**	0	0	0	0	0
6	**3**	1	1	1	**1**	0	0	0	0
7	3	1	1	1	1	**1**	0	0	0
8	**4**	1	1	1	1	1	**1**	0	0
9	4	**2**	1	1	1	1	1	**1**	0
10	**5**	2	1	1	1	1	1	1	**1**
11	5	2	1	1	1	1	1	1	1
12	5	2	1	1	1	1	1	1	1
13	**6**	2	1	1	1	1	1	1	1
14	6	2	1	1	1	1	1	1	1
15	**7**	2	1	1	1	1	1	1	1
16	7	2	1	1	1	1	1	1	1
17	**8**	**3**	1	1	1	1	1	1	1
18	8	3	1	1	1	1	1	1	1
19	**9**	3	1	1	1	1	1	1	1
20	9	3	1	1	1	1	1	1	1
21	9	3	**2**	1	1	1	1	1	1
22	**10**	3	2	1	1	1	1	1	1
23	10	3	2	1	1	1	1	1	1
24	**11**	**4**	2	1	1	1	1	1	1

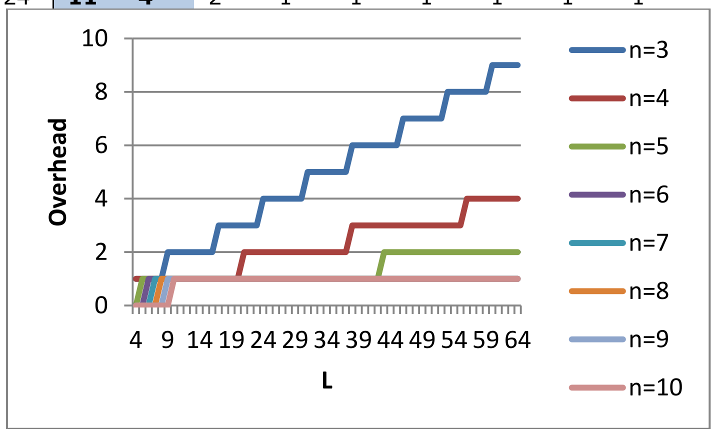


Fig. 3. Plotting using data from Table VI

If we rearrange and extend the data of TABLE VI, we can get the following table:

TABLE VII. MINIMUM L FOR SPECIFIC N AND H

h	n=2	n=3	n=4	n=5	n=6	n=7	n=8	n=9
1	2	3	4	5	6	7	8	9
2	4	9	21	43	88	177	355	710
3	6	17	38	82	171	348	702	
4	8	24	56	122	254	519		
5	10	31	74	161	337	690		
6	13	38	92	201	420	861		
7	15	46	110	240	503			
8	17	53	127	279	586			
9	19	60	145	319	669			
10	22	68	163	358	753			
11	24	75	181	397	836			
12	26	82	199	437	919			
13	28	89	216					
14	31	97	234					
15	33	104	252					
16	35	111	270					
17	38	118	288					
18	40	126	305					
19	42	133	323					

This table is slightly different from Table V. We can use similar procedure and get

$$h = \frac{L}{1.009e^{0.75n} - 2.171} + 1 \quad (5)$$

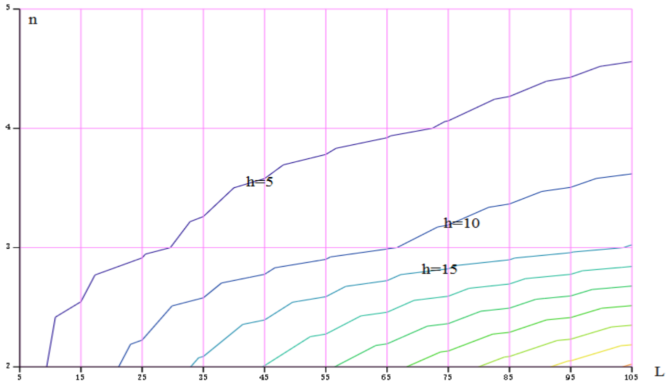


Fig. 4. Minimum overheads versus message length L and number of consecutive “0”s to be avoided

IV. Resulds

We can check the accuracy of (4) by different check point:

TABLE VIII. Comparison of Theoretical and Calculation N by Fixed L and H

Point(L, h)	(44, 19)	(60, 8)	(145, 8)	(437, 11)	(918, 11)	(861, 5)	(702, 2)
Actual n	3	4	5	6	7	8	9
Approximate According (4)	2.99	4.01	5.00	5.97	6.92	7.87	8.81
deviation	0.33 %	0.33 %	0.04 %	0.55 %	1.14 %	1.65 %	2.12 %

From this table, we can see, when n is between 3 and 9, (4) is accurate enough to determine the minimum consecutive *“0”s* can be avoided for fixed L and m.

We can also check the availability of (5) by different check points in TABLE IX:

From this table, we can see, when n is between 3 and 6, this formula is accurate enough to determine the minimum required overhead bits h to avoid n consecutive “0”s for fixed message length L.

TABLE IX. Comparison of Theoretical and Calculation N by Fixed L and H

L	88	82	56	74	92	46
n	6	5	4	4	4	3
Actual h	2	3	4	5	6	7
h from (5)	1.992597	3.013121	4.094725	5.089458	6.084191	7.214431
deviation	-0.37%	0.44%	2.37%	1.79%	1.40%	3.06%

V. Conclusion

We have considered the problem of L to m mapping to avoid a set of n consecutive “0”s. We derived two formulas to calculate (i) the minimum number of consecutive “0”s that can be avoided for fixed L and m (4) and (ii) the minimum overhead required to avoid a given number of consecutive “0”s with fixed L (5). We found the exact values for small values of L, m and n (Table IV and Table VII). For very long messages, we used the empirical results and combination of several tables to arrive at a formula that will give the desired answer with close approximation.

One may think of splitting a long code into smaller codes and using the results for small values to obtain the parameters for the long code. For example, the splitting of 56-bit code into 8*7-bit codes can simplify the calculation but will not work since a potential problem can occur: Even if all 8 7-bit codes have no “000”, when the frame size is more than 7 bits (e.g. 64 bits), there can exist consecutive “000” in the end of a 7-bit code and the start of another consecutive 7-bit code.

The results obtained can find applications in coding and communication where the synchronization of the transmitter and receiver is of primary concern.

Acknowledgment

This work was supported in part by the NSF EPSCoR grant under Award # EPS-IIA-1301726.

References

[1] V.Sneha Latha et al. “Performance Evaluation of Different Line Codes”.Indian Journal of Computer Science and Engineering (IJCSE). Vol. 2 No. 4, 2011

[2] Stallings, W. “Data and Computer Communications”. 10th Ed. 2013.

[3] J. Saadé et al. “Low Overhead, DC-Balanced and Run Length Limited Line Coding” IEEE 19th Workshop on Signal and Power Integrity (SPI), 2015.

[4] P.A. Franszek and A.X. Widmer, “A DC-Balanced, Paritioned-Block, 8B/10B Transmission Code”, IBM Journal of research and development, Volume 27, Number 5, September 1983.

[5] J. G. Proakis, Digital Communication, 4th ed., Mc-Graw Hill, New York, 2001.

[6] K. W. Cattermole, “Principles of digital line coding,” Int. J. Electron., vol. 55, no. 1, pp. 3–33, July 1983.

Appendix

```
int numOfOccurrences(int m, int n)
{
    if (m < n)          // m < n
        return 0;
    else if (m == n)    // m = n
        return 1;
    else if (m <= 2*n)  // n < m ≤ 2n
    {
        int result;
        result = twoToPowerOf(m-n) + (m-n) * twoToPowerOf(m-n-1);
        return result;
    }
    else                // m > 2n
    {
        int result;
        int *s = new int[m-n+1];
        s[0] = twoToPowerOf(m-n) + (m - n) * twoToPowerOf(m-n-1);
        result = s[0];
        for (int i = 1; i <= m-n; i++)
        {
            s[i] = numOfOccurrences(i-1, n) * twoToPowerOf(m-n-i);
            result -= s[i];
        }
        return result;
    }
}
```

C++ Code for function f (named numOfOccurrences)

Note: The codes shown in this report emphasize on basic idea to implementation. The actual result from the code may be incorrect due to limitation in range in of int data type. Also, a cached table may be necessary to reduce repeated calculations and to improve performance.

twoToPowerOf is a simple function that returns the power of n without importing the C Math library.

```
int twoToPowerOf(int n)
{
    if (n < 0)
    {
        std::cout << "Cannot calculate negative or fractional powers"
                  << std::endl;
        exit(1);
    }
    else
    {
        int result = 1;
        result = result << n;
        return result;
    }
}
```

Code for function twoToPowerOf(n)

Programs to check number of codeword with n consecutive *"0"s*

```
Enter m (Q to quit):10
Enter n:3

# of "000s": in 10-bit space:
520
```

Program outputs for $m = 10, n = 3$

```
bitmapping — a.out — 80×24
Min-MBP:bitmapping min$ ./a.out
Enter the string to avoid
Ex.) 0000 or 1111

        avoid string = 000

Please input the bit mapping values for x and y
Ex.) For a 9bit to 10 bit mapping; x = 9, y = 10

        x = 9
        y = 10


Results:
--------------------------
alpha = 520
threshold = 512
avoid string = "000"
Exeuction Time = 0.002977 seconds

        9B to 10B mapping is IMPOSSIBLE
        Since alpha > threshold

--------------------------
```

Program outputs using code provided by Edgar Solorio [8]

```
// this function calculates the minimum number of consecutive zeros
we can avoid
// by mapping l-bit codes to m-bit codes.
int minimumAvoidableConsecutiveZeros(int L, int m)
{
    if (L >= m)
    {
        return -1; // you cannot avoid any consecutive zeros
        // moreover, if m < L, you cannot even map from L-bit to m-bit
anyway.
    }
    int minZeros =L;
    // starting from 2 zeros "00"
    for (int i = 2; i < L; i++)
    {
        if (twoToPowerOf(m) - numOfOccurrences(m, i) >
twoToPowerOf(L))
        {
            minZeros = i;
break;
        }
    }
    return minZeros;
}
```

Code to solve question 1

Code to solve Question 2

```
#define MAXIMUM_ALLOWED_OVERHEAD 5
#define MAXIMUM_ALLOWED_PERCENT_OVERHEAD 30
// this function calculates the minimum overhead by mapping l-
bit code
int minimumOverheadToAvoid_n_Zeros(int L, int n)
{
    if (n < 2)
        return -1;  // error, n must be at least 2

    int maximum_m = L + MAXIMUM_ALLOWED_OVERHEAD;
    int max_MP = (m *
(100+MAXIMUM_ALLOWED_PERCENT_OVERHEAD) / 100);
    if (max_MP > maximum_m)
        maximum_m  = max_MP;

    ULL numOfLBitCodes = twoToPowerOf(L);

    for (int i = L+1; i < maximum_m; i++)
    {
        if (twoToPowerOf(i) - numOfOccurrences(i, n) >
numOfLBitCodes)
        {
            return (i-L);
        }
    }
    return -1;  // reached maximum allowed overhead
}
```

A Structural Equation Model (SEM) of Governing Factors Influencing the Implementation of T-Government

Sameer Alshetewi
School of Computer Science, Engineering & Mathematics
Flinders University
Adelaide, Australia

Faten Karim
School of Computer Science, Engineering & Mathematics
Flinders University
Adelaide, Australia

Dr Robert Goodwin
School of Computer Scince, Engineering & Mathematics
Flinders University
Adelaide, Australia

Dr Denise de Vries
School of Computer Scince, Engineering & Mathematics
Flinders University
Adelaide, Australia

***Abstract*—Governments around the world have invested significant sums of money on Information and Communication Technology (ICT) to improve the efficiency and effectiveness of services been provided to their citizens. However, they have not achieved the desired results because of the lack of interoperability between different government entities. Therefore, many governments have started shifting away from the original concept of e-Government towards a much more transformational approach that encompasses the entire relationship between different government departments and users of public services, which can be termed as transformational government (t- Government). In this paper, a model is proposed for governing factors that impact the implementation of t-Government such as strategy, leadership, stakeholders, citizen centricity and funding in the context of Saudi Arabia. Five constructs are hypothesised to be related to the implementation of t-Government. To clarify the relationships among these constructs, a structural equation model (SEM) is utilised to examine the model fit with the five hypotheses. The results show that there are positive and significant relationships among the constructs such as the relationships between strategy and t-Government; the relationships between stakeholders and t-Government; the relationships between leadership and t-Government. This study also showed an insignificant relationship between citizens' centricity and t-Government and also an insignificant relationship between funding and t-Government. document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet.**

Keywords*—*t-Government; e-Government; Strategy; Stakeholders; Citizens' Centricity; Funding

I. Introduction

Many nations have put in place e-Government applications to enhance the efficiency of the public sector and streamline government systems to enable creating connections between different government organisations. To create further efficiencies, it has been recommended that countries move towards a goal of transformational government [1, 2]. Nowadays, the concept of a transformational government (t-Government) needs to be viewed on an international scale. In order to facilitate more centrally-connected and citizen-centric-government services, and put the needs of individuals and businesses at the center of online processes, many governments have started shifting away from the original concept of e-Government towards a much more transformational approach that considers the entire relationship between different government departments and users of public services, which can be termed as transformational government (t- Government) [3-5].

This paper identifies and analysis the governing factors (strategy, leadership, stakeholders, citizens' centricity and funding) that influence the implementation of t-Government in Saudi Arabia. The proposed model is empirically tested through an analysis data from 217 surveys from various organisations in Saudi Arabia using Structural Equation Modelling (SEM) technique and utilized AMOS version (22) tools. This study provides insights for government officials and decision- makers to understand and improve the level of t-Government implementation in government organisations.

The organization of this paper is as follows: Section 2 gives an overview of the governing factors. Section 3 provides the methodology for this study that includes the research model, hypotheses and data collection. Section 4 presents the data analysis and results. Section 5 discusses the results, and Section 6 provides the conclusions of the study.

II. Background

There is still insufficient research regarding the critical factors that affect transformational government (t-Government) such as technical, organisational and governing factors [6-8]. This section discusses some of the popular governing factors that have been identified (strategy, leadership, stakeholders, funding and citizens centricity) as affecting t-government in more details as follows:

A. Strategy

Strategies and Regulations are important in any development of e-Government, and one of the most important elements involved in the implementation of t-Government [9], as its establishment inevitably requires major changes in strategic direction, and to modes of paradigm thinking [10, 11]. Thus a comprehensive e-Government strategy becomes a major factor when collaboration between government agencies is necessary to ensure the successful presentation of an online service [11-13]. Appropriate regulations are also important for the successful implementation of t-Government, as the absence of such a regulatory framework is one of the more onerous challenges hindering its effective establishment [14-16]. Strategies and regulations are measured through: the importance of e-Government strategy, strategy's plan, strategy's goal, strategy's vision, and strategy's commitment.

B. Stakeholders

A key factor in the implementation of t-Government projects will be stakeholders: their presence within any e-Government initiative is fundamental to the success of its interoperability [17-19]. Governments need to identify their stakeholders are, and what they want, to succeed in e-government implementation, and achieve the maturity of e-Government [20]. Rowley [21] argued that the classification of stakeholders are often implicit in categorizations of e-Government such as (G2G), (G2C), (G2B), and (G2E), or they offered by a number of authors. For example, Heeks [22] classified stakeholders to Non-profits, other agencies, citizens/customers, businesses, communities, government. Orange, Burke [23] classified stakeholders to Politicians, staff, public, project managers, design developers, other government agencies. UnitedNations [24] classified stakeholders to Public administrators, programmers, end-users, politicians. As this study only focuses on (G2G) e-Government categorization, it concern only about internal stakeholder in government organisations. According to Al-Rashidi [25] Internal stakeholders are political stakeholders, organizational stakeholders, and technological stakeholders such as Public administrators (employees), Other government agencies, Politicians ,E-Government project managers, Design and IT developers, and naturally, co-operation between these categories of stakeholder is critically important to the successful implementation of t-Government projects [26, 27]. Hu, Cui [28] considers that cross-agency cooperation has the potential to transform the way that governments work, share information, and deliver services to external and internal clients. According to [29], implementing interoperability involvers many stakeholders at different government level and coordination between these stakeholders is important. In this study, participating stakeholders are measured by: stakeholder's involvement, stakeholder's management, and stakeholder's coordination.

C. Leadership

Leadership always plays a significant role to any group endeavour, and is thus the quality of the leadership employed will have a significant impact on the successful implementation of t-Government, particularly in e-Government projects requiring a high level of interoperability [17, 30]. t-Government projects are long-term: consequently the quality of leadership is crucial [31]: strong leaders are required to overcome the inherent challenges involved. Researchers have identified both vision and leadership as the main drivers of successful e-Government [10, 19, 32], consequently, and according to many studies, [27, 33, 34], Effective leadership of this kind is a major contributory factor to the successful implementation of t-Government. The effectiveness of any leadership derives from its quality. Altameem, Zairi [16] and Prybutok, Zhang [34] state that effective leaders express more complex and contradictory behaviour than ineffective leaders. Zairi [35] claims that "Nowadays leadership is considered as a must for survival. It comes from the level of inspiration, commitment generated and corporate determination to perform". Thus a particular challenge to government as top manager of a project is the selection of a strong political leader with both IT and management skills capable of leading the project to a successful completion [10, 19, 32]. In this study leadership is measured by: leadership support, leadership style, and the influence of strong leadership.

D. Citizens' Centricity

A citizen centric involves the provision of services from the end-user's point of view rather than the perspective of the government department [36-40]. A t-Government project should have a citizen centric perception. However, citizen-centric service delivery is a complex issue with many perspectives that need to be considered at the very beginning of a transformation project. The provision of citizen-centric service has been identified by some researchers as a critical success factor [38]. Therefore, t-Government should be provided in a way that enable citizens to easily access information and complete their transactions. In order to assess the level of success of the provision of citizen centric services end-user satisfaction should be regularly measured [36-39]. Citizens' centricity is measured in this study by: citizen centric legislation, citizen's measurement and citizen's satisfaction.

E. Funding

e-Government initiatives being long-term, they require long-term financial support from the Government, which can become a major challenge if that funding has to come from a government where political influence may interfere with decisions taken by high level officials [17, 41]. Moreover, Gottipati [42] argues that the way e-government projects are being reviewed and funded in the Arabian gulf is that such projects appear to seen as budget-based instead of seeing those projects as project-based budgets. Thus funding is inevitably a crucial factor in the implementation of t-Government, as a lack of adequate, consistent financial backing will become a major challenge to the successful implementation of any e-Government project[43]. Adequate funding supports the necessary integration of government organisations by facilitating the development of infrastructure (such as building, technology, human resources) required to implement t-Government, and ensures that goals and targets are met on time. Its impact on e-Government interoperability is also positive [17]. Fund is measured in this study through fund amount, measurement mechanism, fund management, and fund controlling.

III. Methodology

A. Research model

The objective of this study is to identify governing factors that influence the implementation of t-Government in Saudi Arabia. The research model of the study is presented below in FIGURE I. The constructs included in the research model, were drawn from the variables used from a review of the relevant literature TABLE I.

TABLE I. Governing Factors Influencing T-Government Implementation from the Literature

Constructs	Items	References
Strategy & Regulations	e-Gov strategy	[9, 44]
	Commitment	[11, 45]
	Plan	[11, 31]
	Goal	[11, 17, 26, 46]
	Vision	[11]
Leadership	Support	[11, 17, 46-48]
	Style	[10, 17, 30, 31, 48, 49]
	Strong leader	[17, 48, 50, 51]
Stakeholders	Involvement	[26, 48, 52-55]
	Identification	[21, 48, 54, 55]
	Management	[55, 56]
	Cooperation	[10, 28, 48, 55, 57]
Fund	Amount	[43, 44, 46, 58-60]
	Measurement mechanism	[26, 49]
	Management	[49, 59, 61]
	Controlling	[17, 47, 49]

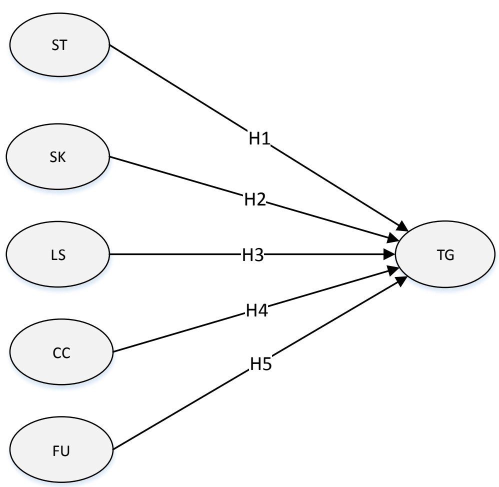


Fig. 1. The hypotheses model

B. Research Hypothesis

TABLE II. Research Hypotheses

No	Hypotheses
H1	Strategy positively influence the implementation of t-Government
H2	Stakeholders positively influence the implementation of t-Government
H3	Leadership positively influences the implementation of t-Government
H4	Citizens' Centricity positively influences the implementation of t-Government
H5	Funding positively influences the implementation of t-Government

C. Data Collection

This study uses a quantitative research approach with methods that include a survey. The survey was divided into different sections to enable respondents to very understand it and a Likert scale with six levels of possible answers was used (from Strongly Agree to Strongly Disagree). Survey was conducted in Saudi Arabia from December 2013 to April 2014. 1194 survey questionnaires are distributed to 166 organisations. From this 917 people have viewed the survey, and 477 people have responded. From the balance, 258 responses are incomplete (more than 70% of the questions are unanswered) and hence unusable. The number of completed responses is 217.

IV. Data Analysis and Result

Statistical Packages for Social Science (SPSS) version (22) was used to analyse the data collected through the surveys. The study applied the Structural Equation Modelling (SEM) technique and utilized AMOS version (22) tools to test the hypotheses among the variables in the model. Structural equation modelling (SEM) is a statistical technique that allows the researcher to examine multiple interrelated dependence relationships in a single model [62-64]. SEM is a popular approach in social science research [65]. It is popular analysis technique because it has flexibility for interpreting the theory to be tested and the sample data [66]. The following section will illustrate the analysis of the study in more detail.

A. Descriptive Statistical Perspective

In the questionnaire 5 demographic questions were used to capture demographic information: age, education level, occupation, organisation size, and the number of G2G services. More than (72.9 percent) of the respondents were between 31 and 45 years old and more than (88 percent) have hold a Bachelor's or postgraduate degree. More than (41 percent) of the respondents are manager, more than (55 percent) of the respondents come from organisation have more than 500 employee, and (53 percent) of the respondents come from organisation have from (1 – 5) G2G services. The demographic information about the respondent is summarised in TABLE III.

TABLE III. THE DEMOGRAPHIC INFORMATION

Demographic Variables		Frequency	Percent %
Age group	20 - 25	7	3.2
	26 - 30	26	12.0
	31 - 35	52	24.0
	36 - 40	36	16.6
	41 - 45	37	17.1
	46 - 50	25	11.5
	51 - 55	26	12.0
	More than 55	8	3.7
Education level	High school	3	1.4
	Diploma	24	11.1
	Bachelor	87	40.1
	High Diploma	13	6.0
	Master	74	34.1
	Doctorate	16	7.4
Occupation	Manager	26	12.0
	Dept Manager	64	29.5
	System-analysis	60	27.6
	Technician	13	6.0
	Others	54	24.9
Employee number	Less than 100	31	14.3
	101 – 500	65	30.0
	501 – 1000	35	16.1
	1001 – 2000	25	11.5
	2001 - 5000	20	9.2
	More than 5000	41	18.9
Numbers of G2G services	1 - 5	115	53
	6 - 10	61	28.1
	11 -15	5	2.30
	16 - 20	6	2.76
	None	30	13.82

B. *Reliability Verification*

This study used Cronbach's alpha to measure the internal consistency. (Straub 1989), outline that internal consistency is a commonly used technique to assess the reliability by using Cronbach's alpha. Also, (Hair Jr, Black et al. 2010) suggests that the reliability test should be conducted before the construct validity analysis is commenced. Constructs are considered reliable when Cronbach's alpha is .70 or higher (Hair Jr, Black et al. 2010).

As shown in TABLE IV, all estimated values of the constructs were above the recommended value (0.70) which indicates that there is a strong reliability and high internal consistency in measuring relationships in the model.

TABLE IV. CRONBACH ALPHA RELIABILITY RESULTS

Construct	No of Items	Cronbach's Alpha
ST	5	0.947
SK	3	0.946
CC	3	0.779
LS	3	0.975
FU	4	0.850
TG	3	0.920

C. *Validity Test*

All the constructs were shown to have a composite reliability greater than the threshold level of 0.60 and their Average Variance Extracted (AVE) is also greater than the threshold level of 0.50 [62] as shown in TABLE V. Also, all the indicators (factors) had significant loadings greater than 0.50 ($p < 0.001$) on their respective constructs as shown in TABLE VI.

Since their factor loadings were meaningful and highly significant, they were retained in the measurement model. Moreover, Average Variance Extracted (AVE) is greater than the squared correlation of constructs show TABLE VII. These results indicated that the measurement model possessed substantial convergent validity and unidimensionality [62].

TABLE V. CONVERGENT VALIDITY RESULTS

	CR	AVE
SK	0.948	0.859
ST	0.947	0.783
CC	0.824	0.621
LS	0.976	0.932
TG	0.921	0.797
FU	0.927	0.762

TABLE VI. CONSTRUCTS FACTOR LOADING

Factors	Items	Estimate
SK	STAK_4	0.981
	STAK_2	0.897
	STAK_1	0.899
ST	LE_2	0.855
	LE_1	0.94
	STA_3	0.874
	STA_2	0.871
	STA_4	0.882
CC	C_3	0.644
	C_2	0.678
	C_1	0.992
LS	L_3	0.984
	L_2	0.915
	L_1	0.996
FU	F_1	0.979
	F_2	0.816
	F_3	0.860
	F_4	0.830
TG	I_1c	0.934
	I_1b	0.821
	I_1a	0.917

TABLE VII. DISCRIMINANT VALIDITY: AVERAGE VARIANCE EXTRACTED (AVE) WITH SQUARED CORRELATION OF CONSTRUCT

	SK	ST	CC	LS	TG	FU
SK	0.927					
ST	0.511	0.885				
CC	0.175	0.110	0.788			
LS	0.404	0.257	0.138	0.966		
TG	0.260	0.025	0.033	0.247	0.893	
FU	0.045	0.012	0.168	-0.061	0.053	0.873

V. HYPOTHESIS TESTING RESULTS

Testing of hypotheses aims to determine which independent variables provide a statistically meaningful relationship to the dependent variables[62]. This study tested the hypotheses using SEM. SEM provides information about the hypothesized impact both, directly from one variable to another and via other variables positioned between the other two. This study was conducted using AMOS 22.0.

The statistics of the model based on the SEM output are: RMSEA 0.053, Chi-square/df 1.61, CFI 0.975, TLI 0.971. These values are within the threshold limits prescribed by [62, 63].TABLE VIII represents the results of testing the current study hypotheses. The 'Result' column indicates whether that hypothesis was: supported or not supported depending on the result of the p value.

TABLE VIII. HYPOTHESIS TESTING RESULTS

Hypotheses	Path coefficient	C.R.	P value	Result
TG<---ST	.15	2.151	0.03*	Supported
TG<---SK	.25	3.753	***	Supported
TG <---LS	.19	2.866	0.004*	Supported
TG<---CC	.03	0.514	N.S	Not supported
TG<---FU	.06	0.955	N.S	Not supported

* p <0.05; ** p <0 .01; *** p < 0.001

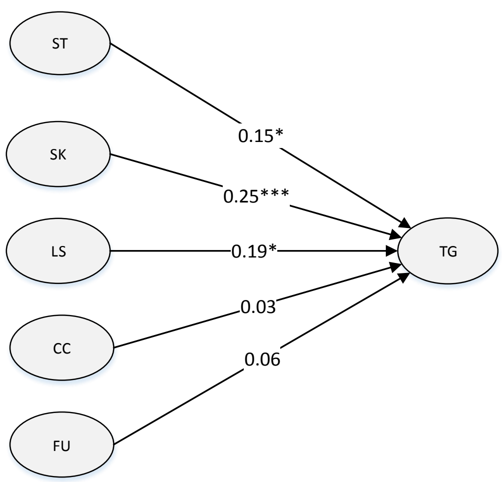


Fig. 2. Path Coefficients for the Proposed Structural Model

VI. DISCUSSION

The aim of this study is to test how governing factors namely strategy, stakeholders, leadership, citizens' centricity and funding to the implementation of t-Government among (n=217) from different organisations in Saudi Arabia. The SEM analysis data supported the analysis and results of the SEM were summarized as follows:

Strategy positively influences the implementation of t-Government

Strategy indicated a significant positive impact on the change towards a t-Government. This hypothesis is supported in not only this study but other studies conducted previously. [9-11, 17, 26, 44, 45, 52, 60] reported the impact of the strategy to the implementation of t-Government. Based on this result, decision-makers, and e-Government officials should pay attention to the e-Government strategy for the transformation to a t-Government. It should be translated into an effective and clear roadmap that can be easily interpreted and committed to by all government organisations.

Stakeholders positively influence the implementation of t-Government

This hypothesis is supported in this study and is consistent with previous studies [17, 26, 52-54, 60] found the impact of stakeholders on the implementation of t-Government. Success of t-Government implementation projects need involvement and collaboration between all stakeholders. Based on this result, it is clearly understandable that decision-makers and e-Government officials should pay attention to the existence of involvement, commitment, management and cooperation of the stakeholders toward the success of the implementation of t-Government.

Leaderships positively influences the implementation of t-Government

This hypothesis is supported in this study and is consistent with previous studies [11, 17, 26, 67, 68]. There is a support from the custodian of the two holy mosques King Abdullah [69]. Although, government organisations need strong leadership to speed up and follow up these kind of projects until the end. In addition, the knowledgeable leadership who understand the technology, the legislations and the policy goals plays a vital role in the outcome of any e-Government project. Therefore, decision-makers, and e-Government officials should pay more attention to leadership support, interaction and cooperation with sufficient knowledge and style on the implementation of t-Government.

Citizens' Centricity positively influences the implementation of t-Government

This hypothesis is not supported in this context and is inconsistent with previous studies [37-40] pointed out the impact of Citizen Centric on the change towards a t-Government. The reason for that could be the lack of regular reviewing and measuring of citizen satisfaction, and participation issues have not been adequately addressed. According to [2], the United Nations e-Government Survey uses a three stage model to assess the e-participation index (EPI). The first stage: 1) e-information measures how participants access public information and information upon demand, 2) e-consultation measures the engagement of contributions of people on public policies and services and 3) e-decision-making to measure the empowerment of people on designing the service delivery. The e-participation index (EPI) indicates Saudi Arabia achieved 85.19% in stage 1, 27.27% in stage 2, and 11.11% in stage 3 which means there is a gap in citizen centric focus. Therefore, decision-makers, and e-Government officials should pay more attention to citizen satisfaction, service quality and should be regularly measured.

Funding positively influences the implementation of t-Government.

This hypothesis is not supported in this context, and it is inconsistent with previous studies [43, 44, 46, 58-60] who pointed out that there is an impact of funding towards t-Government. The reason for that could be the support from the custodian of the two holy mosques King Abdullah to the transformation to e-Government projects [69].

Therefore, funding is not an obstacle by itself the implementation of t-Government, and more effort should be considered to create plans and strategies to fund e-Government projects and to further monitor this funds.

VII. Conclusion

The aim of this study is to identify and test empirically the factors that affect implementation of t-Government from multiple perspectives. It provides a model to assess the success factors to the implementation of t-Government in Saudi Arabia. Also, it identifies the most important factors in this context, regarding evaluating the success of the change to t-Government in Saudi Arabia.

Based on the data collected and the results of the analysis, this study showed positive and significant relationships between factors such as the relationships between strategy and t-Government (H1); the relationships between stakeholders and t-Government (H2); and the relationships between leadership and t-Government (H3). Also, it indicated that there is an insignificant relationship between citizens' centricity and t-Government (H4); and an insignificant relationship between funding and t-Government (H5).

The results of this study generate some useful implications for decision-makers, and government officials in developing countries in general and especially Saudi Arabia, with important guidelines in understanding how different factors affect the transition to a transformational government. Since the Saudi government and many other governments face the problem of a low level of e-government integration [2, 11, 17], it is hoped that the results this study will support decision-makers, and government officials in increasing the level of success of t-Government.

To summarise, this study has attempted to offer a better understanding of the relevant factors that may influence the change toward t-Government through an analysis of the literature and a survey. This study is based on 217 respondents and due to its limited size and exclusive focus on Saudi Arabia future studies can be conducted in future to verify the results and also tests the factors in other cultural contexts. Also, future research can expand this study by including the effect of different factors such as organisational and technical factors that may impact the implementation of t-Government in Saudi Arabia.

References

[1] Zukang, S., E-Government Survey 2012: E-Government for the People. United Nations report, 2012.

[2] Hongbo, W., E-GOVERNMENT SURVEY 2014: E-GOVERNMENT FOR THE FUTURE WE WANT. United Nations report, 2014.

[3] Sundberg, H.P. and K.W. Sandberg, Towards e-government: a survey of problems in organisational processes. Business Process Management Journal, 2006. 12(2): p. 149-161.

[4] Al-Khouri, A.M., An innovative approach for e-Government transformation. arXiv preprint arXiv:1105.6358, 2011.

[5] Alshetewi, S., R. Goodwin, and D. de Vries, Organisational, Technological and Governance Factors Influencing T-government. International Journal of Advanced Research in Computer Science & Technology (IJARCST). 2012. 1(1).

[6] Solli-Sæther, H. and L. Flak. Interoperability, maturity and benefits in e-government. in Norsk konferanse for organisasjoners bruk av informasjonsteknologi (Nokobit), Universitetet i Nordland, Bodø. 2012.

[7] Marc, N. and L. Grace A. , Interoperability in the e-Government Context. 2012, Software Engineering Institute: Carnegie Mellon University. p. 1-35.

[8] Van Veenstra, A.F., B. Klievink, and M. Janssen, Barriers and impediments to transformational government: insights from literature and practice. Electronic Government, an International Journal, 2011. 8(2-3): p. 226-241.

[9] Ray, D., et al., A critical survey of selected government interoperability frameworks. Transforming Government: People, Process and Policy, 2011. 5(2): p. 114-142.

[10] Elnaghi, M., et al., Instigating Transformational Government at a Municipality Level: A Case Study. Handbook of Research on ICT-Enabled Transformational Government: A Global Perspective, 2009: p. 72.

[11] Lam, W., Barriers to e-government integration. Journal of Enterprise Information Management, 2005. 18(5): p. 511-530.

[12] Sang, S., J.-D. Lee, and J. Lee, A Study on the Contribution Factors and Challenges to the Implementation of E-Government in Cambodia. Journal of Software, 2009. 4(6): p. 529-535.

[13] Pilling, D. and H. Boeltzig. Moving toward e-government: effective strategies for increasing access and use of the internet among non-internet users in the US and UK. in Proceedings of the 8th annual international conference on Digital government research: bridging disciplines & domains. 2007. Digital Government Society of North America.

[14] Carter, L. and F. Belanger, The influence of perceived characteristics of innovating on e-government adoption. Electronic Journal of E-government, 2004. 2(1): p. 11-20.

[15] Sahli, N., S. Mellouli, and N. Jabeur. Survey on e-government approaches: which approach best fits my government? in Proceedings of the Second international conference on Global Information Infrastructure Symposium. 2009. IEEE Press.

[16] Altameem, T., M. Zairi, and S. Alshawi. Critical success factors of e-government: A proposed model for e-government implementation. in Innovations in Information Technology, 2006. 2006. IEEE.

[17] Pardo, T.A., T. Nam, and G.B. Burke, E-government interoperability: Interaction of policy, management, and technology dimensions. Social Science Computer Review, 2011: p. 0894439310392184.

[18] Cohen, S.G. and D. Mankin, Complex collaborations in the new global economy. Organizational Dynamics, 2002. 31(2): p. 117-133.

[19] Ke, W. and K.K. Wei, Successful e-government in Singapore. Communications of the ACM, 2004. 47(6): p. 95-99.

[20] Chircu, A.M., E-government evaluation: towards a multidimensional framework. Electronic Government, an International Journal, 2008. 5(4): p. 345-363.

[21] Rowley, J., e-Government stakeholders—Who are they and what do they want? International journal of Information management, 2011. 31(1): p. 53-62.

[22] Heeks, R. Understanding and measuring eGovernment: international benchmarking studies. in UNDESA workshop,"E-Participation and E-Government: Understanding the Present and Creating the Future", Budapest, Hungary. 2006.

[23] Orange, G., et al. Care: an integrated framework to support continuous, adaptable, reflective evaluation of eGovernment systems: A research note. 2006. European and Mediterranean Conference on Information Systems.

[24] UnitedNations, E-government survey 2008; from e-government to connected governance. . 2008, United Nations: New York.

[25] Al-Rashidi, H., The role of internal stakeholders and influencing factors during the phases of e-government initiative implementation. 2013, Brunel University, School of Information Systems, Computing and Mathematics.

[26] Soares, D. and L. Amaral, Information systems interoperability in public administration: identifying the major acting forces through a Delphi study. Journal of theoretical and applied electronic commerce research, 2011. 6(1): p. 61-94.

[27] Jaeger, P.T. and K.M. Thompson, E-government around the world: Lessons, challenges, and future directions. Government Information Quarterly, 2003. 20(4): p. 389-394.

[28] Hu, P.J.-H., D. Cui, and A.C. Sherwood. Examining cross-agency collaborations in e-government initiatives. in System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on. 2006. IEEE.

[29] Ebrahim, Z. and Z. Irani, E-government adoption: architecture and barriers. Business Process Management Journal, 2005. 11(5): p. 589-611.

[30] Pardo, T.A. and G.B. Burke, Government worth having: A briefing on interoperability for government leaders. Center for Technology in Government, University at Albany, SUNY Oct, 2008.

[31] Malinauskienė, E., Conceptual framework for context-based e-government interoperability development. Social Technologies, 2013(01): p. 68-84.

[32] Seifert, J.W. and G.J. McLoughlin. State e-government strategies: Identifying best practices and applications. 2007. DTIC Document.

[33] Ndou, V., E-government for developing countries: opportunities and challenges. The electronic journal of information systems in developing countries, 2004. 18.

[34] Prybutok, V.R., X. Zhang, and S.D. Ryan, Evaluating leadership, IT quality, and net benefits in an e-government environment. Information & Management, 2008. 45(3): p. 143-152.

[35] Zairi, M., Leadership in TQM implementation: some case examples. The TQM Magazine, 1994. 6(6): p. 9-16.

[36] Zhou, X., E-Government in China: A Content Analysis of National and Provincial Web Sites. Journal of Computer-Mediated Communication, 2004. 9(4): p. 00-00.

[37] Karim, M.A., Technology and improved service delivery: learning points from the Malaysian experience. International Review of Administrative Sciences, 2003. 69(2): p. 191-204.

[38] Themistocleous, M., Z. Irani, and P.E. Love. Developing e-government integrated infrastructures: a case study. in System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on. 2005. IEEE.

[39] Irani, Z., et al., T-government for benefit realisation. 2007.

[40] King, S. and S. Cotterill, Transformational government? The role of information technology in delivering citizen-centric local public services. Local Government Studies, 2007. 33(3): p. 333-354.

[41] Eyob, E., E-government: breaking the frontiers of inefficiencies in the public sector. Electronic Government, an International Journal, 2004. 1(1): p. 107-114.

[42] Gottipati, M. Managing e-government portals, Eighth GCC e-government. in Internet & telecommunications forum, May 28. 2002.

[43] Huang, Z. and P.O. Bwoma, An overview of critical issues of e-government. Issues of Information Systems, 2003. 4(1): p. 164-170.

[44] Dos Santos, E.M. and N. Reinhard, Electronic government interoperability: Identifying the barriers for frameworks adoption. Social Science Computer Review, 2011: p. 0894439310392196.

[45] Ramaswamy, M., e-Government implementation in transition countries. Handbook of Research on ICT-Enabled Transformational Government, 2009: p. 441-451.

[46] Tripathi, R., M. Gupta, and J. Bhattacharya, Effect of organizational factors on interoperability adoption for Indian portals. Transforming Government: People, Process and Policy, 2013. 7(3): p. 285-308.

[47] Klischewski, R. and H.J. Scholl. Information quality as a common ground for key players in e-government integration and interoperability. in System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on. 2006. IEEE.

[48] Luk, S.C.Y., The impact of leadership and stakeholders on the success/failure of e-government service: Using the case study of e-stamping service in Hong Kong. Government Information Quarterly, 2009. 26(4): p. 594-604.

[49] Scholl, H.J. and R. Klischewski, E-government integration and interoperability: framing the research agenda. International Journal of Public Administration, 2007. 30(8-9): p. 889-920.

[50] Bekkers, V., The governance of back office integration in e-government: Some dutch experiences, in Electronic government. 2005, Springer. p. 12-25.

[51] Hossan, C.G., M.W. Habib, and I. Kushchu. Success and Failure Factors for e-Government projects implementation in developing countries: A study on the perception of government officials of Bangladesh. in Poster session presented at the proceedings of 2th European Conference on Mobile Government. Retrieved March. 2006. Citeseer.

[52] Klischewski, R., Architectures for Tinkering?: Contextual Strategies towards Interoperability in E-government. Journal of theoretical and applied electronic commerce research, 2011. 6(1): p. 26-42.

[53] Kamal, M., V. Weerakkody, and Z. Irani, Analyzing the role of stakeholders in the adoption of technology integration solutions in UK local government: An exploratory study. Government Information Quarterly, 2011. 28(2): p. 200-210.

[54] Janssen, M. and A.M. Cresswell, An enterprise application integration methodology for e-government. Journal of Enterprise Information Management, 2005. 18(5): p. 531-547.

[55] Scholl, H.J., Involving salient stakeholders Beyond the technocratic view on change. Action Research, 2004. 2(3): p. 277-304.

[56] Ezz, I.E. and A. Papazafeiropoulou. Inter-organisational collaboration towards process integration in the public sector. E-government collaboration in Egypt. in System Sciences, 2006. HICSS'06. Proceedings of the 39th Annual Hawaii International Conference on. 2006. IEEE.

[57] Janssen, M. and H.J.J. Scholl. Interoperability for electronic governance. in Proceedings of the 1st international conference on Theory and practice of electronic governance. 2007. ACM.

[58] Weerakkody, V., Moving from E-Government to T-Government: A Study of Process Reengineering Challenges in a UK Local Authority Context. International Journal of Electronic Government Research,, 2008.

[59] Ray, D., et al., Interoperability and constituents of interoperable systems in public sector. Handbook of Research on ICT-enabled Transformational Government: A Global Perspective, IGI Global, Hershey, PA, 2009: p. 175-95.

[60] Elliman, T., O.D. Sarikas, and V. Weerakkody, Realising integrated e-government services: a UK local government perspective. Transforming Government: People, Process and Policy, 2007. 1(2): p. 153-173.

[61] Landsbergen Jr, D. and G. Wolken Jr, Realizing the promise: Government information systems and the fourth generation of information technology. Public Administration Review, 2001: p. 206-220.

[62] Hair Jr, J., et al., Multivariate Data Analysis; a global perspective (ed.): Pearson Education Inc. New Jersey, USA, 2010: p. 5.

[63] Kline, R.B., Principles and practice of structural equation modeling. 2011: Guilford press.

[64] Byrne, B.M., Structural equation modeling with AMOS: Basic concepts, applications, and programming. 2013: Routledge.

[65] Mueller, R.O., Structural equation modeling: Back to basics. Structural Equation Modeling: A Multidisciplinary Journal, 1997. 4(4): p. 353-369.

[66] Chin, W.W., Commentary: Issues and opinion on structural equation modeling. 1998, JSTOR.

[67] Department of Finance and Deregulation, D., National Government Information Sharing Strategy, A.G.I.O. (AGIMO), Editor. 2009, http://www.finance.gov. au/publications/national-government-information-sharing-strategy/index.html.

[68] Ray, D., et al., Interoperability and constituents of interoperable systems in public sector. Handbook of Research on ICT-enabled Transformational Government: A Global Perspective, 2009: p. 175-95.

[69] Yesser. E-government Program. 2015 [cited 05/08/2015; Available from: http://yesser.gov.sa.

User Interface Design of E-Learning System for Functionally Illiterate People

Asifur Rahman
Graduate School of Information Science and Electrical Engineering, Kyushu University
Fukuoka, Japan

Akira Fukuda
Faculty of Information Science and Electrical Engineering
Kyushu University
Fukuoka, Japan

***Abstract*—Among different type of illiterate people, the print illiterates suffer most from getting crucial information passed around the society. Many print illiterate people are found in the developing countries and in many cases they live in the remote areas working as farmers. These people are deprived of the knowledge generated from the latest scientific researches. This research makes some recommendations related to developing user interface especially suitable for the print illiterate people. In this regard, a user interface is developed based on the recommendations from the previous researchers. The authors find the recommendations insufficient and develop another user interface based on the improvements proposed by the authors. Later both the user interfaces are tested by two different groups of print illiterate people in a remote village in Bangladesh. The test data shows that the proposed improvement contributes significantly to make the user interface more usable to the target population. 13 out of 15 users could complete the assigned task successfully using improved user interface. Whereas only 8 out of 14 users could do the same with the other user interface. Among the successful users, the improved user interface took 26% less time than that of the other user interface. Finally some recommendations to develop user interface for the functionally illiterate people are made based on the results and observations of this research.**

***Keywords*—User Interface; Illiterate people; e-Learning**

I. Introduction

In the backdrop of evolving different technologies, more complex set of illiteracy is being observed in the society. Those include computer illiteracy, media illiteracy, visual illiteracy, etc. [1, 2]. These illiterate people can read and write on a specific language (usually in their native language), however, cannot respond to the respective technology. Inability of reading and writing is termed as print illiteracy. Most of the people in this world are print illiterate except a few languages. There is another type of illiteracy observed among the people migrated from one country to another and are unable to read and write the local language [3]. These people might be very efficient in their native language, though. These types of migrated people usually fall into the category of print illiterate. It can be summarized that every individual belongs to some groups of illiteracy. The extent of literacy also varies from highly skilled to functionally illiterate.

Among different types of illiterate people, print illiterates are the group who suffers most from getting information passed around the society. The largest group of print illiterate people is found in the developing countries. These illiterate people live both in the urban area and in the remote area. Illiterate people living in the remote area suffer more from getting information than ones living in the urban area. Many researches are conducted to introduce innovative ICT-based services to provide information to the people living in the remote areas in the developing countries [4-5]. The researchers also discussed the constraints and opportunities in details. Many of the researches are found not attaining the planned goals because the concerned users cannot use the system properly and are not happy with the benefits getting from the system [6]. Therefore, the target population, despite its usefulness, abandons the system. This research aims to develop a user interface for the print illiterate people.

It is obvious that the access to knowledge and information is the primary means of empowering the people living in the remote areas. They possess the same potential like that of a literate urban people [7]. They are surviving for generations and fighting against all odds without the help of latest scientific research findings. They live based on the skills and techniques learned from their surroundings. The authors intend to develop a platform for them to accelerate their learning considering existing infrastructural limitations, like lack of electricity supply, poor internet connectivity, print illiteracy, poverty, and so on. The learning process is optimum when it is made personalized [8]. Therefore, the researchers aim to develop an e-Learning system [9] that utilizes multimedia and internet technologies to facilitate the users to access services and resources as and when needed.

The existing digital library and e-Learning systems are using multi-media extensively. The audio was used in some projects as technological means to disseminate information to the illiterate people [10]. Non-computerized systems used audio and video cassettes in this regard [11-13]. These are convincing examples that demonstrate that information can be distributed, accessed and created via appropriate technologies that do not need print literacy. The authors decided to use computerized technology to store videos demonstrating the usage of different skills that can empower people in the remote areas and distribute those through a system that does not require print literacy to use. Therefore, developing a video-based e-learning system using an innovative user interface especially suitable for illiterate people. Use of such user interface that can be used by the illiterate people without any help from others makes the application to be self-operated and low cost.

Usability of an application depends more on its ease of use for the target population than its features. There are evidences that many high-quality applications were abandoned because the target users could not use them properly. Considering the target population, the print illiterate people, it is important that the user interface is suitable enough for them. Besides, the authors assume that the application would be available in the remote areas and be operated by the users themselves without any help from others. Therefore, the focus was on developing a user interface that an illiterate person can operate independently.

The rest of this paper is organized as follows: Section 2 contains the related works on developing user interface for the illiterate people. The existing works are categorized based on the technologies used. The limitations of those user interfaces and usability in specific situations are also mentioned. Section 3 details the prototype developed in this work. Section 4 contains the primary design of the user interface. It also describes the primary evaluation of the user interface, findings, and observations. Section 5 describes the improvements that the authors proposed on the existing recommendations to make the user interface more effective to the illiterate people. Section 6 details the experiment conducted and analysis of data. Section 7 contains the discussion, and section 8 concludes with a list of future works.

II. Related Work

User interface for print illiterate people became a subject of interest for the researchers for quite a long time. Different approaches have already been tested to develop a user interface and improve its usability for the print illiterate people. All these attempts might be categorized as: i) innovative use of text and numbers; ii) Use of extensive graphics and minimum text; and iii) Use of audio in addition to text and graphics.

In the early days of computer application development, command line based user interface was the only means of man-machine interaction. Soon after the computers had more processing capabilities, text-based menu was introduced. Text-based menu was shortly replaced by graphics-based menu. During this time it was not considered that poor illiterate people living in the developing countries would use computers. In the early of 21st century, when the price of computers came within the range of common people in the developing countries, the governments and the international development agencies took initiatives to develop application for the people living in the rural areas of the developing countries. However, it was always remain as a challenge to make the rural illiterate and semi literate people using those. Ample of evidences were found where many potential software applications were abandoned because the target illiterate and semi literate users could not use those. This observation encouraged the researchers to work extensively to develop suitable user interface for the illiterate people.

Researches have shown that the illiterate people can remember numbers more easily than words. Therefore, some experiments used more numbers to make it easy for the illiterate users to use the applications [14-18]. However, these recommendations based on usage of number against text are more relevant for mobile phone interface and other man-machine interface where high quality multimedia capabilities are not available.

Some other researchers used speech recognition techniques to solve the problem of illiteracy [19-23]. In these applications, the researchers used short word based commands to activate tasks. In India, China, Bangladesh and some African countries, local dialects varies from district to district within the same state. It would be very difficult to train the speech recognition system to work for a considerably large segment of people. However, the systems work satisfactorily with very limited command set used by people from only small number of villages.

In the recent days, some researchers introduced new techniques using the recent improved hardware and software features [24-28]. Researchers have also tested the effect of different types of graphics like sketch, photo, animated cartoon and video. In different situation the result varies significantly. The researchers also recommended some guidelines for developing effective user interface for the illiterate users.

Considering availability of rich hardware features and software platforms the authors choose to use the recommendations for developing the user interface of the intended e-Learning system for the print illiterate people using text, graphics and intelligent use of audio annotation. Details of the design considerations are explained in a separate section.

III. E-Learning System Prototype for Illiterate People

The authors are intending to build an e-learning system for the illiterate people living in the remote areas of developing countries to distribute knowledge and empowering them. The knowledge includes farming skills, public health, sanitation, child nutrition, maternity care, etc. Technical and non-technical limitations are considered on construction of this system.

The system is organized as shown in Fig. 1. The content server works as the central repository of all data. In addition to the e-Learning system and its related database, database replication management module and DTN module are also available in the content server. The database replication management module ensures that all the local server databases are synchronized properly with the content server periodically. The internet bandwidth in remote areas of developing countries is not suitable yet to transfer bulky multimedia data. Therefore, the bulk data transfer is arranged from content server to the local server over Delay Tolerant Network (DTN) [29]. DTN is a hop-to-hop store and forward data transfer protocol useful in the challenged infrastructure scenario. DTN uses vehicles like Bus, Steamer, Satellites, etc. to transfer data physically from one place to another. In this work, data is transferred from the content server to the local server over internet and DTN using the interim facilities: DTN relay center and DTN node in bus. The local server contains the e-Learning system, database replication management module, and DTN module. The local servers are installed in village information centers (VIC). Here, VIC is a shared ICT service access facility owned by a local entrepreneur, that offer computers and internet connectivity to the villagers to use against money. The details of the DTN module, DTN relay center, and DTN node in bus is out of the scope of this paper.

The content server contains a database that stores video contents and related data. The database replication management module decides which content to be replicated in which local server. It is obvious that the central server database is a huge one. Local servers located in different areas do not need to contain each and every content. For example, mango fruit is not cultivated everywhere in Bangladesh. It is very unlikely that a villager, living in a village where mango fruit is not cultivated, will search for a video related to mango cultivation. Therefore, while replicating the database, contents are selected for a local server based on socio-economic, demographic, agricultural, environmental, and other information related to that area. The database replication management module also ensures that all the local servers are synchronized appropriately periodically.

The content server is accessible over internet. The users of this system may be either registered or anonymous. The administrators and the content providers are the registered users. Both the registered and anonymous users can access the system at any time. Anonymous users can only search for and play video contents. Content providers represent organizations those are interested to produce and share the videos to help the remote villagers. A content provider can upload videos and fill up corresponding data at any time over the internet. He/she can also edit or delete his/her own contents.

The villagers visit village information center (VIC) to use the system. They do not need any registration and can only search for and play the videos.

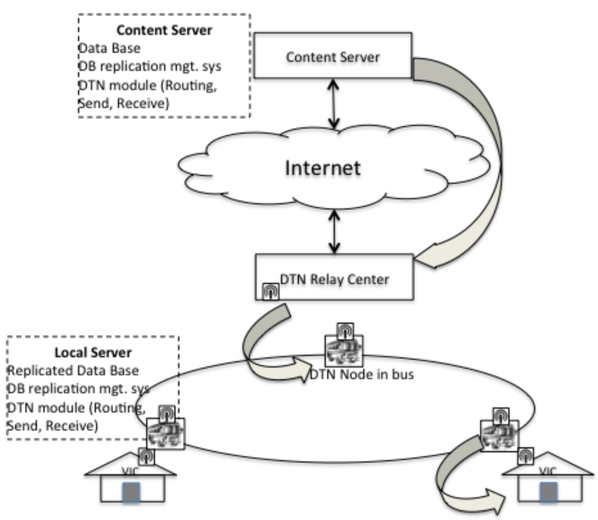


Fig. 1. System Arrangement

A. Features of the system:

The system maintains a repository of video contents that teaches different skills and techniques useful for the target population. It also contains information about new crops, livestock, etc. The features of the systems are briefly described as follows:

a) User Registration: People who are literate can fill up a form to become a registered user. Then the user input his identification information. The registration process completes with verification and approval from the administrator. The registered users can upload video content, edit, delete data or administer the system based on their granted privilege.

b) Content uploading: The registered users may upload a video content after login the system properly. During uploading the video, the content provider provides information like the category, subcategory, language, country, etc. of the video. The content uploader also need to enter a short description of the video and audio of the text description. The process is made such to ensure accountability against the uploaded contents.

c) Content editing: Content providers and the administrators can use this option to edit information related to any specific content. A content provider can only edit data of the contents he uploaded.

d) Content deleting: Content providers and administrators can use this feature to remove any content from the system. A content provider can only remove the contents he uploaded.

e) System configuration: The configuration privilege is only granted to the administrator. Configuration includes content server configuration, data replication criteria setting for different local servers, parameter setting for data transportation from content server to local server and so on.

f) Content searching: This feature can be used by both registered and anonymous users. The literate users can search using a text based user interface and the illiterate users can search using a different user interface especially suitable for illiterate people.

g) Content playing: The user can search the desired video using the search option. Then he/she can play the video to watch using content playing feature.

Most of these features are for administrators or content providers who are literate and possess high IT aptitude. Considering the scope of this paper, Content Searching, and Content Playing should be discussed extensively.

Use case name:	Search for Content
Summary:	Takes input from the user and search for matching video contents from the database and list.
Description: 1. User specifies to search contents. 2. User specifies the category, subcategory, language, country, and full or part of the title of the contents to be part of the list. User may input all or some of the criteria mentioned above. 3. The system searches for the matching contents. 4. The system lists the matching contents.	

Fig. 2. Content Searching use case

B. Content Searching:

Content searching is the most frequently used feature. Administrators and the content providers use it to edit or delete contents. They may also use it to play contents. Anonymous users always use it to find out the desired content before playing it. Fig. 2 shows the use case.

C. Content Playing:

After the user finds the desired content using the content searching feature, he/she will use content playing feature for watching. It is another frequently used feature. Fig. 3 contains the use case related to content playing feature.

Use case name:	Play content
Summary:	Plays the selected video content
Description:	

1. User specifies the video content that he/she wants to play
2. The system starts playing the video content
3. User watches the video content
4. The system finishes playing the video content, moves the playing position to the beginning of the video content, and returns into the pause mode

Fig. 3. Content Playing use case

IV. Primary Design of User Interface

For the purpose of developing the user interface, the authors followed the recommendations made by the previous researchers for developing user interface of similar applications [24-27]. The recommendations are:

a) Minimizing the need of hierarchical navigation. It requires higher level of aptitude to remember the combinations that emerge after a few steps. The situation gets worse when the number of branches in each level becomes higher.

b) Maintaining natural workflow as much as possible. It reduces the load of learning new concept.

c) Use of graphics where users need to select an item. Text does not bear any meaning to an illiterate person. Most of the cases, a picture explains the meaning more clearly than words.

d) Use of Audio annotation to clarify confusing situation. There are situations when it becomes very difficult to develop two distinctly different pictures for two very similar situations. For example, in Fig. 4(a) is a photo of pests and Fig. 4(b) is a photo of insects. Both the photos are collected from Internet. These photos are too similar to differentiate as two different categories pest and insect. Audio annotation in such cases helps significantly.

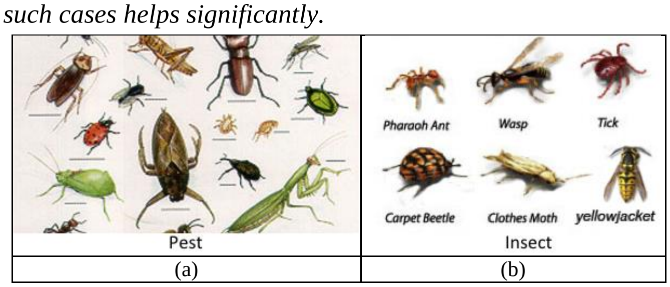


Fig. 4. Photos for categories pest and insect

e-Health was used as an application domain to test the design principals in the aforementioned work. In this work, the authors followed the same recommendations to design the user interface for e-Learning system.

The authors developed a user interface for the illiterate people following the aforementioned recommendations. Fig. 5 shows the user interface, which is refined from the use case description of Content Searching use case. The details of the user interface is described below:

Fig. 5. User interface of Content Searching use case: Fig. (a) shows the user interface for literate users; (b) shows the categories an illiterate user can select: fruit, flower, vegetable, etc.; (c) shows the sub categories an illiterate user can select: mango, jack fruit, orange, strawberry, etc., which are the subcategories of the category fruit; (d) shows the countries an illiterate user can select: Bangladesh, India, Sri Lanka, Indonesia, and Nigeria; (e) shows the languages an illiterate user can select: Bengali, Hindi, Tamil, Sinhala, English, and Bahasa Indonesia; (f) shows the list of contents match the criteria specified in the previous steps

1) The user specifies to search for contents by clicking the search content button available in the header of all the webpages. This search content button is active in all the webpages. The webpage shown in Fig. 5(a) appears.

2) The user specifies the category, subcategory, language, country, and full or part of the title of the contents to be part of the list by selecting one from the list box or filling the text field in the webpage shown in Fig. 5(a). For the print illiterate user, all the text is supplemented by audio annotation. Moreover, when the user makes the mouse cursor hover on a text filed or a button for two or more seconds, the webpage plays the audio annotation explaining function and purpose of the underlying object.

Since it is full of text and not suitable for illiterate users, another webpage of the same features for the illiterate people is also provided. By clicking the bottom right button of the webpage shown in Fig. 5(a), another version of the webpage shown in Fig. 5(b) appears.

The webpage is graphics based and more suitable for the print illiterate people. The webpage shows pictures of fruit, flower, vegetable, etc., which represent categories the user can select. The user can select a category by clicking on the photo.

The selected category would be marked with a black border. Moreover, if the user makes the mouse cursor hover over any picture, the webpage tells the name of the category in voice. Similarly, the print illiterate people can select a subcategory, a country, and a language in the webpages shown in Fig. 5(c), (d), and (e), respectively, which appear subsequently.

Moreover, the system also explains the purpose of any button if the user makes the mouse cursor hover over it.

3) The system searches the database for the matching contents.

4) The system lists the matching contents in the webpage shown in Fig. 5(f). The print illiterate user can identify exactly which one of the matching contents the user wants to play by making the mouse cursor hover over the content description and listening it in voice.

After the user specifically identifies which content the user wants to play, then he/she would click on the video to start playing. This would start execution of Content Playing use case and the user interface would look like Fig. 6.

Fig. 6. User interface for Content Playing use case

Like the previous user interface page, all the text and buttons of this page are also supplemented by audio annotation. The user would put the cursor on different text and buttons to know the purpose and use of the same if necessary.

After developing the user interface the performance was tested with a trial evaluation. Some Indonesian and Japanese students, who do not understand Bangla language, used the system as illiterate users. Although the real system uses audio annotation in Bangla language, English and Japanese annotation were used for Indonesian and Japanese students respectively, so that the users can understand the audio annotation. Each user was given some time to learn the system and then was asked to find a video based on a description. It was expected that the users would use the system independently. Every user could find the desired video. However, the number of wrong attempts in different pages was quite high.

Following are the observations of the trial performance test:

a) All the users are familiar with use of computers. Therefore, they are very conversant to the user interface and the workflow concept. However, they could not understand the text in Bangla Language. Still it took long time for them to find out which button to choose at what stage.

b) Users kept listening to the audio annotation of different elements until the appropriate option was found.

c) Sometimes, the users cannot concentrate on the audio descriptions because the number of prospective buttons is quite large. That leaded to errors.

d) Users feel discouraged to try too many options to find the correct button or text.

The authors identified following limitations of the user interface developed:

a) Users get frustrated if there are many options available to choose in a user interface to do different tasks and users repeatedly make wrong guess. Therefore, available or probable options for the users are to be minimized.

b) Sometimes users cannot recall from where to start and the purpose of the page. Because he/she cannot read the interface texts or button labels. Therefore, a proactive briefing might be useful for the users.

V. IMPROVED DESIGN OF USER INTERFACE

Based on the aforementioned observations and findings, following improvements were made in the user interface design:

a) To minimize the number of options to choose, the authors intended to guide the users to choose from a limited number of most probable options. Among different buttons available in the page, the authors identified the most likely to be used buttons to go to the next options. There are also some text information that the majority users need to know. Those buttons and texts are made blinking to attract attention from the users.

b) To help users recall the purpose of the elements of each page, a proactive audio description of each page was added. The description was played after the page is loaded. Users may listen to the description before taking action in case he/she forgot the purpose of that specific page.

Fig. 7 shows the user interface for the Content Searching use case after the aforementioned improvements. In Fig. 7(a) the text-based user interface is for the literate users and they don't need any assistance from the user interface. Illiterate users may choose one from two options: i) they will move forward to search with photos; ii) they will abandon the search. Therefore, the corresponding two buttons are made blinking. To listen to the purpose of each of these buttons, users may hover over the buttons and choose the appropriate one. There are 13 dropdown list, textboxes and buttons available in the page. This effort indicates that these two blinking buttons are the most probable option for an illiterate user to choose. Similarly in all the pages, some buttons or textboxes are made blinking, so that the illiterate users can understand that these are the probable buttons or textboxes to choose at this stage.

(a) (b)

(c) (d)

(e) (f)

Fig. 7. User interface for Content Searching use case after improvement. The figures are same as Fig. 7 (a)-(f) except in each page few buttons and textboxes are blinking

As mentioned earlier the proposed second improvement, proactive audio description for each page is added. The description is played immediately after the page is loaded. The purpose of the page, action invoked by clicking each blinking button, information content in each blinking textbox is explained in the audio description. If the illiterate users forget the purpose of the buttons and textboxes available in the page, this proactive description allows them to recall and start using the page.

The aforementioned changes are made in every page. The new user interface for the use case Content Playing is shown in Fig. 8. While playing content, the user may want to finish watching the video and go back to the previous list or start searching another content. Therefore, two buttons corresponding to the mentioned actions are made blinking.

Fig. 8. User interface for Content Playing use case after improvement

VI. Experiment and Data Analysis

Purpose of this work is to develop an effective user interface for the illiterate people. The authors identified some limitations of the existing proposals for the same and improved the existing proposals to develop a new user interface. Finally compared the performance of two versions of user interface, one developed based on the existing proposal and another developed based on the proposed improvements.

A. Evaluation Process:

The first version, (henceforth referred as UI-1), which is developed based on the existing recommendations, was evaluated by a group of 14 functionally illiterate people in a village in Bangladesh and the improved version, (henceforth referred as UI-2), was evaluated by another group of 15 functionally illiterate people from the same village. The participants of both the groups are selected randomly among the male residents of the same village. The age of the participants were uniformly distributed among different age groups. All of them are functionally illiterate. Some of them can read and write though. The participants have experience of using mobile phone. Although some of them saw computers, however, none of them used computer. The workflow of the evaluation process is as follows:

1) The evaluation process starts with a detailed description of the system to the user. The user also shown how to use the system. The purpose of evaluation, the role of the user in the process is also described in details. It is categorically mentioned to the user that the purpose of this evaluation is to measure the usefulness of the user interface only. The personal data, like name, age, educational qualification, etc. are also collected during this time.

2) The user is allowed to practice using the system for about an hour, so that they can learn how to use it. During this practice session, they were guided as and when needed.

3) The user is given a list of tasks to perform. The example of a task might be "find a video that describes the scientific process of raising a cow for milk production." The user performs the tasks and his activity during this period is recorded. Later this recorded video clip is used to collect data like time to complete the task, mistakes during search, etc.

4) Oral feedback about the system is collected from the user.

B. Results:

The participants were given a list of tasks to complete during the evaluation. Example of a task can be "find a video that describes the scientific process of raising a cow for milk production." The successful completion of the aforementioned task means that the participant could find the correct video. While performing the task, a participant might find a wrong video or after repeated choice of wrong steps a participant may abandon the search and starts from the beginning.

a) Successful completion: 14 participants used "UI-1" and 15 participants used "UI-2". The performance of the participants of "UI-1" and "UI-2" is shown in Table I. The data shows that 8 out of 14 participants could complete the task using "UI-1". On the other hand, 13 out of 15 participants could complete the task using "UI-2". Therefore, more people can use "UI-2" successfully than "UI-1".

TABLE I. Number of Participants Completed Tasks Using Both the User Interfaces

	UI-1	UI-2
Total Participants	14	15
Failed to complete	6	2
Completed without hints	6	9
Completed with hints	2	4

b) Completion time comparison: The number of tasks attempted and completed by the participants is not same for all participants. The number of attempts depended on interest of the participant on the system and time spent on learning the system. Most of the participants could complete at least one task successfully. Time spent for the unsuccessful attempts is ignored during calculating the average completion time. For the participants, who completed multiple tasks, the authors considered average time of completion as their completion time. Average time of completion for all participants using "UI-1" is 6:32 minutes whereas average time of completion for all participants using "UI-2" is 5:10 minutes. Therefore, it can be concluded from the data of this experiment that, in similar situation a participant can use "UI-2" 26% faster than "UI-1".

c) Age wise performance: It is known that younger people can respond to the new technology faster and older people suffer from techno phobia, therefore, fails to use new technology. However, in this work it is found that people from different age groups could use the system similarly. The participants were selected almost equally among different age groups. Table II shows the number of participants became successful and unsuccessful in different age groups. The data shows that combining all participants of "UI-1" and "UI-2" every age group has 2 unsuccessful persons.

TABLE II. SUCCESSFUL AND UNSUCCESSFUL PARTICIPANTS IN DIFFERENT AGE GROUP

Age group	Successful	Unsuccessful
Under 20	5	2
Under 30	5	2
Under 45	6	2
Rest	5	2

VII. DISCUSSION

The analysis of the data collected from the experiment shows that:

1) *More people can use "UI-2" successfully than "UI-1".*

2) *Users can find the desired video faster in "UI-2".*

3) *The user interface is equally useful to all different age groups.*

It indicates that the recommendations made by previous researchers for similar applications are also applicable in the present scenario. However, the proposed improvements affect significantly in the situation under consideration of this research.

A. Recommendations:

Based on the observations and data analysis, following design principals can be recommend for the UI development for the functionally illiterate people:

1) Use of images and graphics instead of text: People, who cannot read, can understand photos. It is important that the photo emphasizes a specific message only. The background or other object present in the photo does not get prominence. It is found in the previous works that illiterate people can understand sketch better than photos. The background and other objects available in the photo drags their attraction that makes them confused sometimes. Although photos were used in this experiment, users were not confused. The photos used in the user interface are selected carefully to emphasize only one message.

2) Use of audio annotation: In many cases it becomes very difficult to represent two very close subjects with two distinct graphics images. Therefore, it requires using audio annotation in addition to graphics representation. In this research, audio annotation during hover action found the most effective.

3) Adding context sensitive suggestions to the users: Guided and controlled users action reduces errors and mistakes. The illiterate users should always be guided to the possible actions for a specific situation. However, the designer should be careful to select most likely options for a specific situation. Too many suggestions may confuse the users.

4) Proactive briefing: In this research, it is found that a proactive description of the features and possible use of different active elements is very effective for the users to select the correct button or option. Users with less orientation to the technology cannot remember all the steps accurately. Besides, due to lack of experience they cannot make good guess. Therefore, early briefing would help them to make fewer mistakes. Long verbal description might be avoided by some graphics based animated description in order to make it interesting.

B. Observations:

After completion of using the system the users were requested for their comments about the user interface and the system. This session was their first experience of using computers. Therefore, they could not articulate their opinion very clearly. The authors are listing their observations acquired through the whole activity:

1) Orientation to computer usage is an essential part of using any computer-based system. Whatever smart user interface is developed, users cannot use it unless they learn basic computer usage. Therefore, the users need basic computer training for two weeks or more.

2) Age was not a factor for learning and using the system. Appropriate use of graphics, audio annotation, and proactive and innovative guidance were the reasons behind the success.

3) Schooling does not influence learning process. There are people who have gone higher classes in the school but cannot respond well. Some people who can read and write a little still failed to ensure good learning of the system.

4) All participants could understand the audio instruction but it was difficult for a first time user to remember the workflow after listening the audio instruction.

5) Audio annotation is essential during training period. However, it should be brief and to the point. Long description and repetition should be avoided. After the users get used to the system then long description of each page would be very annoying.

6) Animated guidance used in "UI-2" is very effective.

VIII. Conclusion and Future Works

This research demonstrates that it is possible to develop a user interface that allows functionally illiterate people to learn and use a system without any help from others. This success would encourage governments and development organizations to introduce more and more new systems to help the people living in the remote areas to improve their quality of living. A web-based system would further allow the service provider to maintain and monitor the system remotely from the cities.

This study was conducted for a computer-based system. However, the results and recommendations can also be used in other situations for man-machine interface. Consider a situation when a person is moved to another country where the local language is different. In that case, that person becomes a newly illiterate person. While this person needs to use vending machines, he faces the same situation like that of an illiterate person faces in the remote areas with a computer system. The recommendation can be used to develop a system that is easily usable.

This research put some recommendation that demands more resources to the system. For advanced and complex systems, the additional resources could become very exhaustive. In future, the authors would extend the features of the e-Learning system to make it more useful to the users. and would test with the real users for a considerable period of time. This would allow to study the performance of the proposed user interface more accurately. Finally, the authors would like to come to a conclusion for the optimum level of details for the audio annotation, proactive guidance, etc.

References

[1] D. Bawden, “Progress in Documentation, Information and Digital Literacy: a Review of Concepts”, in Journal of Documentation, vol. 57, no. 2 pp 218-259, 2001.

[2] M. Walton, G. Marsden and V. Vukovic, “‘Visual Literacy’ as challenge to the Internationalization of Interface: A study of South African Student Web Users” in Proceedings of the CHI-SA Conference 2001, Pretoria, South Africa, pp. 530-531, 2001

[3] Y. L. Theng, N. Mohd Nasir, H. Thimbleby, G. Buchanan, M. Jones, *et al*, “Children as Design Partners and Testers for a Children’s Digital Library,” in Proceedings of the 4th ECDL Conference, Springer Verlag, pp. 249-258, 2000

[4] T. Almarabeh and A. AbuAli, “A General Framework for E-Government: Definition Maturity Challenges, Opportunities, and Success,” in European Journal of Scientific Research, Vol. 39 No. 1 pp. 29-42, EuroJournals Publishing, Inc. 2010.

[5] G. Sharma, X. Bao and L. Peng, “Public Participation and Ethical Issues on E-governance: A Study Perspective in Nepal” in Electronic Journal of e-Government Volume 12 Issue 1, pp82-96, 2014.

[6] R. Heeks, "Information systems and developing countries: Failure, success, and local improvisations." In The information society vol. 18, no. 2, pp. 101-112, 2002.

[7] N. P. Sullivan, “You Can Hear Me Now: How Microloans and Cell Phones are Connecting the World's Poor To the Global Economy”, John Wiley & Sons, 2007

[8] M. Gell-Mann, “A commentary to R Schank. The third culture; beyond the scientific revolution.” New York: Touchtone Books, pp. 167-180, 1996.

[9] EC [European Commission], Communication from the commission to the council and the European parliament: the e-Learning action plan, Brussels, 28.3, 2001

[10] A. Mutshewa, “Disseminating Environmental Information in Rural Botswana: the Case of Nata Village,” in Information Development vol. 15, no. 2 pp. 96-102, 1999

[11] A. P. du Plooy, “Community Information for the Functionally Illiterate in South Africa by Means of Audio-Cassettes,” in Mousaion vol. 6, no. 1 pp. 2-20, 1988.

[12] M. Rahnema, “The Sound Library, a Simple but Revolutionary Tool for Development,” in Unesco Journal of Information Science, Librarianship and Archives Administration vol. 4, no. 3 pp. 151-158 1982.

[13] P. Manuel, “Cassette Culture: Popular Music and Technology in North India,” University of Chicago Press 1993.

[14] Z. Lalji and J. Good, “Designing new technology for illiterate population: A study in mobile phone interface design”, in Interecting with compiters vol. 20, no. 6, pp. 574-586, 2008.

[15] I. Medhi, S. Patnaik, E. Brunskill, S. N. N. Gautama, W. Thies *et al.* “Designing Mobile Interface for Novice and Low-Literate Users”, in ACM Transactions on Computer-Human Interaction, Vol. 18, No. 1, Article 2, 2011

[16] T. S. Parikh, “HISAAB: An Experiment in Numerical Interfaces”, Media Lab Asia Panel Discussion, Baramati Initiative on ICT and Development., MIT, 2002

[17] T. S. Parikh, K. Ghosh, and A. L. Chavan, “Design Considerations for a Financial Management System for Rural, Semi-literate Users,” in the procedings of CHI ’03 Extended Abstracts on Human Factors in Comouter Systems, Ft. Lauderdale, Florida, USA, pp. 824-825, April 2003

[18] T. S. Parikh, K. Ghosh, and A. L. Chavan, “Design Studies for a Financial Management System for Micro-credit Groups in Rural India,” in the Procedings of the 2003 conference on Universal Usability,Vancouver, British Columbia, Canada, pp. 15-22, November 2003.

[19] M. Plauche and M. Prabaker, “Tamil market: a spoken dialog system for rural India” in the proceeding CHI '06 Extended Abstracts on Human Factors in Computing Systems Pages, pp. 1619-1624, 2006.

[20] M. Plauché, U. Nallasamy, J. Pal, C. Wooters and D. Ramachandran, “Speech Recognition for Illiterate Access to Information and Technology,” in the procedings of International Conference on Information and Communication Technologies and Development, 2006, pp. 83-92, 2006.

[21] A. Lerer, M. Ward and S. Amarasinghe, “Evaluation of IVR Data Collection UIs for Untrained Rural Users,” in the Proceedings of the First ACM Symposium on Computing for Development, December, 2010, London, United Kingdom, p. 2, 2010.

[22] K. Bali, S. Sitaram, S. Cuendet, and I. Medhi, “A hindi speech recognizer for an agricultural video search application,” in the procedings of the 3rd ACM Symposium on Computing for Development, January 2013, P. 5, 2013.

[23] S. Cuendet, I. Medhi, K. Bali, E. Cuttrell, “VideoKheti: Making Video Content Accessible to Low-Literate and Novice Users,” In the procedings of the SIGCHI conference on Human Factors in Computing Systems, 2013, pp 2833-2842, 2013

[24] I. Medhi, M. Lakshmanan, K. Toyama and E. Cutrell, “Some Evidence for the Impact of Limited Education on Hierarchical User Interface Navigation,” in the procedings of the SIGCHI conference on Human Factors in Computer Systems, 2013, pp. 2813-2822, 2013.

[25] I. Medhi, and R. Kuriyan, “Text-Free UI: Prospects for Social-Inclusion,” in the procedings of International Conference on Social Implications of Computers in Developing Countries. Brazil, 2007

[26] I. Medhi, B. Pitti, and K. Toyama, “Text-Free User Interfaces for Employment Search,” in the Procedings of Asian Applied Computing Conference, Nepal, 2005.

[27] I. Medhi, A. Prasad, and K. Toyama, “Optimal Audio-Visual Representations for Illiterate Users of Computers” in the procedings of 16th International conference on World Wide Web, Albarta, Canada, pp 873-882, 2007.

[28] I. Medhi, and K. Toyama, “Full-Context Videos for First-Time, Non-Literate PC Users,” in the procedings of Information and Communication Technologies and Development, 2007, India, PP 1-9, November 2007.

[29] A. Rahman, T. Nakanishi, A Fukuda, "Delay Tolerant Network for Developing Countries," in the Proceedings of International Conference on Informatics, Electronics & Vision, Dhaka, Bangladesh, May 2013.

Acoustic Emotion Recognition Using Linear and Nonlinear Cepstral Coefficients

Farah Chenchah
LR-SITI Laboratory
National Institute of Applied Science and Technology,BP.676 centre urbain cedex
Tunis, Tunisia

Zied Lachiri
LR-SITI Laboratory
National Institute of Applied Science and Technology,BP.676 centre urbain cedex
Tunis, Tunisia

***Abstract*—Recognizing human emotions through vocal channel has gained increased attention recently. In this paper, we study how used features, and classifiers impact recognition accuracy of emotions present in speech. Four emotional states are considered for classification of emotions from speech in this work. For this aim, features are extracted from audio characteristics of emotional speech using Linear Frequency Cepstral Coefficients (LFCC) and Mel-Frequency Cepstral Coefficients (MFCC). Further, these features are classified using Hidden Markov Model (HMM) and Support Vector Machine (SVM).**

Keywords—Mel Frequency Cepstral Coefficients (MFCC); Linear Frequency Cepstral Coefficients (LFCC);Hidden Markov Model (HMM); Support Vector Machine (SVM); emotion recognition

I. Introduction

Emotion is an important aspect of human interaction that needs to be further investigated. Its understanding becomes essential for understanding human communication. Studies on emotion involve several field of research such psychology, neuroscience, philosophy, physiology, computer science and in several other areas. This melting pot of discipline studying emotion gives to emotion recognition all its importance.

Emotions are a whole-body phenomena that are reflected through several cues such facial expression, body gesture and speech. In this context, advances in automatic speech recognition (ASR) have consumed tremendous effort and have reached a level of maturity which results may be widely used in recognizing emotion.

Speech processing techniques provide an extensive array of feature extraction methods that may be used to extract emotional characteristic in human voice. These features can be divided into two main classes: prosodic and spectral features. Prosodic features include but are not limited to Pitch, Energy, Formant frequency[1], Jitter, Shimmer[2], Zero Crossing Rate (ZCR)[3]. Among spectral features, we can list Linear Predictive Coding (LPC)[4], short-time coherence method (SMC) [5] and Mel-Frequency Cepstral Coefficients (MFCC) [6].

Furthermore, extensive work on emotion recognition has been carried out using different classifier such us neural networks[7], Support Vector Machine (SVM)[8], Gaussian mixture models (GMM) [9] and Hidden Markov Model (HMM) [10].

The aim of this paper is to determine which of Hidden Markov Models (HMM) and Support Vectors Machines (SVM) as classifier and MFFC and LFCC, as feature extraction method can be used to derive an efficient system of emotion through vocal channel. The rest of this paper is organized as follows. In Section 2, system design and selected corpora are presented. Then feature extractions are introduced in section 3and classification models are described in Section 4. Experiments and results are presented in Section 5. Finally, Section 6 gives the conclusions.

II. System Design and Selected Corpora

A. System design

The proposed emotion recognition system can be divided into two main parts, feature extraction and emotion classifier. In the feature extraction, we extract all the acoustical features from both of training and testing speeches. Step classification is performed using Hidden Markov Models (HMM) and Support vector Machines (SVM) to identify the emotion class of a speech utterance. System description is illustrated in figure1.

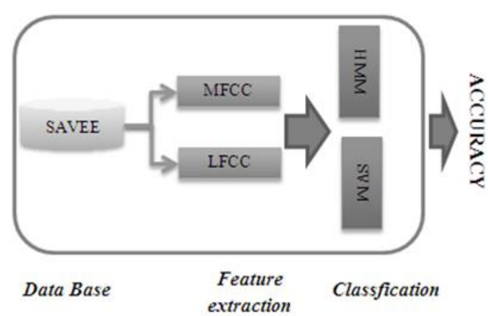


Fig. 1. System description

B. Selected corpus(SAVEE database)

The Surrey Audio-Visual Expressed Emotion (SAVEE) database is a freely available audiovisual data set [11]. This English-language corpus consists of 480 phonetically balanced utterances spoken by four native British male speakers (DC, JE, JK, DC) in seven different emotions (fear, anger, disgust, sadness, surprise, happiness, neutral).

Recordings consisted of 15 TIMIT sentences per emotion (with additional 30 sentences for neutral state).The emotion assessment of recordings was performed by subjective evaluation under audio, visual and audio-visual scenarios. Speech data were labeled at phone-level in a semi-automated way. The sampling rate used for audio data is 44.1 kHz.

III. FEATURE EXTRACTION

Extracting suitable features from signal is an important step in emotion recognition system. Significant descriptors can carries large emotional information about the speech signal; they affectively increase the performance of classifiers. Several researches have shown that effective parameters to distinguish a particular emotional state with potentially high efficiency are spectral features such as Linear Frequency Cepstral Coefficients (LFCC) and Mel Frequency Cepstral Coefficients (MFFC).

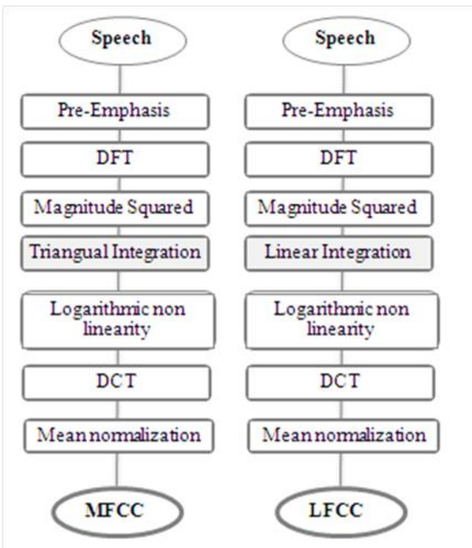


Fig. 2. Mel and Linear filter banks

A. Linear Frequency Cepstral Coefficients (LFCC)

In feature extraction process, [12] introduce a feature method called Linear Frequency Cepstral Coefficients.

The computation of LFCC features can be described; firstly, Fast Fourier Transform (FFT) is applied to windowed signal for converting each frame of N samples from the time domain into the frequency domain. After the FFT block, the power coefficients are filtered by linear frequency filter banks. Finally, the log Mel spectrum is converted into time using Discrete Cosine Transform (DCT).

B. Mel Frequency Cepstral Coefficients (MFCC)

Mel Frequency Cepstral Coefficients (MFCC), introduced by [13], are based on human hearing perceptions which cannot perceive frequencies over 1Khz.

As presented in figure 2, compute MFCC is similar to LFCC with only difference in the spacing of the filter bank For MFCC, after the FFT block, the power coefficients are filtered by a triangular band pass filter bank also known as Mel-scale filter. The Mel-frequency scale is linear frequency spacing below 1 kHz and logarithmic spacing above 1 kHz.

IV. CLASSIFICATION

In general, there are two approaches to develop classifiers: a parametric approach, and a nonparametric approach. This study uses two nonparametric approaches which are Support vector machines (SVM), often called kernel-based methods and Hidden Markov Model (HMM).

A. Hidden Markov Models (HMM)

A Hidden Markov Model is a doubly stochastic modeling appropriate for no stationary stochastic sequence [14]. HMMs lie at the heart of emotion recognition through vocal channel systems.

HMM is a variant of a finite state machine defined by a (i) set of hidden states, (ii) a transition probabilities distribution, (iii) observation symbol probability distribution in each state and (iv) initial state probability distribution.

The aim of the training phase of the HMM is to decide which one of the H MMs is more likely to have emitted the received sequence. For this purpose, the parameters describing an HMM are estimated. To this end, one or several observation sequence that has been generated by the corresponding stochastic process is used to estimate the unknown parameters.

B. Support Vector Machines (SVM)

Support Vector Machine is a very efficient and simple classifier algorithm which is widely used for pattern recognition.

SVM is a non-linear classifier by transforming the original input set into a higher dimensional feature space by using kernel mapping function, it searches fort the linear optimal separating hyper plane [15].

The most frequently used SVM kernel functions are linear kernel, polynomial kernel and Radial Basis function (RBF) kernel. Considering data from two different classes, an SVM attempts to solve an optimization problem that finds a hyper plane that separates the data with maximum margin. The binary class problem is extended to multiclass classification, methods such as One-Against All (OAA) and One-Against-One (OAO) can be applied.

OAA is the earliest and simplest approach. It involves k binary SVM classifiers, one for each class. Each binary SVM is trained to separate one class from the rest. The winning class is the one that corresponds to the SVM with highest output. OAO involves $k(k-1)/2$ binary SVM classifiers. Each classifier is trained to separate each pair of classes. There are different strategies used to combine these binary classifiers. The crucial widely used strategy is a majority voting.

V. EXPERIMENTAL SETUP AND RESULTS

A. Experimental setup

In our experimental studies, we collect all the available sentences which are classified in four emotional states that we examine: angry, happy, neutral and sad.

The utterances are expressed by 4 male actors. The sampling frequency of each recording is 16KHz.Signals samples are segmented into frames with 50% overlap.

The feature vector of MFCC and LFCC consists of 13 coefficients. Extraction of cepstral coefficients from emotional speech was performed using LFCC-RASTAMAT toolbox. To compare the effectiveness of these features, step classification was performed using HMM and SVM.

The first step consist of varying the frame length in the range of {50ms, 100ms, 250ms, 500ms, 750ms, 900ms, 1s}. Data was tested using MFCC as feature vector and HMM as classifier. The best frame length obtained will be used for the remaining experiments.

For HMM classifier, we evaluate the topology by varying the number of mixture components and the number of states. HMM models are built for four emotions individually. The HMM classification is done using the Hidden Markov Toolkit (HTK) [16].

For SVM classifier, two Kernel functions are used, polynomial and gaussian, with multiclass strategies, OAA and OAO. To select suitable parameters for each Kernel (C,σ), a cross-validation algorithm was performed by varying the regularization parameter C in [1,100] and Gaussian width σ in [1,10]. The SVM classification is done with the SVM-KM Toolbox for Matlab [17].

B. Results

Speech emotion recognition is implemented using MFCC and LFCC features, we evaluate the system of recognizing emotion state with two classifiers using HMM and SVM.

TABLE I. CLASSIFICATION ACCURACY USING HMM AND MFCC USING DIFFERENT FRAMES

Frame	0.05	0.1	0.25	0.5	0.75	0.9	1
DC	85.00%	65.00%	70.00%	65.00%	70.00%	50.00%	50.00%
JK	50.00%	50.00%	60.00%	60.00%	50.00%	55.00%	55.00%
JE	60.00%	60.00%	65.00%	60.00%	45.00%	55.00%	60.00%
KL	50.00%	60.00%	15.00%	30.00%	45.00%	45.00%	35.00%
Average	**61.25 %**	**58.75 %**	**52.50 %**	**53.75 %**	**52.50 %**	**51.25 %**	**50.00 %**

Table 1 presents classification accuracy using MFCC feature with HMM as classifier. The aim of this step is to set the optimal frame length to be used. Results show that the best accuracy is obtained with a frame length of 50ms with an average recognition rate of 61.25%.

In the following experiments samples will be segmented into frames of 50ms each with 50% overlap.

Table 2 shows the classification results for the four speakers obtained from SVM OAA method applied to MFCC features. These results were run using polynomial and Gaussian kernel. The classifier gives accuracy for testing data are in the range of [39% 53%] with an average of 45.83% for the polynomial Kernel, and in the range of [45% 67%] with an average of 54.58% for Gaussian kernel.

TABLE II. CLASSIFICATION ACCURACY USING SVM/OAA AND MFCC

	Testing	Training	Kernel	C	σ	SV
DC	49.17%	100.00%	poly	1	5	66
	66.67%	99.58%	gaussian	11	11	203
JE	39.17%	100.00%	poly	11	3	79
	55.83%	97.50%	gaussian	1	10	211
JK	53.33%	100.00%	poly	1	6	81
	50.00%	100.00%	gaussian	11	9	218
KL	41.67%	100.00%	poly	1	9	88
	45.83%	96.67%	gaussian	1	10	220
AVERAGE	**45.83%**	**100.00%**	**poly**			
	54.58%	**98.44%**	**gaussian**			

TABLE III. CLASSIFICATION ACCURACY USING SVM/OAO AND MFCC

	Testing	Training	Kernel	C	σ	SV
DC	65.00%	97.91%	poly	1	10	110
	69.17%	95.42%	gaussian	21	91	72
JE	45.00%	100.00%	poly	11	2	45
	52.50%	100.00%	gaussian	21	10	108
JK	53.33%	100.00%	poly	1	5	53
	55.00%	100.00%	gaussian	21	10	112
KL	46.67%	100.00%	poly	1	9	60
	43.33%	99.17%	gaussian	11	10	114
AVERAGE	**52.50%**	**99.48%**	**poly**			
	55.00%	**98.65%**	**gaussian**			

The same data sets with same features applied to a SVM OAO are presented in table 3 which shows a testing data average classification rate of 52.50% for polynomial kernel with a minimum of 45% and a maximum of 65%, and 55% for Gaussian with classification rate between 43% and 69%.

We can remark that for both methods Gaussian kernel gives better results, and that One Against One method of Support vector machine are slightly better than One Against All.

TABLE IV. CLASSIFICATION ACCURACY USING SVM/OAA AND LFCC

	Testing	Training	Kernel	C	σ	SV
DC	52.50%	100.00%	poly	1	10	79
	51.67%	99.58%	gaussian	11	5	202
JE	35.00%	53.33%	poly	1	1	159
	40.83%	93.33%	gaussian	1	7	198
JK	45.83%	49.58%	poly	21	1	234
	50.00%	84.16%	gaussian	1	10	203
KL	39.16%	100.00%	poly	1	7	112
	36.67%	88.75%	gaussian	1	10	210
AVERAGE	**43.12%**	**75.73%**	**poly**			
	44.79%	**91.46%**	**gaussian**			

We can remark from table 4 that both kernels used gives an important range of classification rate between minimum and maximum figures giving a difference of 17.5% for polynomial and 15% for Gaussian. The results of Gaussian kernel are better in term of accuracy than polynomial kernel.

Table 4 presents results is obtained from applying LFCC feature to OAA strategy of Support Vector Machine with same kernels used for MFCC.

TABLE V. CLASSIFICATION ACCURACY USING SVM/OAO AND LFCC

	Testing	Training	Kernel	C	σ	SV
DC	46.67%	100.00%	poly	41	9	46
	51.67%	99.16%	gaussian	21	10	90
JE	35.00%	62.50%	poly	1	1	76
	45.83%	96.25%	gaussian	21	8	92
JK	42.50%	53.33%	poly	11	1	98
	50.00%	81.67%	gaussian	1	8	120
KL	39.16%	100.00%	poly	1	8	61
	40.00%	100.00%	gaussian	41	5	110
AVERAGE	**40.83%**	**78.96%**	**poly**			
	46.88%	**94.27%**	**gaussian**			

The same experiment conducted using One Against One strategy is described in table 5 which shows that for all speakers Gaussian Kernel gives better results with an average accuracy for testing data of 46.88%.

Among the four result tables using SVM as classifier with its two strategies, we can remark that:

- very similar trends can be observed: the best performance is achieved by Gaussian kernel:
- for almost results with polynomial kernel, One against one strategy is better than One Against All:
- MFCC describes better emotional trends in speech signal.

TABLE VI. CLASSIFICATION ACCURACY USING HMM

	MFCC	LFCC
DC	85.00%	50.00%
JE	60.00%	45.00%
JK	50.00%	60.00%
KL	50.00%	25.00%
Average	61.25%	45.00%

Table 6 reflects the classification rates obtained from HMM classification method applied to MFCC and LFCC features. In this table, it is obvious that MFCC performs better than LFCC when for three of the four speakers MFCC feature, and in average, the first method gives better recognition rates.

The results demonstrate that HMM is better classifier than SVM with its two strategies with an average accuracy of 61.25%. One reason for this might be that HMM can model dynamic changes of acoustic features in given emotional state. Moreover, MFCC proves to be most descriptive than LFCC, the results obtained using this feature is steadily better than those from LFCC features.

The classification dispersion between speakers reaching 35% can be presented as the field in which improvement can be made. This may be due to the fact that it is difficult even for human subjects to determine the emotion of some recorded utterance.

VI. CONCLUSION

In this work, a classification methods using SVM and HMM was designed by empirical guidance. These methods were applied to SAVEE data base using a set of features including MFCC and LFCC.

Experiment results demonstrate that our method can serve as a viable approach for the classification of emotions from speech with a recognition rate reaching 61.25%. Besides, we have been able to conclude that MFCC describes better emotional state in speech than LFCC, and that HMM is better classifier than SVM for the used set of data.

Many future modifications can be integrated within this framework. We can for instance develop the used methods with a larger emotional speech databases with reasonably large numbers of speakers in order to improve the generalization of the classification results.

REFERENCES

[1] D.Ververidis, C. Kotropoulos, and I. Pitas, “Automatic emotional speech classification”, IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '04,vol. 1, pp. 593-596, Montreal, May 2004.

[2] J. Kreiman and B. R. Gerrat, “Perception of aperiodicity in pathological voice”, Acoustical Society of America, vol.117, pp. 2201-2211, 2005.

[3] L. R. Rabiner and M. R. Sambur, “An algorithm for determining the endpoints of isolated utterances,” Bell System Technical Journal, vol. 54, no. 2, pp. 297-315, February 1975.

[4] L. Rabiner, R. Schafer, “Digital Processing of Speech Signals””,Pearson Education, 1978.

[5] R. Le Bouquin, Enhancement of noisy speech signals: application to mobile radio communications”, Speech Commun., 18 (1) (1996), pp. 3–19.

[6] Specht, D. F., “Probabilistic neural networks for classification, mapping or associative memory”, Proceedings of IEEE International conference on Neural Network, Vol. 1, pp.525-532, Jun. 1988.

[7] V. Petrushin, Emotion recognition in speech signal: experimental study, development and application, in: Proceedings of the ICSLP 2000, 2000, pp. 222–225

[8] B. Schuller, G. Rigoll, M. Lang, Speech emotion recognition combining acoustic features and linguistic information in a hybrid support vector machine-belief network architecture, in: Proceedings of the ICASSP 2004, vol. 1, 2004, pp. 577–580.

[9] M. Slaney, G. McRoberts, Babyears: a recognition system for affective vocalizations Speech Commun., 39 (2003), pp. 367–384

[10] Inanoglu, Z., & Caneel, R. (2005). Emotive alert: Hmm-based motion detection in voicemail messages. In Proceedings of the 10th international conference on intelligent user interfaces (IUI’05), no. 585. San Diego, California, USA: ACM Press.

[11] S. Haq, P. Jackson, and J. Edge, “Audio-visual feature selection and reduction for emotion classification”, In Proc AVSP, pp. 185-190, 2008.

[12] X.Zhou, D.G.Romero, R.Duraiswami, C.E.Wilson, S.Shamma, “Linear versus mel frequency cepstral coefficients for speaker recognition”, IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU), pp.559-564,2011.

[13] S.B. Davis, P.Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”, IEEE Transactions on Acoustics speech and Signal Processing, Vol ASSP-28, No 4, pp 357-366,August 1980.

[14] L.R Rabiner, “A tutorial on hidden markov models and selected applications in speech recognition”,Proceedings of the IEEE, Vol 77, No 2, February 1989.

[15] V.N.Vapnik V.N, “Statistical Learning Theory”, Wiley-Interscience, New York, 1998.

[16] S. J. Young, G. Evermann, M. J. F. Gales, T. Hain, D. Kershaw, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, P. C. Woodland, The HTK Book, version 3.4.2006.

[17] S. Canu, Y. Grandvalet, V. Guigue, A. Rakotomamonjy, SVM and Kernel Methods Matlab Toolbox, Perception Systèmes et Information, INSA de Rouen, France,2008.

Proactive Software Engineering Approach to Ensure Rapid Software Development and Scalable Production with Limited Resources

A. B. Farid
Information Systems Department
Faculty of Computers & Information, Helwan Univ.
Cairo - Egypt

***Abstract*—Nowadays, the need for building scalable systems in narrow time window is needed. While the efforts and accuracy usually required for building high scale systems is not simple, the agile nature of system requirements spawn a need for enhancing some software engineering practices. These practices should be integrated together in order to help software (SW) development teams to build, and test scalable systems rapidly with a high confidence level in their scalability.**

This research explains the proposed Proactive Approach, which presents a set of software engineering practices that could help in producing scalable system while minimizing the wasted time within the production cycle. This set of practices have been validated, verified and tested through building 46 releases of one of the most important, mission critical and scalable systems. Applying these practices succeeded to enhance average response time of web pages by %1921.5, test code churn by more than % 5000, time to release by % 300, and succeeded to produce a system that could stand against 95375 users with % 99.921 scalability ratio.

Keywords—***Software Engineering; Load Testing; Test Analysis; ISO 29119; Continuous Integration; Static Analysis; Stress Analysis; Application Scalability; Building High Scalability System; Build Verification Test; Software Configuration Management; Version Control; Source Control***

I. Introduction

Working in a project that targets building scalable system with a limited man-power is a common request those days. Using Agile practices in an organized way may lead to eliminating a notable portion of wasted time [1], [2], [3]. Many previous research work have spoken about how to develop a high quality scalable system. Unfortunately, all previous work efforts assumed the availability of enough resources to accomplish the mission conveniently [4],[5],[6],[7],[8]. Going through a software (SW) development project that has limited amount of different types of resources including; budgets, time, and man-power resources, needs to implement an enhanced approach that presents a set validated set of enhanced engineering practices. This set of practices should assure, not only, building the system with the limited set of resources, but also it should help in assuring the scalability of the system when this is needed. This paper is targeting bridging this gap through designing a proposed approach of engineering practices that target achieving three objectives:

1) Achieving the development tasks rapidly without wasting time in fixing code through:

a) Minimizing the Inability of development teams to comply with the code writing best practices standards (e.g. writing secure code best practices, writing a reliable multi-threaded code, code naming conventions,...etc.)

b) Designing the version control change- sets taxonomy in a way that helps in extending the code with newer releases, batches, and fixes in a way that minimizes code churn, and code rewriting.

c) Establishing a proactive Quality controls that make sure that the code units are performance-friendly units.

2) Automating many SW engineering tasks that may need some technical staff. This could help in minimize the dependency on man-power thus, minimizing the amount of needed man-power.

3) Assuring acceptable scalability levels of systems through passing reliable set of load tests.

In order to verify the proposed practices, a 4.5 year research study has been conducted on one of the highly scalable systems that has load of 95,375 simultaneous users with a 55 million potential users. At the early stages of this study, Microsoft has published a case study about the engineering practices that have been developed overtime, and marked those practices as successful [9]. It is important to mention that this research results could help under the following assumptions:

1) Team size doesn't exceed 9 members including all roles

2) Time to release is limited relative to the amount of required work items.

3) Computation resources are limited.

4) The required system should be scalable to huge amount of users.

This paper will directly goes through the proposed SW engineering approach's practices. First, it will introduce the recommended design of the version control workspaces for the code stored during the development phase. Second, the paper will show how these practices could assure proactively the code quality during the development and before moving to the quality control. Third, the paper will explain the recommended steps towards conducting a reliable load testing for assuring system scalability. This part will talk also about analysing the load test result and applying corrections. Then, the paper will explain more details about the study that has been conducted to verify these practices and the results that has been shown out of it. The paper ends with showing the future work and conclusion.

II. Lifecycle Configuration Management's Controls

The name of the game for building a reliable solution rapidly is to design the development process in a way that helps in accelerating the development lifecycle while maintaining high quality levels for the written code. The proposed set of practices assumes a group of pre-set controls. These controls are:

- Version Control (VC) Server has to be used[1].
- Before Uploading code to the Version Control Server the code has to automatically pass through static analysis code reviews.
- Before being uploaded to the server some unit tests has to be passed by the uploaded code.
- To assure high level of continuous integration, the version control has to test the written code before accepting it.

III. Configuring the Version Control Branches

The VC server such as Microsoft Team Foundation Server (TFS), or IBM Rationale [10] has to be configured in a way that helps in automating the code uploading process that is usually called 'Checkout'. It is important to organize source codebase in a way that simplify any development, and maintenance of the application's source code.[11] Fig. 1 and Fig. 2 depicts how the source code could be organized in a way that simplifies applying fixes, and releases.

The *Main* codebase workspace branch could be created over version control as baseless branch from the older version [12]. Fig.2 depicts the taxonomy of branches for version 2. This simple model provides an easy and consistent VC taxonomy for utilizing Forward Integration (FI) and Reverse Integration (RI) models between the *Main* and *Dev* branches, yet allows for increasing complexity with the addition of future development branches when needed. A development branch has been taken from the Main branch of code to drop all required components and write new replacing ones.

Multiple development areas are supported by creating additional development branches from *Main*. These are peers to each other and children of *Main*.

Any additional releases are supported by creating additional release branches for each product release. Each release branch is a child of *Main* and a peer to each other (e.g. release2.0 branch is peer to release3.0 and both are children of *Main*). Once the release branch is created *Main* and the Dev branches can start taking changes approved for the next product release. After the first Reverse Integration of the code (RI1) the first final release build takes place and this generates Version 1.0. After this point a new branch is being created which is the *Release* child branch. While running the release in production, the cycle of load testing begins (for more information concerning what are the proposed steps for conducting a reliable load test , please check section V). According to the issues that will be discovered during load testing, some fixes are expected to be applied over the release branch that has updated the *M*ain branch through Reverse Integration RI2 that in turn updates the *D*ev branch through the Forward Integration FI2.

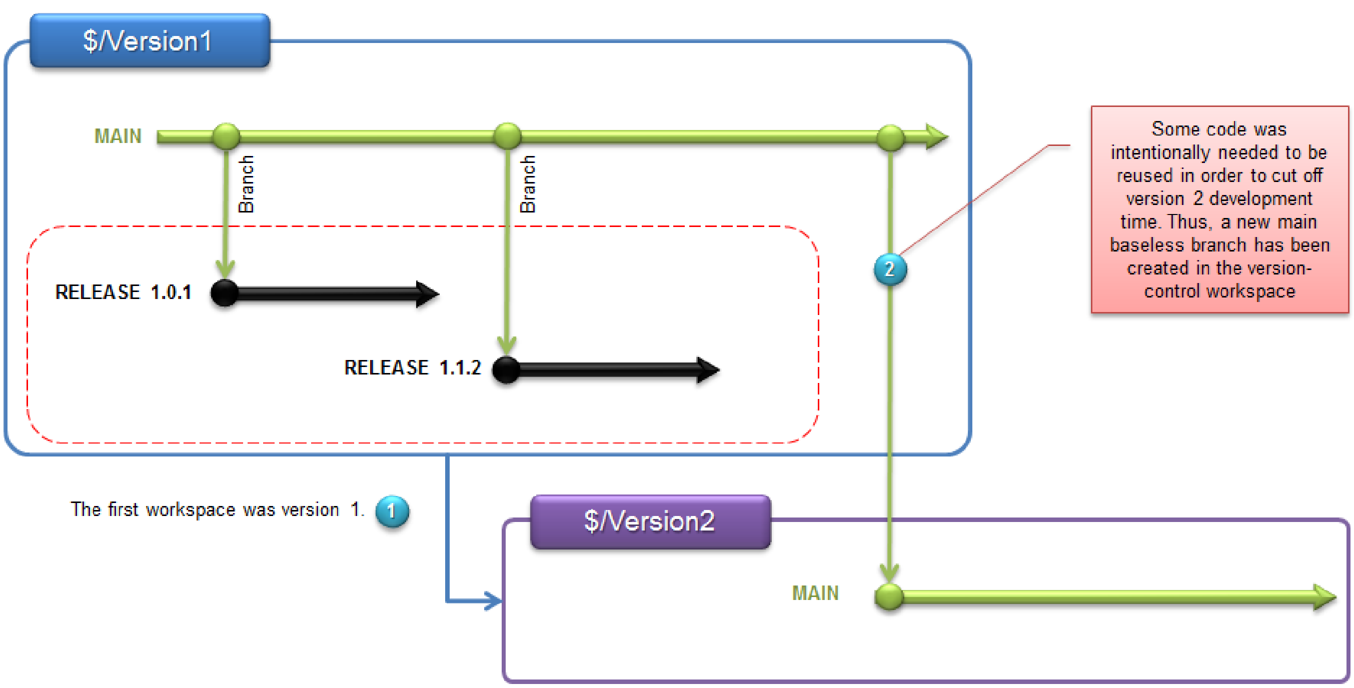


Fig. 1. The Design of the Version Control Workspace based on a previous available version

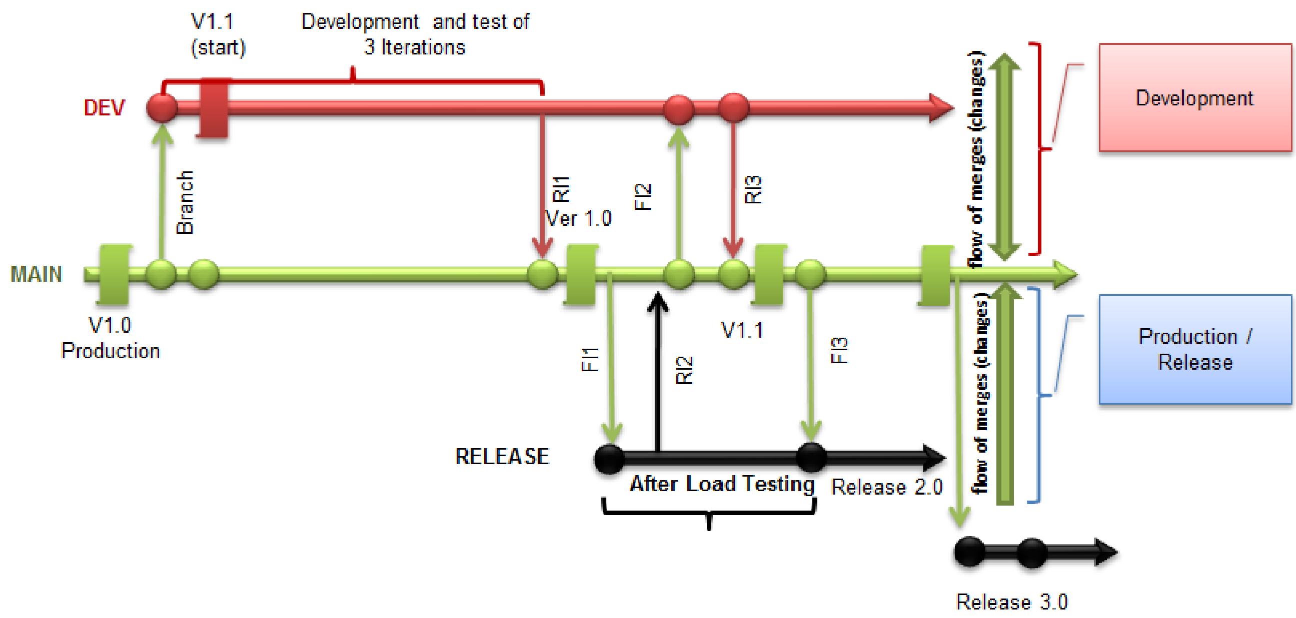


Fig. 2. A Proposed Branching Taxonomy of Version Control Workspace

After fixing some minor final bugs in the *D*ev branch, the main branch should be updated back through Reverse Integration RI3. This version of *M*ain is built to generate version 1.1 that updates *R*elease 2 branch through the Forward Integration 3 (FI3). A new branch could be created at this point that is release 3. Release 3 could be managed as a major final release. This release is assured to be scalable enough and complying with the code writing conventions and standards!

IV. Proactive Quality Control Check-Up

Configuring Version Control (VC) code uploads (check-ins) in a way that helps in controlling the quality in a proactive mood is, of a great importance for the team. This could save the time and staff that are needed to review the code in a manual mood with humanitarian effort. This applied specially when the available time is too tight, and no room is available for discovering the bugs even after regular nightly builds are accompanied with Build Verification Tests (BVTs).

When a developer checks-in a new code that breaks the build, the result could be a significant hassle for the teams. The cost to larger teams can be even more expensive when measured by lost productivity and schedule delays. To guard the code base against these problems, a fake build Server could be configured. This fake build server could be an auxiliary build server. Before checking the code into the VC server, the VC server sends first the code to the fake build server where specific build definition along with its Build Verification Tests (BVTs) are being applied. The build definition doesn't really build the code. Instead, it helps in determining whether the source code that is required to be checked-in, will most probably pass the BVTs, and the build when they take place, or not? This minimizes the number of build failures thus, minimizing the lost time, while preserving the flexibility of Continuous Delivery (CD) [13]. To achieve this, the VC server (e.g. TFS) could be configured through writing a program that acts as a coded Check-In rule. Each time a piece of source code is requested to be checked-in to the VC server, this Check-in rule triggers that verification build definition in order to run over the fake build server before checking-In the source code. The build definition in turn, triggers the associated Build Verifications Tests (BVTs) that has to be applied automatically over the source code. Part of these tests and checks was about checking the architecture rule and constraints. These collections of constraints have put some mandates on the architecture of the system. For instance, a rule that prohibits any developer from checking in a code that directly access the System's database (DB) from presentation tier, or application tier components. If the code passes this set of BVTs and checks; the fake build server notifies the VC server that the source code is acceptable to be checked-in otherwise, it notifies the VC server to reject checking-in the source code. Fig. 3 explains these steps in a graphical way.

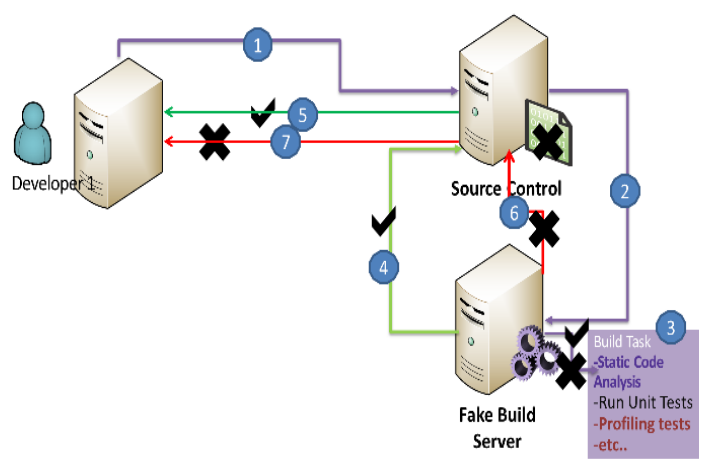


Fig. 3. Proactive quality control check-up steps

V. LOAD TESTING SCALABILITY OF THE SYSTEM

Building a scalable system that is intended to serve for instance millions of users is not an easy task. Unless the system is scalable enough the system will be useless and all investments that have been spent into building it will vanish. When the system has a strategic importance (e.g. mission critical systems, financial institutions systems, or national-level), the financial loss is nothing compared to the loss it might be caused to the national security. That is why, it is crucial to assure that the system of this effort will scale to the required amount of users [14]. That is why it is important to have a clear load test plan for assuring the scalability of the system throughout the application lifecycle [15],[16]. The plan is highly recommended to be based on ISO 29119 Part 2 that describes the test process, 29119 part 3 that describes test documentation, and 29119 part 4 that describes the test techniques [17], [18], [19], [20]. This roadmap should include the following ordered tasks:

- Envisioning, and planning for the system scalability.
- Conducting an assurance Strategy.
- Planning for a load test.
- Conducting the load test.
- Analysing results
- Taking corrective actions.
- Conducting Isolation retesting.
- Reviewing lessons learned.

A. *Envisioning, and planning for the system scalability*

During envisioning, it is important to understand the vision, and the required scale of the system. During the planning of the system, it is important to calculate the maximum expected number of users. Sometimes, it could be a fixed expected number (e.g. the total number of citizens of a country for national systems, or total number of employees for enterprise systems,..etc.). This is the planned number that the team has to put in consideration while planning for system's capacity. Understanding the potential load of a given system is crucial towards understanding the required appropriate system architecture to be designed.

B. *Conducting an Assurance Strategy*

At first place, it is important for the team to assure that the code performance on a single user mood will be acceptable enough, and no design or code writing performance unti-patterns will be made. Afterwards, it is important to make sure that the whole application will be scalable enough against the expected users load, and no architectural mistakes will be committed. This should be done through configuring the proactive check-ups during checking the code into the VC server (for more information about proactive check-ups, please check The Proactive Quality Control Check-ups). These check-ups assures in certain way that the written code complies with the code writing performance friendly best practices that are requested by the team's leadership. It is phenomenal fact that the overall system scalability couldn't be assured for huge amount of users unless performance suitability level could be assured initially in a single user level. That is why part of the proactive pre-Check during conducting tests should be; the Performance tests. If the page or unit of code will not pass the predesigned QC proactive checks through achieving certain performance threshold levels, the VC server refuses to accept checking in this source code. This assures the quality of the code during the code writing phase, and before transferring it to any Quality Control (QC) team.

C. *Planning for a Load Test*

After building the first release of the system, it is important to conduct a comprehensive load test over the system. Conducting this load test is one of the most crucial tasks during the development lifecycle in order to check the system resources' behaviour against the expected users load. This should include the following steps:

1) Preparing the Testing Environment: In order to conduct a realistic load test, it is important to have a test environment that typically imitates the actual production environment from the resources capacity point of view (i.e. network bandwidth, storage, memory, Input-Output (IO) speed, and processing power) capacity. That is why deploying this environment over physical machines is highly recommended and preferred than using a collection for virtual machines.

2) Defining the duration of the load test: Defining a suitable duration period of the load test is crucial towards receiving accurate results. According to this study, the suitable load test duration should be defined according to the amount of time that you expect to have a peak load in. For instance, if it is expected to have a three day special offer on an e-commerce system, it expected to have a peak load during these days. In this case the suitable load test duration length should be 72 hours. Choosing a shorter period of the load test may result in experiencing a crash of the system when being put in production for a longer period. According to the experiences that have been gained through this study, part of the load is coming not only from the number of users, but also from the amount of this number that is pressing on the system's resource for certain amount of time.

3) Calculating the actual load size:

Calculating the expected users load means calculating the simultaneous users that use the system at any point of time. Three factors always affect calculating the expected user load; the total number of users, duration of load, and the major usage scenario that mostly unveils the peak time of load. If the expected users have specific maximum number then, you have to have an assumption on how much time it is expected to find all users visiting the system (e.g. one day, one week, one months,…etc.). This means it is needed to calculate the peak load based on that. Additionally, you need to calculate the average time that each user will consume while working on the system based on the main user story. For instance; let us assume that the total number of users is 54 million users over 48 hours with an average usage time of five minutes per user. Based on that, it is possible to calculate the maximum expected simultaneous users as follows:

Total Number of targeted users (TNV) = 54 Million.

Peak Duration (*PDM*) in Min. =48 hrs. * 60

Average Time/Usage Scenario (*AVU*) = 5Min.

Number of Simultaneous Users Load (*SUL*).

$$SUL = AVU\left(\frac{TVN}{PDM}\right) \qquad (1)$$

SUL = 5 (54,000,000/2880)

SUL = 93750 Simultaneous Users Load.

After adding a safe Margin of 5%

The Final User Load (FSUL) = 98500 Users

4) Setting an architecture for the load test Architecture

Fig. 4 depicts a the proposed load test servers' architecture. Based on this architecture, there is only one Test Controller (TC) server that executes and controls the test, and finally sends back to the client machine the load test counters' values. In another hand this TC server manages and calls one or more Test Agent (TA) servers that imitate the actual users load.

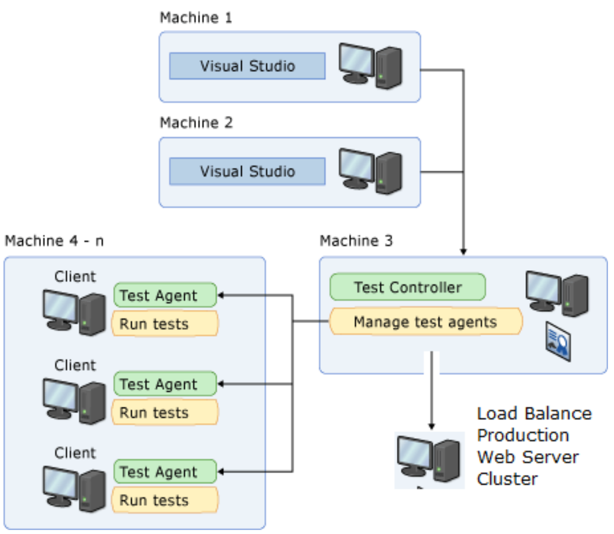


Fig. 4. Load Test Architecture

Each Test agent is responsible of generating number of certain number of users. The main challenge is to define the suitable number of agents that can generate the desired number of users. One good technique to define this number is the Goal Based Load Test. This gives the number of generated users when the TA resources reach certain threshold level. Assuming that the amount of users load that could be smoothly generated by the TA is η. Then the number of required TAs NTA to conducted a load test that could give reliable could be calculated as follows:

$$\mathrm{NTA} = \frac{SUL}{\eta} \qquad (2)$$

5) Risk Mitigation is an important part of any test. Based on the previous assumption of resources limitations, the whole test load test process is in a risk of experiencing any hardware failure. If hardware crashed, there will not be enough resources available including money, and time to replace these resources. According to ISO 29119, it is important to have a risk mitigation plan. This mitigation plan should be based on using the cloud infrastructure instead of using the on premise deployment. This may provide a more appropriate economical, and fast to gain solution especially when the amount of servers needed is not attainable due to resources limitation. Working over the cloud is another effort that will be extended in the near future.

D. Conducting a Load Test

In order to have a good analysis to what is going on; the load test should be conducted through a gradual step by step process. According to the study, Scalable load tests should go gradual since they need to run for long times. Since it needs a serious amount of resource to be available, it is not practical to start a 48 hours intensive load test all of a sudden. This should take place as follows:

Step 0: Running a constant load of the designated potential users load (previously referred to as SUL) for 10 minutes.

It is important to begin with a short period of test to make sure that the deployed architecture could stand against the potential load or not. Sometimes, the resources shows inability to stand against the load at all, and the load crashes at the early beginning or malfunctions so, running the test for this short amount of time could be a good start as just try.

Step 1 Run the test for an hour.

Running the test for one full hour gives more confidence in the available resources and their ability to cope with the load test itself. While the remaining steps are directed towards testing the system itself, steps 0 and 1 are directed towards testing the load test itself. Fig. 5 shows an example of the intense of errors that could arise when the resources are not enough. The message alert says that the agent has failed.

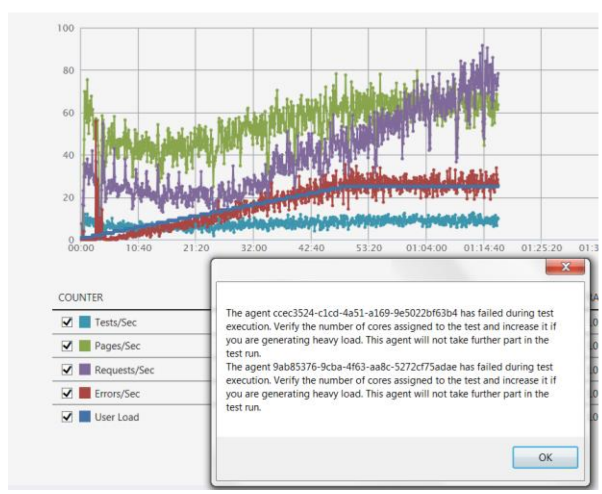


Fig. 5. shows an excerpt of erros after 1 hour

Step 3: Run same load test ¼ the total load test duration with full load size:

For instance, if the system is expected to work with the maximum load for 48 hours, accomplish this load test step for 12 hours. This gradual load test duration is important to have when there is a large scale of users that is expected for a constant relatively long period. In relatively long tests that exceeds 24 hours, it is important to accomplish the load test gradually so, if your test spawns number of errors that exceeds the predefined threshold you don't need to wait till the end of the long load test. This saves resources, time, and increase the speed of reaching a more stable system.

Fig. 6 shows one of these experienced cases during the study, more than 60 errors have been discovered during the first 2 hours. This is enough to stop the first run to analyse and trace logs. When such experience happens, it is highly recommended to trace the log files to check the root cause of the resulting errors. One of the most common root causes is the capacity of the Input-Output storage capacity. Due to the huge load, load agents fail to continue applying the load test due to problems with Input-Output storage unit that writes the test log. Whatever the problem is, it should be solved and the test should be restarted smoothly gain with no issues. Running the test successfully for ¼ of the total test length period (12 hours if more than 24 hours) could be the first positive sign that the architecture of the system is scalable enough to stand against such huge load for certain amount of time. In another hand, it couldn't assure that the system will perform well for the full required duration.

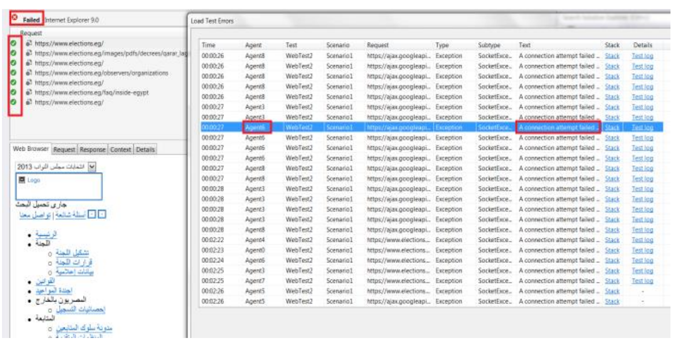


Fig. 6. Reviewing load Agents' status and log files

Step 4: Run Load Test of a half duration length.

Running a test for 1/2 duration gives more confidence to the test analyst through showing a clear image about the resources required to accomplish the actual full duration test. Usually, running the test with this period will give some lessons that could be helpful towards conducting the full-length test. Fig. 7 shows an excerpt of a sample run of 24 hours (1/2 duration length in the study) that shows a smooth load run. The figure shows that the number of errors has been dropped significantly after 24 hours (less than 10 during 24 hours).

Step 5: Run the predesigned full test duration period.

This is the actual planned load test. After solving the load test issues gradually over different durations, through the previously conducted load test steps, the test has to pass the test duration successfully. This doesn't always mean running the test with no errors; it means that the number of errors shouldn't exceed a threshold that is defined during the test planning phase.

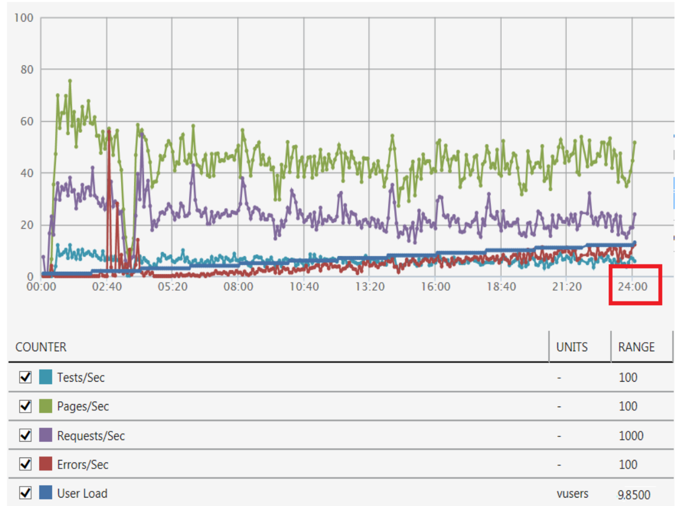


Fig. 7. An excerpt of the Load test analysis for the 24 hours load test run

E. *Analysing Test Results and Applying Corrections*

Some errors/warnings may appear at the end of the load test that was related to the storage speed, processing power, and a minor warning for the network bandwidth.

According to the conducted study, after investigating the possible causes for the load tests for 46 releases under the verification study, the root cause of the insufficient average response-time may be the processor clock speed and storage IO speed. Enhancing the response time with a significant enhancement, while consuming the least possible resources could be achieved through two logical options:

1) Option -1 Increasing the Processing Clock Speed

Firstly it is important to decide whether it is more feasible to enhance the performance through increasing the number of processors, or to enhance the performance through enhancing the IO storage speed? If the recourses availability is infinite then, there will be no issues however; the truth is that always in software development, it is important to trade-off decisions with the available resources. This enforces tackling the available budget for availing resources. In order to answer this question it is important to find out the percentage enhancement based on the dollars spent in that. If the dedicated budget would increase the number of CPUs 4 times from 8 to 32, and giving in considerations that 60% of the system's code is written as a parallel code. Based on that Amdahl's law [21] could be used to calculate the expected performance enhancement if CPUs increased from 8 to 16, from 16 to 32, and from 32 to 64. Table 1 shows Amdahl's Law calculation for 10% parallel code [21]. This law calculates the performance enhancements due to CPU increase knowing the percentage of existing parallel code as follows:

Amdahl's Law: (2)

$$x = \frac{1}{\propto + \frac{1-\propto}{p}}$$

Where:

x : The maximum speedup that could be achieved.

$\propto$: The percentage of the sequential code in the system.
p: Number of processors used

TABLE I. PERFORMANCE INCREASE DUE TO PROCESSOR COUNT INCREASE IN CASE OF 10% PARALLEL CODE

Trial #	p	x	Performance Increase	Cumulative Increase
1	8	1.09589	N/A	N/A
2	16	1.10345	0.69%	0.69
3	32	1.10727	0.35%	1.04%

Fig. 8 shows the achieved speedup for a system with 50%, and 60% parallel code. These numbers shows that the maximum possible speedup that could be achieved with a reasonable amount of cores (510 Core) is less than 1.5. This means that if it is needed to increase the speed of a given system. Adding additional processing power will not result into a huge difference. This leads to searching for another option for making an evolutionary speedup enhancement.

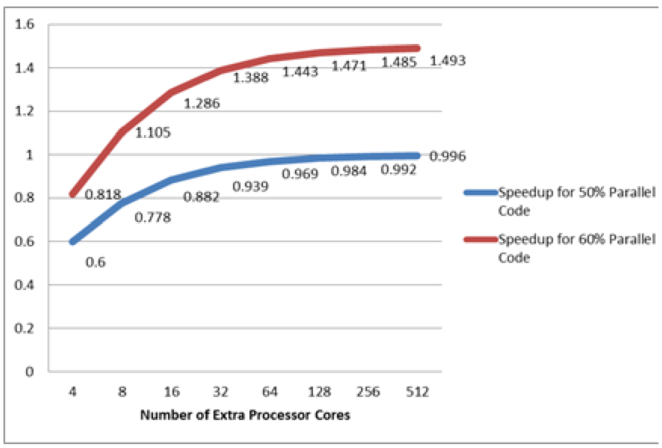


Fig. 8. System Speedup according to No. of processing Cores when parallel code percentages are 50%, and 60%

2) Option -2 increasing the Storage speed

Usually, in a limited resources environments system are being deployed on storage regular storage Hard disk drives (HDDs), or HDDs enrolled in RAID array. Increasing the speed of the system may need increasing the IO speed. According to many studies and benchmarks that has been applied and tested during this research moving all database files to a solid stage device (SSD) storage media could enhance the speed of the overall system up to 2030.34%. SSD technology has been compared to Flash SSD technology. It is very well known for everybody that SSD storage technology is faster than regular SAS HDD storage devices. According to DELL [22], and IBM [23] using Flash SSD is highly recommended for READ/Random-Access intensive systems specially when the read ratio is not less than 85%

According to a study that has been conducted by DELL [22] on random access, Read intensive systems, Flash SSD storage can perform up to 59.46 times better than SAS HD, with a price(USD)/IOPS (IO Operations Per Second) ratio of only 33% [22]. This means that the performance gain is strongly justifying the price difference! This shows that enhancing the storage speed (Maximum enhancement could be 5946%) will give better results than enhancing the processing power (maximum performance enhancement could be 1.04%).

According to some studies that has been conducted previously, and to the research that has been for done this study, moving the DB indexes to SSD drives enhance the system performance with less than 10% extra cost [23].

VI. PRACTICES VERIFICATION

In order to measure the ability of the previous software engineering practices in enhancing the speed of building a scalable system, a four years study has been conducted as part of the process of building two of the national information systems with a potential total load size of 55 million users and 95375 simultaneous users [9]. The study began on March 2011, and finished on September 2015. During this study, six full development cycles/Major versions have been conducted with 46 different releases [24][25], [26].

The first Major version was managed without applying any part of the above proposed practices to be the reference sample version for any changes that could happen after applying the above proposed engineering practices. In 2012 the system has been fully rebuilt with a new version 2.0 while developing, applying and verifying the proposed engineering practices. Then, another four releases of both systems has been produced (Version 3.0, Version 4.0, Version 5.0, and version 6) through separate four development major versions. During each version out of the five (Ver. 2 to Ver. 6), some lessons have been learned and the practices have been enhanced to help in enhancing the next release production. Table 2 shows a comparison between Version 1.0 practices situation and version 6.0 including; the group of proposed practices that have been applied in Version1 and version 6.0 Fig. 9 shows the final load test result of the final release (release 46) of this study with no load errors.

Fig 10 shows the enhancements that happened in the Code Churn metrics due to applying the proposed practices. It is clear that the effect of applying the proposed practices has led to a significant improvement in the No. of lines of codes that are deleted, and modified.

TABLE II. SUPPORTED PRACTICES IN VER.1.0 AND VER. 6.0

Practice	Ver. 1.0	Ver. 6.0
Using Proposed Branching Taxonomy	No	Yes
Using Proposed Proactive architecture check-up	No	Yes
Using Proposed Proactive Static code analysis	No	Yes
Using Proposed Proactive Code Performance Test	No	Yes
Using Proposed Proactive build checks using fake build server	No	Yes
Using Proactive BVT	No	Yes
Using Proposed Goal Based Load Test to define required number of test agents	No	Yes
Applying Gradual Load Test.	No	Yes
Storing Database Indexes of the system data on a SSD drive	No	Yes
Update proactive check-ups based on lessons learned	No	Yes

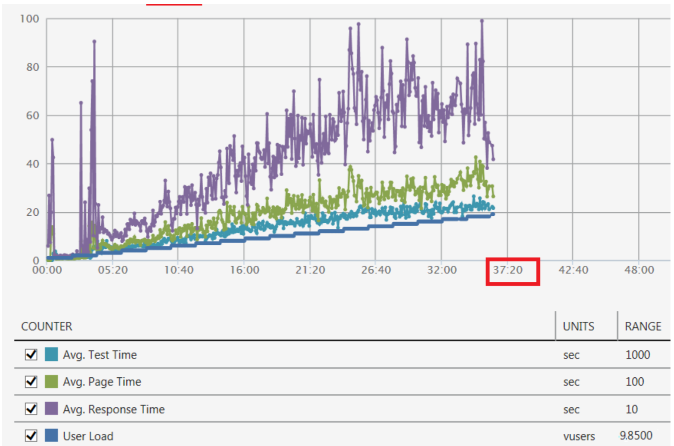


Fig. 9. the results of the load test with no errors after 37 hours of the Load test run

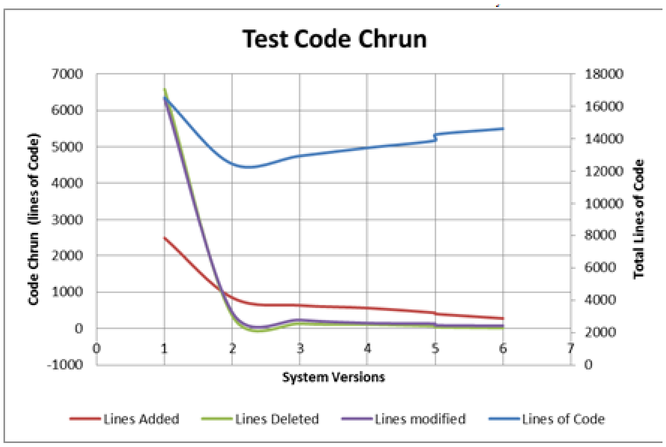


Fig. 10. Enhancements in Code Churn metrics across versions

The numbers of lines that have been added are basically depending on the new set of requirements that have been requested over the six major versions of the system under development during the study. The difference between the Ver.1 where no practices have been applied and other practices is clear to be noticed.

According to the Fig.10, applying the proposed SW engineering practices has enhanced the test code churn more than % 5000. The Average response time of the system pages has been enhanced from 73 seconds to 3.8 seconds by % 1921.5 enhancement ratio. The time to release has been cut down from 15 weeks though 5 sprints to 5 weeks through two sprints with % 300 enhancements. Additionally, it succeeded to reach the required availability percentile of 99.92 after being 17.886% only. At the same time the system scaled up from 1843 simultaneous users to 95375 with 5170% scalability enhancement rate. Table 3 concludes the key enhancements that could be measured between version 1.0 that used regular agile practices, and version 6.0 that applied the Proactive Quality Approach. Table 3 summarises these results with the enhancement achieved in each results between Ver.1.0 that has used none of the proactive approach's practices, and Ver. 6.0 that has used all practices of the approach. Fig.11 summarises how are these practices distributed over the different cycle activities. According to the figure, it clear that the lessons learned during the different phase over the cycles should lead to updating the proactive test rules that are being applied on the uploaded code to the version control server in the next releases.

TABLE III. COMPARISON BETWEEN USING REGULAR AGILE PRACTICES IN VER.1.0 AND USING THE PROACTIVE QUALITY APPROACH IN VER. 6.0

Comparison Aspect	Ver. 1.0	Ver. 6.0	Enhancement %
Code churn(Lines modified)	6321	75	842.8
Average page response time	73 Sec.	3.8 Sec.	1921.5
Time to release (Per sprint)	15 Weeks	5 Weeks	300
Availability	17.886	99.92	558.6
User Scalability	1843	95375	517.5

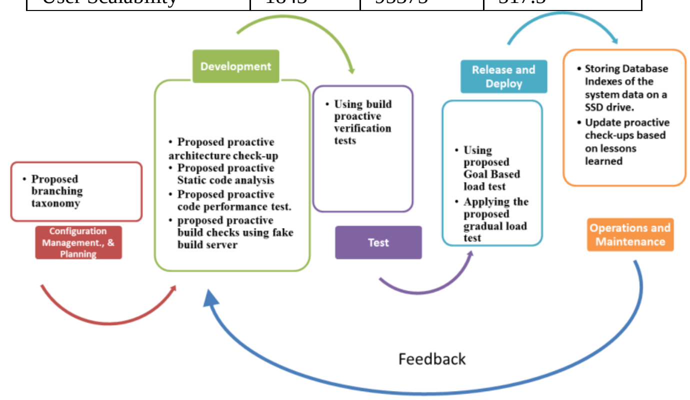


Fig. 11. Major practices of the Proposed Proactive approach across the different cycle activities

VII. FUTURE WORK

In the future this research study will be repeated over cloud-based platform in order to test the effect of using the cloud vs. using the on premise deployment within a limited resources based project. Some extra Application Lifecycle Management (ALM) maybe used to enhance the overall quality of the system.

VIII. CONCLUSION

This research study has shown the planning of the SW engineering process design for developing a resource-limited high-scale, and mission-critical system. The study showed how does version control could be utilized to streamline any changes that may arise in the middle of the project, and how continuous integration could be mixed with some proactive check controls that can assure the compliance of the checked-in code with the predefined quality assurance measures.

The study explained the proposed gradual load testing process that has to be conducted to assure the scalability of the system to the expected amount of transactions and users. Having a clear load testing strategy that complies with the major business requirements is a major success factor for the whole system. This is compliant with ISO standard number 29119 that is concerned with SW testing. Conducting the load test in a right way is important, however, analysing the resulting errors and taking corrective actions could make the load testing more valuable since it helps in enhancing the overall scalability of the system. Some good lessons have been learned and elaborated at the end of this study. Applying the proposed practices has led to enhance many indicators with a notable percentage.

REFERENCES

[1] Otvio Augusto Lazzarini Lemos, Sushil Bajracharya, Joel Ossher, Paulo Cesar Masiero, and Cristina Lopes, "A test-driven approach to code search and its application to the reuse of auxiliary functionality", *Information and Software Technology, No. 53, P. 294-306, 2011, , ElSevier.*

[2] F.G. Wilkie , "An Investigation of coupling, reuse, and maintenance in a commercial C++ application", *Information and Software Technology, No. 53, P. 801-812, 2001: ElSevier,*

[3] Y. M. Helmy , A. B. Farid, and A. S. Abd elghany, "Simplifying CMMI Version 1.3 Implementation By Using Agile Practices: An Empirical Study", *International Journal Computing, and Information Science",Vol. 14, No. 4, P.31-45, Oct. 2104.*

[4] M. Rizwan Jameel Qureshi, and Isaac Sayid,"Scheme of Global Scrum Management Software",, *Information Engineering and Electronic Business, Vol. 7, No. 2, P.1-7, 2015*: MECS

[5] Saja Al Qurashi, M. Rizwan Jameel Qureshi, "Scrum of Scrums Solution for Large Size Teams Using Scrum Methodology", *Life Science Journal.Vol. 11 No. 8, 2014.*

[6] Samireh Jalali, Claes Wohlin, "Agile Practices in Global Software Engineering - A Systematic Map", International Conference on Global Software Engineering (ICGSE),2010. Princeton,NJ – USA.

[7] Del Nuevo, E., Piattini, M., Pino, F.J., "Scrum-based Methodology for Distributed Software Development", Global Software Engineering (ICGSE), 6th IEEE International Conference, 2011.

[8] Jeong Ah Kim, Seung Young Choi, Sun Myung Hwang, "Process Evidence Enable to Automate ALM (Application Lifecycle Management", Parallel and Distributed Processing with Applications Workshops (ISPAW) 9th IEEE International Symposium, 2011.

[9] Microsoft, "Egyptian Ministry Facilitates Transparently Open Elections with New Applications: a Solution Case Study".[Online].Available:https://customers.microsoft.com/Pages/Download.aspx?id=15111

[10] T. E. Murphy et al., "Gartner, magic quadrant for application development life cycle management", 19 November 2013.

[11] Steven St. Jean et al, "Professional Team Foundation Server 2013", Wrox, May 19 2014.

[12] Mickey Gousset et al, "Professional application life cycle management with Visual Studio 2013", Wrox, March 31 2014.

[13] Eric Sink, "Version control by example", Pyrenean Gold Press, 2011.

[14] Larry Brader, and Roberta Leibovitz, Jose Luis Soria Teruel "Building release pipeline using Team Foundation Server 2012 (TFS 2013 Editions)", Microsoft Patterns and Practices. Feb 2014.

[15] Larry Brader, Howie Hilliker, Allan Cameron Wills, "Testing for continuous delivery using Visual Studio 2012", Microsoft Patterns & Practices, March 2, 2013.

[16] Satheesh Kumar N, , and Subashni S, "Software Testing Using Visual Studio 2013", Packt Publishing, July 26 2013.

[17] *Software Testing: Concepts & Definitions, ISO/IEC/IEEE Std. 29119-1:2013*, September 2013.

[18] *Software Testing: Test Processes, ISO/IEC/IEEE Std. 29119-2:2013,* September 2013.

[19] *Software Testing: Test Documentation, ISO/IEC/IEEE Std. 29119-3:2013*, September 2013.

[20] *Software Testing: Test Techniques, ISO/IEC/IEEE Std. 29119-4 FDIS Draft*, March 2014.

[21] Amdahl, Gene M.. "Validity of the Single Processor Approach to Achieving Large-Scale Computing Capabilities". *AFIPS Conference Proceedings* (30) P. 483–485. (1967).

[22] Vamsee Kasavajhala, "Solid State Drive vs. Hard Disk Drive Price and Performance Study", DELL, May 2012.

[23] Aamer Sachedina and, Mathew Huras, "Best Practices Database Storage", IBM, January 2012.

[24] Jiang, Z.; Hassan, A., "A Survey on Load Testing of Large-Scale Software Systems," in *Software Engineering, IEEE Transactions on* , vol.PP, no.99, pp.1-1 doi: 10.1109/TSE.2015.2445340.

[25] Y.M. Helmy , A. B. Farid, and A. S. Abd Elghany, "Simplifying CMMI Version 1.3 Implementation By Using Agile Practices: An Empirical Study", in International Journal of Intelligent Computing and Information Science *IJICIS, Vol.14, No. 4 OCTOBER 2014.*

[26] Ahmed Bahaa Farid, Enas M. Fathy, Mahmoud Abd Ellatif, "Towards Agile Implementation of Test Maturity Model Integration (TMMI) Level 2 Using Scrum Practices", *(IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 9,P. 230-238. 2015.*

Protein Sequence Matching Using Parametric Spectral Estimate Scheme

Hsuan-Ting Chang*, Hsiao-Wei Peng, Ciing-He Li
Photonics and Information Laboratory
Department of Electrical Engineering
National Yunlin University of Science and Technology
Douliu Yunlin

Neng-Wen Lo
Department of Animal Science and Biotechnology
Tunghai University
Taichung, Taiwan

***Abstract*—Putative protein sequences decoded from the messenger ribonucleic acid (mRNA) sequences are composed of twenty amino acids with different physical-chemical properties, such as hydrophobicity and hydrophilicity (uncharged, positively charged or negatively charged amino acids). In this paper, the power spectral estimate (PSE) technique for random processes is applied to the protein sequence matching framework. First, the twenty kinds of amino acids are classified based on their hydrophobicity and hydrophilicity. Then each amino acid in the protein sequence is mapped to a corresponding complex value. Consider the various Hidden Markov chain orders in the complex valued sequences. The PSE method can explore the implicit statistical relations among protein sequences. The mean squared error between the power spectra of two sequences is determined and then used to measure their similarity. The experimental results verify that the proposed PSE method provides the consistent similarity measurement with the well-known ClustalW and BLASTp schemes. Moreover, the proposed PSE can show better similarity relevance than ClustalW and BLASTp schemes.**

Keywords—protein sequence; amino acids; digital signal processing; parametric spectral estimate; hydrophilicity; hydrophobicity; Markov chain

I. INTRODUCTION

In the past two decades, deoxyribonucleic acid (DNA) and protein sequences in various organisms have been massively obtained with the help of high-throughput sequencing technologies [1]. Biologists unravel the functionality and capability of numerous protein sequence domains by understanding their 3-D structures obtained by the x-ray diffraction technique or NMR technology. These procedures require laborious preparations of protein crystals and are extremely time-consuming. Therefore, alternative methods based on digital signal processing (DSP) technique were developed to circumvent the extremely complicated crystallographic tasks. Generally speaking, two types of methods are commonly used to analyze the protein sequences and predict their functions : (1) Statistical methods [2], which apply the well-known mathematical models in stochastic processes to analyze the sequences. (2) Geometrical methods [3], which apply graphs to represent the sequences and then analyze them. Both types of methods first transform the symbolic amino acids to numerical values. The global or local similarity of any two sequences can then be measured according to the differences between the extracted sequence features. High similarity between two sequences may infer two meanings: (1) the two sequences could be homologous; (2) the protein structures and/or their biological functions are similar.

Recently, various methods are proposed to study the DNA and protein sequences. Among them are the DSP-based methods [4]-[7]. Some of the related studies put the focus on the visualization of sequences in various graphic forms [8]-[15]. In DSP techniques, each character in the DNA sequences or each amino acid in the protein sequence is mapped to a numerical value. According to the characteristics of the organisms, the different values used in the numerical mapping can be designed to accordingly fit the physical-chemical properties [16]-[21]. Thus, the comparison method especially utilizing certain properties of the residence in the DNA or protein sequence must be specifically designed. There are two well-known character-based tools for DNA and protein sequence comparison; ClustalW [22] and BLAST [23]. ClustalW is designed by using multiple-sequence alignment based on the meta-heuristics methods. The feature that the arrangement of each amino acid is similar in the evolution of the same species is utilized. On the other hand, BLAST alignment has four components: query, database, program, and search purpose/goal. BLAST is designed to locate the homologous sequence sites between two sequences using a heuristic approach. It compares partial sequences progressively, such that the local alignment results are obtained. Standard protein-protein BLAST (BLASTp) compares an amino acid query sequence against a protein sequence database. It is used for both identifying a query amino acid sequence and for finding similar sequences in protein databases.

In this paper, a parametric spectral estimate (PSE) method based on stochastic signal processing is proposed for protein sequence comparison. The numerical signals are used to represent the protein sequences and then analyzed in the frequency domain. First, a new model of mapping complex values to amino acids according to the physical-chemical characteristics is proposed. Next, the PSE method is used to determine the power spectrum density (PSD) of each numerical protein sequence. Finally, the mean squared error (MSE) values between two power spectra under various Markov orders are determined and served as a metric for sequence similarity measurement. As compared to the ClustalW and Blastp methods, our experimental results show that the proposed method provides an alternative way to efficiently distinguish the differences between two protein sequences. The remaining part of this paper is organized as follows: Section 2 describes the proposed PSE method. The experimental results under different perspectives and their discussions are provided in Section 3. Finally, the conclusion is drawn in Section 4.

II. Methods

Figure 1 shows the block diagram of the proposed method for protein sequence comparison. There are three parts in this method: (1) numerical mapping; (2) PSE with different Markov orders; (3) MSE estimations, which are described in the following subsections.

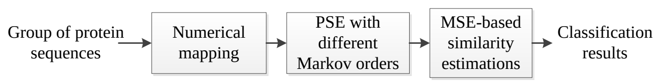


Fig. 1. The block diagram of the proposed method

A. *Numerical mapping*

Table 1 shows the standard one-letter abbreviation of the 20 amino acids and their properties on charge, hydrophilicity or hydrophobicity, which are important for protein structure and protein-protein interaction. [24, 25].

According to the physical-chemical properties, twenty kinds of amino acids can be represented by twenty complex values which locate at the different positions on the unit circle in the complex plane. Figures 2(a)-2(d) show the four arrangements, Methods 1-4, of mapping the symbolic amino acids to the numerical ones. The twenty complex values are distributed on the unit circle whose center is at the origin of the complex plane. As shown in Fig. 2(a) (Method 1), the hydrophilic amino acids are distributed on the upper part of the circle. The first amino acid H is assigned to be 45。 with respect to the real axis of the circle. The other amino acids {R, K, E, D, Q, N, Y, C, T, and S}, which have 9。 separation from each other, are assigned after the amino acid H. The hydrophobic amino acids are distributed on the lower part of the circle. The first amino acid G is assigned to 234。 with respect to the real axis on the circle. The other amino acids {A, V, L, I, P, M, F, and W}, which have 9。 separation from each other, are assigned after the amino acid G. In addition to the hydrophobic and hydrophilic properties, we assign the positions according to their basic structures and the general chemical characteristics in their side chain (R) groups [26]. According to the position of the amino acids on the circle, the mapping is established so that every amino acid is adequately separated.

In addition to Method 1 shown in Fig. 2(a), three other mapping methods shown in Figs. 2(b)-2(d) are also proposed for performance comparison and evaluation. Table 2 shows the complex values corresponding to the coordinates of the 20 amino acids on the circle. In addition to Method 1, Methods 2, 3, and 4 are proposed to verify the effects of amino acid properties in this study by changing the positions on the circle. In Methods 2 and 3, the 20 amino acids still conform the rules of the characteristics of hydrophilicity and hydrophobicity, but the position can be exchanged in the random and horizontally reversed ways, respectively. In Method 4, the mappings are random, and thus none of the rules of the characteristics of hydrophilicity and hydrophobicity is obeyed.

B. *Parameter spectrum estimation (PSE) method*

Consider a stochastic process with the random variable (RV) $\{X_n,\ n = 1, 2, 3, \ldots, m\}$, which describes a protein sequence composed of twenty kinds of amino acids and m is the sequence length. Let a sequence be denoted in Eq. (1), where   represents that the status is i when an amino acid is at the nth position in the sequence.

$$X_n = X_1, X_2, \ldots, X_m, \quad X_n = i \in \{\mathrm{H, R, K, E}, \cdots, \mathrm{W}\} \tag{1}$$

Let P_{ij} denote a transition probability given that the current status is i and the next status is j.

$$\begin{aligned} p_{ij} &= p\{X_n = j \mid X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \cdots, X_2 = i_2, X_1 = i_1\} \\ &= p\{X_n = j \mid X_{n-1} = i_{n-1}\}, \quad \forall n > 1. \end{aligned} \tag{2}$$

Note that Eq. (2) denotes a Markov chain process, in which the probability of the current amino acid X_n is only dependent on the previous amino acid X_{n-1}. A one-step transition probability matrix of order 20×20 is obtained by letting the first–order Markov chain model corresponding to the possible transitions between two amino acids in a protein sequence and is shown in Eq. (3).

$$P = [p_{ij}] = \begin{bmatrix} p_{HH} & p_{HR} & p_{HK} & \cdots & p_{HW} \\ p_{RH} & p_{RR} & p_{RK} & \cdots & p_{RW} \\ p_{KH} & p_{KR} & p_{KK} & \cdots & p_{KW} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ p_{WH} & p_{WR} & p_{WK} & \cdots & p_{WW} \end{bmatrix},$$

$$p_{ij} \ge 0, \sum_j p_{ij} = 1, i \in \{\mathrm{H, R, K, E}, \cdots, \mathrm{W}\}. \tag{3}$$

Let the amino acids in a protein sequence be denoted as a discrete signal source where the occurrence probability of each element is p_i, i=1, 2,…, m. Equation (4) defines the information amount $I(x_i)$ of xi for an event which occurs with a probability p_i.

$$I(x_i) = \log_2(\frac{1}{p_i}), \quad \sum_{i=1}^{m} p_i = 1 \tag{4}$$

The average information or entropy of X is defined in Eq. (5):

$$H_2(X) = E\{I(x_i)\} = \sum_X p(x_i)\log_2(\frac{1}{p(x_i)}) \tag{5}$$

The conditional entropy of the current RV, X, given the m previous RVs: $x_1, x_2, \ldots, x_m$, is defined in Eq. (6):

$$\begin{aligned} &H_m(X \mid x_1, x_2, \cdots x_m) = \\ &\sum_X p(x_i \mid x_{i1}, x_{i2}, \cdots x_{im}) \log_2 \frac{1}{p(x_i \mid x_{i1}, x_{i2}, \cdots x_{im})} \end{aligned} \tag{6}$$

The entropy of a first-order Markov process is defined in Eq. (7):

$$H_1(X) = \sum_{X^2} p(x_{i1}, x_{i2}) \log_2 (\frac{1}{p(x_{i2} \mid x_{i1})}) \qquad (7)$$

where $p(x_{i1},x_{i2})$ is the probability when the two values x_{i1} and x_{i2} occur together. The entropy of a second-order Markov process is defined in Eq. (8):

$$H_2(X) = \sum_{X^3} p(x_{i1}, x_{i2}, x_{i3}) \log_2 (\frac{1}{p(x_{i3} \mid x_{i2}, x_{i1})}) \qquad (8)$$

The homologous gene sequences have similar entropies when a higher-order Markov process is used [27]. In a lower order, however, each set of homologous gene sequences have various entropy values. The q^{th}-order Markov model is shown in Eq. (9)

$$X[n] = \sum_{k=1}^{q} a_k X[n-k] + W[n] \qquad (9)$$

where both RVs X and W are zero-mean and the variance $\mathrm{Var}\{W[n]\}$ of W is denoted as σ_W^2. In the estimation of random variables [28], the q-dimensional vector is denoted as $X \triangleq (X[n-1], X[n-2], \ldots X[n-q])^{\mathrm{T}}$. The Equation $E[Y \mid \mathrm{X}] = \sum_{i=1}^{N} a_i X_i$ provides linear prediction estimate of the scalar random variable $Y = X[n]$ in terms of the $\mathbf{a_k}$. The coefficients $\mathrm{a} \triangleq (a_1, a_2, \ldots, a_q)^T$ are determined as the solution of the orthogonal equation, and the optimum value of $\mathbf{a}$, denoted as $\mathbf{a}_o$ is shown in Eq. (10).

$$\mathbf{a}_o^T = \mathbf{k}_{Y\mathbf{X}} \mathbf{K}_{\mathbf{XX}}^{-1} \qquad (10)$$

Equations (11) and (12) define the cross covariance vector and the covariance matrix of $X[n]$, respectively.

$$K_{Y\mathbf{X}} = \begin{bmatrix} E\{X[n]X[n-1]\} \\ E\{X[n]X[n-2]\} \\ \vdots \\ E\{X[n]X[n-q]\} \end{bmatrix} = \begin{bmatrix} K_{XX}[1] \\ K_{XX}[2] \\ \vdots \\ K_{XX}[q] \end{bmatrix} \qquad (11)$$

$$K_{\mathrm{XX}} = \begin{bmatrix} E\{X[n-1]X[n-1]\} & E\{X[n-1]X[n-2] \cdots E\{X[n-1]X[n-q]\} \\ E\{X[n-2]X[n-1]\} & E\{X[n-2]X[n-2] \cdots E\{X[n-2]X[n-q]\} \\ & \vdots \\ E\{X[n-q]X[n-1]\} & E\{X[n-q]X[n-2] \cdots E\{X[n-p]X[n-q]\} \end{bmatrix}$$

$$= \begin{bmatrix} K_{XX}[0] & K_{XX}[1] & \cdots & K_{XX}[q-1] \\ K_{XX}[-1] & K_{XX}[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & K_{XX}[1] \\ K_{XX}[-(q-1)] \cdots & K_{XX}[1] & K_{XX}[0] & \end{bmatrix}. \qquad (12)$$

To obtain a simple PSD estimate, the covariance function $k_{XX}[\tau]$ is replaced as shown in [28], and the solution yields parameter estimates $\hat{a}_1, \hat{a}_2, \ldots, \hat{a}_q$. Finally, the PSD estimate is defined in Eq. (13):

$$\hat{S}(\omega) = \frac{\hat{\sigma}_W^2}{|1 - \sum_{k=1}^{q} \hat{a}_K \exp(-j\omega k)|^2} \qquad (13)$$

where $\hat{\sigma}_W^2$ denotes the variance and $\hat{a}_k$ is a parameter of RV. The covariance matching property when the PSD function is regarded to as an auto-regressive model is defined in Eq. (14).

$$\mathrm{IFT}\{\hat{S}(\omega)\} = \hat{R}_x[\tau], \quad |\tau| \leq q', \qquad (14)$$

where IFT{} denotes the inverse Fourier transform, τ is the power spectral order, q' is the highest order, and $\hat{R}_x$ is the auto-correlation function of $\hat{S}(\omega)$.

C. MSE determination

When the PSD of each sequence in the q^{th} Markov order is determined by the use of PSE, the PSD values are normalized within the range [0, 1]. Next, the MSE defined in Eq. (15), is used to compare the similarity between two protein sequences y_i and y_j under the q^{th} Markov order.

$$\mathrm{MSE}(y_i^q, y_j^q) = \sqrt{\frac{\sum_{k=1}^{N} [y_i^q(k) - y_j^q(k)]^2}{N}}, \qquad (15)$$

where the parameter N is the length of protein sequence, y_i^q and y_j^q are the normalized PSD values under the q^{th} order transformed from one and the other protein sequences, respectively. If the lengths of two sequences are different, their PSDs are not of the same length, either. Thus the MSE cannot be directly computed. To solve this problem, two methods are used to make two sequences the same length. First, the shorter sequence is interpolated to be of the same length with the longer one and is denoted as y_{sl}. Second, the longer sequence is down-sampled to be of the same length with the shorter one and is denoted as y_{ls}. These two methods are determined by Bi-linear interpolation shown in Eqs. (16) and (17).

$$y_{\mathrm{sl}}(j'_n) = \alpha_n \times y_{j+1} + (1 - \alpha_n) \times y_j, \quad j = 1, 2, \cdots, N_{\mathrm{s}}, \qquad (16)$$

for $j'_n = \begin{cases} j_1, & \text{if } n = 1 \\ \frac{N_{\mathrm{s}} - 1}{N_{\mathrm{l}} - 1} + j'_{n-1}, & \text{if } n > 1 \end{cases}$, $n = 1, 2, \ldots, N_{\mathrm{l}}$.

$$y_{\mathrm{ls}}(j'_n) = \alpha_n \times y_{j+1} + (1 - \alpha_n) \times y_j, \quad j = 1, 2, \cdots, N_{\mathrm{l}}, \qquad (17)$$

for $j'_n = \begin{cases} \frac{N_{\mathrm{l}} - 1}{N_{\mathrm{s}} + 1} + j_1, & \text{if } n = 1 \\ \frac{N_{\mathrm{l}} - 1}{N_{\mathrm{s}} + 1} + j'_{n-1}, & \text{if } n > 1 \end{cases}$, $n = 1, 2, \ldots, N_{\mathrm{s}}$.

Here N_{s} and N_{l} are the lengths of the shorter and longer sequences, respectively, j'_n is between j and j+1, and $\alpha_n = (j'_n - j)$. In the proposed methods, the MSE based on short sequences, MSE_s, and the MSE based on long sequences, MSE_l, under a certain Markov order q, are determined, respectively, to obtain the average value MSE_{final} shown in Eq. (18).

$$MSE_l = \sqrt{\frac{\sum_{j=1}^{N_l}[y_{sl}(j) - y_l(j)]^2}{N_l}}, \quad MSE_s = \sqrt{\frac{\sum_{j=1}^{N_s}[y_{ls}(j) - y_s(j)]^2}{N_s}},$$

$$MSE_{final} = \frac{MSE_l + MSE_s}{2}. \tag{18}$$

TABLE I. THE STANDARD ONE-LETTER ABBREVIATION OF THE 20 AMINO ACIDS AND THEIR PROPERTIES ON POLARITY, CHARGE, AND HYDROPHILICITY OR HYDROPHOBICITY

Amino acid	**1-letter**	**Side-chain polarity**	**Side-chain**	**Hydropathic index**
Alanine	A	Nonpolar	Neutral	1.8
Arginine	R	Basic polar	Positive	−4.5
Asparagine	N	Polar	Neutral	−3.5
Aspartic acid	D	Acidic polar	Negative	−3.5
Cysteine	C	Nonpolar	Neutral	2.5
Glutamic acid	E	Acidic polar	Negative	−3.5
Glutamine	Q	Polar	Neutral	−3.5
Glycine	G	Nonpolar	Neutral	−0.4
Histidine	H	Basic polar	Positive (10%)	−3.2
Isoleucine	I	Nonpolar	Neutral	4.5
Leucine	L	Nonpolar	Neutral	3.8
Lysine	K	Basic polar	Positive	−3.9
Methionine	M	Nonpolar	Neutral	1.9
Phenylalanine	F	Nonpolar	Neutral	2.8
Proline	P	Nonpolar	Neutral	−1.6
Serine	S	Polar	Neutral	−0.8
Threonine	T	Polar	Neutral	−0.7
Tryptophan	W	Nonpolar	Neutral	−0.9
Tyrosine	Y	Polar	Neutral	−1.3
Valine	V	Nonpolar	Neutral	4.2

TABLE II. THE COMPLEX VALUES OF 20 AMINO ACIDS IN THE PROPOSED FOUR MAPPING METHODS

	Amino acids (property)	Method 1	Method 2	Method 3	Method 4
Hydrophilic amino acid	H (Basic)	0.707+i 0.707	0.309+i 0.951	-0.707+i 0.707	0.707+i 0.707
	R (Basic)	0.588+i 0.809	-0.707+i 0.707	-0.587+i 0.809	-0.588-i 0.809
	K (Basic)	0.454+i 0.891	0.156+i 0.987	-0.454+i 0.891	0.454+i 0.891
	E (Acidic and their Amide)	0.309+i 0.951	-0.309+i 0.951	-0.309+i 0.951	-0.309-i 0.951
	D (Acidic and their Amide)	0.156+i 0.987	0.454+i 0.891	-0.156+i 0.987	0.156+i 0.987
	Q (Acidic and their Amide)	i	-0.454+i 0.891	i	-i
	N (Acidic and their Amide)	-0.156+i 0.987	-0.156+i 0.987	0.156+i 0.987	-0.156+i 0.987
	Y (Aromatic)	-0.309+i 0.951	0.707+i 0.707	0.309+i 0.951	0.309-i 0.951
	C (Hydroxyl or Sulfur-containing)	-0.454+i 0.891	i	0.454+i 0.891	-0.454+i 0.891
	T (Hydroxyl or Sulfur-containing)	-0.587+i 0.809	-0.587+i 0.809	0.588+i 0.809	0.454-i 0.891
	S (Hydroxyl or Sulfur-containing)	-0.707+i 0.707	0.588+i 0.809	0.707+i 0.707	-0.707+i 0.707
Hydrophobic amino acid	G (Aliphatic)	-0.588-i 0.809	-i	0.588-i 0.809	0.588+i 0.809
	A (Aliphatic)	-0.454-i 0.891	-0.454-i 0.891	0.454-i 0.891	-0.454-i 0.891
	V (Aliphatic)	-0.309-i 0.951	0.156-i 0.987	0.309-i 0.951	0.309+i 0.951
	L (Aliphatic)	-0.156-i 0.987	-0.588-i 0.809	0.156-i 0.987	-0.156-i 0.987
	I (Aliphatic)	-i	-0.156-i 0.987	-i	i
	P (Cyclic)	0.156-i 0.987	-0.309-i 0.951	-0.156-i 0.987	0.156-i 0.987
	M (Hydroxyl or Sulfur-containing)	0.309-i 0.951	0.588-i 0.809	-0.309-i 0.951	-0.309+i 0.951
	F (Aromatic)	0.454-i 0.891	0.454-i 0.891	-0.454-i 0.891	-0.587+i 0.809
	W (Aromatic)	0.588-i 0.809	0.309-i 0.951	-0.588-i 0.809	0.588-i 0.809

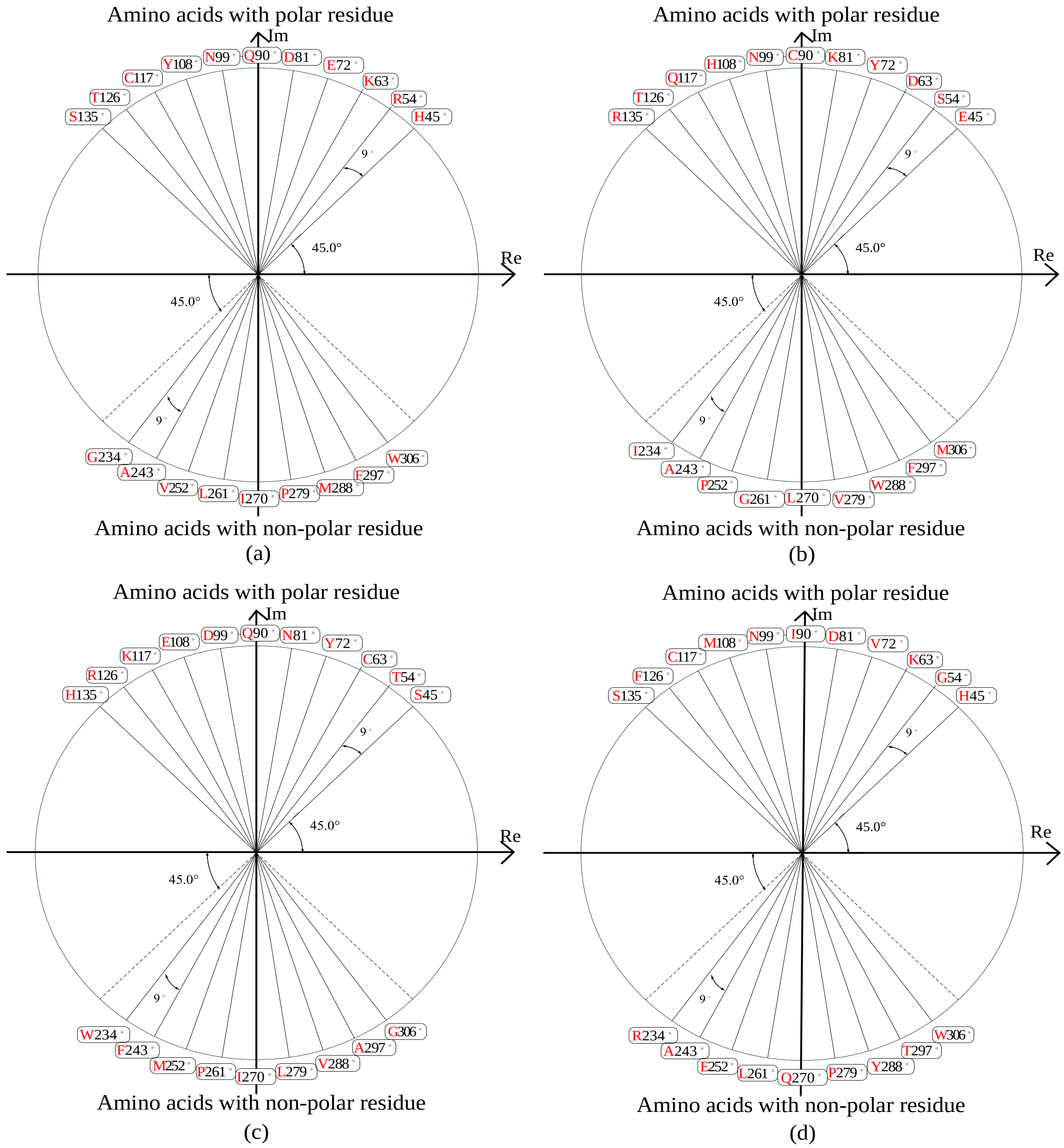


Fig. 2. Graphs of the mapping between the 20 amino acids and their corresponding positions on the circle in the proposed four mapping methods: (a) Method 1, (b) Method 2, (c) Method 3, and (d) Mehtod 4

III. Experimental Results

Eighteen proteins sequences extracted from NCBI databases shown in Table 3 are considered in our experiments. According to their amino acid (aa) numbers, these sequences are categorized into six groups: Group 1 (<200 aa), Group 2 (300-400 aa), Group 3 (401-800 aa), Group 4 (801-1200 aa), Group 5 (1201-2000 aa), and Group 6 (2001-3000 aa) sequences. In each group, two sets of different homologous sequences are tested. The nine MSE values corresponding to the Markov orders from one to nine are determined in each sequence pairs. Then, the average value is regarded the final result in each experiment. Four experiments according to the four mapping schemes (Methods 1-4) shown in Table 2 are designed. Method 1 encounters the properties of hydrophilic and hydrophobic amino acids sorted by the physical-chemical properties. Method 2 only encounters the properties of hydrophilic and hydrophobic. The amino acids are randomly distributed at hydrophilic part with the property of hydrophilic and hydrophobic part with the property of hydrophobic. Method 3 is a similar mapping with Method 1, but the amino acids are distributed in a horizontally reversed way. Finally, the mapping in Method 4 encounters none of the physical-chemical properties of the amino acids. The comparison results with the well-known ClustalW (Pairwise Alignment: SLOW/ACCURATE, Weight Matrix: BLOSUM, Gap Open Penalty: 10, Gap Extension Penalty: 0.1) and BLASTp (Weight Matrix: BLOSUM62) schemes are provided to validate the proposed methods.

Based on Method 1 shown in Table 2, the PSDs of the protein sequences in these six groups are determined. Figure 3 shows the power spectra of the various Markov orders obtained by using the PSE method for the sequences in Group 1. The blue and red lines represent the normalized power spectral values of two sequences. The x-axis denotes the angular frequency ($0 - 2\pi$), while the y-axis denotes the normalized PSD. For each pair of protein sequences, five power spectra corresponding to five Markov orders q=1-5 are determined. If these two lines are close to each other in the same Markov order, we may infer that the two sequences are significantly related. For example, the PSDs in Figs. 3(a), 3(b), 3(f), 3(m), 3(n), and 3(o) are similar in each Markov order because all the sequence pairs are homologous protein sequences, while in Figs. 3(c), 3(d), 3(e), 3(g), 3(h), 3(i), 3(j), and 3(l) are obviously different because the sequence pairs are non-homologous.

Based on Methods 1-4 shown in Table 2, Tables 4 to 9 show the comparison results of the MSE values between various pairs of the sequences in the six groups, respectively. A smaller MSE value represents less difference between two sequences. In these six tables, all the MSE values of the homologous sequence pairs are smaller than 0.2. On the contrary, the MSE values of the non-homologous sequence pairs are greater than 0.2. This can be observed in the other parts in these tables as well. The proposed method contributes the classification of protein sequences and thus can serve as an alternative for the sequence comparison task.

Figures 4(a)-4(f) illustrate the comparison results of the proposed method, ClustalW, and BLASTp in Tables 4 to 9. The horizontal axis denotes the set of experiments (sequence pair), while the vertical axis denotes the normalized difference value. In order to compare the sequence similarly in each method accordingly, the BLASTp scores are replaced by the value 1-BLASTp to correspond to the same numerical characteristics with that in the proposed methods and ClustalW. In Table 3, some results of BLASTp method are shown as NF, which means that the two protein sequences have no similarity found. Here, we set the NF value as 0 and then the value of 1-BLASTp is 1, which denotes the maximal difference. In Fig. 4(a), the depicted lines of MSE1, MSE2, MSE3, MSE4, and ClustalW have the similar rising and descending trends for the short sequences. The scores in all the six methods basically can be used to distinguish the homologous and non-homologous sequences. However, in Figs. 4(b) and 4(c), MSE2 and MSE4 are not consistent to MSE1, ClustalW, and 1-BLASTp for the medium and long sequence pairs. In Figs. 4(b) and 4(f), the 1-BLASTp scores are quite different from other scores and the variations among these methods are larger than that in the other figures. Note that the MSE4 values are higher than other values for Sequence pairs 14, and 15 in Figs. 4(d) to 4(f). The experimental results show that the mapping methods while encountering characteristics of hydrophilic amino acids and the general physical-chemical properties of amino acids can affect the comparison results. In Figs. 4(a) to 4(f), the MSE1 and MSE3 are nearly the same because the separation between each two amino acids is the same even if the mapping positions are horizontally reversed.

According to the results of the proposed methods, ClustalW, and BLASTp, the following observations are obtained: (1) The numerical mapping according to the physical-chemical characteristics of amino acids affect the results of comparison. In the experimental results, the more characteristics of protein are considered and arranged, the more correct results are obtained. (2) The MSE1 results are similar to the ClustalW scores, which means that the mapping scheme is consistent with the ClustalW method. (3) BLASTp and ClustalW are different methods, especially designed for the global and local sequence comparisons, respectively. The differences shown in the experimental results are especially obvious for the sequences in Groups 2 and 6.

IV. Conclusion

We proposed a new comparative tool for protein sequence comparison utilizing the parametric spectral estimate in stochastic processes to analyze protein sequences. The concepts of hydrophobicity in the amino acid physical-chemical properties are used to transform amino acids to numerical values. The experimental results show that the proposed methods effectively achieved the consistent comparison results with the well-known ClustalW and BLASTp. This research provides a new insight for the biologists as to how protein sequences can be analyzed. In our future work, more protein sequences will be tested by the proposed method. The problems encountered by two protein sequences with large difference in length will be tackled as well.

Acknowledgment

This research is partially supported by Ministry of Science and Technology (MOST), Taiwan under the contract number MOST 103-2221-E-224-035-MY2.

TABLE III. The Three Groups of the Test Protein Sequences in Our Experiments

	No.	Index	Class	Length (aa)
Group 1 (<200)	1	XP_933607.1	PREDICTED: hypothetical protein	90
	2	XP_001129788.1	PREDICTED: hypothetical protein	90
	3	XP_001129824.1	PREDICTED: hypothetical protein	90
	4	NP_001463.1	G antigen 2 [Homo sapiens]	116
	5	NP_001468.1	G antigen 7B [Homo sapiens]	117
	6	NP_036328.1	G antigen 8 [Homo sapiens]	116
Group 2 (300~400)	7	CAI46074.1	hypothetical protein [Homo sapiens]	329
	8	EFB16212.1	hypothetical protein PANDA_018121 [Ailuropoda melanoleuca]	329
	9	CAH89386.1	hypothetical protein [Pongo abelii]	323
	10	AAB51177.1	human RAD23A homolog [Homo sapiens]	363
	11	AAI33283.1	RAD23A protein [Bos taurus]	362
	12	AAH84695.1	RAD23 homolog A [Rattus norvegicus]	351
	13	AAH33781.1	PAXIP1 protein [Homo sapiens]	757

Group 3 (401~800)	14	CAD98066.1	hypothetical protein, partial [Homo sapiens]	675
	15	AAB91434.1	CAGF28, partial [Homo sapiens]	744
	16	AAP04006.1	NIMA-family kinase NEK8 [Homo sapiens]	692
	17	NP_001179652.1	serine/threonine-protein kinase Nek8 [Bos taurus]	698
	18	DAA19021.1	TPA: NIMA-related kinase 8-like [Bos taurus]	703
Group 4 (801~1200)	19	NP_031375.3	PAX-interacting protein 1 [Homo sapiens]	1069
	20	EHH17890.1	hypothetical protein EGK_14374, partial [Macaca mulatta]	1049
	21	AAH77588.1	K14 protein [Xenopus laevis]	1320
	22	CAC35387.1	suppression of tumorigenicity 5 [Homo sapiens]	1137
	23	AIC55177.1	ST5, partial [synthetic construct]	1137
	24	JAA49668.1	Putative ras signaling inhibitor st5 [Desmodus rotundus]	1137
Group 5 (1201~2000)	25	BAA13389.1	KIAA0259, partial [Homo sapiens]	1550
	26	AFJ70237.1	DNA topoisomerase 2-binding protein 1 [Macaca mulatta]	1527
	27	AAI51238.1	Topoisomerase (DNA) II binding protein 1 [Homo sapiens]	1435
	28	AAI12162.1	Tumor protein p53 binding protein 1 [Homo sapiens]	1972
	29	BAE06107.1	TP53BP1 variant protein, partial [Homo sapiens]	1984
	30	JAB19186.1	tumor suppressor p53-binding protein 1 isoform 1 [Callithrix jacchus]	1970
Group 6 (2001~3000)	31	BAA83718.1	RNA binding protein [Homo sapiens]	2752
	32	NP_057417.3	serine/arginine repetitive matrix protein 2 [Homo sapiens]	2752
	33	BAA20782.3	KIAA0324 protein [Homo sapiens]	2800
	34	AAK39635.1	DNA polymerase theta [Homo sapiens]	2724
	35	CAI56770.1	hypothetical protein [Homo sapiens]	2149
	36	ELK18077.1	DNA polymerase theta [Pteropus alecto]	2597

TABLE IV. THE COMPARISON RESULTS OF THE PROPOSED, CLUSTALW, AND BLASTP METHODS FOR GROUP 1 (<200 AA) SEQUENCES

Set of experiments	**Sequence pairs**	**MSE_1**	**MSE_2**	**MSE_3**	**MSE_4**	**ClustalW**	**Blastp score**	**1-Blastp score**
1	(1,2)	0.0919	0.0623	0.0919	0.0946	0.0111	0.98	0.02
2	(1,3)	0.0166	0.0348	0.0166	0.0182	0.0056	0.99	0.01
3	(1,4)	0.3502	0.2343	0.3504	0.2682	0.4667	NF	1
4	(1,5)	0.3472	0.2273	0.3475	0.2545	0.4667	NF	1
5	(1,6)	0.3525	0.2346	0.3527	0.2605	0.4667	NF	1
6	(2,3)	0.0864	0.0572	0.0864	0.1043	0.0056	0.99	0.01
7	(2,4)	0.3755	0.2532	0.3751	0.2887	0.4667	NF	1
8	(2,5)	0.3718	0.2448	0.3715	0.2714	0.4667	NF	1
9	(2,6)	0.3770	0.2537	0.3767	0.2815	0.4611	NF	1
10	(3,4)	0.3540	0.2414	0.3540	0.2697	0.4667	NF	1
11	(3,5)	0.3505	0.2332	0.3507	0.2580	0.4667	NF	1
12	(3,6)	0.3564	0.2417	0.3564	0.2613	0.4667	NF	1
13	(4,5)	0.0567	0.0684	0.0567	0.0504	0.0086	0.97	0.03
14	(1,2)	0.0171	0.0055	0.0171	0.0247	0.0043	0.99	0.01
15	(1,3)	0.0585	0.0653	0.0586	0.0441	0.0129	0.97	0.03

TABLE V. THE COMPARISON RESULTS OF THE PROPOSED, CLUSTALW, AND BLASTP METHODS FOR GROUP 2 (300 – 400 AA) SEQUENCES

Set of experiments	**Sequence pairs**	**MSE_1**	**MSE_2**	**MSE_3**	**MSE_4**	**ClustalW score**	**BLASTp score**	**1- BLASTp score**
1	(7,8)	0.0409	0.1644	0.0409	0.0719	0.0304	0.94	0.06
2	(7,9)	0.1331	0.1834	0.1331	0.2182	0.2307	0.54	0.46
3	(7,10)	0.3324	0.3035	0.3325	0.1600	0.4863	0.71	0.29
4	(7,11)	0.3119	0.3181	0.3120	0.1860	0.4863	0.71	0.29
5	(7,12)	0.2822	0.2702	0.2823	0.1572	0.4742	0.83	0.17
6	(8,9)	0.1383	0.2212	0.1383	0.2146	0.2260	0.54	0.46
7	(8,10)	0.3309	0.3022	0.3312	0.1467	0.4620	0.71	0.29
8	(8,11)	0.3129	0.3130	0.3132	0.1736	0.4666	0.71	0.29
9	(8,12)	0.2862	0.2926	0.2864	0.1375	0.4681	0.71	0.29
10	(9,10)	0.3268	0.2768	0.3270	0.2615	0.4598	0.41	0.59
11	(9,11)	0.3183	0.3027	0.3186	0.2810	0.4721	0.41	0.59
12	(9,12)	0.2835	0.2320	0.2838	0.2584	0.4861	0.30	0.70
13	(10,11)	0.1175	0.2194	0.1175	0.0466	0.0221	0.93	0.07
14	(10,12)	0.1957	0.2301	0.1958	0.0421	0.0627	0.94	0.06
15	(11,12)	0.1653	0.2967	0.1654	0.0598	0.0698	0.91	0.09

TABLE VI. THE COMPARISON RESULTS OF THE PROPOSED, CLUSTALW, AND BLASTP METHODS FOR GROUP 3 (401 – 800 AA) SEQUENCES

Set of experiments	Sequence pairs	MSE1	MSE2	MSE3	MSE4	ClustalW score	BLASTp score	1-BLASTp score
1	(13,14)	0.0823	0.0272	0.0823	0.0330	0.0007	0.99	0.01
2	(13,15)	0.1413	0.2633	0.1413	0.3128	0.0894	0.96	0.04
3	(13,16)	0.3831	0.4798	0.3832	0.2793	0.4523	0.28	0.72
4	(13,17)	0.3687	0.4880	0.3688	0.2105	0.4592	0.27	0.73
5	(13,18)	0.3770	0.4945	0.3771	0.2209	0.4595	0.28	0.72
6	(14,15)	0.1570	0.2747	0.1569	0.3338	0.1089	0.96	0.04
7	(14,16)	0.3769	0.4823	0.3769	0.2879	0.4511	0.28	0.72
8	(14,17)	0.3666	0.4910	0.3666	0.2097	0.4578	0.28	0.72
9	(14,18)	0.3702	0.4976	0.3720	0.2203	0.4578	0.28	0.72
10	(15,16)	0.4245	0.4603	0.4246	0.3161	0.4798	0.44	0.56
11	(15,17)	0.4021	0.4617	0.4022	0.3442	0.4807	0.33	0.67
12	(15,18)	0.4044	0.4676	0.4044	0.3478	0.4808	0.33	0.67
13	(16,17)	0.0702	0.0816	0.0703	0.1614	0.0260	0.94	0.06
14	(16,18)	0.0750	0.0798	0.0751	0.1547	0.0318	0.94	0.06
15	(17,18)	0.0264	0.0214	0.0264	0.0279	0.0064	1.00	0

TABLE VII. THE COMPARISON RESULTS OF THE PROPOSED, CLUSTALW, AND BLASTP METHODS FOR GROUP 4 (801 – 1200 AA) SEQUENCES

Set of experiments	Sequence pairs	MSE_1	MSE_2	MSE_3	MSE_4	ClustalW score	BLASTp score	1- BLASTp score
1	(19,20)	0.1147	0.1193	0.1148	0.0734	0.0186	0.96	0.04
2	(19,21)	0.1811	0.1667	0.1814	0.1116	0.1240	0.84	0.16
3	(19,22)	0.3297	0.3615	0.3298	0.3311	0.4528	0.16	0.84
4	(19,23)	0.3326	0.3628	0.3327	0.3313	0.4511	0.16	0.84
5	(19,24)	0.3586	0.3763	0.3587	0.3442	0.4602	0.29	0.71
6	(20,21)	0.1925	0.1113	0.1928	0.0605	0.1239	0.82	0.18
7	(20,22)	0.3396	0.3470	0.3397	0.3527	0.4561	0.40	0.60
8	(20,23)	0.3452	0.3498	0.3452	0.3518	0.4557	0.40	0.60
9	(20,24)	0.3677	0.3710	0.3677	0.3787	0.4576	0.29	0.71
10	(21,22)	0.3011	0.3714	0.3011	0.3738	0.4705	0.37	0.63
11	(21,23)	0.3154	0.3738	0.3154	0.3721	0.4705	0.30	0.70
12	(21,24)	0.3266	0.3821	0.3265	0.4029	0.4776	0.26	0.74
13	(22,23)	0.0319	0.0235	0.0319	0.0319	0.0022	0.99	0.01
14	(22,24)	0.0813	0.1339	0.0813	0.1763	0.0299	0.94	0.06
15	(23,24)	0.0881	0.1329	0.0881	0.1842	0.0312	0.94	0.06

TABLE VIII. THE COMPARISON RESULTS OF THE PROPOSED, CLUSTALW, AND BLASTP METHODS FOR GROUP 5 (1201 – 2000 AA) SEQUENCES

Set of experiments	Sequence pairs	MSE_1	MSE_2	MSE_3	MSE_4	ClustalW score	BLASTp score	1- BLASTp score
1	(25,26)	0.0298	0.0390	0.0298	0.0811	0.0111	0.98	0.02
2	(25,27)	0.0377	0.0456	0.0376	0.0716	0.0000	1.00	0.00
3	(25,28)	0.3208	0.2858	0.3208	0.4020	0.4452	0.30	0.70
4	(25,29)	0.3238	0.2841	0.3238	0.4076	0.4445	0.30	0.70
5	(25,30)	0.2909	0.2970	0.2910	0.4384	0.4287	0.28	0.72
6	(26,27)	0.0594	0.0562	0.0594	0.1174	0.0098	0.98	0.02
7	(26,28)	0.3218	0.2835	0.3218	0.3992	0.4601	0.23	0.77
8	(26,29)	0.3245	0.2829	0.3245	0.4070	0.4430	0.23	0.77
9	(26,30)	0.2892	0.2974	0.2892	0.4289	0.4280	0.26	0.74
10	(27,28)	0.3166	0.2866	0.3166	0.3912	0.4408	0.33	0.67
11	(27,29)	0.3198	0.2868	0.3198	0.3970	0.4401	0.33	0.67
12	(27,30)	0.2905	0.3018	0.2906	0.4190	0.4300	0.26	0.74
13	(28,29)	0.0138	0.0120	0.0138	0.0325	0.0005	0.99	0.01
14	(28,30)	0.0922	0.0694	0.0922	0.1595	0.0251	0.95	0.05
15	(29,30)	0.0877	0.0698	0.0877	0.1715	0.0244	0.95	0.05

TABLE IX. THE COMPARISON RESULTS OF THE PROPOSED, CLUSTALW, AND BLASTP METHODS FOR GROUP 6 (2001 – 3000 AA) SEQUENCES

Set of experiments	Sequence pairs	MSE$_1$	MSE$_2$	MSE$_3$	MSE$_4$	ClustalW score	BLASTp score	1- BLASTp score
1	(31,32)	0.0026	0.0026	0.0026	0.0012	0.0002	0.99	0.01
2	(31,33)	0.0059	0.0094	0.0059	0.0023	0.0004	0.99	0.01
3	(31,34)	0.3391	0.4435	0.3391	0.5081	0.4670	0.88	0.12
4	(31,35)	0.3242	0.4118	0.3241	0.5117	0.4591	0.88	0.12
5	(31,36)	0.3251	0.3371	0.3251	0.4379	0.4704	0.44	0.56
6	(32,33)	0.0061	0.0077	0.0061	0.0023	0.0002	0.99	0.01
7	(32,34)	0.3384	0.4443	0.3384	0.5079	0.4670	0.88	0.12
8	(32,35)	0.3233	0.4127	0.3233	0.5115	0.4591	0.88	0.12
9	(32,36)	0.3245	0.3375	0.3245	0.4378	0.4704	0.44	0.56
10	(33,34)	0.3391	0.4485	0.3391	0.5078	0.4662	0.88	0.12
11	(33,35)	0.3241	0.4172	0.3240	0.5114	0.4581	0.88	0.12
12	(33,36)	0.3254	0.3385	0.3254	0.4373	0.4704	0.44	0.56
13	(34,35)	0.0998	0.1003	0.0999	0.1355	0.0016	0.99	0.01
14	(34,36)	0.1142	0.254	0.1142	0.3358	0.09261	0.81	0.19
15	(35,36)	0.1394	0.2791	0.1394	0.3236	0.10051	0.80	0.20

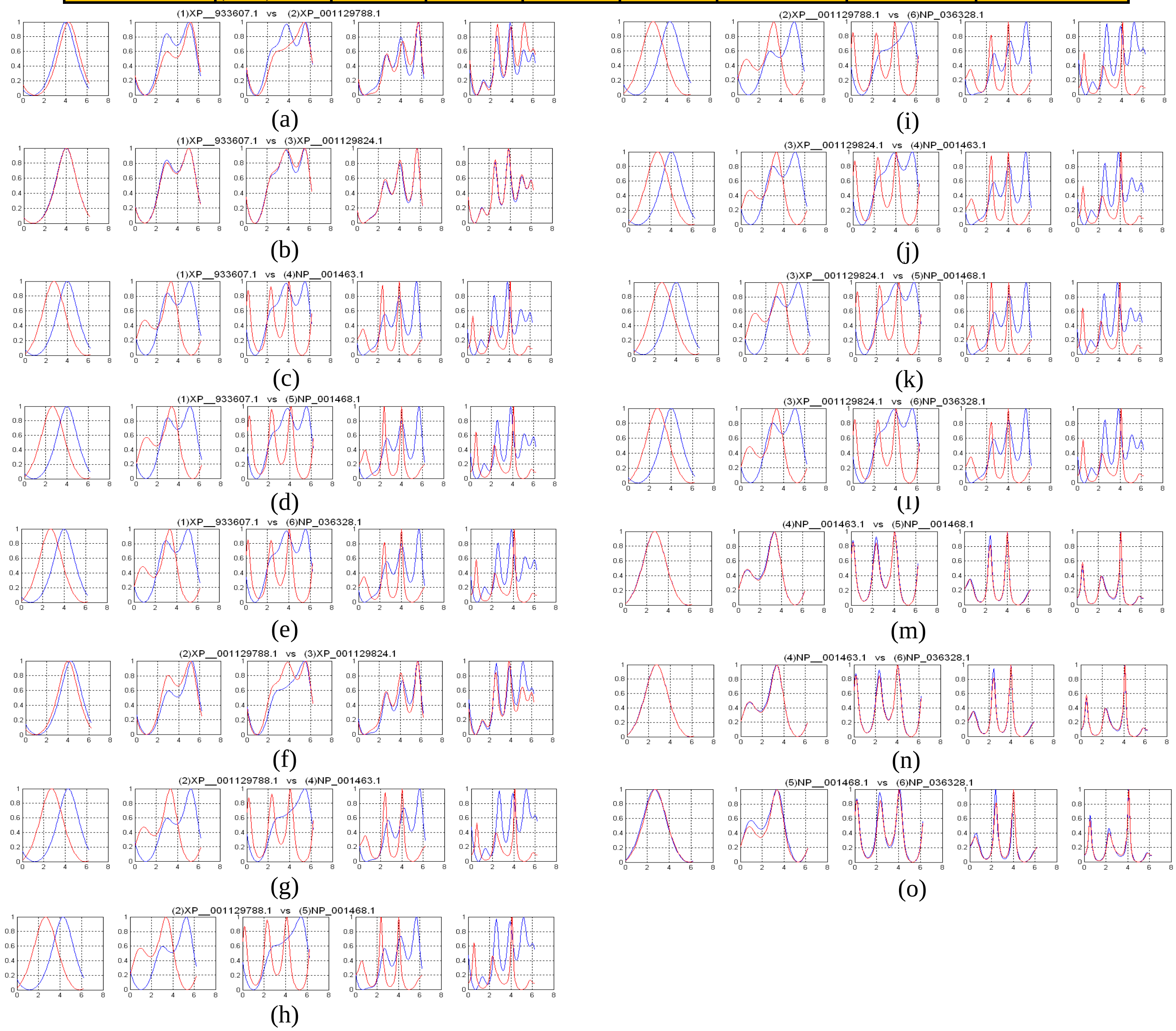


Fig. 3. The PSE comparison results between each sequence pair in Group 1 under various Markov orders q=1-5

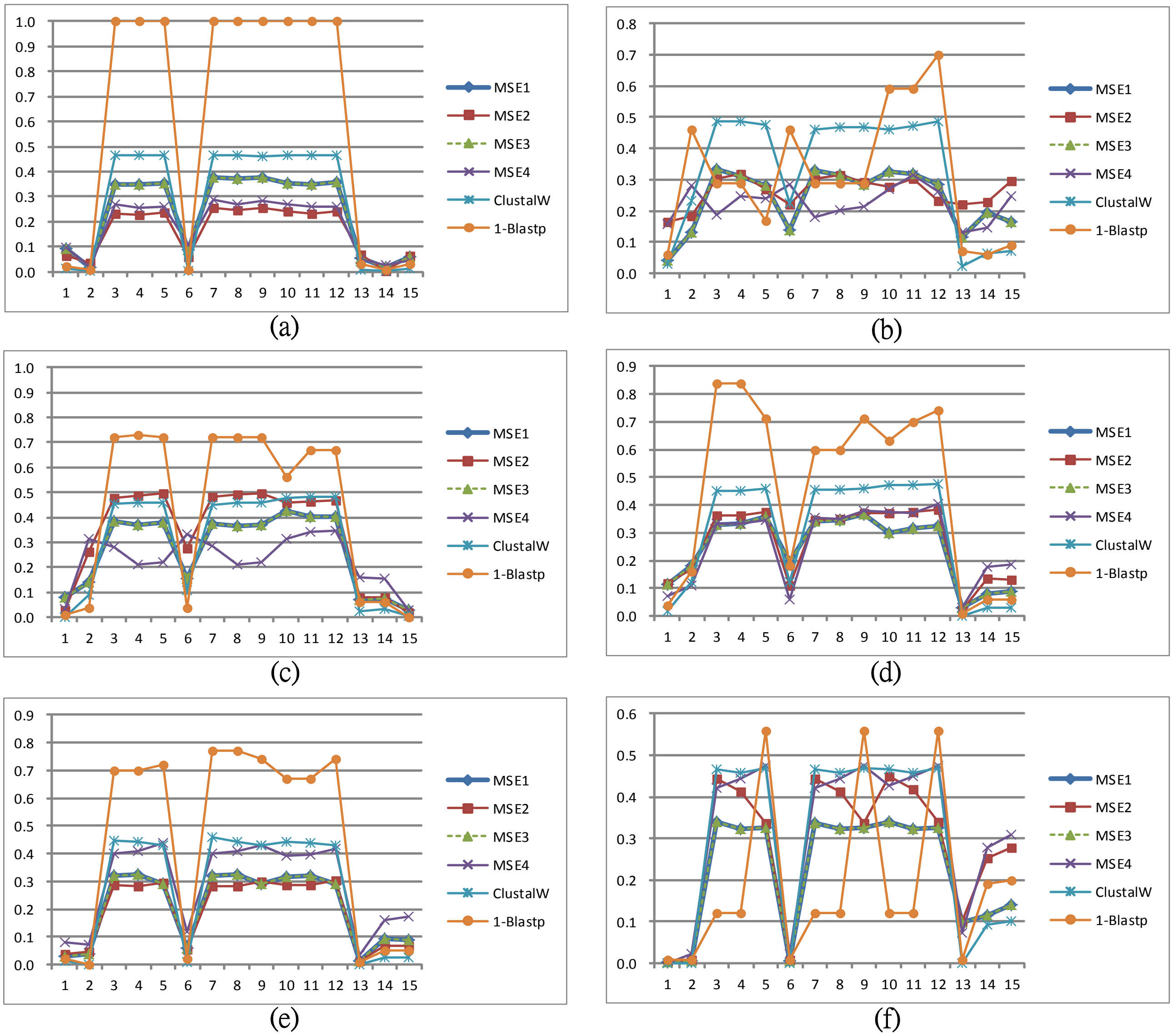


Fig. 4. The graphical reprsentations of the comparison results on MSE1, MSE2, MSE3, MSE4, ClustalW, and 1-BLASTp in (a) Table 3; (b) Table 4; (c) Table 5; (d) Table 6; (e) Table 7; and (f) Table 8

REFERENCES

[1] M. Guarnaccia, G. Gentile, E. Alessi, C. Schneider, S. Petralia, and S. Cavallaro. “Is this the real time for genomics?” Genomics, vol. 103, no. 2-3, pp. 177-182, 2014.

[2] G. A. Price, G. E. Crooks, R. E. Green, and S. E. Brenner, “Statistical evaluation of pairwise protein sequence comparison with the Bayesian bootstrap,” Bioinformatics, vol. 21, pp. 3824-3831, 2005.

[3] Z. G. Yu, V. Anh, and K. S. Lau, “Chaos game representation of protein sequence based on the detailed HP model and their multifractal and correlation analyses,” Journal of Theoretical Biology, pp. 341-348, 2004.

[4] E. A. Cheever, D. B. Searls, W. Karunaratne, and G. C. Overton, “Using signal processing techniques for DNA sequence comparison,” Proceedings of the 1989 Fifteenth Annual Northeast Bioengineering Conference, pp. 173-174, 1989.

[5] D. Anastassiou, “DSP in genomics: processing and frequency-domain analysis of character strings,” IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 2, pp. 1053-1056, 2001.

[6] P. Cristea, “Genetic signal analysis,” Sixth International Symposium on Signal Processing and Its Applications (ISSA 2001), vol. 2, pp. 703-706, 2001.

[7] H. T. Chang, C.J. Kuo, N.-W. Lo, and W.-Z. Lv, “DNA sequence visualization and comparison based on quaternion number system,” International Journal of Advanced Computer Science and Applications, vol. 3, no. 11, pp. 39-46, 2012.

[8] D. Wu, J. Roberge, D. J. Cork, B. G. Nguyen, and T. Grace, “Computer visualization of long genomic sequence,” Proceedings of IEEE Conferences on Visualization, pp. 308-315, 1993.

[9] E. H.-H Chi, P. Barry, E. Shoop, J. V. Carlis, E. Retzel, and J. Ried, “Visualization of biological sequence similarity search results,” Proceedings of IEEE Conferences on Visualization, pp. 44-51, 1995.

[10] H. H. Chi, J. Riedl, E. Shoop, J. V. Carlis, E. Retzel, and P. Barry, “Flexible information visualization of multivariate data from biological sequence similarity searches,” Proceedings of IEEE Conferences on Visualization, pp. 133-140, 1996.

[11] H.T. Chang, N.W. Lo, W.C. Lu, and C.J. Kuo, “Visualization and comparison of DNA sequences by use of three-dimensional trajectory,” The First Asia-Pacific Bioinformatics Conference (APBC 2003), vol. 19, pp. 81-85, Adelaide, Australia 2003.

[12] H.T. Chang, S. W. Xiao, and N.-W. Lo, “Feature extraction and comparison of TDTs: an efficient sequence retrieval algorithm for

genomic databases," The Third Asia-Pacific Bioinformatics Conference (APBC 2005), pp. 86, 2005.

[13] N.-W. Lo, H.T. Chang, S.W. Xiao, and C.J. Kuo, "Global visualization of DNA sequences by use of three-dimensional trajectories," Journal of Information Science and Engineering, vol. 23, no. 6, pp. 1723-1736, Nov. 2007.

[14] H.T. Chang, "DNA sequence visualization," in Advanced Data Mining Technologies in Bioinformatics , Chapter 4, pp. 63-84, Edited by Dr. Hui-Huang Hsu, Idea Group Publishing, ISBN 195140864-4, 2006.

[15] H.T. Chang, S.W. Xiao, and C.H. Lee, "Feature extraction for fast data retrieval for visualized DNA sequence," National Symposium on Telecommunications, Kaoshiung Taiwan, 2006.

[16] W. Wong and D. H. Johnson, "Computing linear transforms of amino acidic signals," IEEE Transactions on Signal processing, vol. 50, pp. 628-634, 2002.

[17] P. D. Cristea, "Conversion of nucleotides sequences into genomic signals," Journal of Cellular and Molecular Medicine, vol. 6, no 2, pp. 279-303, 2002.

[18] P.J.S.G. Ferreira, V. Afreixo, and D. Santos, "Spectrum and amino acid distribution of nucleotide sequences," Digital Signal Processing, vol. 14, pp. 523-530, 2004.

[19] R. Gupta, A. Mittal, and S. Gupta, "An efficient algorithm to detect palindromes in DNA sequences using periodicity transform," Signal Processing, vol. 86, pp. 2067-2073, 2006.

[20] X. Meng and V. Chaudhary, "A high-performance heterogeneous computing platform for biological sequence analysis," IEEE Transactions on Parallel and Distributed Systems, vol. 21, pp. 1267-1280, 2010.

[21] N. Maillet, C. Lemaitre, R. Chikhi, D. Lavenier, and P. Peterlongo, "Compareads: comparing huge metagenomic experiments," BMC Bioinformatics, vol. 13, (Suppl 19): S10, 2012.

[22] http://www.clustal.org/clustal2/

[23] http://blast.ncbi.nlm.nih.gov/

[24] R. E. Hausman and B. M. Cooper, The cell: a molecular approach. Washington, D.C: ASM Press. pp. 51. ISBN 0-87893-214-3, 2004.

[25] J. Kyte and R. Doolittle, "A simple method for displaying the hydropathic character of a protein," Journal of Molecular Biology, vol. 157, no. 1, pp. 105–32, May 1982

[26] https://www.boundless.com/biology/definition/r-group/ (R group)

[27] H.T. Chang, Y.-L. Liu, C.J. Kuo, C.H. Lee, and N.-W. Lo, "Analysis of genomic DNA sequence by use of parametric spectral estimate," Bioinformatics in Taiwan (BIT2005), pp. 68, Tainan Taiwan, Sep. 2005

[28] H. Stark and J. W. Woods, Probability and Random Process with Applications to Signal Processing, Chapter 9. Applications to Statistical Signal Processing, pp. 553-619, Third Edition, Prentice Hall, ISBN-10: 0130200719, 2001.

Testing the Use of the Integrated Model in Designing the Management Information Systems by Using the Mathematical Probability Theories

Mohammad M M Abu Omar & Khairul Anuar Abdullah
Limkokwing University of Creative Technology, Malaysia

***Abstract*—The integrated model is a new model that is recently developed to decrease from the classical approach weaknesses and problems in building the management information systems (MIS's) that are used to solve the management problems in the practical life. The use of this integrated model needs to be tested, to prove how efficiently and successfully the model works. To achieving this objective, this paper uses the mathematical probability theories to implement an internal test of the integrated model work before using it in the practical life. The paper uses the qualitative research method in its methodologies.**

Keywords*—*Integrated Model; Management Information System; MIS; Classical Approach; Information System Life Cycle; ISLC; Simple Random Sampling; Probability Theory

I. Introduction

A. What is the classical approach

Classical or (traditional) approach is one of the information systems methodologies that are used to build the information systems. One of these information systems is a management information system (MIS), which is a computer-based system that is used in solving management problems [12],[13].

Classical approach has the variety of approaches according to its use; since it adopts the information system life cycle (ISLC) on its work. There are many different ISLC models, but each one of them consists of a series of defined steps or phases, which should occur when building the MIS [10].

These five main steps or stages are:

1) Planning Stage: it is the first stage in the information system life cycle. The responsibilities of this stage are:

- Defining the problem and collecting the required information about the problem that the system will solve it [2], [3].
- Determining the user's requirements, which the developed system will solve.. [2],[3].
- Determining the estimated budget and time to accomplish the system [2], [3].
- Suggesting a solution or list of solutions to the problem [2], [3].

2) Analysis Stage: it is the second stage in the information system life cycle. In this stage, the system analyzer will study each solution in the list of the suggested solutions that is obtained from the previous stage (planning stage) and then choose the best solution [2],[3].

3) Design Stage: it is the third stage in the information system life cycle. Here, the designer's team will provide all the necessary design requirements such as input screens, output screens, reports, database and system algorithms [2], [3].

4) Development Stage: it is the fourth stage in the information system life cycle. Here the system will be programmed and operated [2],[3].

5) Test and Maintenance Stage: it is the fifth (final) stage in the information system life cycle. Here, the system will be tested to verify whether it includes errors or if it needs improvements to be better and effective in achieving the users' requirements [2],[3].

B. What is the drawback of the classical approach

The information system life cycle of the classical approach (ISLC) consists of five main stages that are: planning, analysis, design, implementation, and maintenance or test stages. The classical approach uses these stages in its work in building the management information systems that will be used to solve the management problems. The weakness and problem of the classical approach arises from the permanent use of all of these five stages in all cases; because there are many cases that actually need to use some of these stages but not to all of them, and as a result the classical approach suffers from the weakness and problem in consuming more time and additional cost when it is used to build the management information systems (MIS's) in order to solve the management problems [1],[2],[3].

C. What is the integrated model, why there is need for this model

The integrated model is a recently developed model by Mohammad Abu Omar and Khairul Anuar Abdullah [1], which uses the classical approach as one of the systems development methodologies to build the management information systems (MIS's) which are used to solve the management problems.

The integrated model aims to improve the weakness of the classical approach that consumes additional time and cost. The integrated model has accomplished its mission by through the following actions:

a) Finding a new classification of the management problems, that may be solved by building the MIS's.

b) Developing new sub- approaches from the information system life cycle (ISLC) that is adopted by classical approach. These sub-approaches don't use the all stages of the (ISLC) of the classical approach, but only the necessary ones.

c) Implementing a compatibility between the newly developed sub-approaches and the different types of the management problems, to achieve suitable and effective use of the sub-approaches in solving the management problems by building the MIS's.

The integrated model will avoid the classical approach in the permanent use of the all stages in its life cycle, by choosing a suitable sub-approach to a suitable management problem.

II. THE BACKGROUND OF THE INTEGRATED MODEL STRUCTURE AND WORK

The integrated model develops a new classification of the management problems, as follows [1]:

1) The first category management problems, which include the following problems:

a) First order management problem

b) Second order management problem.

2) The second category management problems, which includes the following problems:

a) First level management problem

b) Second level management problem

3) The combined management problems, which include the following problems:

a) First combined management problem

b) Second combined management problem

c) Third combined management problem

d) Fourth combined management problem

Meanwhile, the integrated model develops the sub-approaches from the information system life cycle (ISLC) that is adopted by classical approach. These sub- approaches are:

1) *MIS new approach (1)*
2) *MIS new approach (2)*
3) *MIS new approach (3)*

And finally, the integrated model appoints a suitable sub-approach for solving the related management problem, as in the following procedures [1]:

- **The first procedure**: (MIS new approach (1) for solving the first order management problem):

The first order management problem has a direct solution; which is clear, and it is usually one. One need only to collect the required information about this problem in order to implement it, and therefore there is no need to find multiple solutions to the problem and having to choose the best solution. This means that there is no need to consume more time, effort and cost in using the analysis stage, which is the second stage in the information system life cycle that is adopted by classical approach [1],[3].

In this case, the integrated model develops a new sub-approach which is defined as: MIS new approach (1), which skips using the analyses stage in building the management information system by using the classical approach [1],[3].

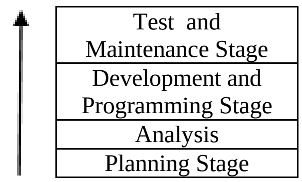


Fig. 1. The MIS New Approach (1)

Thus, the skipping of the analysis stage by the MIS approach (1) will decrease from the consumption of additional time and cost through using the classical approach in designing the MIS's to solve the management problems.

- **The second procedure**: (MIS new approach (2) for solving the first level management problem):

The first level management problem needs software, which can be found sin off-the shelf software packages. This means that management problem doesn't need to build and develop new software, but it can use off-the shelf software packages which can be purchased directly in the markets [1], [2].

According to the information system life cycle (ISLC) adopted by the classical approach, and explained in the introduction of this paper, we will find that the building of the computer program (software), which the management information system (MIS) will use to solve the management problems, is related directly with the third stage (Design Stage) in the information system life cycle. This is because the building of the computer program (software) includes the building of: algorithms, database, input and output screens, reports,….etc, and these are the responsibilities of the (Design Stage) in the information system life cycle [1],[2].

In this case, the integrated model develops a new sub-approach which is defined as: MIS new approach (2), which skips the use of the design stage in building the management information system by using the classical approach [1],[3].

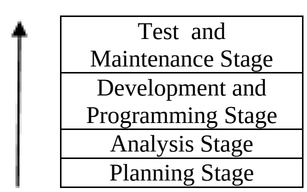


Fig. 2. The MIS New Approach (2)

Thus, the skipping of the design stage by the MIS approach (2) will decrease from the consumption of additional time and cost through using the classical approach in designing the MIS's to solve the management problems.

- **The third procedure**: (MIS new approach (3) for solving the first combined management problem):

The first combined management problem is formed by the combination of the first order management problem and the first level management problem [1], [2].

The first combined management problem is a first order management problem and a first level management problem at same time. Consequently, this problem inherits the skipping of the analysis stage from the first order management problem, and also inherits the skipping of the design stage from the first level management problem. Meanwhile, the integrated model develops a new sub-approach which is defined as: MIS new approach (3), which skips the using of the both design and analysis stages in building the management information system by using the classical approach [1],[3].

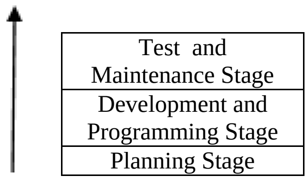


Fig. 3. The MIS New Approach (3)

Thus, the skipping of the analysis and design stages by the MIS approach (3) will decrease more from the consumption of additional time and cost through using the classical approach in designing the MIS's to solve the management problems.

- The fourth procedure:

The integrated model uses the whole information system life cycle (ISLC) that is adopted by classical approach without skipping any stage in the other management problems, which are defined in the management problems classification and which aren't used in the previous three procedures. These problems are:

- Second order management problem
- Second level management problem
- Fourth combined management problem

The nature and properties of these management problems make the integrated model use all the stages of the (ISLC) of the classical approach under the name of: the MIS classical approach [1], [2], [3].

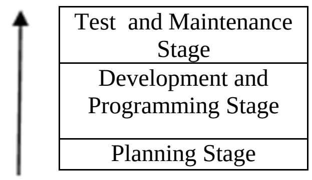


Fig. 4. The MIS Classical Approach

The following figure shows the integrated model structure and work [1]:

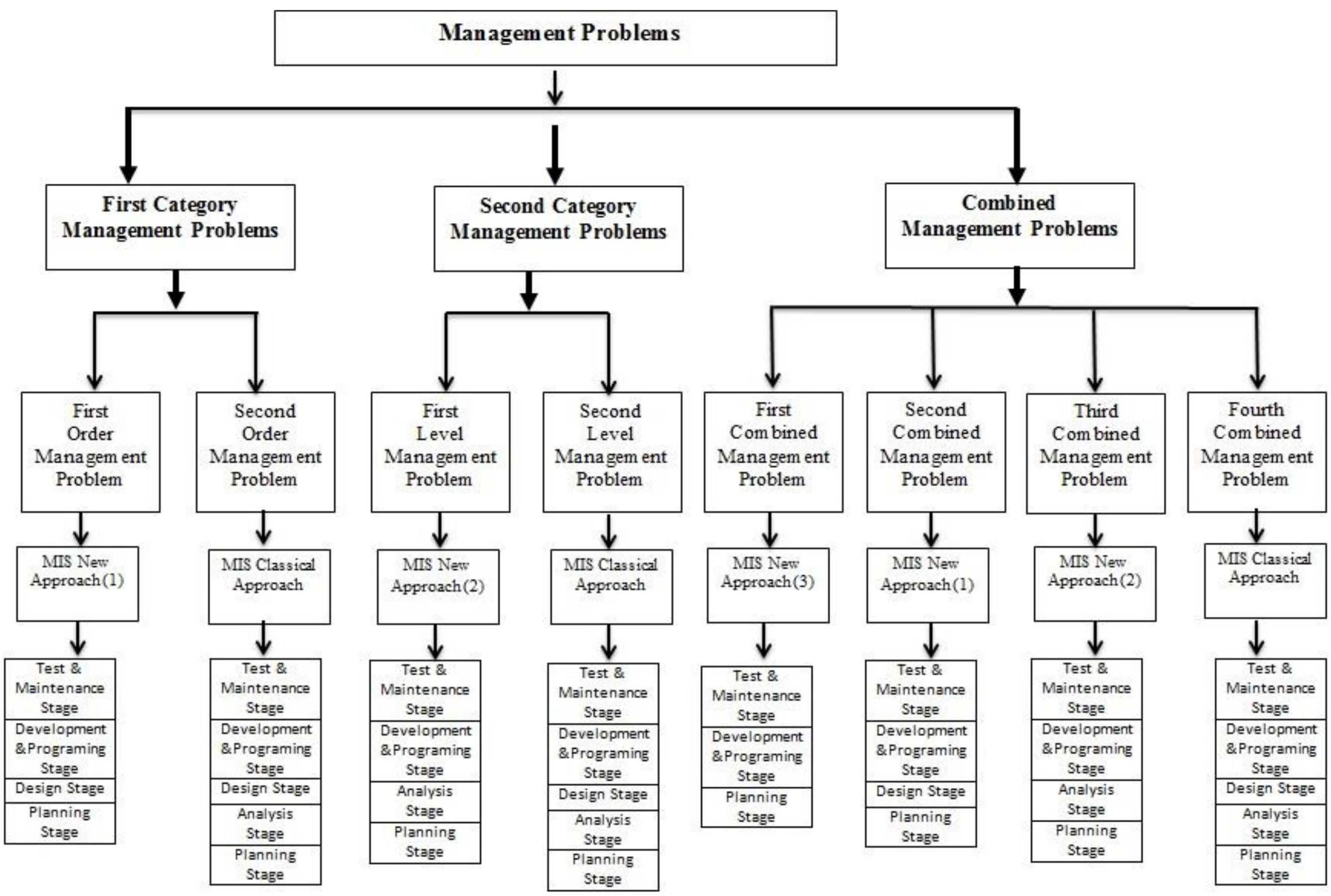


Fig. 5. The Integrated Model: Structure and Work

III. Why Testing of the Integrated Model

According to the previous procedures of the integrated model, some of the integrated model's sub- approaches achieve the integrated model goal in preventing the classical approach from the consumption of additional time and cost while building the MIS's, because these sub- approaches skip some stages in the ISLC of the classical approach which are not necessary in building the MIS's, at the same time, there are some cases that the integrated model can't skip any stage in the ISLC of the classical approach while building the MIS's. From here, the integrated model needs to the test, in order to see the overall evaluation of its work, to determine if it is valid or not in achieving its goal in solving the classical approach problem which is the additional time and cost consumed while building the MIS.

A. Testing of the Integrated Model

As it is mentioned in this paper, the integrated model classifies the management problems as in the following:

First Category Management Problems
➢ First Order Management Problems
➢ Second Order Management Problems
Second Category Management Problems
➢ First Level Management Problems
➢ Second Level Management Problems
Combined Management Problems
➢ First Combined Management Problems
➢ Second Combined Management Problems
➢ Third Combined Management Problems
➢ Fourth Combined Management Problems

Fig. 6. The management problems types in the integrated model

Now, the research will determine the management problems which will be solved without consuming additional time and cost, by using one of the developed MIS sub-approaches, as in the fig 7.

This research uses the mathematical probability theories to implement the testing of the integrated model. This testing acts as internal test of the integrated model before using it in the practical life.

The testing procedure includes two tests, which are:

1) Test (A): The Percentage Law Test

2) Test (B): The Simple Random Sampling (SRS) Probability Method Test

No	**Management Problem Type**	**MIS Developed Approach**	**Saving Time & Cost During the building of MIS to Solve the Management Problem**
1	First Order Management Problem	MIS Approach (1)	✓
2	Second Order Management Problem	MIS Classical Approach	
3	First Level Management Problem	MIS Approach (2)	✓
4	Second Level Management Problem	MIS Classical Approach	
5	First Combined Management Problem	MIS Approach (3)	✓
6	Second Combined Management Problem	MIS Approach (1)	✓
7	Third Combined Management Problem	MIS Approach (2)	✓
8	Fourth Combined Management Problem	MIS Classical Approach	

Fig. 7. The harmonization between the MIS sub-approaches and all types of management problems in the integrated model

B. Test (A): The Percentage Law Test

It is a general test which aims to measure the percentage of management problems which will be solved through the integrated model without consuming additional time and cost:

By returning to the figures: 6 & 7, we will obtain the following:

- The number of all management problems types in the integrated model = (8).
- The number of management problems types which will be solved by the integrated model without consuming additional time and cost = [The number of management problems types which will be solved by using one of these developed MIS sub-approaches: MIS (1), MIS (2), MIS (3)] = (5).
- The number of management problems types which will be solved by the integrated model without saving time and cost =[The number of the management problems types which will be solved by using the MIS classical approach] = (3).

Now, test (A) will use the percentage law, as in the following:

Number of All Management Problems Types in the Integrated Model (which equals 8) $\longrightarrow$ 100

Number of management problems types which will be solved by the integrated model without consuming additional time and cost (which equals 5) $\longrightarrow$ X ??

Then, the result of applying the previous percentage law is follows:

The value of $X = \dfrac{5 * 100}{8}$

Thus, the percentage of $X = \dfrac{5 * 100}{8}$ %

Hence, we can find the following:

a) The percentage of the management problems which will be solved by the integrated model without consuming additional time and cost, will be calculated as follows:

$$\frac{5 * 100}{8} \% = 62.5 \%$$

b) The percentage of the management problems which will be solved without saving time and cost is:

$$100 \% - 62.5 \% = 37.5\%$$

1) Analysis of the Test (A) Results:

Test (A) gives positive results of using the integrated model, since the results of test (A) shows that (62.5%) of the management problems in the integrated model will be solved without consuming additional time and cost, while (37.5%) of the management problems will be solved without saving time and cost, thus, test (A) confirms the success of using the integrated model.

The results of the test (A) are clarified through the following chart:

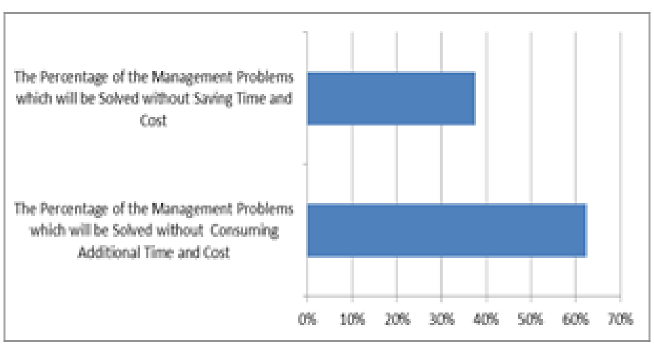


Fig. 8. The test (A) results

C. *Test (B): The Simple Random Sampling (SRS) Probability Method Test*

Test (B), is a determined test and it aims to measure the expected probability of using each developed sub-approach: [MIS approach (1), MIS approach (2), MIS approach (3)] which will solve the management problems without consuming additional time and cost. Accordingly, the expected probability of using the MIS classical approach that will solve the management problems without saving time and cost will also be measured. and Finally, the comparison between the two results will be created, in order to prove the validity or invalidity, of using the integrated model.

This test is implemented by using the simple random sampling (SRS) probability method, as in the following experiment (Test (B) Experiment):

a) Test (B) Experiment

The test (B) experiment describes the integrated model as in the following figure:

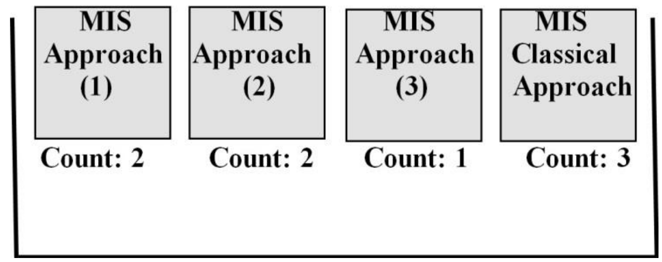


Fig. 9. Test (B) experiment

b) Test (B) Experiment Ingredients:

According to the simple random sampling probability method (SRS), the test (B) experiment ingredients will be as follows:

- **The Sample Space (Ω):** it is the set of all possible outcomes of the experiment.

In test (B) experiment, the sample space (Ω) will be as follows:

$$\Omega = \{ \textit{MIS Approach } (1), \textit{MIS Approach } (2), \textit{MIS Approach } (3), \textit{MIS Classical Approach} \}$$

- **The Number of Sample Space Elements**:

As it is shown in figure (9), the test (B) experiment includes:

- Two from MIS approach (1).
- Two from MIS approach (2).
- One from MIS approach (3).
- Three from MIS classical approach.

So, the number of all (Ω) elements, is (8) elements in the test (B) experiment.

- **The Event:**

The event (E) is a set of outcomes of an experiment (a subset of the sample space) to which a probability is assigned [20].

In test (B) experiment, the events will be as follows:

E1: {MIS approach (1)}.

E2: {MIS approach (2)}.

E3: {MIS approach (3)}.

E4: {Classical MIS approach}.

- **The Probability Theory:**

The simple random sampling (SRS) probability method defines the probability of an event E [P (E)], as the number of outcomes favorable to the event, divided by the total number of outcomes [19]:

$$P(E) = \frac{Number\ of\ Elements\ of\ E}{Total\ Number\ of\ Elements\ in\ \Omega} = \frac{Favorable\ Outcomes}{Total\ Outcomes}$$

Each element in the sample space (Ω) has an equal probability of selection and each combination of elements has an equal probability of selection.

Now, test (B) will apply the probability theory of the simple random sampling method to all experiment events, and thus the probability of using of each sub-approach in the integrated model will be calculated, as follows:

- Probability of using the MIS approach (1) is indicated as: P (MIS approach (1)), and then:

$$\text{P (MIS approach (1))} = \frac{2}{8} = 0.25$$

-Probability of using the MIS approach (2) is indicated as: P (MIS approach (2)), and then:

$$\text{P (MIS approach (2))} = \frac{2}{8} = 0.25$$

-Probability of using the MIS approach (3) is indicated as: P (MIS approach (3)), and then:

$$\text{P (MIS approach (3))} = \frac{1}{8} = 0.125$$

-Probability of using the MIS classical approach is indicated as: P (MIS classical approach), and then:

$$\text{P (MIS classical approach)} = \frac{3}{8} = 0.375$$

c) Analysis of the Test (B) Results

The previous results of test (B) experiment show a high probability of using the MIS developed approaches which will limit the consumption of additional time and cost in building the MIS's in order to solve the management problems. These approaches and the probabilities of using each one of them are as follows:

The MIS Developed Approach	Probability of the Using
MIS approach (1)	0.25
MIS approach (2)	0.25
MIS approach (3)	0.125
Total of the Probabilities: 0.625	

Fig. 10. The positive results of test (B)

The probability of using the MIS classical approach that will not limit the consumption of additional time and cost in building the MIS's in order to solve the management problems, is (0.375). Thus, the results of test (B) experiment prove the validity and success of using the integrated model.

The results of the test (B) experiment are clarified through the following chart:

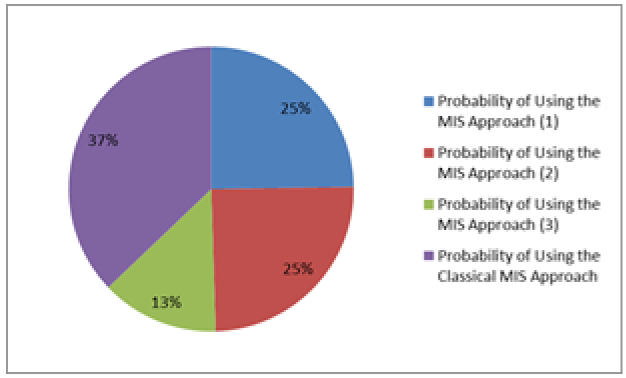


Fig. 11. Test (B) Results

IV. Conclusion

This paper uses the mathematical probability theories to test the validity of the integrated model in achieving its work and mission in limiting the consumption of additional time and cost through using of the classical approach in building the MIS's in order to solve the management problems. The paper implements two main internal tests that are test (A) and test (B), and the tests' results confirm the success of the integrated model in achieving its work and mission, since the tests' results show that the integrated model can solve 62.5% from the management problems without consuming additional time and cost, while 37.5% of the management problems will be solved without the saving of time and cost.

V. Future Work

After the success of the integrated model in the internal test that implemented in this paper, the integrated model needs now an external test in the practical use. This test will increase the confidence of system-developers in the integrated model.

References

[1] Mohammad M M Abu Omar, Khairul Anuar Abdullah "*Developing a New Integrated Model to Improve the Using of Classical Approach in Designing Management Information Systems"* , International Journal of Advanced Computer Science and Applications- IJACSA, September 2015,Volume 6 – Issue 9.

[2] Mohammad M M Abu Omar, "*A New Approach to Increase the Efficiency of Classical Approach in Designing Management Information Systems(MIS's)*", International Journal of Research in Computer and Communication Technology-IJRCCT, November 2014, Volume 3– Issue11.

[3] Mohammed M M Abu Omar, "*Developing New Methods in Designing Management Information Systems by Using Classical Approach to Solve Management Problems*", International Journal of Computer Applications-IJCA, September 2014, Volume 101.

[4] Mathias Riechert, *"Research Information Standardization as a Wicked Problem: Possible Consequences for the Standardization"*, in Proceedings of the International Conference of Current Research Information Systems, at Rome, 2014, Volume 12.

[5] Samson Yusuf Dauda, Jongsu Lee, "*Technology adoption: A conjoint analysis of consumers' preference on future online banking services*", Elsevier Journal of Information Systems, 2015, Volume 53.

[6] Ram L Kumar, Antonis C Stylianou, "*A process model for analyzing and managing flexibility in information system"*, European Journal of Information Systems, 2013.

[7] Abubucker Samsudeen Shaffi , Mohaned Al-Obaidy "Analysis and Comparative Study of Traditional and Web Information Systems Development Methodology (WISDM) Towards Web Development Applications", International Journal of Emerging Technology and Advanced Engineering-IJETAE, November,2013,Volume 3-Issue 11.

[8] Valdis Vizulis, Edgars Diebelis, "*Self-Testing Approach and Testing Tools"*, Journal of University of Latvia, 2012, Volume 787.

[9] Lionel Briand, Yvan Labiche, "*A UML-Based Approach to System Testing*", Journal of Carleton University, 2002, Version 4.

[10] Hardcastel, E., Business Information Systems, Ventus Publishing, 2008.

[11] Langer, A. M., Analysis and Design of Information Systems, Springer, 3rd Edition, 2008.

[12] Laudon, Kenneth C., and Laudon, Jane P., Management Information Systems, Prentice –Hall International, 9th Edition, 2006.

[13] Laudon, Kenneth C., Essentials of Management Information Systems, Prentice –Hall International, 6th Edition, 2005.

[14] Lejk, M. and Deeks, D., An Introduction to Systems Analysis Techniques, 2nd Ed., Pearson Education Limited, 2002.

[15] Satzinger, J., Systems Analysis and Design, 2nd ed., Thomson Learning, 2002.

[16] Dennis, A. et al, Systems Analysis and Design: An- Object-Oriented Approach with UML, John Wiley &Sons Inc., 2002.

[17] Maciaszek, A., Requirments Analysis and System Design, Developing Information Systems with UML, Addison-Wesley, 2001.

[18] Griffin, R. W., Management, 7th Edition, Houghton Mifflin, 2001.

[19] Ash, R. B., Basic Probability Theory, 1st Edition, Dover Publications Inc., New York, 2008.

[20] *http://www.mathgoodies.com/*

[21] . Ziya Aktas, A., Structured Analysis and Design of Information Systems, Prentice- Hall, 1987.

[22] Mayes, T. R. and Shank, T. M., Financial Analysis with Microsoft Excel, South-Western College Publishing, 6th Edition, 2011.

[23] Gitman,L.J., Principle of Managerial Finance, Pearson Education Inc., 2003.

[24] Christian Nielsen, Morten Lund (Eds.), The Basics of Business Models, bookboon.com., 2nd Edition, 2014

About the Authors

Mohammad M M Abu Omar is currently pursuing his PhD in Information Systems at Limkokwing University of Creative Technology, Malaysia. Abu Omar has published many papers related to the area of information technology and management, in impact international peer reviewed journals. He is a member in many technical committees and editorial boards in international peer reviewed journals. Abu Omar is also a member of the scientific committees of many international conferences throughout the world including: Switzerland, Poland, Malaysia, Japan, and India.

Dr. Khairul Anuar Abdullah is currently a faculty member in the Post Graduate Center of Limkokwing University of Creative Ttechnology, Malaysia. He received his PhD from the University of Newcastle, Australia., He is a supervisor of many masters and PhD thesis in Limkokwing University of Creative Technology.

An Embedded Modbus Compliant Interactive Operator Interface for a Variable Frequency Drive Using Rs 485

Adnan Shaout
The Electrical & Computer Engineering Department
The University of Michigan – Dearborn, MI U.S.A.

Khurram Abbas
The Electrical & Computer Engineering Department
The University of Michigan – Dearborn, MI U.S.A.

***Abstract*—The paper proposes the architecture and software design of a Modbus Compliant Operator Interface Panel (MCOIP) for a high speed Variable Frequency Drive (VFD) – a state of the art embedded design that offers several key advantages over the existing proprietary industrial models in use today. The use of serial Modbus RTU communication over RS485 allows an economically feasible, open source, vendor neutral, feature laden, robust and safe operating model. Through the use of an ARM based RISC microcontroller, the low response time of the design makes the human machine interface more real-time and interactive.**

Keywords*—*Modbus RTU; Variable Frequency Drive Operator Panel; Modbus Master VFD

I. Introduction

A Variable Frequency Drive is an embedded system that controls the speed and frequency of a motor. The system consists of a drive Operator Interface Panel along with the main drive controller as depicted in Figure 1. A high speed VFD typically requires the user of a remote human machine interface that can control and monitor the drive. Typically, this is done through the use of a panel either mounted on the chassis or remotely through a very short shielded cable. These panels allow a human operator limited functionality mainly run, stop and speed control. The panels usually have a primitive display to indicate drive speed or consist of light emitting diodes to indicate status of the drive. The drive parameters are factory programmed according to the specifications of the customer's motor ratings as indicated by the motor name plate. To operate the VFD with different rated motors, the control board of a drive would have to be shipped back to factory. Furthermore, control of the VFD through applications such as a LabVIEWTM environment to build a sub Virtual Interface is not easy for machine integration due to the proprietary communication interface developed by VFD manufacturing vendors.

High speed VFDs also known as AC Motor Drives are used in very specialized industrial applications such as very high precision grinding and milling. An exhaustive review of commercially available VFD brochures was completed to ensure the need for this design approach. Existing VFDs require an operator to be in close proximity which often poses a safety concern with the acoustic noise and the inhalation of fine particle dust [1, 2].

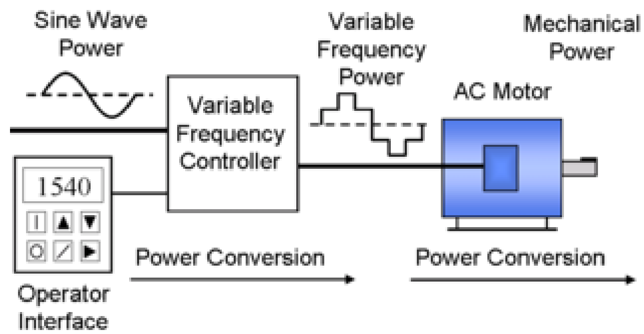


Fig. 1. High level illustration of a Variable Frequency Drive (VFD)

This limitation is mainly due to the fact that existing Operator Interface Panels on VFDs use primitive serial communication such as RS232 and the standard puts restrictions on the distance between the operator interface panel and the actual VFD [3]. The proposed design uses RS485 that can allow the remote Operator Interface Panel to be placed several hundred meters away reliably alleviating many of the safety concerns by allowing the operator to control the VFD from a safe distance or isolated premises.

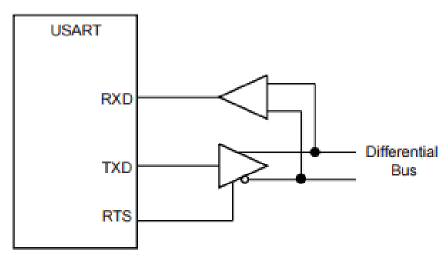


Fig. 2. Connection to a RS485 Interface

The control board of a VFD contains hundreds of drive parameters which define the way the drive manipulates the frequency and voltage to control the motor. An extensive literature review and survey of commercially available VFDs reveals that for a consumer to use a differently rated motor would require factory reprogramming [4, 5, and 6]. A convenient way to modify these drive parameters on the fly would make motor control more customized. Furthermore, the design also opens up the realm of providing access to the complete parameter table stored in the drive.

The afore mentioned RS485 communication illustrated in figure 2 also offers higher bandwidth than RS232 in present day VFD Operator Interface Panels to make this increased flow of bidirectional data between the VFD and the operator station. This would require adding more peripherals namely a graphics display and a full keypad. The use of specialized data structures to handle the read and write requests between the VFD and the Operator Interface Panel is also proposed in this design. This is achieved using Modbus at the application layer of the Open Source Interconnection Model which is an open source industrial protocol suited for this application. This is particularly beneficial as the biggest challenge with VFDs is the lack of standardization and use of proprietary software for communication. It also makes it easy for the user of a VFD to replace an Operator Interface Panel with prepackaged Modbus ready VIs for LabVIEW™ integration. In applications where the VFD powers up a high speed motor to several thousand RPMs in a matter of seconds the added responsiveness of the system makes the speed display (and other parameters) more real time. This is a desired change as in existing VFD Operator Interface Panels which use typical seven segment displays for speed this real time ramp up of speed is not observed.

The paper is organized as follows: section 2 will present requirements for the Modbus Compliant Operator Interface (MCOIP) Panel for a high speed Variable Frequency Drive, section 3 will present the design of the MCOIP, section 4 will present simulation and results, section 5 will present the performance of the system and section 6 will present conclusions.

II. Requirements

A. Functional Requirements

Firstly, although an Operator Interface Panel for a VFD is primarily a communication device it also serves other purposes. The ability to read/write parameters adds complexity to the main loop of the embedded device. It requires the use of various peripherals of the microcontroller to interface with the graphics display, keypad matrix etc. Furthermore, the requirement to meet the responsiveness of the VFD is further constrained if the main loop encompasses other tasks such as display refresh, keypad scanning etc. in addition to the communication routine.

The programmability of the design will allow the system to be used with a wide range of different motor ratings. Depending on the application, there could be well over 200 drive parameters that could be modified. This feature would require interactive communication with the VFD through the use of the operator interface to modify drive motor parameters (individually on a per parameter basis or in bulk).

The design allows a single operator interface to be used with various master units (multiple drives/multiple motors). Various industrial applications require one operator unit and multiple VFD units (e.g. high precision grinding/milling applications). Presently, this is done through daisy chaining and signal mirroring which requires the use of additional hardware. Modbus is suited for such single slave multiple master implementation and the software design techniques to accomplish this will be described in greater detail in the design.

Since the operator panel is always communicating with the VFD in an open control loop, various drive parameters are always being retrieved by the operator interface. This is to ensure that the human operator is always visibly aware of the drive status while navigating through various different menus of the operator interface. Modbus has special commands to read large registers in one go which are purposely placed contiguously.

Lastly, a requirement of the design is to allow the VFD operator to be able to connect the drive several thousand feet away from the actual drive [7]. Existing drives systems use RS232 which imposes restrictions on the distance between the operator interface and the drive controller. This limitation can potentially add risk to the operator in certain applications where high speed, high voltage AC motors are used. Using RS485 which uses a differential communication model can hence offer more safety by allowing the operator interface to be located farther away with very little error rate [8]. Modbus standards limit this distance to 1000m [9]. Furthermore, Modbus also defines the type of connector to be used along with the pin out.

This paper discusses the various aspects of the project progress in conjunction with the waterfall model by describing the requirements, design, implementation, verification (system performance).

B. Non-Functional Requirements

The functional requirements stated add stringent real time constraints on the embedded system. To overcome some of these timing constraints the design would incorporate a phase locked loop with an aptly selected crystal oscillator. The embedded coding would also utilize a highly optimized compiler for code efficiency while keeping in mind the space limitations to keep the cost down.

The use of timer and peripheral interrupts can delay execution of tasks within the main system loop and the execution of the operator interface will require a robust scale of reentrancy.

The paper aims to put forward a design which should be compatible with practically any Modbus compliant controller by simply plugging the operator interface into the port. This non-functional requirement would require some robustness in the design.

This embedded system should leverage the use of a watch dog timer to prevent deadlocks and also offer error handling to ensure safe drive operation. The system model and state design machine should ensure all process deadlines are met. A brownout circuit will also be explored in case of loss of power detection.

The operator interface being a slave Modbus system should be interchangeable with any Modbus compliant control methodology. The software inside the embedded device should be designed such that it can be ported and used in a National Instruments LabVIEW™ environment on a computer.

Table I summarizes the key advantages that the proposed design offers over existing VFD operator panels.

TABLE I. KEY ADVANTAGES OF MCOIP

Features	MCOIP
Interactivity	• Full keypad array (numeric). • Fast response time (more real-time).
Programmability	• Through keypad or any Modbus compliant slave. • High data transmission rates.
Reliability	• Increased noise immunity. • Reliable communication.
Flexibility	• Remote operation up to 1000m. • Vendor neutral.

III. DESIGN

The ATMEL AT91SAM7S series of RISC Processor was chosen for this embedded system. It incorporates the ARM7TDMI ARM Thumb Processor and provides 32 bit RISC architecture. The RISC architecture is suited for Modbus in an industrial environment as this provides the design immunity to electromagnetic interference [10]. Figure 3 shows the peripheral block diagram of the series and illustrates the outlined peripherals to be used this design [11]. This was determined in lieu of the prior requirements set forth while following the software waterfall model.

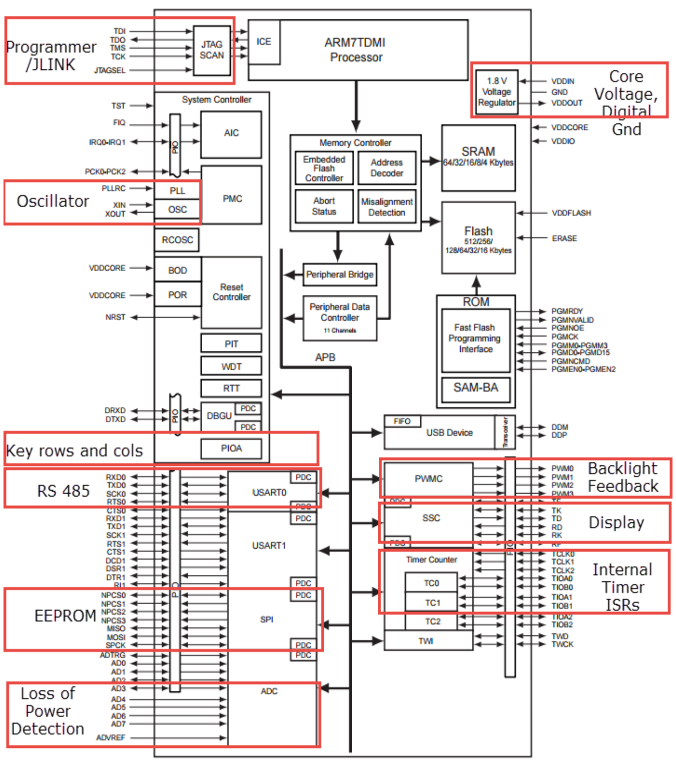


Fig. 3. AT91SAM7S Block Diagram and Pheripherals Used

A readily available oscillator/crystal with a frequency of 18.432MHz can be used [11].

$$Clock\ Frequency = 31.334MHz$$

$$1\ instruction = {}^{1}/_{31334000\ HZ} = 31.91 \text{ ns}$$

$$1\ instruction = {}^{1}/_{31334000\ HZ} = 31.91 \text{ ns}$$

$$Max.Instructions = {}^{50ms}/_{31.91ns} = 1{,}566{,}906$$

The design in this paper focuses primarily on the main functional aspects of the Operator Interface Panel for the VFD although some of the other functions which may be outside the scope of the paper are mentioned. The design was suited to fit the needs of the AT91SAM7S and its available multiplexed peripherals along with a prototype printed circuit board (PCB). However, it has flexibility and portability to be applied to other microcontrollers and a different schematic implementation of a hardware design. Conceptually, the crux of the proposed state of the art design described next is what makes the implementation unique and should be adhered to as much as possible.

Modbus protocol was originally developed and published for industrial PLC communication in the 70s. It supports RS323, RS422, RS485 communication interfaces as well as Ethernet interface (TCP/IP). Figure 4 illustrates the place of Modbus pertaining to the OSI model. Modbus facilitates communication between devices and dictates how devices send requests, responses, handle errors and records. It uses the “big-Endian” for representation of Big-Endian data item addressing i.e. most significant byte sent first if more than a single byte transferred/received.

Register addressing scheme in Modbus will be used to accomplish the latter. A VFD can design the parameter table in any way desired as long as it adheres to Modbus registering format [12]. This makes the Operator Interface Panel (Master) universally compatible with any Modbus Slave VFD.

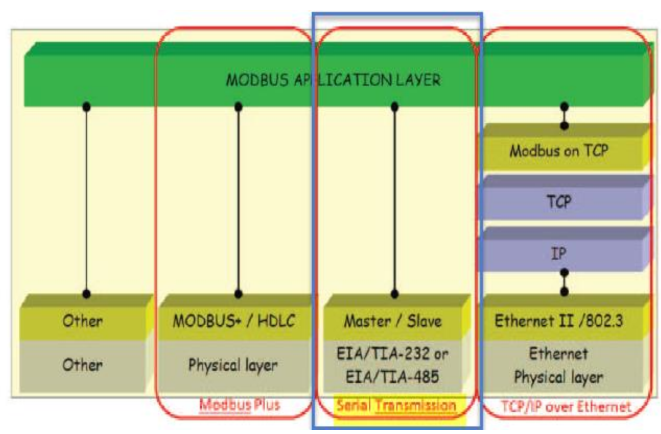


Fig. 4. Modbus and the OSI Layer

The data field is closely related to the function code and can vary with what is sent as shown in Table II.

Modbus also defines the physical connectivity for the serial RTU implementation [9]. It requires the use of RJ45 serial connector as shown on the right on Figure 5. The pin out and the color coding of the signals of the twisted pairs is also standardized as shown and implemented.

TABLE II. MODBUS FUNCITON CODE

Function code	Name	Function
01	Read Coils	Read the current status of coils
02	Read Input Discrete	Read the current status of discrete input
03	Read Multiple Registers	Read the contents of holding registers
04	Read Input Register	Read the contents of Input registers
05	Write Single Coil	Write a single output to either ON or OFF
06	Write Single register	Write a binary number to single holding register

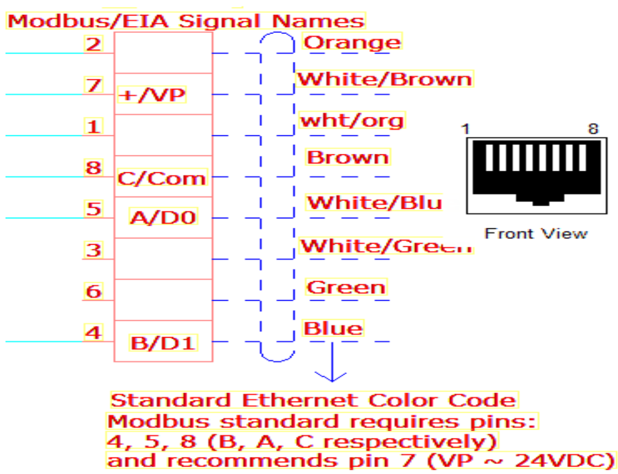


Fig. 5. RJ45 Jack EIA/Modbus Signal Names

This controller allows formation of an asynchronous state machine which features a main polled loop. Figure 6 depicts the high level state machine in terms of a flow chart. Per the non-functional requirements of this design the loop should execute within 50ms. The design is completely embedded and autonomous. It powers up and loads the Parameter Table from the drive. If that process fails the operator interface loads the default Parameter Table from code. In the main loop it performs a check to see if a key was pressed, carries out Modbus communication and refreshes the graphics display. This main polled loop is allotted a maximum total execution time of 50ms. A loop execution time of larger than 50ms will result in the design losing a responsive feel and will render the design non-real time.

A. Interrupts

- A timer interrupt set to occur every 1ms is used to implement key scanning to scan and check key entry against every row column combination. It is also used to enough time has passed (50ms) to flash the data entry cursor and then to process the key to determine if a key was pressed by the operator and which one. Figure 7 is a flow chart depicting the ISR for Timer 1.

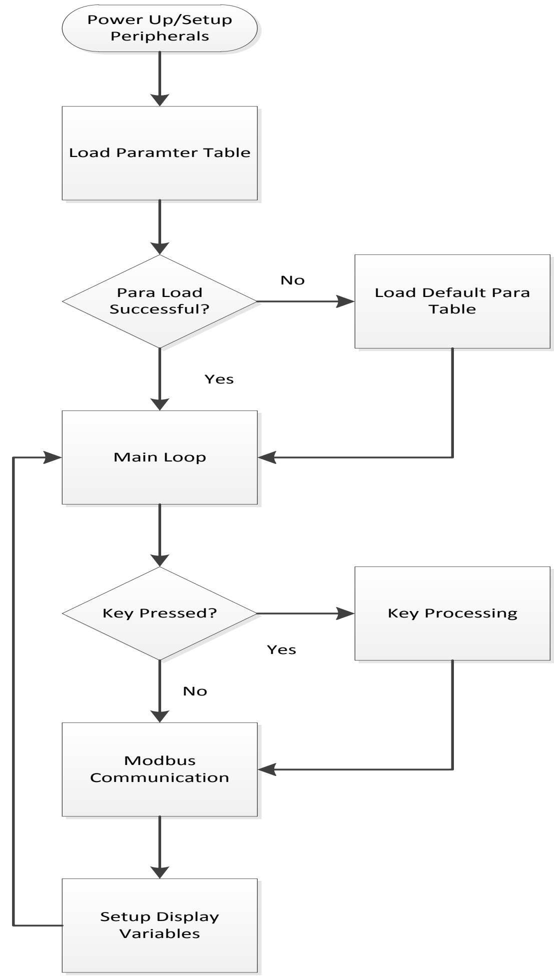


Fig. 6. Operator Panel Interface Main Call Graph

- Another timer interrupt is set to trigger every 250usec. Inside this handler a counter called Frame Silent Interval Counter is used to keep track of a 3.5 character delay. Before transmitting via Modbus this is the counter which is checked to see if that delay has not yet been met. Another counter keeps track of whether enough time has passed to determine whether a message response timeout has occurred or not (meaning the Modbus slave i.e. VFD control board in the VFD did not reply). Figure 8 is a flow chart depicting the ISR for Timer2.
- A Universal Serial Asynchronous Receiver Transceiver timer is set to trigger every time a character is received. The ISR first checks to see if the transmission for this character is complete. It would then go ahead and put the character in the Modbus Receive Buffer provided the buffer size has not exceeded 256 Bytes in conjunction with Modbus stipulations on maximum buffer size [12]. The ISR then resets the counters tracking the response timeout and the frame silent interval. The routine then checks to ensure that a 1.5 character timeout has not occurred again per Modbus stipulations indicating the frame has been received [12]. If it has occurred the interrupt is disabled and the

response time counter is reset along with the frame silent interval counter. An end of frame flag is then set to true indicating successful receives of the Modbus frame. Figure 9 is a flow chart depicting the ISR for the UART Receive.

Since the design primarily happens to be a communication device the highest priority is assigned to the USART Receive Interrupt. An incoming reception from the Modbus Slave (the VFD) is the absolute critical task which should be services immediately. The only exception could potentially be a Analog to Digital Interrupt which could potentially be triggered on a drop in voltage of the main power indicating the loss of power. This can be used in conjunction with a loss of power indication circuit to immediately service some tasks that require immediate attention such as locking EEPROM to prevent corruption, halting master requests gracefully or turning graphics display backlight off. However, these are beyond the scope of the proposed design and in normal operation the Modbus reception of a frame through the use of the USART receive interrupt has the highest priority. Table III depicts the interrupts associated with the design along with their priorities and rate/triggers.

TABLE III. INTERRUPT PRIORITIES

Interrupt	Priority (5 being highest)	Rate/Trigger
USART0	5	Receive character (1 byte)
TIMER0	4	Every 1ms
TIMER2	3	Every 250usec

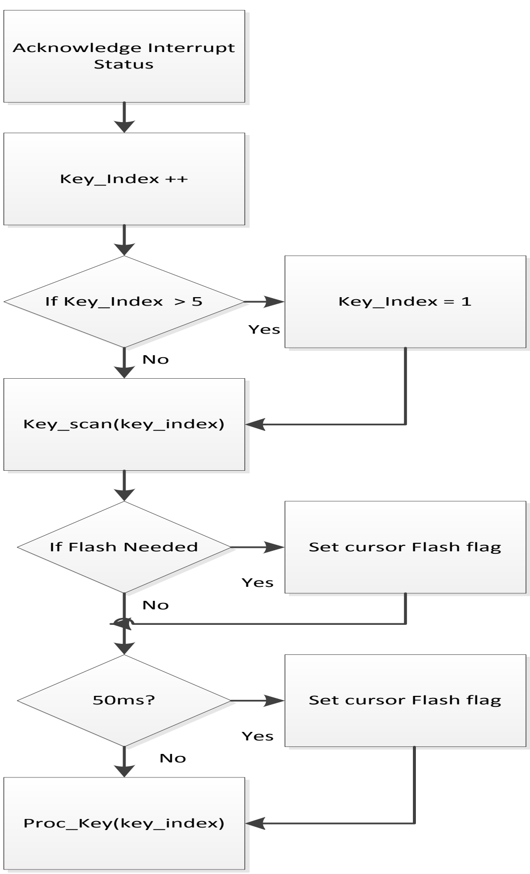


Fig. 7. Timer 1 (1ms) – Interrupt Service Routine

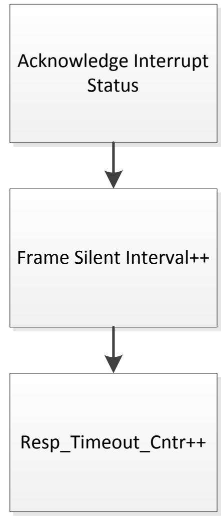


Fig. 8. Timer 2 (250usec) – Interrupt Service Routine

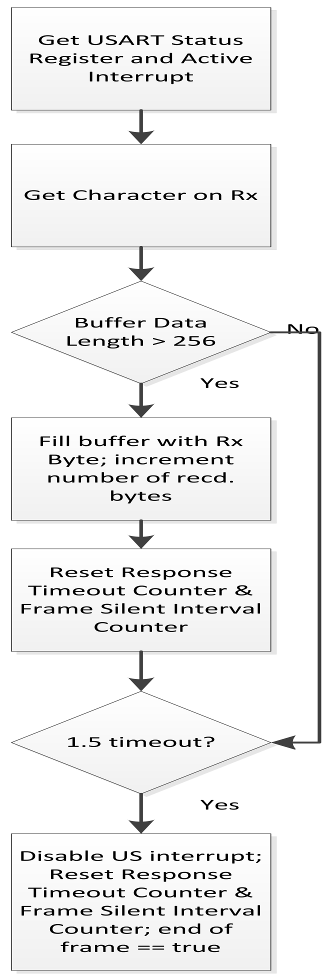


Fig. 9. USART Receive Interrupt Service Routine

B. *Modbus Communication*

Modbus RTU serial communication is carried out using RS485. This requires using the Universal Synchronous Receiver Transceiver (USART) in RS485 mode. As defined in the Modbus specifications a stop bit, even parity, 8 bit mode is used [12]. In order to meet stringent requirements Direct Memory Access (DMA) is used for transmit. DMA with Modbus receive cannot be used because a received frame can be variable length and indeterminate. The Modbus frame is received through the use of a USART ISR described previously.

The Modbus communication routine strictly adheres to the requirements set forth by the standards both in terms of timing and data structures [12]. Certain status parameters indicate critical information such as the drive run statistics, speed, fault etc. depending on what section of the menu a user is in. These are purposely placed contiguously in the Modbus memory block for easy of retrieving. Modbus uses Cyclic Redundancy Check to ensure error free reception and transmission of frames. This is done by calculating the CRC value that uses a standard 16 bit generated polynomial to check out 16 bit-check code for any length information fields [12]. Any Modbus frames received are also checked for CRC to ensure errors have not occurred [9, 12]. Modbus frames prepared for transmissions are also appended with the checksum calculation to satisfy Modbus requirements and to ensure the slave receives error free frames.

For timing the communication routine ensures that before it sends out a new frame the silent interval has passed. There are certain drive parameters that the Operator Interface Panel is always reading in every execution of the main poll loop which executes the Modbus Communication block. Figure 10 shows the Modbus Slave (VFD) state diagram while Figure 11 shows the Modbus Master (Operator Panel Interface) state diagram.

In case of a transmission error the frame retransmission is tried thrice with a delay of 10ms to avoid bus contention with the Master. If there is a problem that persists then the slave should do back off retries i.e. introducing an extended delay between each retry to reduce processing on the slave (VFD) side. In such a scenario the display would display no data for the status or the parameter being checked. As soon as communication is reestablished the data would appear on the display. The user is still able to browse through the menus and do limited operations on the Operator Interface Panel in the event of no communication from the Slave (VFD). This is a stark improvement to the traditional VFDs in the industry today where an operator panel with lack of interactivity to begin with becomes completely unresponsive until communication is reestablished.

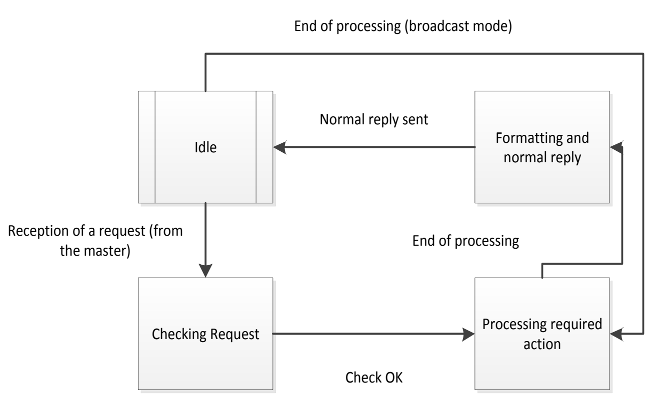


Fig. 10. Modbus Slave State Diagram

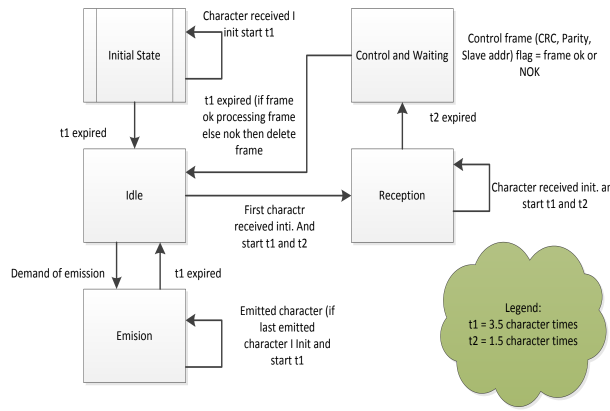


Fig. 11. Modbus Master Transmission Mode State Diagram

C. *Keyscanning*

A requirement of the Operator Interface Panel required a user to be able to do data entry in order for drive parameter programming. This is made possible through the addition of a 5 x 5 key matrix using tactile switches. This is introduced to meet the requirements to make the Operator Interface Panel programmable and fully interactive with rich human-machine interfacing when it comes to drive parameters.

TABLE IV. KEYPAD ENTRIES

Key	Function
Status	Go to Status Screen
Parameter	Go to Parameter Screen
Setup	Go to Setup Screen
Fault	Go to Fault Screen
Control Panel	Go to Control Panel Screen
0 to 9	Numeric Data Entry
Decimal	Decimal Entry
Clear	Clear Entry
Enter	Accept Entry
Escape	Cancel Data Entry
+/Up	Scroll Up
-/Down	Scroll Down
Start (I)	Start Command
Stop (O)	Stop Command
F1	Future Use
RST	RESET Microcontroller

Key scanning is implemented by putting a software routine to scan the 5 columns rows inside a 1ms timer as indicated in Figure 7. The general purpose I/O pins of the AT91SAM7S translating to 5 rows and 5 column can be used for this purpose and should be defined as internal pull ups. The 5 column pins should be defined as inputs only. The un-driven row pins should be weakly pulled up (open source).

It is assumed that only one key is to be pressed at a time. If there is a second key pressed either at the same time (theoretically) or right after the first one, it will be ignored by the key scanning and key processing routines. In case of a simultaneous press the order of the key scanning process dictates what key takes precedence. Table IV highlights the available options of the 25 keys. This can be customized based on the needs of Control Interface Panel.

The timer interrupt triggering every 1ms shown in Figure 7 does key scanning and processing. The key scanning cycles through the rows and columns to determine what is pressed where as the key processing routine implements de-bouncing and goes ahead and aptly stores the pressed keys into a “First In First Out” (FIFO) buffer. A gear count is also used in case the same key is repeatedly pressed to speed up scrolling processes. The main poll loop checks to see if a key is pressed by reading the FIFO as indicated in Figure 6. There are certain menu screens of the Control Interface Panel where pressing a key can potentially trigger a data entry routine to meet the requirement of making the design programmable. This could be a parameter entry. Hence, every key is handled on a case scenario and special consideration of the state of the Control Interface Panel keeps track of the data entry.

IV. Simulations & Results

A. Results

To test the design the Modbus Slave Simulator is programmed with dummy variables. The Modbus Slave Simulator is set up using the communication settings as shown below which are on par with the design code on the USART peripheral settings i.e. 115Kbps Baud Rate, 8 Data Bits, Even Parity and 1 Stop Bit.

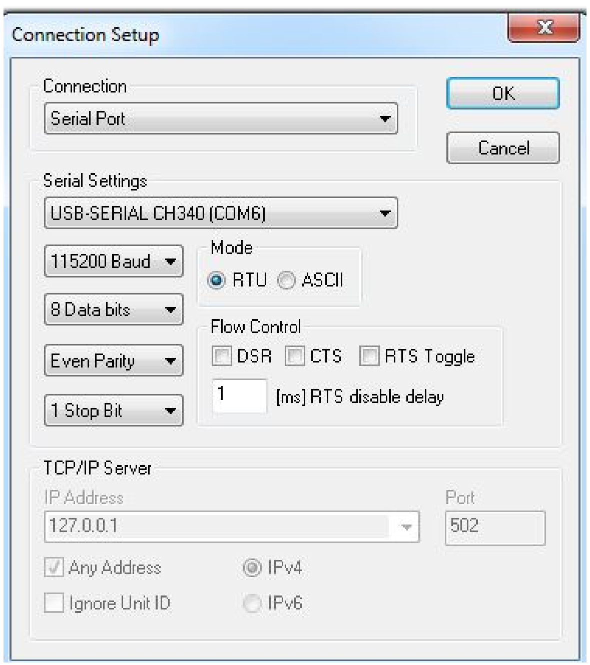


Fig. 12. Slave Communicating Settings

As an example parameter 2.04 of the drive is chosen arbitrarily. From the parameter table 2.04 in figure corresponds to Register Number 40097 as shown in Figure 13.

Modbus	Index	Address	#	Parameter Description	Unit
				Motor Control [All Motors]	
40093	46.5	0x5D	2.01	Voltage Gain	
40095	47.5	0x5F	2.02	Scaling Factor (Current)	
40097	48.5	0x61	2.03	Electrical Frequency	Hz
40099	49.5	0x63	2.04	Frequency Gain	

Fig. 13. Parameter Table Entry of Parameter 2.04

The electrical frequency of a motor is typically a floating point value. To optimize our design floating point variables were not used in the design. Floating point values were handled by using Binary 16 and Binary 32 variables with scaling. For this parameter, entries were defined in the parameter table; a snippet shows in Figure 13. The store radix indicates that the value would be using a 3rd decimal place. The maximum and minimum are merely for a local master bound check for this particular application and do not necessarily represent the actual data structure’s maximum and minimum. Hence an out of bound value was retrieved by the master is capped to the maximum applicable value as defined by the parameter table from Figure 13 for this parameter.

A dummy register entry was made in the Modbus Slave Simulator in the register 40097 as shown in Figure 14. The value entered is 23 (0x17).

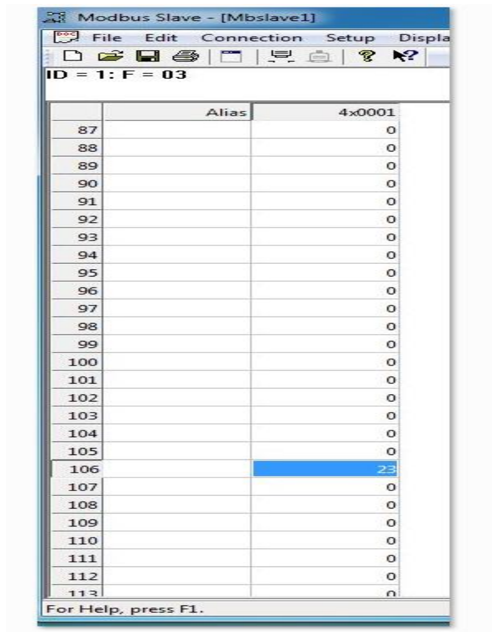


Fig. 14. Slave Communicating Settings

The Modbus poll request for this frame would be:

01 03 00 60 00 02 45 D7
Device Address: 0x01
Function: 0x03 (Read Holding Register)
Point Address: 40097
Point Code: 2
Checksum: No Error

This is verified by the communications log monitoring of the Modbus Slave shown in Figure 15 (or through the use of a Logic Analyzer connected to the RS485 transceiver). The Modbus response to this poll would be:

01 03 04 00 00 00 17 BA 3D
Device Address: 0x01
Function: 0x03 (Read Holding Register)
Point Count: 2 (Implied by byte count of 4)
Point Index 1: 0x00
Point Index 2: 0x17
Checksum: No Error

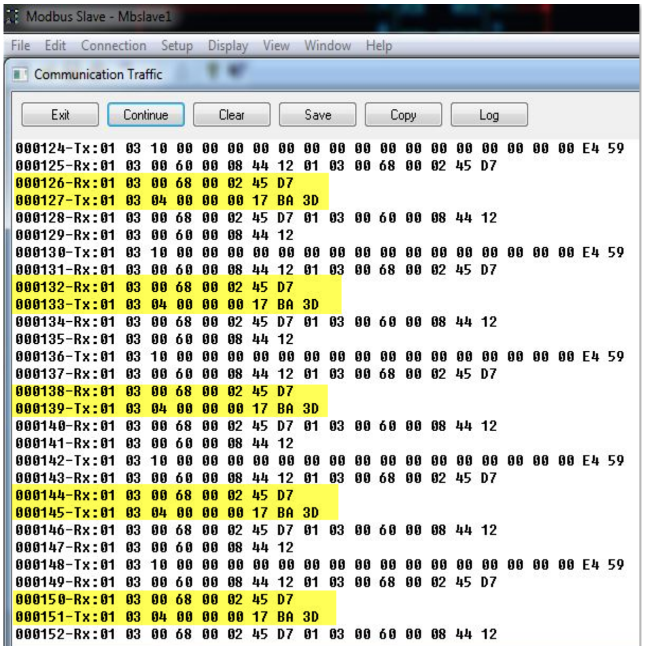


Fig. 15. Modbus Slave Communication Log

This value is translated using the data structure definition in Figure 14 by the case check for the display parameter routine and displayed on the screen as shown in Figure 16 which shows an actual snippet while this was performed.

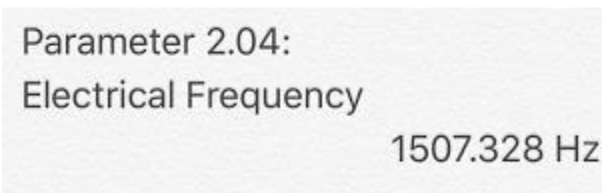


Fig. 16. Parameter 2.04 value retrieved from Slave

B. Simulation

Since Modbus limits the length of the RS485 cable to 1000m a test was done to ensure the design meets the requirements of being remotely deployed. A 1000m CAT5 Ethernet Cable could be used with the design and a count of errors through the internal error log as well as the Modbus Slave Simulator could be performed over time to count CRC errors. Since a commercially available 1000m Ethernet cable is expensive, a signal integrity simulation was performed using Mentor Graphics Hyperlynx to assess the effect of Bit Rate and Frame Delay Variation (also known as Jitter) vs. cable length. This was done through the use of exporting the schematic model available from the Mentor Graphics PADS Schematic to the Hyperlynx simulation product. The schematic is shown in Figure 17.

A differential voltage threshold of -100mV and +100mV was used to demonstrate a practical (and not a perfect) receiver) between pins A/D0 and B/D1. The simulation charted in Figure 18 show found to be within acceptable levels for up to 850 meters which falls slightly short of the 1000m for this design as set forth by Modbus (highlighted in yellow in Figure 18).

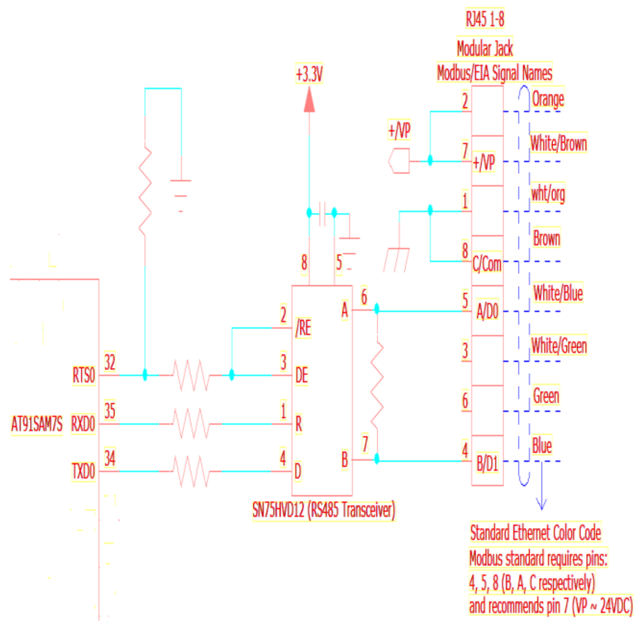


Fig. 17. Schematic of the RS485 circuit

This is a limitation in the Hyperlynx Simulation Tool however the design specifications of Modbus guarantee operation up until 1200m. The Baud Rate used in this design (115Kbps) is also highlighted in purple.

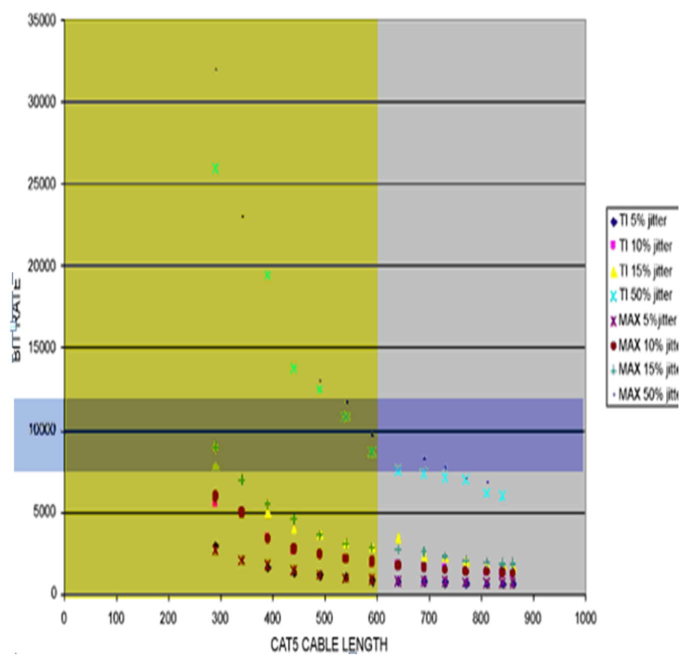


Fig. 18. Chart showing frame delay variation with Bit Rate (Kbps) vs. Cable Length (m)

V. System Performance

Due to the nature of the application the timing requirement of 50ms discussed in the requirements section is critical. Many techniques were used both to measure and modify the system response time to ensure that all the tasks complete without the required time. With the use of these techniques it was found that the main loop on average operation on a typical screen of the Operator Interface Panel was measured to be:

Worst case execution time for Main Poll Loop = 32ms

Worst case CPU utilization (no retransmissions) = 64%

Some of these techniques to evaluate and improve the system performance of the device are listed below.

A. Probing

Various routines inside the main polled loop were individually measured to determine the compositions. This was easily implemented through the use of probing an unused I/O pin. The multiplexed I/O pin could be set before the start of the routine and then cleared after the routine is finished. A digital oscilloscope (a logic analyzer could be used here as well) was used to determine the high time of the wave form indicating the time taken for the routine. The response times of the interrupts were also evaluated to ensure no unnecessary latency exists.

Decisions to use the SSC peripheral for updating the display and DMA for USART Transmit were primarily based on the results of the probing. SSC and DMA access the RAM independently without bogging down the processor allowing plenty of headway into the 50ms main polled loop.

B. Communication Response Times

Modbus communication response times from a simulation Modbus Slave Tool are exported and sent over to an open source software to perform an analysis [13]. Status 1 screen has the most amount of communication and represents the worst case scenario. For this reason it is chosen for testing response times and results depicted in Table V for the average and maximum response times (all under 50ms).

TABLE V. STATUS 1 RESPONSE TIMES

τ	Avg. RT (ms)	Max. RT (ms)
1	43.238	45.642
2	41.875	42.327
3	41.056	43.293
4	40.240	43.549

C. Inline Assembly Code

In scenarios instances involving initialization of peripherals where exact timing is necessary; inline assembly was used to ensure complete control over timing through the use of modules based on no operations instructions.

D. Optimized Compiler

An IAR Embedded Workbench Compiler for ARM processors was used to compile, debug and build the code. A Segger J-Link JTAG was used to flash the microcontroller. A balanced approach was taken with regards to optimization for speed vs. code space.

E. Dissassembly

In order to ensure optimized compiler behavior was not causing undesired results the disassembly was used as a means to verify correct operation. This was especially useful in the ISR to ensure it is short and simple and does not increase latency.

VI. CONCLUSION

The paper has presented the architecture and a software design of a Modbus Compliant Operator Interface Panel (MCOIP) for a high speed Variable Frequency Drive (VFD). The use of serial Modbus RTU communication over RS485 allows an economically feasible, open source, feature laden, robust and safe operating model. The use of specialized data structures to handle the read and write requests between the VFD and the Operator Interface Panel was proposed in this design. This is achieved using Modbus at the application layer of the Open Source Interconnection Model which is an open source industrial protocol suited for this application. The programmability of the design allowed the system to be used with a wide range of different motor ratings through the use of an ARM based RISC microcontroller. The low response time of the design made the human machine interface more real-time and interactive. Based on this proposal, future implementations can be designed to allow functionality with other variants of Modbus (namely TCP/IP) and other vendor neutral open source communication protocols.

REFERENCES

[1] N. Hashemi, L. R. Lisner, D. Holmes. “Accoustic Noise Reduction for an Inverter-Fed Three-Phase Induction Motor,” Conference Record of the IEEE, 2004, pp. 2030-2035 vol 3.

[2] “Dust and its control,” Occupational Safety and Health Administration.“https://www.osha.gov/dsg/topics/silicacrystalline/dust/chapter_1.html (last accessed 10/27/2015).

[3] B. Carlson. “Maximum transmission distance for RS-232 Computer Cables,” Pulp and Paper Ind. Tech. Conf., 1988, pp. 86-93.

[4] J. Kimbrell. “Top 10 Tips for Specifying VFDs,” Design Engineering, 2014, pp. 30-33 vol 60 issue 1.

[5] [12] B. MB and Vishwanath, “Hegde Energy Conservation using Variable Frequency Drives of a Humidification Plant in a Textile Mill”, 2015 International Conf. on Power and Advanced Control Engineering (ICPACE).

[6] [13] G. Morris and B. Weiss, “Driving Energy Efficiency with Design Optimization of a Centrifugal Fan Housing System for Variable Frequency Drives”, Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm), 2012 13th IEEE Intersociety Conf.

[7] G. Guarese, F. Sieben, T. Webber, M. Dillenburg, C. Marcon. “Exploiting Modbus Protocol in Wired and Wireless Multilevel Communication Architecture,” Brazilian Symp. on Comput. Sys. Eng., Natal, 2012, pp. 13-18.

[8] V. Kumar. “Overcoming Data Corruption in RS485 Communication,” IEE Int. Conf. on Electromagnetic Interference and Compatibility, 1995, pp. 9-12.

[9] “Modbus over Serial Line Specifications and Implementation Guide V1.02,” www.modbus.org/docs/Modbus_over_serial_line_V1_02.pdf (last accessed 10/27/2015).

[10] K. Wang, D. Peng, L. Song, H. Zhang. “Implementation of Modbus Communication Protocol Based on ARM Cortex-M0,” IEEE Int. Conf. on Systems Sci. and Eng., Shangai China, 2014, pp. 69-73.

[11] “AT91SAM ARM-based Flash MCU,” ATMEL Corporation, http://www.atmel.com/images/doc6175.pdf (last accessed 10/24/2015).

[12] “Modbus Application Protocol Specifications V1.1b3,” www.modbus.org/docs/Modbus_Application_Protocol_V1_1b3.pdf (last accessed 10/24/2015).

[13] G Kunzel, C. Pereira. “A Tool for Response Time and Schedulability Analysis in Modbus Serial Communications,” IEEE Int. Conf. on Ind. Informatics, 2014, pp. 446-451.

A Sleep Monitoring System with Sleep-Promoting Functions in Noise Detection and Sound Generation

Lyn Chao-ling Chen
Graduate Institute of Networking and Multimedia
National Taiwan University
Taipei, Taiwan

Kuan-Wen Chen
Intel-NTU Connected Context Computing Center
National Taiwan University
Taipei, Taiwan

***Abstract*—Recently, there has been a growing demand and interest in developing sleep-promoting systems for improving sleep condition. Because sleep environments are various, and sensitivity to noise differs individually, it is difficult for current sleep-promoting systems to provide an adoptable solution. This paper develops a non-invasive sleep monitoring system with adaptive sleep-promoting sound according to sleep environments and sleep habits. For people who fall asleep in a quiet environment, a constant sound playing probably affects their sleep. The proposal is designed to distinguish the noise disturbances, and a sleep-promoting sound is triggered automatically. A device with multiple sensors: an infrared depth sensor, a RGB camera, and a four-microphone array, is used to detect sleep disturbances. When a noise is detected, an ambient sound is playing to cover the noise automatically. Besides, it also applies to people who are used to sleep with sound by providing additional sound playing from the beginning of their sleep. Moreover, from the input of depth signals and color images, the scores are calculated from the sleep information, and are record for sleep quality evaluation. An overnight experiment was carried out, and the results show the efficiency of the proposed system in diverse sleep environments. The adaptable method is feasible for individuals, and it is also convenient and cost-effective to be used in home context.**

Keywords—ambient sound; image sequence analysis; noise detection; non-invasive sleep monitoring; sleep promotion

I. INTRODUCTION

This paper develops a system for generating sleep-promoting sound automatically based on noise detection in sleep monitoring by using a device with multiple sensors.

Currently, sleep-promoting system are popular in modern society, such as smartphone applications and commercial products. The numerous sleep promotion systems show the need of people to create good sleep condition. To better understand an ideal sleep environment, some researches in social science has indicated the importance of tranquility to most of people [1][2][3]. Furthermore, darkness is also an important sleep promoter that low light condition endogenous melatonin production [4]. However, it is hard for current sleep-promoting technology to take whole control of the lighting factor in home context. Considering other condition, the comfortable temperatures of people for sleep are not all the same, and a suitable firmness bed is also different by shapes and sizes of people [5][6]. Besides, the effect of smells on sleep experience of people has not been extensively study. Sound-based sleep promotion has become the majority in current sleep-promoting systems, for the purpose of creating a relaxing atmosphere. Hence, sound condition will be addressed in this study.

A sound-based sleep promoting system is absolutely necessary for people to create good sleep condition. The sleep-promoting sound is also closely related to sleep environments and sleep habits. To our best knowledge, this problem has been seldom investigated in the literature. Noises from outside traffic, or people talking in the street, that causes difficult sleeping. But a snoring partner who sleeps in the same room is a more serious noise problem [7]. On the contrary, some people probably have difficult to fall asleep, because it is too quiet. They prefer sleeping with a sleep company, such as turning on a radio or television, raining and wind whistling form outside of window, or playing a soft music in the background. From the observation, the sensitivity to noise differs individually. Because sound is the only sensory system that remains truly active during sleep, and a brain naturally craves sensory input [8]. A quiet night also enhances the stimulus from the sleep environment, because any random noises probably active the restless brain. In both cases, the evidence has shown that white noise and rhythmic sound may help to mask other noises in a noisy room [9]. Besides, it also gives a brain a tonic signal that dampens its own internal system in a quiet room. From the mechanism of a brain working, sound is the major factor to interrupt sleep. On the contrary, it also can be used to mask environmental noises [10]. People who fall asleep easily in a quiet room only need sleep-promoting sound when noises happening, that a constant sound playing probably causes negative effect to their sleep. However, to those usually sleep with sound, the sleep-promoting sound should play continuously until they fall asleep. Besides, they also need the sound mask when noises happening during sleep. Moreover, these sleep-promoting sound should be adjust to moderate volume in sleep condition. Some researches has shown that overly loud white noise significantly reduces REM (Rapid Eye Movement) sleep in a variety of animals [8]. Hence, the sound-based sleep promotion helps to improve sleep condition of people. Besides, for the variety of sleep environments and the diversity of sensitivity to noise of people, an adaptive method is in an urgent need.

It is possible to provide adaptive sleep-promoting sound if noise disturbances can be detected. Then when the noises happen, they are covered by sleep-promoting sound immediately. Currently, there is not any automatic method in sleep promotion systems to detect sleep disturbances.

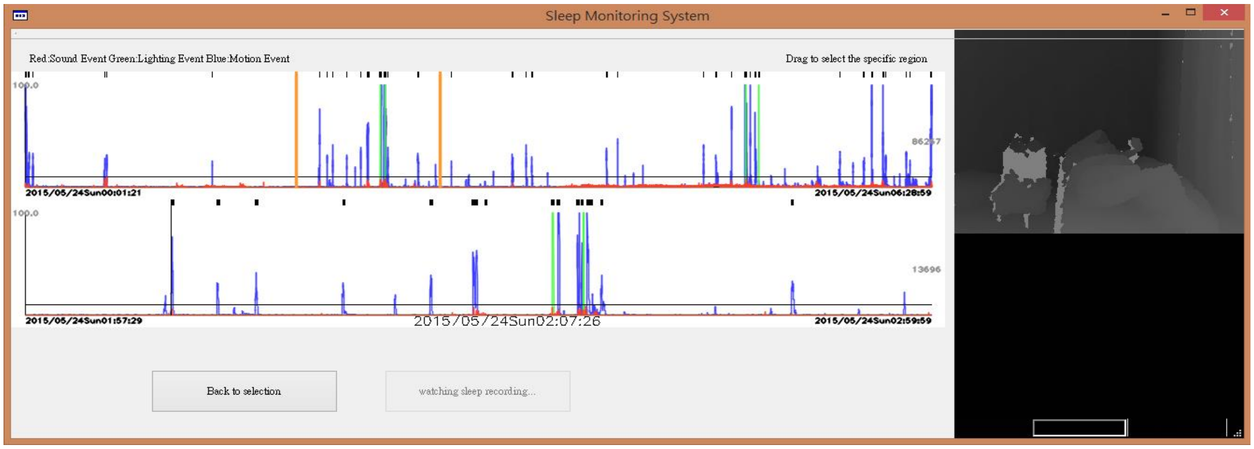


Fig. 1. Interface of sleep diagram for sleep examination. The black marks on the top indicate index of sleep disturbances recording. Users select the specific region with marks to browse the corresponding video clips. Meanwhile, the selected region is expanded to show more details

In previous work, a method has been proposed for distinguishing occurrences of disturbances from a sleep environment [11]. Furthermore, the method can be used to identify the appropriate time for sleep sound promotion.

In this paper, a sleep monitoring system with adaptive sleep-promoting sound by an epoch approach method has been developed to detect the occurrence of sound events. To our best knowledge, the proposed system is the first adaptive sound-based sleep promotion system, which is triggered by the time estimation from the noise detection in sleep monitoring. Given a stream of audio signals, the audio signals are scored in a constant interval by an epoch approach method [11]. When the noise is detected, an ambient sound is playing to mask the noise automatically. Meanwhile, the motion behavior and lighting changes are scored, and the depth images and color images are recorded simultaneously. Finally, an interface with sleep diagram is provided for sleep examination, which specific section can be expanded to browse the corresponding video clips showing in Fig.1. For people who fall asleep in quiet sleep condition, the proposal is designed to distinguish noise disturbances, instead of constant sound playing which may affects their sleep. Besides, it also applies to people who are used to sleep with sound by providing additional sound playing from the beginning of their sleep. The results of an overnight experiment show the efficiency and flexibility of the system.

This paper is organized as follows. Section II reviews the related works. Section III presents the proposed adaptive sleep-promoting method. The experiment and evaluation are discussed in section IV. Section V provides the conclusion and future work.

II. Related Research

In the section, the related research is reviewed. Objective of this research is to propose a method for sleep promotion and sleep quality improvement toward heathy people or some with sleep problems. Sleep disorder is one of the biggest modern plague in the twenty-one century, and is harm to health and quality of life of people. Not only people who with sleep disorders, but also heathy people may have difficulties falling asleep occasionally [12] [13]. Under the promise that total sleep time is not reduced, even a mild sleep disturbance suppress deep sleep that is sufficient to affect memory performance of healthy people [14]. Therefore, the World Sleep Day has been named on March 14 by WASM (World Association of Sleep Medicine) to arise the attention from public to sleep issues. In spite of the medicine treatment is still widely used, a common non－pharmaceutical treatment is using sound as sleep promoter. Because sound is the only sensory system that remains active in sleep [8]. Hence, it is feasible to provide sound-based sleep promotion system for sleep experiences improvement.

A. Sound-based Applications in Sleep Promotion

A large number of mobile applications with sleep-promoting sound had been proposed, and sounds from nature environmental materials are widely used in current systems. Sleepmaker Rain collected real rain sound recording from World Heritage Listed Forests in Tasmania, Australia. It allows users to choose their favorite rain sound to sleep with, such as drizzling rain, pouring rain, or thunder rain. Nature Sound Relax and Sleep collected more all natural sounds without any music composition, like song of birds and winter snowing. aSleep considered both natural sounds and background music with adjustable speed.

Moreover, Relax Melodies, and Relax and Sleep provided users with large amount of ambient sounds to organize their own playing lists. Some researches started to focus on the correlation between binaural beats frequencies and brainwaves. The evidence show that binaural beats frequencies slow brainwaves, and that help people to fall asleep [15]. From this point, both Relax Melodies and Pzizz considered binaural beats frequencies. In addition, Pzizz generated new harmonious soundtrack through neuro-linguistic programming with every time use. Proactive Sleep Alarm Clock expanded sound database by using the iTunes songs of users as sleep promoter, and also provided a sleep report for sleep examination. Besides, it not only considered an objective measurement to monitor body movements during sleep through an accelerator of a cellphone. But also a subjective measurement to record daytime activities of people for proactive prediction. In addition, all these sleep promotion applications can be used for meditation or napping at daytime, or for relaxing sleeping at night.

TABLE I. COMPARISON OF SOUND-BASED SLEEP PROMOTION SYSTEMS

Sound-based Sleep Promotion	Functions				
	Sound material	***Adjustment***	***Wakening***	***Addition***	***Evaluation***
Sleepmaker Rain	Real nature sound	X	X	X	X
Nature Sound Relax and Sleep	Nature sound	X	X	X	X
aSleep	Ambient sound Background music	Speed	Alarm	X	X
Relax Melodies	Ambient Sound Music	Playing list	X	X	X
Relax and Sleep	Ambient sound	Playing list	Alarm	X	X
Pzizz	Synthesized soundtrack	Voice on/off 3D effect	X	X	X
Proactive Sleep Alarm Clock	Ambient sound iTunes music of users	X	Music alarm Snooze feature Viberate wakening Gradual wakening	Motion detecion Sleep diary	Sleep report
SleepPhones	Ambient sound Soundtracks from users	X	X	X	X
Philips Wake-up Light	Natural sound Radio show	X	Alarm Snooze feature	Lighting theropy	X
Lee's Sleep monitoring system [16]	X	X	X	Temperature Posture detection	Sleep report
The proposed system	Natural sound Soundtracks from users	Adaptive promotion	X	Motion detection Lighting detection Noise detection	Sleep report

B. Sound-based Commercial Products in Sleep Promotion

Some commercial products using multimedia as sleep promoter has been announced. SleepPhones used a soft headband with removable speakers, and the firm fabric helps to block the noises from an environment (http://www.sleepphones.com/). The headphone can be plugged in devices of users to play their own music, and the devices can be hidden under their pillows. In addition, a commercial CD with ambient sound selling on their website also provided a good option. This attached product needs people to wear to bed, some people may have uncomfortable sleeping. Moreover, Philips Wake-up Light created a bedside lamp that combined with natural sounds and lighting control to stimulate an environment with nature sunlight (https://www.usa.philips.com/c-p/HF3520_60/wake-up-light). The evidence has shown the correlation between lighting and melatonin production, and that is related to sleep qualities of people [4]. It gradually decreases sound and light for providing a relaxing sleep environment, then gradually turns into a bright warm light to wake up users naturally in the morning. Both products are convenient for people to use at home for sleep promotion.

C. Disturbances Detection in Sleep Monitoring

Many studies had aimed on an activity-based technique for monitoring a subject in bed. However, only few researches emphasized environmental factors in sleep condition. In previous work, a sleep monitoring system has been developed to identify three kinds of events in a sleep environment: sound events, lighting events, and motion events [11]. Another monitoring system also considered the temperature and humidity changes in a room, and sleep postures of a subject, but noises factor was disregarded [16]. The major problem of the system is that the providing posture detection is based on the skeleton recognition from Microsoft Kinect, it is hard to detect a correct posture if a subject sleep with quilt covering. That is hard to apply to real sleeping.

Table I compares the functions of related works and the proposed system in the paper. These works used sound as sleep promoter to improve sleep experiences of people. The providing sound for sleep promotion can be classified into three types: natural sounds, background music, and synthesized soundtracks. Except SleepPhones is a wearable product, the rest of them do not need to attach on the bodies of people during sleep. In addition, most of them have an alarm to end up sleeping. Because the purpose of the proposed system is to record normal sleep of users, it is not offering an alarm to interrupt their sleeping. Only sleep monitoring systems considered the influences of environmental factors on people. Besides, both the two sleep monitoring systems and Proactive Sleep Alarm Clock provided a sleep report for sleep examination. From the comparison on Table I, the proposed system in this paper is superior to these sound-based sleep promotion systems by providing an adaptive promotion solution.

In this section, the related research has been reviewed. From the discussion, various sound-based system had been developed for sleep promotion. However, they did not well meet users need. In this paper, both the technologies of monitoring and promotion in sleep condition are considered, to identify the appropriate time for sleep promotion. Moreover, a sleep report with sleep diagram and sleep quality evaluation is also provided for users to examine their sleep.

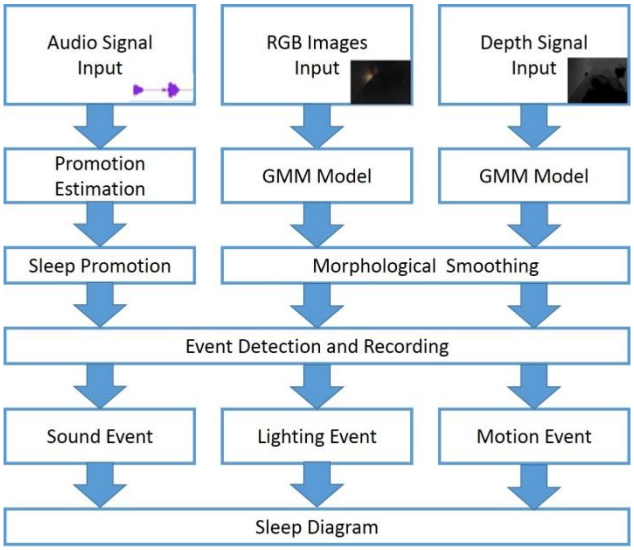


Fig. 2. Flow chart of development of the adaptive sound promotion system

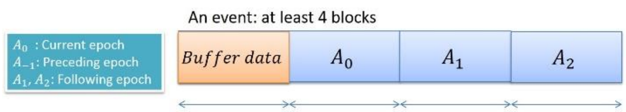


Fig. 3. The components of an event. The recording consists of data from the previous epoch, and data from the following three epochs to compeletely record occurerence of an event

III. System Design

In this section, the system flow of the sleep monitoring system with adaptive sleep-promoting sound will be described. This section reviews the core components of the proposed system, the method for promotion estimation, the automatic sleep-promoting sound, and the sleep quality evaluation. Fig. 2 shows the flow chart of the proposed system. In the following, an epoch approach method is introduced to estimate appropriate time of sleep promotion, which is based on noise detection in sleep monitoring. Then a sleep-promoting sound is triggered automatically to cover noises in an environment according to the estimation. Finally, a sleep/wake identification method is adopted to calculate motion information from the sleep recording for sleep quality evaluation.

A. An Epoch Approah Method for Promotion Estimation

This section explains the method for promotion estimation that can be used to identify appropriate time for sleep promotion during sleep. Given a stream of audio signals, the audio signals are scored and normalized to a scale from 0 to 1. In preliminary stage, the empirical threshold was set from audio sensor data. An epoch approach method was proposed to determine a continuous event at a constant interval [11]. Using an epoch length of 3 seconds, once an audio score in the interval is over the empirical threshold, a noise is detected.

There are two different aspects in sleep promotion and sleep monitoring , and two empirical thresholds are set in audio signals to classify degree of noises: audio promotion threshold and audio detection threshold. The audio promotion threshold (0.15) is used to detect sounds over 40 dB. The table from ASHA (American Speech-Language-Hearing association) declares the decibel levels of common environmental sound (http://www.asha.org/public/hearing/Noise/). The 40dB sounds is in moderate degree, such as talking in low voice. In general, a sleep environment under 40dB sounds is quiet that the sound do not wake up people. However, people may have turning behaviors, small twitches during sleeping under the feeling of noise disturbances. Therefore, the audio promotion threshold is used to estimate the appropriate promotion time when a noise over 40dB is detected. Another threshold (0.6) is used to detect remarkable sound over 60dB. The 60dB sound is about the volume of normal conversation voices and that has already may startled people from sleep.

The audio detection threshold is used to identify and record specific events for sleep examination. Once a sound event is triggered, the audio signals and the corresponding depth images and color images are recorded simultaneously to generate video clips. The length of the event is at least with 4 units of epoch lengths (12 sec.) showing in Fig.3. The continuous event completely record the occurrence of an event. Besides, the report from ASHA also mentioned the long-term effect of noise problem in sleep condition, that people easily have
cardiovascular disease sleeping under 50dB sound. Moreover, to sleep under 70dB sound will increase 30% incidence of MI (Myocardial infarction). Hence, the effect of noise disturbances should not be ignored in a sleep environment.

To consider other disturbances in a sleep environment, the epoch approach method is also applied to determine a motion event and a lighting event. It iteratively aggregates data in a constant interval in (1) [11].

$$\begin{cases} if\ A_{-1} < A_0, start\ of\ an\ event \\ if\ A_0 \leq A_1, the\ event\ continues \\ ;\ otherwise, end\ of\ the\ event \end{cases} \quad (1)\ [11]$$

where A_{-1}, A_0, A_1 = the activity count of the preceding epoch, the current epoch, and the following epoch. The empirical thresholds from depth sensor and color sensor data was set in the preliminary stage. Given streams of depth signals and color images, one GMM background model is established in depth signals, and another GMM background model is established in RGB images [17]. After morphological smoothing, the noises of segmented foreground objects are reduced, then the area of foreground objects are scored and normalized in the same scale. The depth score and color score are considered as magnitude of movements, and magnitude of lighting changes respectively. After that, the epoch approach method iteratively calculates number of scores over the empirical thresholds. Then recording of motion events and lighting events are triggered when activity count of current epoch is bigger than activity count of the previous epoch. Through this way, the sparse score value is discarded for that is not able to be considered as an event. Hence, a sleep diagram is provided with various sleep information for sleep examination.

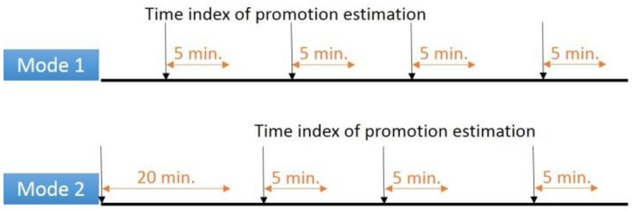


Fig. 4. Two playing modes of automatic sleep-promoting sound

Algorithm 1 shows the steps of promotion estimation and event detection. Given a set of scores from audio signals stream, if the audio signal is over the sound promotion threshold, a promotion time is marked. In addition, an sound event is triggered if the audio signal is also over the sound detection threshold. In other word, the appropriate time for sleep promotion includes the time index of sound events happening. Meanwhile, give a set of scores from depth signals stream, and a set of scores from color images stream, if the activity count of current epoch is bigger than that of the previous epoch, an event is triggered and starts a recording. The recording of the event continues while the activity count of the following epoch is equal or bigger than the activity count of the current epoch, otherwise the recording of the event is ended.

B. *Automatic Sleep-promoting Sound*

This section explains two modes of sleep-promoting sound and the providing sound materials. From the promotion estimation, the appropriate time for sleep-promoting sound playing has been identified. Moreover, for people have different sensitivity to noises, two modes are provided for general people and some who prefer to sleep with sound showing in Fig. 4.

1) People sleeping with quiet: The sleep-promoting sound is only playing according to the estimation time from the poposed epoch approch method. The sleep-promoting sound plays 5 minutes according to the time index recording to cover and smooth noises.

2) People sleeping with sound: To people who are used to sleep with sound, the steps of sound promotion as follows:

- At the beginning of sleep, the sleep-promoting sound plays 20 minutes to help them fall asleep.
- During sleep, the promoting sound plays 5 minutes according to the time index recording from the promotion estimation method to cover and smooth noises.

Besides, ambient sounds from natural materials are provided by default, such as calm sea waves, dripping rains, pouring rain or thunderstorms, and users are also allowed to sleep with their own soundtracks. Hence, the adaptive sleep promotion system is realized through the method of promotion estimation by both considering sound-based sleep promotion and sound detection in sleep monitoring.

Algorithm 1 Promotion Estimation and Event detection
Input: Given a set of scores from depth signals S_d, color images S_c, and audio signals S_a;
Output: a set of promotion estimation E_p;
a set of motion events E_m, lighting events E_l, and sound event E_s;

```
1: initial audio promotion threshold A_P and audio detection
   threshold A_D;
   initial audio score A_s;
   initial count of current epoch A_d0, A_c0 and preceding
   epoch, A_d-1, A_c-1;
2. if(A_s > A_P)
   {
        update E_P;
        start playing;

        if(A_s > A_D)
          start recording;
   }
   else
   {
        if(A_d0 > A_d-1 || A_c0 > A_c-1)
          start recording;
   }
3. if(start recording)
   {
      do
      {
         update E_m, E_l, E_s;
      }while (repeat<=4);
   }
4. update activity count of preceding epoch, A_d-1,
   A_c-1 and current epoch A_d0, A_c0;
5. repeat 2 to 4.;
```

C. *Motion Information for Sleep Quality Evaluation*

A historical activity-based method, actigraph, had been proposed by Tryon for sleep quality evaluation [18]. It is a statistical model from sampling activity data to identify sleep or wakefulness state. At first, a device based on the method, actigraphy, is a wearable device with accelerator and light sensors to monitor activities of people during sleeping. The correlation between motions and sleep states has been validated in clinical field. There is a strong correlation of 0.81 between measurement of total sleep time from actigraphy and that from PSG (Polysomnography) [19].

The three major methods of sleep/wake identification are listed in Table II, and they can be compared with each other for they all have similar parameters in their equations, and two of them using the same adjustment rules [20] [21] [22]. The accuracy of Webster algorithm determined by comparison sleep/wake scores with EEG (Electroencephalography) data, and the accuracies of Cole algorithm and Sadeh algorithm agreed with PSG data. Nowadays, many wearable devices use the sleep/wake identification algorithm for sleep quality evaluation. Only the sampling data of Webster had been eliminated for the improvement of device, the algorithms of Cole and Sadeh are still popular used in clinical diagnosis. In this paper, Sadeh algorithm is adopted to evaluate sleep quality for the highest accuracy and the simplest rescoring rule.

TABLE II. COMPARISION OF SLEEP AND WAKEFULNESS IDENTIFICATION METHODS

Sleep/Wake Identification	*Year*	*Wrist activity*	*Epoch length*	*Scoring interval*	*Rescoring rule*	*Accuracy*
Webster algorithm [20]	1982	Dominant hand	2 sec.	1 min.	5	94.46%-94.74%
Cole algorithm [21]	1992	Non-dominant hand	2 sec.	1 min.	Using 5 rules of Webstrer	87.93%-88.25%
Sadeh algorithm [22]	1994	Dominant hand Non-dominant hand	30 sec.	1 min.	1	91.37%-92.77%

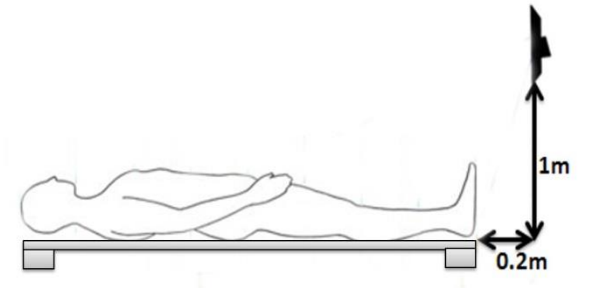


Fig. 5. An illustration of device setting in the experinment

The concept of the algorithm calculates activity count within a 30 seconds epoch length, and considers the current epoch, the preceding 5 epochs, and the following 5 epochs showing in (2).

$$PS = 7.601 - 0.065 \cdot Mean\_W\_5\_min - 1.08 \cdot NAT - 0.056 \cdot SD\_last\ 6\ min - 0.07 \cdot LOG\_Act \quad (2)[22]$$

where Mean_W_5_min is the average number of activity counts during the scored epoch, and the five epochs preceding and following it; NAT is the number of epochs with activity level equal to or higher than 50 but lower than 100 activity counts of the scored epoch, and the five epochs preceding and following it; SD_last 6 min is the standard deviation of the activity counts during the scored epoch, and the five epochs preceding it; finally LOG_Act is the natural logarithm of the number of activity counts during the scored epoch, and pluses 1. If PS (probability of sleep) is zero or over zero, the specific epoch is scored as sleep state, otherwise it is scored as wake state.

One rescoring rule with 1-minute scoring interval is used to improve the accuracy. There are two 30 seconds epochs in the interval of one minute. If both sleep state and wake state present in the same minute, then the minute is rescored as wake state. There is high possibility to be wakefulness if wake state occurs in this short period, and that helps to avoid the false wake problem. Moreover, the algorithm has been verified the tolerance of device sensitivity by adding a constant value or multiplying a constant ratio with the activity count value, and that did not affect the accuracy of the result. Hence, the method can be adopted with a totally different device. From the motion information of the device, the sleep or wakefulness state of sleeping is identified, then the SE (sleep efficiency, total sleep time/total time in bed) is calculated for sleep quality evaluation.

D. Summary

In summary, this section described the method to realize adaptive sleep promotion in the proposed sleep monitoring system. Moreover, an activity-based method to calculate sleep efficiency is applied to evaluate sleep quality. In next section, two different kinds of sleep conditions are discussed: a quiet sleep environment and a noisy sleep environment. The calculation of sleep quality is used to examine the effect of sound-based sleep promotion in the overnight experiment.

IV. EXPERIMENT

This section evaluates the proposed method with a realistic scenario. A device (Kinect for Windows) with a depth sensor, a RGB camera, and a microphone was used in the experiment, with 8-bit VGA resolution (640 × 480 pixels) of RGB images, 11-bit VGA resolution (640 × 480 pixels) of depth signals, and 24-bit resolution of audio signals. Both of depth signals and RGB images are at a frame rate of 30 Hz, and the sampling rate of audio signals is 16 kHz. In addition, a region-of-interest (ROI) (320 × 350 pixels) region is selected on both depth signals and color images for reducing signal noises. The device was placed at a distance of 0.2 meter and in a height above the bed of 1 meter showing in Fig. 5. It set in front of a bed at a depression angle of 27° that the camera view is in the center of a bed covering the entire body of a participant.

A. Experiment Settings

The samples of 5 subjects includes 4 men and 1 woman, ranging in age between 34 to 40 years (mean age=35.6; SD=2.88). The samples includes 5 normal adults and 1 patient who has been diagnosed with sleep apnea (age: 40). All participants were volunteers to participate in the study, for access the effects of an adaptive sleep promotions and discover the disturbances in their sleep environments.

Experiment space was set in home scenario that all participants were asked to sleep on their own beds. It avoids the effect from an unfamiliar environment which usually causes trouble sleeping, and also helps us to collect various sleep information from different sleep environments. All participants were allowed to keep their sleep habits, such as turning on a small lamb. There is no fixed sleep time in the experiment setting, and they were asked to get to bed and wake up as usual. For testing the effect of the adaptive sleep-promoting sound, three conditions were tested in the experiment as follow:

1) Participants sleeping with original sleep condition: All participants sleep in their usual sleep envionmtents without any interventions by deault.

2) Participants sleeping with sleep-promoting sound: All participant sleep with sleep-promoting sound when the noises are detected in their sleep enviornments.

3) Participants sleeping with sleep-promoting sound and noise interuptions: All paritciapnt sleep with sleep-promotion sound when the noises are detected in their sleep environment. In addiotn, noise interruptions is setting in a constant interval during their sleeping to test the effect of sleep-promotiong sound, especially for the participants whose rooms are usually quiet with few noise problems. Four kinds of noise intruptions with a runinig truck, going down stairs, road-works constructing, and a howling ambulance were playing in one hour, one and half an hours, two hours, and two and half an hours respectively counting from the start of their sleeping.

The experiment was carried out with blind test. All participants did not know the applying conditions above. The sysetm implemented a program with one of the condtions randomly during the experinment period. In addtion, the time index of sleep sound promotion were recored. During their sleeping, all the participants were asked to answer a question from the system to record the reason of their arousing once they woke up. The exepriment period is from May 20 to July 24 in 2015. There are 36 nights of recording in the experiment. 1 subject spent 9 nights, 2 subjects spent 8 nights, 1 subejct spent 7 nights, and 1 subject spent 4 nights. The subjects slept between 11:22 P.M. to 03:45 P.M.. A total of 13,573 minutes data were collected.

B. Evaluation Results and Analysis

1) Participants sleeping with original sleep condition:

Firstly, the original sleep condition of the participants will be discussed. Table III (Condition: O is original condition, P is promotion sound, and PN is promotion sound and noise interruptions) shows the influences of noise disturbances in their sleep environments, and according to it their sleep environments can be classified into a quit sleep environment or a noisy sleep environment. From the percentage of sound events, participant A, participant C, and participant D had quiet sleep environments with few noise disturbances. But the noise problems existed in both participant B and participant E.

After the sleep experiment, an interview was held with each participant to understand more specific details about the noises sources. From the information, participant B lived near a big street where in the middle of a city, and there existed serious traffic noises even in the midnight. The source of noises were coming from a pump in the basement where participant E lived in the first floor of an apartment. From the sleep diagrams showing in Fig. 6, there are minor waves of sound event curves with participant A, participant C, and participant D. But there are obvious strong sound event curves showing on sleep diagrams of participant B and participant E. More evidences shows on the Table III. From the number of sleep promotion on the Table III, it also reveals the high frequencies of noise problems with participant B and participant E, and that results in the high quantity of sleep-promoting sound triggering.

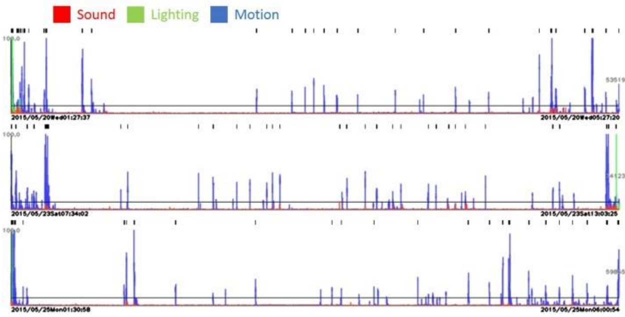
They were also asked to answer a question to the system when they aroused in the middle of sleep, and the system recorded the answer time and answer code (1: toilet, 2: insomnia, 3: noises, 4: sleep partners, 5: alarm, 6: others). From the wakening reason on the Table III, all participants woke up from the middle of sleep occasionally about only one or two days a week, except participant A. Because participant A is a light sleeper, and she usually got up to went to the toilet in the midnight. Besides, she slept with her cat and she also got up to get the door for her cat occasionally (6: others). In addition, participant C was used to sleep in his living room without a curtain, and that causes the obvious waves of lighting events showing in Fig.6. For the reason, the percentage of lighting events with participant C is also obvious higher than the others. The locations of sleep environments of the rest users are listed as follows: participant A lived with her family and owned her room, participant C rented an apartment near an alley, and participant D slept in a student dorm with a roommate.

A sleep diagram is provided for all the participants to see what exactly happened during their sleep showing in Fig.1, because they may not remembered it. Especially, when they waken by noises, they may missed the major reason of wakening and thought they just needed to go to the toilet. The circumstance occasionally happens with participant B and participant E, and they both slept in noisy sleep environments. After browsing their sleep diagrams, they found out that the noise problems were not absent with the interruptions of their sleeping. From the experimental results, all participants agreed that the monitoring system is robust to detect and to record sleep events, and the sleep diagram helps them to recognize disturbances during their sleep efficiently.

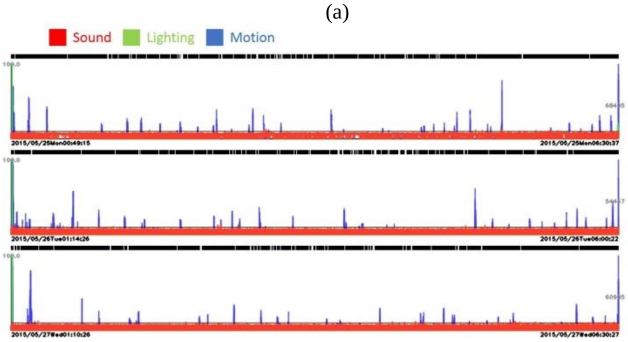
2) Participants sleeping with sleep-promoting sound:

In this section, the condition of sleep-promoting sound will be further discussed. Only participant E chose rain sound to sleep with for three days and sound of sea wave to sleep with for four days in the experiment, and the rest participants chose dripping rain sound to sleep with. Besides, they all chose the same promotion mode with the sound playing from the beginning of their sleep. That means they were aware of the sleep-promoting sound during the experiment period, even the condition of sleep-promoting sound was chosen randomly by the system. They thought that the sleep-promoting sound was provided for helping them fall asleep, but they all did not know the additional function of the adaptive promotion sound for covering noises disturbances during their sleeping. Participant A, participant C, and participant D heard the sleep-promoting sound few times during their sleeping, because their rooms were really quiet. From the number of sleep promotion showing on Table III, there are obvious few times of sleep-promoting sound playing with them. The system only plays the sleep-promoting sound when a noise is detected. However, they all did not notice the occurrences of noises problems in their rooms, and that means they did not find out the mechanism of adaptive sleep-promoting function. Therefore, the sleep-promotion sound well covered and smoothed the noises in their rooms.

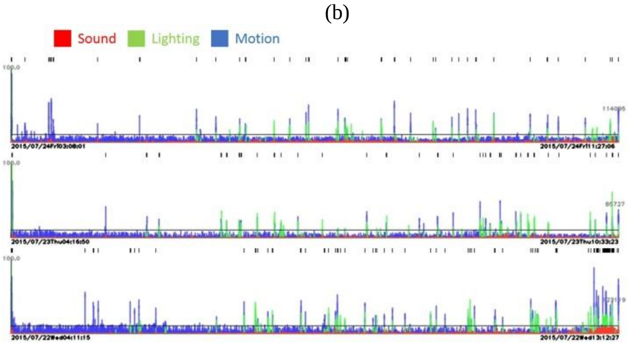
From the time index recording of sleep promotion, the special circumstance happened on both participant B and participant E. For their sleep environments were too noisy, the system played the sleep-promoting sound constantly during their whole night sleep. They both heard the playing sound sometimes during their sleeping, but it is impossible for them to find the mechanism of automatic playing sound for the constant playing. They may thought the sleep-promoting sound was originally set for whole night playing, and their sleep environments were already too noisy that they barely noticed the sleep-promoting sound playing. From the average of sleep efficiency showing on Table III, the sleep qualities in condition of sleep-promoting sound improves slightly with participant B and participant E; and the sleep qualities between the original sleep condition and the sleep-promoting sound condition are similar with participant A and participant C. Because the effect of adaptive sleep-promoting sound is not obvious in a quiet sleep environment for seldom noise disturbances. Participant D was not considered in the discussion for few data collection. From the observation, both participant B and participant E slept in noisy rooms, and the adaptive sleep-promoting sound worked well to cover the noises disturbances, and that also improves their sleep qualities.



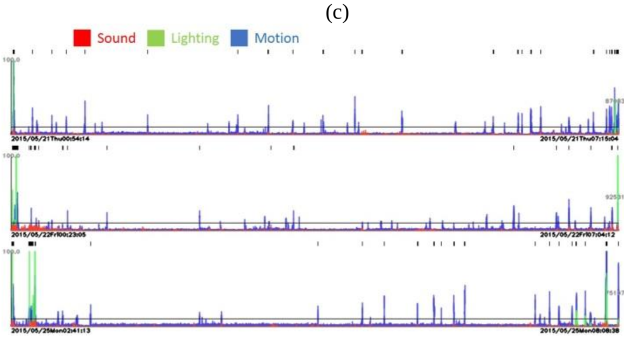
(a)



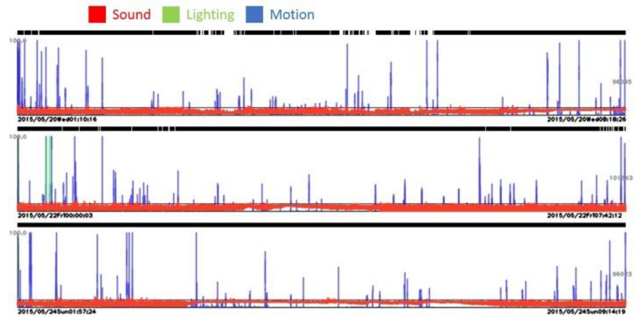
(b)



(c)



(d)



(e)

Fig. 6. Sample Sleep diagrams from participants. (a), (b), (c), (d), (e) were the sleep recording of participant A, B, C, D, E, respectively. The graph consists of red curve of sound events, green curve of lighting events, and blue curve of motion events. (a), (c), (d) have quiet rooms, and both sleeping of (a), (b) indicates the noise problems with these participants

3) Participants sleeping with sleep-promoting sound and noise interuptions:

Finally, the condition of sleep-promoting sound and noise interruptions will be discussed. All participants were not aware of the noise interruptions setting. Almost all participants were occasionally aroused in the sleep condition, except participant C who was a heavy sleeper without interruption of his sleeping during the whole experiment period. From the examination of the time index between the question reply from the participants and the noise interruptions setting, and there are no overlapping with their arousing time and the noise interruptions playing time. In addition, the function of adaptive sleep-promoting sound were also applied in the condition, and the function covering noise disturbances also helps to improve sleep qualities. The correlation between sleep quality and the sleep-promoting sound is further discussed. From the average of sleep efficiency showing on the Table III, the effect of sleep-promoting sound is not obvious with participant B. Because the interruption of his sleeping both happened in the two days sleeping of this experiment condition. However, there are better performances of sleep qualities with the rest participants. From the observation, the adaptive sleep-promoting sound helps to improve their sleep qualities efficiently to bring better sleep experiences.

V. CONCLUSION

This paper developed a system that provides adaptive sleep-promoting sound based on the sound detection in sleep monitoring. An estimation method was proposed to determine the appropriate time for sleep promotion. According to the time index from the estimation method, the adaptive promotion sound is applied to cover the noise disturbances in a sleep environment. From the experiment result, it has clearly demonstrated the efficiency of the adaptive sleep promotion for different individuals, and the reliability of diverse sleep environments. Hence, the adaptive system helps to improve life qualities of people from the improvements of better sleep experiences. Besides, the non-invasive system do not cause uncomfortable sleeping, and the device is also cost-effective and convenient to be used in home context.

TABLE III. SLEEP PARAMETERS OF EACH PARTICIPANT

Subjects	**Sleep Parameters**									
	Cond.	***Total sleep time (min.)***	***Total time in bed (min.)***	***SE (%)***	***Average of SE (%)***	***Number of sleep promotion***	***Sound events (%)***	***Lighting events (%)***	***Motion events (%)***	***Wakening reason***
A	O	193	223	86.55	90.52		0	0.05	2.15	1
		433	461	93.93			0	0	1.44	6
		327	359	91.09			0	0.06	1.9	1, 1, 1
	P	295	318	92.77	90.09	11	0	0.01	1.76	1
		273	309	88.35		12	0	0.02	1.86	6, 6
		222	249	89.16		6	0	0.01	1.83	1, 1
	PN	283	299	94.65	95.67	13	0	0.05	1.31	1, 1
		379	392	96.68		16	0	0.01	1.3	5, 5
B	O	272	299	90.97	87.12		0.15	0.03	1.47	1
		189	227	83.26			0.14	0.05	1.99	X
	P	260	285	91.23	89.68	66	0.11	0.05	1.98	X
		233	254	91.73		61	0.11	0.03	1.15	X
		260	302	86.09		69	0.15	0.1	1.81	X
	PN	194	217	89.4	87.14	53	0.08	0.05	1.89	1
		202	237	85.23		62	0.19	0.53	2.29	1
C	O	492	534	92.13	95.23		0	0.29	2.28	X
		351	357	98.32			0	0.39	0.9	X
	P	482	513	93.96	95.61	33	0	0.44	1.35	X
		462	475	97.26		31	0	0.26	0.92	X
	PN	337	348	96.84	97.49	20	0.12	0.49	1.74	X
		337	343	98.25		19	0	0.49	1.7	X
		408	418	97.61		24	0	0.31	1.49	X
		387	398	97.24		22	0	0.25	1.59	X
D	O	344	386	89.12	89.12		0	0.07	0.33	X
	P	293	363	80.72	80.72	19	0	0.03	0.62	X
	PN	518	525	98.67	98.36	39	0	0.03	0.11	5
		301	307	98.05		24	0	0.03	0.1	X
E	O	555	570	97.37	96.58		0.9	0.01	1.12	X
		409	423	96.69			0.26	0.02	1.09	X
		388	400	97			0.17	0.01	1.03	X
		411	421	97.62			0.33	0.01	1	X
		457	485	94.23			0.84	0.01	1.63	1
	P	393	400	98.25	97.81	83	0.25	0	0.97	X
		531	547	97.07		115	0.38	0.01	1.25	1
		364	371	98.11		77	0.19	0.01	0.93	X
	PN	542	558	97.13	97.13	116	0.15	0.21	1.49	1, 1

For the improvement of portable device, a smartphone with depth camera has announced recently and that is more convenient for users to use the system at home without additional device. It brings the possibilities to develop the system in mobile application. In addition, a sleep posture classification based on image analysis will be developed in the future work.

REFERENCES

[1] H. Urponen, I. Vuori, J. Hasan, and M. Partinen, “Self-evaluations of factors promoting and disturbing sleep: an epidemiological survey in Finland,” Soc. Sci. Med. 26, pp. 443–450, 1988.

[2] M. H. Bonnet, “Sleep deprivation,” In: M. H. Kryger, T. Roth, and W. C. Dement (eds.), Principles and Practice of Sleep Medicine, 3rd edn, pp. 53–71, WB Saunders, Philadelphia, PA, 2000.

[3] H. C. Van Vugt, J. Du, and R. J. E. M. Raymann, “Using qualitative methods to study sleep in home contexts,” Sleep 32 (Suppl.), 1134, 2009.

[4] C. A. Czeisler, C. A. Buxton, and S. B. S. Khalsa, “The human circadian timing system and sleep-wake regulation,” In: M. H. Kryger, T. Roth, and W. C. Dement (eds.), Principles and Practice of Sleep Medicine, 4th edn, pp. 375–394, W.B. Saunders, Philadelphia, PA, 2005.

[5] J. Du, C. E. Postma, H. C. Van Vugt, and E. Zwartkruis, “Understanding the user needs for sleep enhancement,” Philips Internal Report, 2008.

[6] B. H. Jacobson, A. Boolani, and D. B. Smith, “Changes in back pain, sleep Quality, and perceived stress after introduction of new bedding systems,” J. Chiropr. Med, 8(1), pp. 1-8, 2009.

[7] S. Venn, “It’s okay for man to snore: The influence of gender on sleep disruption in couples,” Sociol. Res. 12(5), 2007.

[8] R. A. Velluti, “Interactions between sleep and sensory physiology,” J. Sleep Res. 6, pp. 61–77, 1997.

[9] R. J. Cole, “Non-pharmacologic techniques for promoting sleep,” Clin. Sports Med. 24, pp. 343–353, 2005.

[10] M. Vallet, and J. Mouret, “Sleep disturbance due to transportation noise: ear plugs vs. oral drugs,” Experientia. 40, pp. 429–437, 1984.

[11] L. C. L. Chen, K. W. Chen, and Y. P. Hung, “A sleep monitoring system based on audio, video and depth Information for detecting sleep events,” IEEE International Conference on Multimedia and Expo. (ICME), China, July 2014.

[12] P. Cuartero, E. Estivill, “Sleeping, dreaming and resting: the keys to revitalising sleep,” RBA, Hospes, AIPS, Barcelona, 2007.

[13] E. Bixler, “Sleep and society: an epidemiological perspective,” Sleep Med. 10, pp. 1–4, 2009.

[14] Y. D. Van Der Werf, E. Altena, M. M. Schoonheim, E. J. Sanz-Arigita, J. C. Vis, W. De Rijke, and E. J. W. Van Someren, “Sleep benefits subsequent hippocampal functioning,” Nat. Neurosci. 12, pp. 122–123, 2009.

[15] J.L. Carter, and H.L. Russell, “Method and apparatus for changing brain wave frequency,” US Patent 5,036,858, 1991.

[16] J. Lee, M. Hong, and S. Ryu, "Sleep monitoring system using Kinect sensor," International Journal of Distributed Sensor Networks, 875371, 2015.

[17] C. Stauffer, and W. Grimson, "Adaptive background mixture models for real-time tracking," CVPR, pp. 246–252, 1999.

[18] W. W. Tryon, "Activity measurement in psychology and medicine," New York: Plenum Press, 1991.

[19] M. J. Chambers, "Actigraphy and insomnia: a closer look; part 1," Sleep 1994; 17:405-8.

[20] J. B. Webster, D. F. Kripke, S. Messin, and et al., "An activity-based sleep monitor system for ambulatory use," Sleep, vol. 5(4), pp. 389–399, 1982.

[21] R. J. Cole, D. F. Kripke, W. Gruen, D. J. Mullaney, and J. C. Gillin, "Automatic sleep/wake identification from wrist actigraphy," Sleep, vol. 15, pp. 461–469, 1992.

[22] A. Sadeh, K. M. Sharkey, and M. A. Carskadon, "Activity-based sleep/wake identification: an empirical test of methodological issues," Sleep 3:, pp. 201–207, 1994.

Impact of Heterogeneous Deployment on Source Initiated Reactive Approach

Nonita Sharma *, Ajay K Sharma
Department of Computer Science
National Institute of Technology
New Delhi, India

Kumar Shashvat
Department of Computer Science
Chandigarh Group of Colleges
Mohali, India

***Abstract*—Selection of an optimal number of high energy level nodes and the most appropriate heterogeneity level is a prerequisite in the heterogeneous deployment of wireless sensor network, and it serves several purposes like enhanced network lifetime, finest energy consumption, and optimal sensing coverage. The paper presents the mathematical modeling of cost, energy and sensing range analysis of 2-level, 3-level, and n-level heterogeneous wireless sensor network. An experimental investigation has been carried out to investigate the effect of heterogeneity on a proposed Energy Efficient Source Initiated Reactive Algorithm. Studies on these aspects have been done to find the limitations of the algorithm for homogeneous networks and to find how it enriches sensing range and network lifetime. Based on the simulated experimental and numerical results, a mathematical model is presented to calculate the optimal number of high-level nodes which can simultaneously enhance network lifetime and achieve optimal sensing coverage. The results are compared with the homogeneous network to prove the effectiveness of the stated approach and proposed a model.**

Keywords—Wireless sensor networks (WSNs); Cost Analysis Model; Energy Analysis Model; Sensing Range Model; Optimality

I. Introduction

Wireless Sensor Networks (WSNs) might be the fastest growing technology in our industry today. One of the ramifications of that growth is a dramatic increase in the number of applications exploring this technology and a proportionate surge in the number and types of research opportunities that could be evaluated in the future. The three primary classifications of applications in WSN are: sensor nodes forwarding data ceaselessly to the base station; base station commanding sensors to do something, and early warning systems.

Most of the classical approaches assume that the network deploys similar types of nodes regarding capabilities. However, there is a growing realization that the effect of heterogeneity needs to be explored and evaluated. Heterogeneity refers to a network that deploys nodes with different initial energy nodes. The use of heterogeneity has gained popularity because of their positive impact on network lifetime operation and enhanced energy efficiency. While the effects of heterogeneity can be evaluated after the deployment, pre-computing the optimal number of high-level nodes can make the network operations much efficient. Carrying the motivation of calculating the optimal number of high-end nodes one step further, different levels of heterogeneity are explored in heterogeneous networks. In this paper, three analysis models viz. Cost analysis, energy analysis, and sensing range analysis models for 2-level, 3-level, and n-level heterogeneous network are framed.

Also, if the number of high-end nodes increases in the network, inevitably, it will have a significant impact on the duration of network operation and network life time. But at the same time, increasing the number of high-end nodes in the network will have a direct impact on the overall cost of the network. And almost all applications work under a prescribed cost constraint. So, if there is a need to keep the cost factor same, then increasing the number of high-end nodes will lead to the decrease in the density of low-end nodes that will further impact the sensing coverage area and sensing coverage degree of the network. Sensing coverage area refers to the area that is being monitored by at least one functioning sensor in the network. Sensing degree refers to the number of total functioning nodes controlling the area. If the number of alive nodes that are monitoring the area is more, then more precise values can be generated. Hence, keeping the cost constraint in mind, a heterogeneous network needs to be explored for an optimal number of high-level nodes which can optimize both network lifetime and sensing coverage. Therefore, a mathematical model for optimality is presented to decide the optimal number of high-end nodes.

Further, selection of a particular routing algorithm depends on the characteristics of an application, which can be broadly be classified as event driven, continuous monitoring and query driven. Proactive routing strategy is applicable for continuous monitoring systems while reactive algorithms are the most suitable choices for event driven and query driven applications. In reactive algorithms, a path from sensor nodes sensing the events to the base station is formed only when events are occurring. Rest of the network operation, sensor nodes stay in sleep mode. Sleep mode helps in avoiding the energy consumption on maintaining the paths even during inactivity period. Hence, the proposed three models and mathematical analysis of optimality are simulated on a proposed reactive approach.

In this manuscript, an energy-efficient source initiated reactive (EE-SIR) approach is suggested that selects the parent node by several factors like residual energy, neighboring nodes, distance to the base station, stability values. Afterward, analysis of cost, energy and sensing range for different levels of heterogeneity on proposed EE-SIR routing strategy is presented as it has direct applications in real-life applications. The analysis is done for a direct communication model where every sensor is communicating directly with the sink. The case is presented for uniform distribution of nodes where high-end nodes are evenly distributed in the sensor area. An optimal number of heterogeneous nodes having a direct impact on network lifetime and sensing range is desired. Experiments have been conducted to create a mathematical model to calculate the optimal number of high-level nodes which can enhance network lifetime and achieve an optimal sensing range. Results demonstrate that an optimal number of high-end nodes can significantly improve the network operation, under the prescribed cost constraint. Besides, different heterogeneity levels and temporal correlation show different patterns over time, which will have a substantial effect on network performance and deployment decisions.

II. Related Work

In our research work, authors work on finding the impact of adding different energy levels to a reactive routing protocol that can help to enhance the energy efficiency and sensing coverage under prescribed cost constraint.

As the number of nodes increases, homogeneous WSN is known to exhibit poor performance and reduced scalability. The throughput of a node is $\Theta(1/\sqrt{n \log n})$ where n is the number of nodes in the network [1]. Also, as paths between nodes turn out to be longer, the likelihood of packets being lost gets to be higher. Likewise, as the number of nodes in system advances, the packet delivery rate drops significantly. Experimentation results [2] and test beds [3] have shown a fading performance with increasing number of nodes.

Over the years, the impact of energy heterogeneity on the performance of a WSN has been a topical area of research. As per the definition given in [4], a heterogeneous sensor network is defined as a network of nodes with different functionalities and capabilities. For example, Mhatre et al. propose using two types of nodes; type 0 nodes acting as pure sensor nodes, and type 1 nodes that serve as the cluster head nodes [5]. Adding high capability nodes in the network has been considered to be an efficient and effective way to enhance reliability and network lifetime [6]. Along with, Mhatre et. al presented a survey showing that the heterogeneous WSN is more suitable for real-life applications as compared to their homogeneous counterpart [7]. Inculcating heterogeneity in the network enhanced response time and network lifetime. Heterogeneous HEED [8] showed a significant improvement in the lifetime just by introducing heterogeneity of various levels. Authors recommended nodes with different energy to prolong lifetime and reliability. Most of the time, the achievable lifetime could be optimized using multiple power levels. The paper concluded that the routing strategy should be such that the energy reserves should be the focus to balance the energy consumption; rather than minimizing the absolute consumed power to maximize the lifetime. However, in the current context of issues related to this area, more work can be done to improve energy efficiency.

More to the point, a lot of work regarding three analysis models i.e. cost, energy and sensing coverage analysis of heterogeneous WSN have been discussed in the literature. In paper [9], the cost model for evaluating the cost of heterogeneous network deployment is presented. Also, the authors reached a conclusion that heterogeneous deployment achieved higher coverage rate and lower deployment cost, keeping the number of nodes same. A cost based comparative study between different communication models is done by [10], where, authors described an optimal heterogeneous sensor deployment regarding battery energy and lower cost. Heinzelman et al. [11] provided an analysis of an optimal number of clusters of energy efficient clustering protocol and provided energy consumption model for different consumption schemes. Further, Duarte-Melo and Liu [12] presented the energy consumption model and quantified the optimal number of clusters based on the suggested model. Not only the cost and energy factors are significant, but also sensing coverage remains the primary focus of heterogeneous domain. Huang et al. discuss the worst case and best case coverage problems [13]. Different sensing coverage degree has been reviewed and simulated for different network locations [14]. A coverage configuration protocol proposed by Wang et al. [15] suggested a model for application based coverage. Howard et al. [16] proposed an initial deployment to maximize the coverage. None of the work in the literature has considered a tradeoff between three models. Lee et al. [17] proposed an optimal mixture of low-end and high-end nodes that can simultaneously optimize network lifetime and sensing coverage. The work can be applied to find a solution to the optimality problem between network lifetime and sensing operations for various levels of the heterogeneous network. None of the work in the literature combined the cost, energy and sensing coverage analysis models and provided the optimal number of high-end nodes for various levels of heterogeneity. The novelty of our work lies in finding the mathematical model for multi-objective deployment of heterogeneity.

III. Working Description of Proposed Approach

Reactive algorithms are best suitable for event detection applications. The sensing nodes remain at sleep node till some event occurs, or the base station fires a query to a sensing node. Hence, there is no need to create the paths from the node to the base station in advance as these nodes will be using the paths only when the event is happening. Afterward, nodes will go to sleep mode again. This approach helps in saving the overhead to maintain the routing tables by the sensor nodes and eventually leading to energy savings. EE-SIR proposes the same approach of creating paths on a need basis. EE-SIR follows a similar approach as Source Initiated Reactive (SIR) [18] with better implementation regarding the primary routing strategy used, i.e. energy aware gossiping. EE-SIR allows source nodes to gossip their data (e.g., on event detection) as long as a route is not available. Sensing node selects the parent node by a parameter based on four factors viz. Distance to the base station, residual energy left; no of times that node has already been a parent node earlier, and the total neighboring nodes. These features make EE-SIR more reactive in the presence of dynamic network behavior, but more moderate in scenarios with demanding traffic. The originality of the stated approach is to provide the simulation and performance evaluation of EE-SIR algorithm on different levels of heterogeneous networks. Levels represent the types of nodes with different capabilities a system deploys e.g. 2-level means there are two kinds of nodes; one equipped with lower initial energy and another with higher initial energy. Similarly, 3-level means that nodes with three types of energy levels are considered. Again, n-level means that n- levels of energy are examined. Adding heterogeneity would enhance the uniform dissipation of energy. If a node with higher capability is there in the vicinity of the sensing node, then that node will be selected as a parent node otherwise, one of the same capability nodes will be chosen by other three factors defined in the stated parameter. The working of the stated approach is as under:

A. Initialization Phase

In the first phase, all the nodes are divided randomly. In the initialization phase, the network is simulated for four deployments. First, the homogeneous deployment of the network, where all the similar nodes are deployed randomly in the network. 2-level heterogeneous network, where the network considers two types of nodes viz. Lower capability nodes and higher capability nodes. Similarly for 3-level and further for n-level.

B. Neighboring Node Selection Phase

This step selects the neighboring node of the sensing node. The sensing node selects the adjacent node by several factors like residual energy, neighboring nodes, distance to the base station, stability values. The number of times the node has been a leader node determines its stability value. To ensure the uniform energy drainage amongst the nodes, the role of the leader rotates among all the powerful nodes of the sub-network. The neighboring leader node is selected using a Selection parameter S (l):

$$S(l) = \min_{s,i \in n} d(s,i) + \frac{E_{residual}}{E_{total}} + \frac{K}{1 - K\left(r\left(mod\left(1/K\right)\right)\right)} + T(n) \quad (1)$$

In (1), s is the sink node, T(n) is the number of neighboring nodes, K is the number of rounds that have already been done, and r is the current round. The third factor in the above equation ensures that if a leader node has been elected in the last 1/k round, it will not be selected in this round. After the selection of a neighboring node, the node maintains a routing table, which contains the id of the neighboring node selected. The process continues till the route to the base station is not formed.

The working of EE-SIR approach is as follows:

1) The sensing node that detects an event sends a control packet to all the neighboring nodes.

2) A control packet describing all the four entries i.e. distance to the base station, residual energy, no. of times the node has been selected earlier, no. of neighboring nodes is sent back to the sensing node.

3) Depending on the value of the parameter calculated from the four factors, sensing node selects the parent node to forward and route the sensed data.

4) The selected node is sent the sensed data, sensing node id and a timestamp.

5) The selected parent node in turn will send the control packet to all the neighboring nodes to select its parent, and the process is repeated till the base station is reached

IV. SYSTEM MODEL

Before proceeding to discuss the performance characteristics of the stated approach, a discussion is laid on the underlying system models (viz. Energy Consumption, Cost model and Sensing coverage Mode) which have been used to analyze the same. Correspondingly, the analysis has been presented for 2-level, 3-level, and n-level heterogeneous network. In the case of a 2-level network, two different types of nodes are considered viz. Normal nodes and advanced nodes (having capabilities more than normal nodes). If there is a single advanced node in the network, then it is evident that it will be acting as a base station, and hence, this type of network can be treated as a homogenous network. On the other hand, if the number of advanced nodes is greater that one, then the particular network behaves like a 2-level heterogeneous network. Similarly, for a 3-level network, three types of nodes i.e. normal, advanced and super nodes are considered with their capabilities descending in the same order. If there is only a single sink, and the super node is acting as the sink, then this type of network can be analyzed as the 2-level heterogeneous network. Further, in an n-level network, the generalization for multiple levels of nodes representing various initial energy levels is done.

A. Cost Analysis Model

For the 2-type heterogeneous network, two types of nodes are assumed: Advanced nodes (A) and normal nodes (N). The total number of nodes to be deployed in the network depends on the cost constraint as $Cost_{constraint} \geq n_a C_a + n_n C_n$ where $n_{a,}$ C_a and $n_{n,}$ C_n are the number and cost of advanced nodes and normal nodes respectively.

$$C_{total} = n_a * C_a + n_n * C_n \quad (2)$$

If the number of advanced nodes is one i.e. just the base station and N_n^* is the maximum number of normal sensor nodes that can be deployed in the network with a single advanced node i.e. base station, which can be determined by the following equation:

$$C_{total} = C_a + N_n^* C_n \quad (3)$$

Total number of normal nodes that can be deployed in a 2-level heterogeneous network is:

$$n_n = N_n^* + \frac{C_a}{C_n}(1 - n_a) \quad (4)$$

For the 3-type heterogeneous network, three levels of nodes are deployed: Super Nodes(S), advanced nodes (A) and normal nodes (N). The total number of nodes to be deployed in the network depends on the cost constraint as $Cost_{constraint} \geq n_s C_s + n_a C_a + n_n C_n$ where n_s, n_a and n_n are the numbers of super nodes, advanced nodes, and normal nodes respectively, and C_s, C_a and C_n are the costs of the respective sensor node. Total cost is determined from the following equation:

$$C_{total} = n_s * C_s + n_a * C_a + n_n * C_n \quad (5)$$

If the number of super nodes is one i.e. just the base station, then the total cost is determined by:

$$C_{total} = C_s + N_a^* C_a + N_n^* C_n \quad (6)$$

Where N_a^*, N_n^* is the maximum number of advanced and normal sensor nodes that can be deployed in the network with a single super node i.e. base station, which can be determined from the following equation:

$$N_n^* = n_n + (n_a - 1)\frac{C_a}{C_n} \quad (7)$$

$$N_a^* = 1 + (n_s - 1)\frac{C_s}{C_a} \quad (8)$$

The number of advanced nodes in a 3-level heterogeneous network is determined by:

$$n_a = (1 - n_s)\frac{C_s}{C_a} + N_a^* + (N_n^* - n_n)\frac{C_n}{C_a} \quad (9)$$

The number of normal nodes in a 3-level heterogeneous network is:

$$n_n = N_n^* + (1 - n_a)\frac{C_a}{C_n} \quad (10)$$

For n-type heterogeneous network, where n different energy levels are considered:

$$C_{total} = \sum_{i=1}^{n} n_i * C_i \quad (11)$$

The cost is determined from numerous factors like energy, communication range and sensing range and other parameters which include memory and processing capacity. Under the prescribed cost constraint, if the node at higher end increases, then undeniably the number of nodes at lower end decreases that might have a direct impact on sensing coverage and network lifetime.

B. Energy Analysis Model

In this manuscript, the event detection model is assumed where a temporal correlation of events is considered. Furthermore, temporal correlation of real signals is one of the many statistical features which is efficaciously practiced for event detection in WSN applications. In addition to that, the communication model assumed is single-hop communication model in which every node is directly communicating with the base station. Energy consumption is cumulative of data processing, sensing, communication, amplifier energy that is related to the distance from the sink, and the path-loss exponent is assumed to be 2.The energy model used here in EE-SIR is based upon the goal of minimization of energy, whenever, there is some discrete event field. The model assumes that the main consumption of energy comes from the exchange of data messages, and a fixed amount of energy is used for control messages exchanged. All energy consumption values are collected from source to destination, and all calculations are based on an event field and not based on rounds of simulations as done by other researchers.

For 2-level heterogeneous network, initial energy is given by:

$$E_{total} = n_n * E_0 + n_a E_0 * (1 + \alpha) \quad (12)$$

Effectively, this heterogeneous network has 'αm' nodes amounting to 'αm' times the additional energy in the network.

For 3-Level heterogeneous network, the total energy of the netw network is:

$$E_{total} = n_n * E_0 + n_a * E_0 * (1 + \alpha) + n_s * E_o(1 + \beta) \quad (13)$$

In multilevel heterogeneous WSN, the random energy l of sensor nodes is determined from a close-set *[E, E*(1-α_{max})]*, where '*E*' is the lower bound and 'α_{max}' ascertain the value of the maximum energy. Initially, the node 's_i' is equipped with an initial energy of '$E_0 * (1 + \alpha_i)$', which is 'α_i' times more energy than 'E_0'. Hence, the total network energy is given by:

$$E_{total} = \sum_{i=1}^{n} E_0 * (1 + \alpha_i) \quad (14)$$

$$E_{total} = E_0 * \sum_{i=1}^{n} (1 + \alpha_i) \quad (15)$$

C. Sensing Range Model

For our simulation purpose, the sensing model assumed is the deterministic sensing model [19]. In this model, every sensor is involved in event detection. The event is detected by a node based on the condition that if the received signal is greater than the detection threshold. The event detection is based on the received strength independent of environmental factors and the node specifications. Sensing coverage is calculated as the aggregate information gathered from all the alive nodes in the network. Hence, sensing coverage of the network is defined as the sum of sensing coverage of all the nodes.

$$S_{sum} = \sum_{i=1}^{n} \pi R_i^2 \quad (16)$$

Sensing coverage can be defined as the total information that can be extracted from all the sensor nodes deployed in the network. Sensing coverage covers both sensing coverage area and sensing coverage degree. Sensing coverage area infers the area of the network that is monitored by at least one functioning sensor. Coverage degree represents the overlapping coverage among the neighboring sensor nodes i.e. the average number of sensors that cover any network area.

$$S_{sum} = S_d S_a \quad (17)$$

For 2-level heterogeneous network, under prescribed cost constraint, total sensing coverage can be expressed as:

$$\begin{aligned} S_{sum} &= n_a \pi R_a^2 + n_n \pi R_n^2 \\ &= \pi(n_a R_a^2 + (N_n^* - \frac{C_a}{C_n}(n_a - 1) R_n^2 \\ &= \pi\left(-\left(R_n^2 \frac{C_a}{C_n} - R_a^2\right) n_a + \left(\frac{C_a}{C_n} + N_n^*\right) R_n^2\right) \\ &= \pi(\gamma_a n_a + \gamma_b) \end{aligned} \quad (18)$$

Where R_n and R_a are the sensing range of advanced nodes and normal nodes.

For the 3-level heterogeneous network, the case is presented for three types of nodes viz. super nodes, advanced nodes and normal nodes having ranges R_s, R_a, and R_n respectively.

$$\begin{aligned} S_{sum} &= n_s \pi R_s^2 + n_a \pi R_a^2 + n_n \pi R_n^2 \\ &= \pi\left[n_s R_s^2 + \left[(1 - n_s)\frac{c_s}{c_a} + N_a^* + (N_n^* - n_n)\frac{c_n}{c_a}\right] R_a^2 + n_n R_n^2\right] \\ &= \pi\left[n_s R_s^2 + R_a^2 \frac{c_s}{c_a} - n_s \frac{c_s}{c_a} R_a^2 + N_a^* R_a^2 + (N_n^* - n_n)\frac{c_n}{c_a} R_a^2 + \right] \\ &= \pi(\gamma_a n_s + \gamma_b + \gamma_c) \end{aligned} \quad (19)$$

Where γ_a , γ_b , γ_c are the factors determined from the above equation in order to simplify the equation.

Similarly for n-level heterogeneous network, the equation can be generalized as follows:

$$S_{sum} = \sum_{i=1}^{n} \pi n_i R_i^{\,2} \tag{20}$$

$$S_{sum} = \pi(\gamma_n n_n + \sum_{i=1}^{n-1} \gamma_i) \tag{21}$$

V. NUMERICAL MODEL FOR OPTIMAL NUMBER OF HIGH-END NODES

A. Impact of heterogeneity on network lifetime:

All the three types of nodes are using the deterministic sensing model proposed by Ming *et* al. [22] but the sensing range of super nodes are higher than the advanced nodes. In this model, an event is detected if the received signal strength is greater than the sensing threshold set for event detection. For communication model, first order radio model as proposed by Wendi *et* al. [23] is used. The energy consumption of a single hop network can be calculated from two factors: one is independent of node range from sink and second factor that is entirely dependent on: can be calculated from Data Processing energy that is independent of the distance from the sink and the amplifier energy that is dependent on the distance from sink.

$$E_{total} = E_1 + E_2 d^2 \tag{22}$$

2-level Network

$$T = \frac{\pi E_n n_a}{\pi E_1 n_a + S_a E_2} \tag{23}$$

3-level Network

$$T = \pi\left(\frac{E_a n_s}{(\pi E_1 n_s + S_a E_2)} + \frac{E_n n_a}{(\pi E_1 n_a + S_a E_2)}\right) \tag{24}$$

n-level Network

$$T = \pi \sum_{i=1}^{n} \frac{E_i n_{i-1}}{(\pi E_1 n_{i-1} + S_a E_2)} \tag{25}$$

B. Impact of Heterogeneity on Sensing Coverage:

Sensing coverage can be evaluated as the sensing coverage of all the alive nodes in the network. Hence, the total sensing coverage information can be calculated as:

$$S = \sum_{t=0}^{T} S_{sum}(n_a, t) \tag{26}$$

Where T is the network lifetime and S_{sum} is the total sensing coverage information at any point t, which indicates the sensing and the data gathering from all other sensors. The above equation is based on the assumption that each sensor is collecting the information from its sensing coverage and forwards the data to the base station.

Total sensing coverage can be expressed as:

$$S = \pi(\gamma_a n_a + \gamma_b) \frac{\pi E_n n_a}{\pi E_1 n_a + S_a E_2} \tag{27}$$

To find out the optimal number of heterogeneous nodes that will maximize the sensing coverage, the equation will be differentiated on zero, Hence:

$$n_{a-opt} = \frac{1}{\sqrt{\gamma_a}} \sqrt{\left(\frac{S_a E_2}{\pi E_1}\right)^2 + \gamma_b \frac{S_a E_2}{\pi E_1}} - \frac{S_a E_2}{\pi E_1} \tag{28}$$

3-Level network

Sensing range can be defined as the total of all the sensing coverage of super nodes till the first node depletes its energy:

$$S = \sum_{t=0}^{T} S_{sum}(n_s, t) \tag{29}$$

Similarly as above, the optimal number of heterogeneous nodes for 3-level network can be evaluated by setting the derivative on n_s as zero:

$$n_{s-opt} = \pi(\gamma_a n_s + \gamma_b + \gamma_c)\pi\left(\frac{E_a n_s}{(\pi E_1 n_s + S_a E_2)} + \frac{E_n n_a}{(\pi E_1 n_a + S_a E_2)}\right) \tag{30}$$

N-level network

The sensing range here will be the total of the highest level of energy node. If that is represented as n_i, then the sensing range of the network is:

$$S = \sum_{t=0}^{T} S_{sum}(n_i, t) \tag{31}$$

Similarly as above, the optimal number of highest level of nodes is expressed as:

$$n_{i-opt} = \pi \sum_{i=1}^{n} \frac{E_i n_{i-1}}{(\pi E_1 n_{i-1} + S_a E_2)}\, \pi(\gamma_n n_n + \sum_{i=1}^{n-1} \gamma_i) \tag{32}$$

VI. METHODS AND DISCUSSION

The sensors are built using the design software Proteus created by Lab Center Electronics, which provides simulation for electronic circuits and control the schematic capture as well as PCB design. Few necessary presumptions of our routing protocol as given below:

a) All the available sensor nodes within the network have connectivity to communicate with each other or the base station (BS).

b) Random deployment of sensor node

c) All sensors dissipate their energy resource at the same rate at the time of iteration.

d) The primary factor of the total network lifetime is defined as the time span from the deployment to the instant when the first sensor node expires or when the entire sensor nodes perish.

e) The energy dissipation of sensing data and the energy dissipation for clustering are having negligible values and hence neglected.

f) One iteration round is defined as the time span when the Base Station collects information from all the sensor nodes and cluster head communicating the aggregation of data once. Each sensor node delivers only one sensed data with the same packet size that is defined for the network.

g) The sensor nodes within the network receive the data and combine one or more packets to produce a same-size resultant packet, and the number of data that needs to be sent by radio is reduced, because it is having many correlations between the data sensed by the different sensor nodes.

h) The sensor node energy dissipation of fusing one-bit data is a constant value.

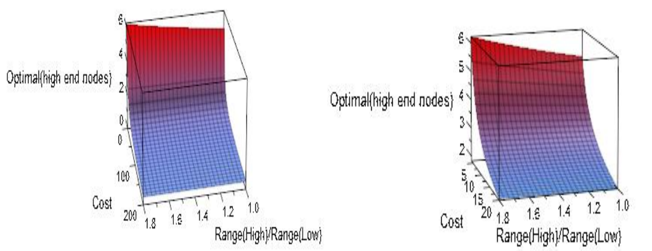
Based on the above-identified problem and scope of work, following scenarios needs to be explored for arriving at a conclusive agreement or disagreement for identifying areas of improvement over the design of the new iterative algorithm.

Homogeneous and heterogeneous are to be simulated. Heterogeneous are defined at three levels: 2-level, 3-level, and n-level.

TABLE I. SIMULATION PARAMETERS

Parameters	Values
N∗ L	500
E2/E1	0 .2
R_L	0.1
R_H	0.2
Sa	3π
C_H/C_L	10

Impact of cost and Range ratio on Heterogeneity

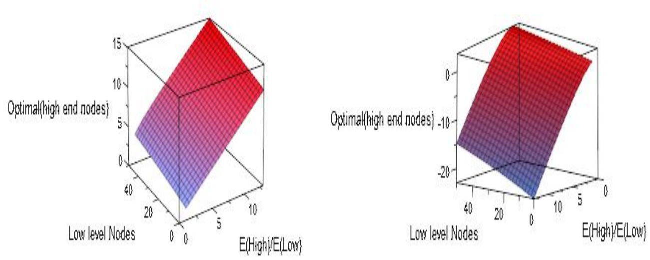


2-level Network 3-level Network

Fig. 1. Impact of Cost and Sensing Range on Optimal number of High nodes

The predictor variables i.e. cost(high nodes)/cost(low nodes) and range(high nodes)/range(low nodes) are displayed on x and y scales and response(z) variable i.e. the optimal number of high end nodes is represented using surface graph. Fig. 1 provides a clearer picture of the optimal number of high end nodes based on the cost ratio and range ratio of 2-level network. As seen in the graph as cost approaches to range ratio, number of optimal high end nodes increases steeply which in a way means that as cost decreases and range increases, it is viable to put more number of high end nodes in order to enhance the network lifetime. On the contrary, if the range ratio is not closer to cost ratio, it will not have a much impact on the heterogeneous deployment. Hence, it can be safely inferred that cost and sensing ratio affect the heterogeneous deployment only when the ratios are closer to each other.

Impact of Low cost nodes and Energy ratio on Heterogeneity



2-level Network 3-level Network

Fig. 2. Impact of Low Cost Nodes and Energy Ratio on Optimal Number

Fig.2 shows the impact of number of low level nodes and energy ratio of advanced and super nodes with respect to the optimal number of advanced nodes. In 2-level heterogeneous network, the optimal number of advanced nodes varies almost linearly with the parameters mentioned. However, in 3-level heterogeneous network, the impact of the parameters is more than the 2-level network. Thus, more number of advanced nodes need to be deployed in a 2-level network as compared to the 3-level network.

Impact of Heterogeneity on Total Energy Spent: Total energy spent is considered to be the sum of energy consumed in sending and energy spent in receiving the packets. An increment of power consumption with the higher number of sources is depicted in the graph. Fig. 3 depicts that maximum energy consumption for homogenous is 61 Joules, which is almost double to 39.9 Joules consumed by the n-level heterogeneous network. 2-level and 3-level stay almost in the range of 50-55 Joules.

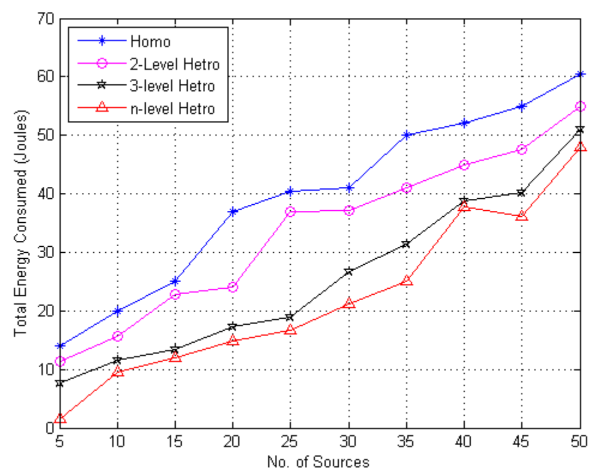


Fig. 3. Total Energy Spent

VII. CONCLUSION AND FUTURE SCOPE

In this manuscript, cost analysis, sensing range, and energy analysis model for different heterogeneity levels in WSNs is proposed. Along with, a numerical model is presented that can provide an optimal number of high-end nodes to optimize network lifetime and sensing coverage simultaneously. This is an original work in the area of understanding the effect of energy heterogeneity and different heterogeneity levels in WSN. To the best of our knowledge, cost analysis for various levels has never been done in the literature. Extending the work to different levels of heterogeneity is the originality of our approach. The proposed approach is a novel approach to find out a mathematical model to calculate optimality for different levels of heterogeneous WSN. Discussed approach is well suited for resource-constrained networks operating with dynamic event detection. As future work, authors intend to explore more of heterogeneity parameters like computation, node capacities. The analysis models need to be applied for proactive strategies to verify the correctness of the other routing approach as well.

REFERENCES

[1] Gupta, P., & Kumar, P. R. (2000). The capacity of wireless networks. Information Theory, IEEE Transactions on, 46(2), 388-404.

[2] Intel heterogeneous sensor networks research group.

[3] Du, Xiaojiang, and Fengjing Lin. "Improving routing in sensor networks with heterogeneous sensor nodes." In Vehicular Technology Conference, 2005. VTC 2005-Spring. 2005 IEEE 61st, vol. 4, pp. 2528-2532. IEEE, 2005.

[4] Xu, Kaixin, Xiaoyan Hong, and Mario Gerla. "An ad hoc network with mobile backbones." In Communications, 2002. ICC 2002. IEEE International Conference on, vol. 5, pp. 3138-3143. IEEE, 2002.

[5] Mhatre, Vivek, and Catherine Rosenberg. "Homogeneous vs heterogeneous clustered sensor networks: a comparative study." In Communications, 2004 IEEE International Conference on, vol. 6, pp. 3646-3651. IEEE, 2004.

[6] Yu, Liyang, Neng Wang, Wei Zhang, and Chunlei Zheng. "Deploying a heterogeneous wireless sensor network." In Wireless Communications, Networking and Mobile Computing, 2007. WiCom 2007. International Conference on, pp. 2588-2591. IEEE, 2007.

[7] Mhatre, Vivek, and Catherine Rosenberg. "Design guidelines for wireless sensor networks: communication, clustering and aggregation." Ad hoc networks2, no. 1 (2004): 45-63.

[8] Kour, Harneet, and Ajay K. Sharma. "Hybrid energy efficient distributed protocol for heterogeneous wireless sensor network." International Journal of Computer Applications 4, no. 6 (2010): 1-5.

[9] Chakrabarty, Krishnendu, S. Sitharama Iyengar, Hairong Qi, and Eungchun Cho. "Grid coverage for surveillance and target location in distributed sensor networks." Computers, IEEE Transactions on 51, no. 12 (2002): 1448-1453.

[10] Lee, Jae-Joon, Bhaskar Krishnamachari, and CC Jay Kuo. "Impact of heterogeneous deployment on lifetime sensing coverage in sensor networks." In Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on, pp. 367-376. IEEE, 2004.

[11] Wendi Rabiner Heinzelman, A. Chandrakasan and H. Balakrishnan. 2000. Energy-Efficient Communication Protocol for Wireless Microsensor Networks. Proceedings of the 33rd Hawaii International Conference on System Sciences (HICSS '00): 1-10.

[12] Duarte-Melo, Enrique J., and Mingyan Liu. "Analysis of energy consumption and lifetime of heterogeneous wireless sensor networks." In Global Telecommunications Conference, 2002. GLOBECOM'02. IEEE, vol. 1, pp. 21-25. IEEE, 2002.

[13] Huang, Chi-Fu, and Yu-Chee Tseng. "The coverage problem in a wireless sensor network." Mobile Networks and Applications 10, no. 4 (2005): 519-528.

[14] Du, Xiaojiang, and Fengjing Lin. "Maintaining differentiated coverage in heterogeneous sensor networks." EURASIP Journal on Wireless Communications and Networking 2005, no. 4 (2005): 565-572.

[15] Wang, Xiaorui, Guoliang Xing, Yuanfang Zhang, Chenyang Lu, Robert Pless, and Christopher Gill. "Integrated coverage and connectivity configuration in wireless sensor networks." In Proceedings of the 1st international conference on Embedded networked sensor systems, pp. 28-39. ACM, 2003.

[16] Howard, Andrew, Maja J. Matarić, and Gaurav S. Sukhatme. "Mobile sensor network deployment using potential fields: A distributed, scalable solution to the area coverage problem." In Distributed Autonomous Robotic Systems 5, pp. 299-308. Springer Japan, 2002.

[17] Lee, Jae-Joon, Bhaskar Krishnamachari, and CC Jay Kuo. "Impact of heterogeneous deployment on lifetime sensing coverage in sensor networks." In Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004. 2004 First Annual IEEE Communications Society Conference on, pp. 367-376. IEEE, 2004.

[18] Sharma, Nonita, and Ajay K. Sharma. "Comparative Analysis of SIR and SID in Wireless Sensor Networks Using Temporal Correlation." network 1, no. 2: 3-4.

[19] Jellali, Z. ; Atallah, L.N. ; Cherif, S., "A study of deterministic sensors placement for sparse events detection in WSN based on Compressed Sensing," Proceedings of International conference on Communication and Networking, COMNET 2014,pp. 1-5, March 2014.

Case Based Reasoning: Case Representation Methodologies

Shaker H. El-Sappagh
Faculty of Computes and Information,
Minia University, Egypt

Mohammed Elmogy
Faculty of Computers and Information,
Mansoura University, Egypt

***Abstract*—Case Based Reasoning (CBR) is an important technique in artificial intelligence, which has been applied to various kinds of problems in a wide range of domains. Selecting case representation formalism is critical for the proper operation of the overall CBR system. In this paper, we survey and evaluate all of the existing case representation methodologies. Moreover, the case retrieval and future challenges for effective CBR are explained. Case representation methods are grouped in to knowledge-intensive approaches and traditional approaches. The first group overweight the second one. The first methods depend on ontology and enhance all CBR processes including case representation, retrieval, storage, and adaptation. By using a proposed set of qualitative metrics, the existing methods based on ontology for case representation are studied and evaluated in details. All these systems have limitations. No approach exceeds 53% of the specified metrics. The results of the survey explain the current limitations of CBR systems. It shows that ontology usage in case representation needs improvements to achieve semantic representation and semantic retrieval in CBR system.**

Keywords—Case based reasoning; Ontological case representation; Case retrieval; Clinical decision support system; Knowledge management

I. Introduction

Clinical Decision Support System (CDSS) that bear similarities with human reasoning and explanation have benefits. They are often easily accepted by physicians in the medical domain [1-4]. Many of the early AI systems attempted to apply pure Rule-Based Reasoning (RBR) as 'reasoning by logic in AI' for decision support in the medical area.

However, for broad and complex domains where knowledge cannot be represented by rules (i.e. IF-THEN), this pure rule-based system encounters several problems. Because medical knowledge evolves rapidly, updating large rule based systems and proving their consistency is expensive. Knowledge acquisition bottleneck is one of the most critical problems in any knowledge-based system. A risk is that medical rule-based systems become brittle and unreliable. One faulty rule may affect the whole system's performance [5]. Case-Based Reasoning (CBR) is a promising AI method that can be applied as "reasoning by experience in AI" for implementing CDSSs in the medical domain since it learns from experience in order to solve a current situation [6]. Readers interested in CBR applications in healthcare can read these reviews [7, 8, 9].

CBR is based on remembering past experiences and using them to solve current situations that are similar to the ones already solved and stored. CBR is especially suitable when domains are difficult to formalize. In CBR, experiences in the form of cases are used to represent knowledge. A case may be a patient record structured by symptoms, diagnosis, treatment and outcome, and clinicians often reason with cases by referring to and comparing with previous cases. Many other AI and statistical methodologies can be used to implement CDSS. CBR is much better compared to RBR, Artificial Neural Network (ANN) and other statistical and machine learning techniques [10]. For example, ANN is a black box and cannot always explain why they arrived at a particular solution. Moreover, they cannot always guarantee a completely certain solution, arrive at the same solution again with the same input data, or always guarantee the best solution. Aamodt and Plaza [11] provided scheme of the CBR working cycle comprising of four phases RETRIEVE, REUSE, REVISE and RETAIN. These phases depend on the existence of the Knowledge Base (KB) in the form of Case Base. Case representation is a critical success factor in CBR because the reasoning capabilities of CBR depend mainly on the structure and content of cases. Cordier [12] and Finnie [13] added a case base building phase to [11] that required case representation process. Cases can be represented as simple feature vectors, or they can be represented using any AI representational formalism such as frames, objects, predicates, semantic nets, or rules. The choice of particular representational formalism is largely determined by the information to be stored within a case.

There is a lack of consensus within the CBR community to determine case contents and representational formalism. However, two pragmatic measures can be taken into account in deciding both the information to be stored in a case and the appropriate representational formalism: the intended functionality and the ease of acquisition of the information represented in the case [14].

Cases are the basis of any CBR system: a system without cases would not be a case-based system. Yet, a system using only cases and no other explicit knowledge (not even in the similarity measures) is difficult to distinguish from a nearest-neighbour classifier or a database retrieval system. In other words, such a system does not exploit the full power of CBR, resulting usually in poor system performance due to inefficient retrieval based upon case-by-case search of the completely case base. Ontologies play an important role in enhancing the capabilities of CBR systems. They improve case indexing and retrieval, case representation and storage in case base, case adaptation and case retention. They solve the problem of knowledge acquisition bottleneck by allowing the case base to be represented as ontology and allowing discovery of cases from existing domain ontologies [15, 16]. It facilitates the creation of Knowledge Intensive CBR (KI-CBR) systems where cases, in some way or another, are enriched with explicit general domain knowledge [17]. The role of the general domain knowledge is to enable a CBR system to reason with semantic and pragmatic criteria, rather than purely syntactic ones. By making the general domain knowledge explicit, the CBR system is able to reason in a more flexible and contextual manner than if this knowledge is compiled into predefined similarity metrics or feature relevance weights [18, 19]. What is more, ontologies can be used for case representation, which enhance the integration between case base and domain knowledge.

Case based reasoning is applied in different fields ranging from non-medical domains [20] to medical domain [21]. Since 1997 over 130 major companies worldwide were fielding CBR applications [22]. As the paper concentrate on medical field and because of space restrictions, most medical case based reasoning systems have been collected in these surveys [8, 9]. Case representation in medical domain applications differ from others in three points (1) the form of used ontology as medical systems use standard ontologies as ICD [23], (2) the integration of Clinical Practice Guidelines (CPG) in case base ontology, and (3) the usage of soft-CBR because medical data are incomplete and vague, in most cases [24].

Although case representation is the most critical decision in building CBR systems, there is a shortage in surveys for this aspect. Bergmann et al. [25] has provided a survey for traditional case representation methods. This survey is old and did not discuss the semantic aspect in case representation using ontologies and rules.

This paper reviews all existing case representation formalisms in CBR concentrating on the logical structure of cases in case base. Cases can be physically stored in databases (relational or object oriented), XML files or even flat files. Cases can be represented using traditional methods as feature vector. Moreover, it can be represented in more intelligently enhanced ways using semantic mechanisms as ontologies and rules. The paper concentrates on the role of ontology in CBR named ontological case based reasoning. The databases ScienceDirect, IEEEXplore, and Springer have been used in our research. Moreover, the paper have done an exhaustive literature search in some proceedings of yearly CBR conferences as the European Conference “Advances in Case-Based Reasoning” and the International Conference on Case-Based Reasoning. Because there is very little researches in the case representation methodologies, our search strategy concentrated on collecting case based reasoning systems between 1990 and 2012 and studying their case representation strategies, evaluating, grouping and comparing them. Screening has based on titles and abstracts containing combinations of keywords “case representation strategy, methodology, model”, “case base organization”, “case based reasoning”, “semantic CBR” + “ontological case base”, “ontology based CBR”, “case base ontology” + “case base storage model, ontology”. The paper will be organized in 5 sections as follows. Section 2 discusses CBR definitions, models and importance of case representation. Section 3 discusses traditional case representation methods. Section 4 discusses semantic case representation methods. Comparison between semantic case representation methods is discussed in section 5. Section 6 discusses semantic retrieval methods. Section 7 provides the CBR challenges. Finally, the conclusion is discussed in section 8.

II. Case Based Reasoning

CBR is a problem solving methodology that aims at reusing previously solved and memorized problem situations, called cases. A case is a concrete problem solving experience. One of the main assets of CBR is its eagerness to learn. Learning in CBR can be as simple as memorizing a new case or can entail refining the memory organization or meta-learning schemes. CBR has developed from these premises, and been found suitable to solve any type of problem, but preferably experimental sciences problems, where cases are readily available in the form of patients, living beings, or natural phenomena. Kolodner [26] defines case as "A case is a piece of knowledge in a particular context representing an experience that teaches an essential lesson to reach the goal of the reasoner." Cases may be kept as concrete experiences, or a set of similar cases may form a generalized case. There are many models for CBR lifecycle such as Hunt’s [27], Allen’s [28], Kolodner and Leake’s [6], and Aamodt [11]. All have nearly similar phases. According to Aamodt [11], CBR working cycle can be described best in terms of four processing stages (R4 model) as shown in Fig. 1 [11]: (a) RETRIEVE the most similar case(s). (b) REUSE the case(s) to attempt to solve the current problem. (c) REVISE the proposed solution if necessary. (d) RETAIN the new solution as a part of a new case.

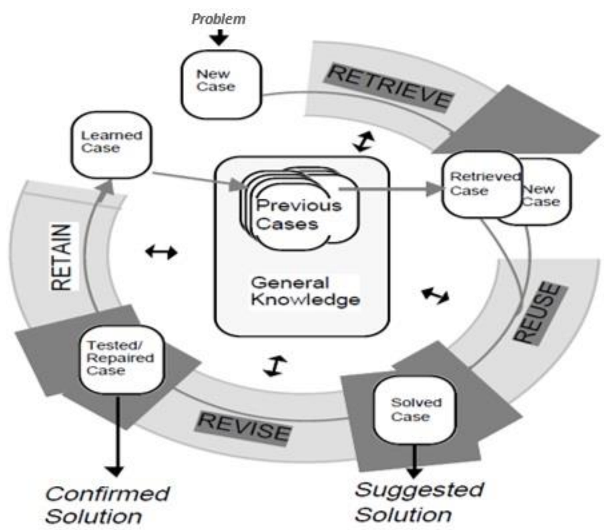


Fig. 1. Case-based reasoning life cycle

Each of these steps can be decomposed in sub-steps as shown in [11]. In fact, the representation of cases is important for CBR because CBR is heavily dependent on the structure and content of its collection of cases. The previous models do not contain a separate phase for case representation and assume that the case base is ready for the first process (case retrieval). Gavin and Zhaohao in [13] propose a model that considers case base building as the first step. In the R5 model, repartition, retrieve, reuse, revise and retain are the main process steps in the CBR. While the other process steps are the same as those in the R4 model, repartition is used to build the case base.

Cordier proposed a model [12] that composed of five stages: (1) preparation, (2) memory retrieval, (3) reuse (adaptation), (4) revision, and (5) memorization (learning). This model also asserts a case base building step, preparation, where a set of cases is capitalized in the knowledge base (base case). This model is used by Maalel et al. [29] to build CBR system to manage railroad accidents. Finally, building the case base is critical to the success of the CBR system. A case representation methodology must be selected which determine the content and structure of the case base. Case structure consisting of the following five parts: (1) a problem description (e.g. symptoms); (2) a solution (e.g. a diagnosis or a therapy) and, sometimes, means of deriving it; (3) outcome (e.g. the result of applying the solution); (4) explanations of results, if necessary and available, of why it might not have worked as well as expected; (5) lessons that can be learned from the experience. Beside the case base, CBR makes use of other types of knowledge as the vocabulary, similarity measures, and adaptation knowledge [30].

Case representation in CBR makes use of familiar knowledge representation formalisms from AI to represent the experience contained in the cases for reasoning purposes. The two main categories of case representation are traditional methods that are discussed in section 3 and semantic methods that are discussed in section 4. Case representation and case retrieval are the main and most important steps in CBR [31]. The efficiency of case retrieval algorithm is determined and affected directly by the case representation method used [32]. As a result, it is more logical to introduce case retrieval methods after surveying the representation methods to link them together. Case retrieval methods will be discussed in section 6.

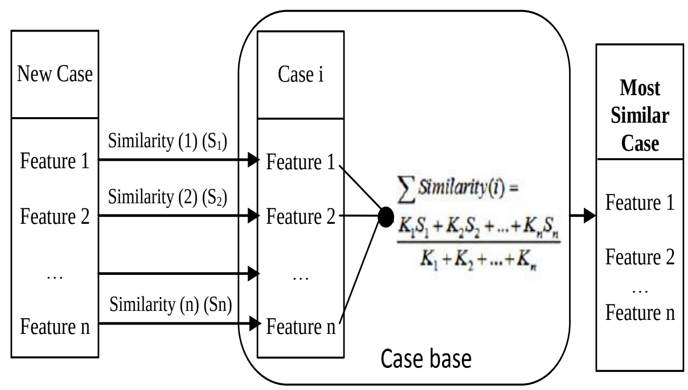
III. CBR Traditional Case Representation Methods

A CBR system should be organized with some basic elements: the knowledge representation, to depict the cases, and the similarity measure to define how much a case is similar to another one. In a CBR system, the knowledge is in the case base. One case contains the knowledge of an experimented episode and the context in which the knowledge can be applied. When considering case representation, this problem needs to be studied from two points of view: first, the conceptual models that are used to design and represent cases, and second the means of implementing cases in the computer [25]. Case representation in CBR contains three issues: defining which attributes describe a case, defining the structure for describing the case content and how to organize the cases in the case base [33]. Case base can take any format to physically store cases. The case base can be a relational or object oriented database, XML files or plaintext files. Cases must be indexed so that the retrieval step can use a structure to have access to cases. Usually there is a separation between the case storage from the indexing structure because indexes can be built without knowing how and where the cases are stored.

Moreover, different indexes can be defined upon the same set of cases to allow the evaluation of different indexing techniques. This work concentrates on the logical structure of cases that can be stored in any format in the case base. A large variety of representation formalisms have been proposed such as feature vector (or propositional) cases, structured (or relational) cases, and textual (or semi-structured) cases.

A. Feature vector representation

This is the simplest form of case representation where each case is represented as a set of features describing the problem (attribute-value) and associated solution (see Fig. 2 [34]). All cases have the same kind and the same number of features [33]. The similarity is straightforward since each feature is compared with its corresponding one. There are no relationships or constraints between features. Moreover, exact similarity is performed only (i.e. no semantic similarities are possible since there is no domain knowledge) [35].We must first have a sufficiently complete conceptual model about the problem, then compare problem features between the new case and past cases in the case-base, and finally get the most similar past case for reuse by similarity measuring. If the conceptual model is known incompletely or ambiguously, then the subsequent steps cannot continue [34].



K1, K2, K_n are features Weights

Fig. 2. the retrieval mechanism of CBR

The PROTOS system [36] uses a feature vector approach for domains with weak or intractable theories. A category is extensionally represented as a collection of cases called exemplars (see Fig. 3 [11]).

A new case is classified into a category if a match can be found between an exemplar and the new case. This matching process is knowledge intensive and tries to build an explanation that connects the features of the new case with an exemplar. Since each explanation is a path constructed inside a semantic net, retrieval is the process of explaining the (similarity) relation between a new case and an exemplar. Unlike most early CBR systems that use feature vector representations, PROTOS already uses a knowledge-intensive similarity measure. Features can be organized in a hierarchy where generic features are in top of tree/graph and specific features in the bottom [37].

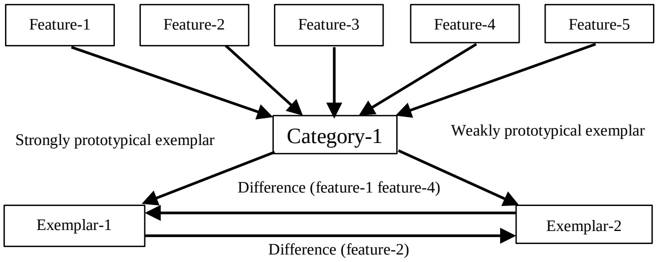


Fig. 3. The structure of categories, features and exemplars

B. *Frame-based representation*

Frames provide a natural way for the structured and concise representation of knowledge. In a single entity, a frame combines all necessary knowledge about a particular object or concept. A frame provides a means of organizing knowledge in slots to describe various attributes and characteristics of the object. Each frame has its own name and a set of attributes, or slots, associated with it [38]. In CBR terminology, a frame can represent a case and each frame slot is a case feature. A slot may contain a primitive value or a pointer to another frame. In the same way, in CBR, features can be primitive (simple) or complex (compound). Cases represented as frames can have semantic relationships because a case may have a feature (slot or attribute) whose value is pointer to another frame. Moreover, as inheritance is an essential feature of frame, a hierarchy of cases connected by IS_A and PART_OF relationships can be formed.

This case hierarchy enhances the semantic retrieval and indexing of cases and adaptation operations. Frame based representations have been (partially) formalized by description logics [39]. The notion of "cases as terms" [40] argues that viewing structured cases as terms in feature logics (a particular brand of description logics) helps in better understand several aspects of case-based reasoning. Domain knowledge can be integrated using a sort hierarchy and the issue of composite cases (cases that group together other objects or sub-cases) is understood by the fact that a sub-term is also a term. Finally, the notion of similarity between two cases is linked to the concepts of subsumption and anti-unification of terms.

C. *Object Oriented (OO) representation*

The feature vector model is not suitable for building a complex case data structure. In this situation, OO representation works much better. The OO method needs less memory storage to represent each case. Furthermore, since OO is a natural way of representing IS-A, HAS-A and PART_OF relationships, case representation is easier for users to understand. Cases are represented as collections of objects, each of which is described by a set of attribute-value pairs. The structure of an object is described by an object class [41]. CASUEL [42] is an early example in plain ASCII, but recent languages are XML compatible. Generally, with the object-oriented structures of the cases, the similarity measures follow the "local-global" principle [43].

D. *Textual representation*

Textual case-based reasoning is a CBR where some or all of the knowledge sources are available in textual format. It aims to use these textual knowledge sources in an automated or semi-automated way for supporting problem solving through case comparison [44]. Many techniques for textual case base representation are available. Burke et al. [45] developed FAQ-Finder, a question–answering system. It starts with a standard Information Retrieval (IR) approach based on the vector space model, where cases are compared as term vectors with weights based on a term's frequency in the case versus in the corpus. In addition, FAQ-Finder includes a semantic definition of similarity between words, which is based on the concept hierarchy in WordNet. Wilson [46] investigated cases that required mixed representations including both textual and non-textual features. Another group of projects focused on developing methods to map textually expressed cases into the kinds of structured representations used in CBR systems such as SPIRE [47]. In [48], textual case representations decompose the text that constitutes a case into information entities (IEs). IE represents any basic knowledge item such as attribute-value pair. As a result, a case consists of a unique case descriptor and a set of IEs linked to it. The case base is a network with nodes for the IEs observed in the domain and additional nodes denoting the particular nodes. An IE is a word or a phrase contained in the text that is relevant to determine the reusability of the episode captured in the case. The set of cases that form the case base is organized in the form of a case retrieval net (CRN), which is a directed graph with nodes representing cases and their IEs. These nodes are linked according to their similarity. Hence, knowledge about similarity is encoded into the strength of the links between the nodes in the CRN. Case retrieval is similar to activation propagation in a neural network: the IEs that occur in the current problem are activated and this initial activation is propagated through the case retrieval net according to the similarity-based link strength. A promising and highly ambitious approach, using natural language processing (NLP) to derive a deep, logical representation, has been proposed for the FACIT project [49]. It derives a first-order representation of the case texts. Weber et al. [50] introduced a semi-automated approach to populate case templates from textual documents. This method is based on knowledge engineering, NLP and data mining. Bag-Of-Words (BOW) representation is introduced by Brüninghaus □□Ashley [51, 52]. They applied text classifiers to automate the mapping from texts to structured case representations. Brüninghaus □□Ashley [53] argues that text representation that combines some background knowledge and NLP combined with a nearest neighbour algorithm leads to the best performance. As a result, textual CBR will mostly require textual descriptions of cases to be mapped onto structural representations that facilitate computationally comparing cases [54].

E. *Hierarchical case representation*

The previously discussed approaches typically represent cases at a single level of abstraction. However, Cases can be represented using multiple representations at different levels of abstraction. The basic idea behind these approaches is to represent a case at multiple levels of detail, possibly using multiple vocabularies. When a new problem must be solved, similar cases at appropriate levels of abstraction are retrieved from the case base, and solutions from these cases will be combined, and these solutions may be refined [55]. Watson [56] asserted that as the problem space increases (number of cases features), it becomes statistically less likely that a close matching case will exist. Thus, the CBR system will return a distant solution (see Fig. 4 [56]).A potential solution to this problem is that, where suitable, a large problem is divided into several smaller sub-problems, each of which can be solved separately using CBR (Fig. 5 [57]). The sub-solutions can then be combined to produce an accurate solution to the entire problem [57].

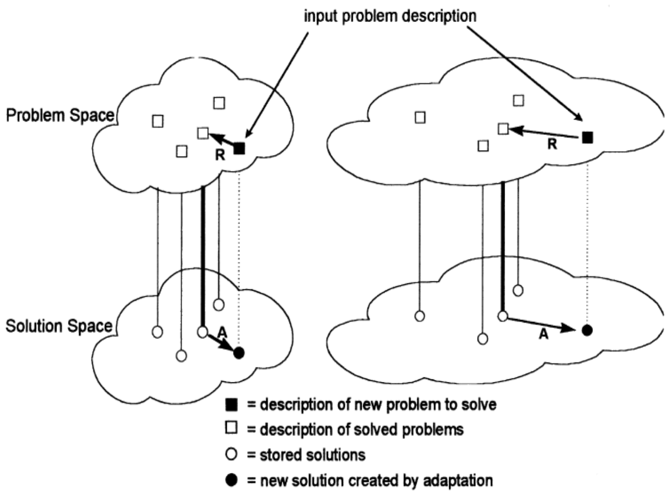


Fig. 4. Small and large problem & solution spaces

The advantage of this approach is that each individual sub-problem is represented by a case-base that is significantly smaller (in terms of problem and solution space size) than if the whole problem were represented by a single case-base. Because each sub-problem space has fewer case features, the theory predicts, that each individual sub-case retrieval distance will be shorter than for the un-decomposed problem. Therefore, the adaptation distance will be shorter and a better sub-solution will be generated. Assuming there are no conflicting constraints, the re-composition of sub-solutions will produce a better solution than would have been obtained by using a single large case-base. One way that has been suggested to reduce constraint problems with solution re-composition is to use contextual information to guide retrieval [58].

F. *Predicate based case representation*

A predicate is a relation among objects, and it consists of a condition part and an action part, IF (condition) and THEN (action). Predicates that have no conditional part are facts. Cases can be represented as a collection of predicates [24]. The advantage of predicate representation is that it uses both rules and facts to represent a case, and it enables a case-base designer to build hybrid systems that are integrated rule/case-based. Although the traditional data models described above are useful to represent and to index cases, in many practical situations when specifying a case, it is often difficult to articulate the feature values precisely. This uncertainty may be caused by incomplete, missing, unquantifiable information, overlapping of the data regions, or user ignorance. Therefore, to make cases more expressive in dealing with such situations, soft computing techniques are introduced [24]. These techniques include fuzzy logic, neural network, rough sets and data mining. These techniques are outside the scope of this paper.

IV. CBR SEMANTIC CASE REPRESENTATION METHODS

The above case representations may be characterized as being knowledge-poor. They do not contain many (or any) structures that describe the relationships or constraints between case features. However, these case representations usually describe relatively simple cases with few indexed features, perhaps in the order of ten to twenty indexed features. In many situations, additional knowledge (background knowledge) is required with the case base to cope with the requirements of an application. In [59], the author integrated two kinds of rules with the case base. The first kind is Completion Rules that infer additional features out of known features of an old case of the query. These rules complete the description of a case. The second type is Adaptation Rules that describe how an old case can be adapted to fit the current query. As in Fig. 6 [59], the general knowledge, in the form of rules, will guide the CBR query and adaptation operation. However, creation of a rulebase, managing its execution in the form of forward or backward chaining, and integrating rules with CBRare challenging.

Knowledge-Intensive CBR (KI-CBR) assumes that cases are enriched and/or coupled with general domain knowledge [17, 60, 33]. In CREEK [18] cases are embedded within a general domain model. It provides a strong coupling between cases and general domain knowledge in that cases are submerged within a general domain model. This model is represented as a densely linked semantic network.

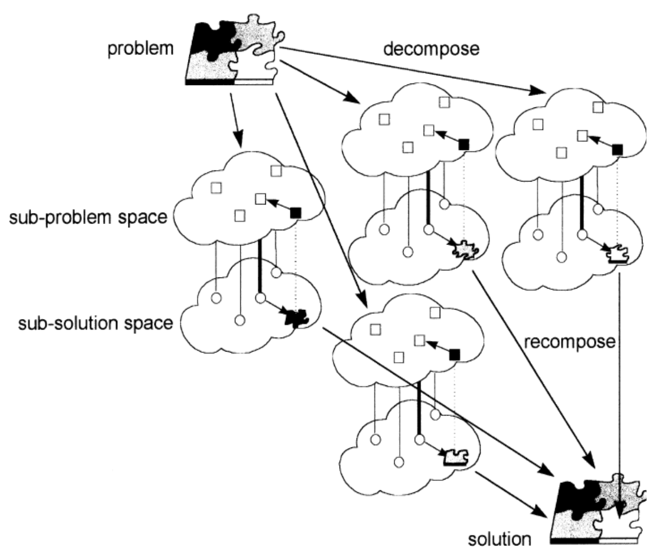


Fig. 5. Problem decomposition and solution re-composition

Concepts are inter-related through multiple relation types, and each concept has many relations to other concepts. The network represents a model of that part of the real world which the system is to reason about, within which model-based reasoning methods are applied. From the view of case-specific knowledge, the knowledge-intensiveness of the cases themselves are also increased, i.e. the cases become more “knowledgeable”, since their features are nodes in this semantic network [61]. Fig. 7 [18] shows the semantic network that integrate cases with the general domain knowledge. It illustrates the three main types of knowledge in CREEK, a top-level ontology of generic, domain-independent concepts, the general domain knowledge, and the set of cases. The retrieva l of relevant cases will based on the semantic and pragmatic criteria, rather than purely syntactic ones. By making the general domain knowledge explicit, the case-based reasoner is able to interpret a current situation in a more flexible and contextual manner than if this knowledge is compiled into predefined similarity metrics or feature relevance weights.

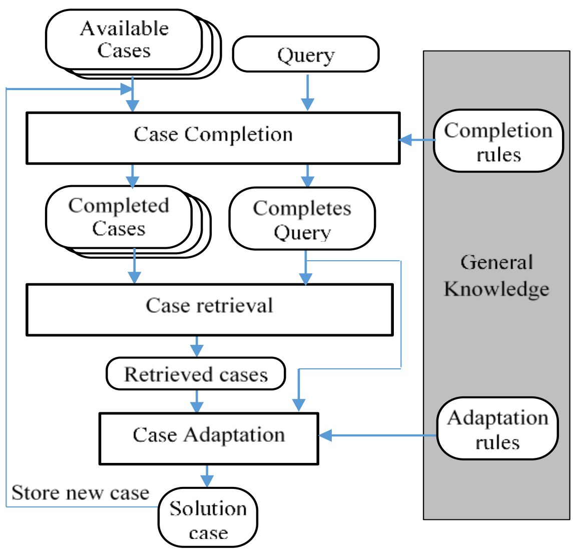


Fig. 6. Architecture for integrating general knowledge in to CBR

Studer et al. [62] defined ontology as "a formal, explicit specification of a shared conceptualization." Ontologies can be useful for designing KI-CBR applications because they allow the knowledge engineer to use knowledge already acquired, conceptualized and implemented in a formal language, reducing considerably the knowledge acquisition bottleneck. It has powerful abilities in knowledge acquisition, representation, and semantic understanding [63]. Moreover, the reuse of ontologies from a library also benefits from their reliability and consistency. Ontologies may help in the creation of complex, multi-relational knowledge structures to support the CBR methods.

In CBR, knowledge is distributed among the four knowledge containers: vocabulary, similarity measures, adaptation knowledge and case base. Ontology plays critical roles in representing all of these knowledge containers. For example, as the vocabulary to describe cases and/or queries, as a knowledge structure where the cases are located, and as the knowledge source to achieve semantic reasoning methods for similarity assessment and case adaptation that are reusable across different domains [64]. Bergmann et al. [65] concluded that ontology-based knowledge management and CBR knowledge management complement each other very well. Most ontology-based systems utilize logic-based deductive inference, while CBR systems provide a search functionality that makes use of similarity measures for ranking results according to their utility with respect to a given query. On the one hand, logic deduction produces only correct and provable results, which are consequences of the ontology and metadata. On the other hand, CBR retrieval suggests results even in the case that no exactly matching answers can be found. As a result, each method solves problems that the other method cannot solve.

A. *Ontology as the CBR's domain vocabulary*

This approach build case base in any traditional methodology as feature vector and store it in relational database, and build ontology for domain knowledge (domain vocabulary).

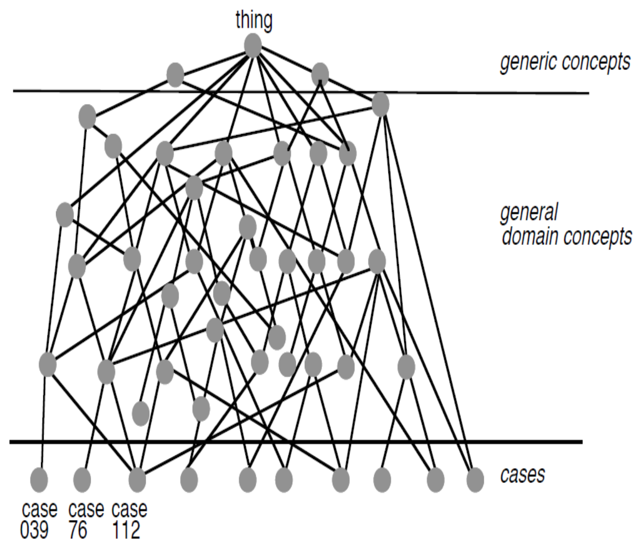


Fig. 7. Integrating cases and general knowledge

The case structure is defined using types from the ontology even if the cases are not stored as individuals in the ontology. There are also simple types like string or numbers that are considered in the traditional way [64].

Regarding the query vocabulary, there are two options to define the queries:

- Using exactly the same vocabulary used in the cases, i.e. the same types used in the case structure definition.
- Using the ontology as the query vocabulary, that allows richer queries and semantic retrieval. The user can express better his requirements if he can use a richer vocabulary to define the query. During the similarity computation, the ontology allows to bridge the gap between the query terminology and the case base terminology.

In this approach, Case base is stored in a SQL database, the retrieval and similarity computation methods are configured as (Nearest Neighbour) NN based on numeric and standard similarity functions, while adaptation is defined as a substitution method that relies on DLs to find suitable substitutes on the domain model. Numerical similarity functions based on ontologies is used where similarity between cases can be divided into two components that are aggregated. The computation of a concept based similarity that depends on the location of the concept in the ontology and the computation of a slot-based similarity that depends on the fillers of the common attributes between the compared objects.

B. Ontologies as case base and domain vocabulary

For better communication between case base and domain vocabulary, Assali et al. [66] created ontology for domain vocabulary and ontology for case base. In [66], it is based on a knowledge base that integrates domain knowledge along with cases in an ontological structure, which enhances its semantic reasoning capacities. Users can describe their cases using instances from the knowledge base. The resulting case base is heterogeneous where cases do not always share the same attributes (dynamic representation of cases). Inspired by Lamontagne and Lapalme [67], COBRA architecture is composed of two main parts (see Fig. 8 [67]): processes and knowledge containers.

- **Processes:** This is the functional part of the system and consists of off-line and on-line processes.
- **Knowledge containers:** As in Richter [19]. COBRA represented case base and domain knowledge in an ontological structure to allow a better communication between the knowledge about the cases and the domain [68].

The domain ontologies are core ontology that contains generic concepts that provide the context and domain ontology that are specializations of other concepts in the core ontology. The case ontology consists of three main parts: a problem description part describing the context of the experience, a failure mode part describing the type of failure, and a cause part describing the different possible causes of this failure, see Fig. 9 [68].

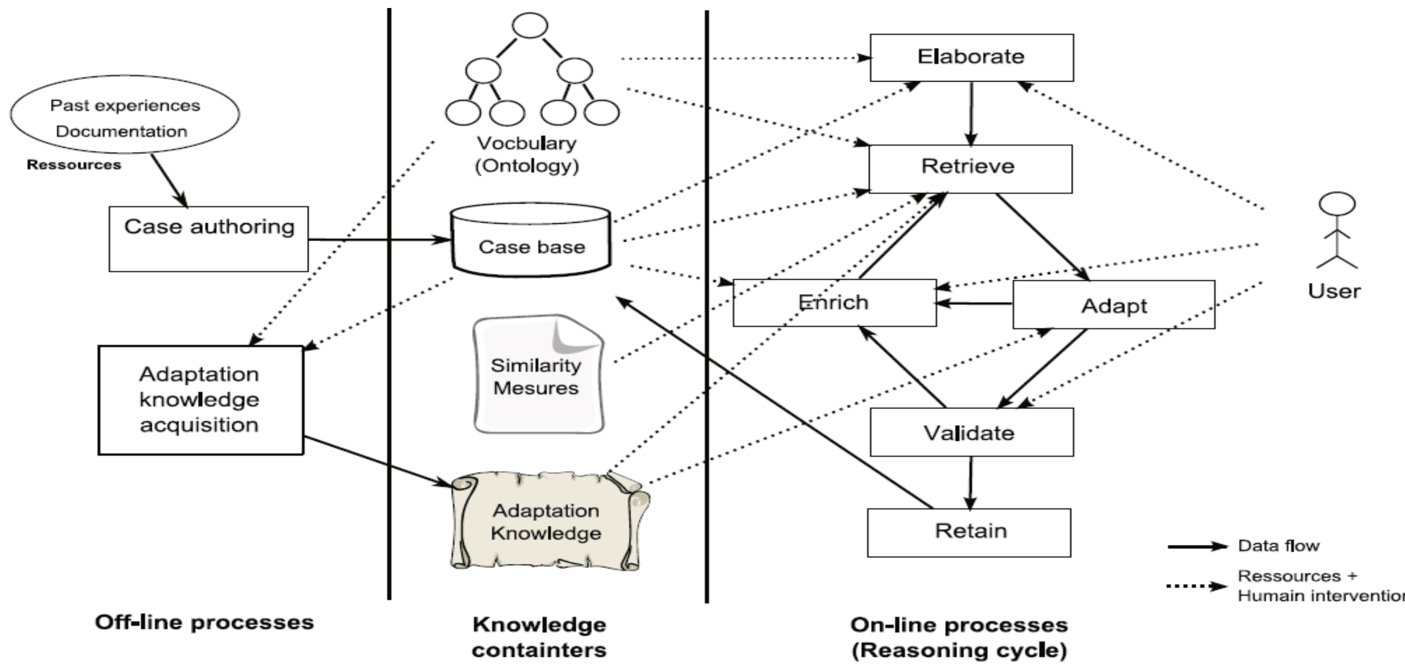


Fig. 8. COBRA architecture

The retrieval is guided by the adaptability [69]; i.e. a case is retrieved if its solution can be reused to construct a solution for the target problem. To determine adaptable cases, given a target problem, each attribute of the query must be compared to its corresponding attribute in each source case. In homogeneous case bases, all cases share the same predefined structure, and thus, corresponding attributes are already identified. On the other side, heterogeneous case bases contain cases that do not share a predefined structure (in terms of attributes), which complicates the retrieval process. The problems of heterogeneity are solved by case alignment [66]. This approach gets similarities or mapping between cases attributes of query and target cases using the support of domain ontology, and using the notions of similarity regions and roles of attributes. The same scenario is followed by Maalel et al. [70] to develop an ontological CBR system for railroad accidents application. Their methodology depended on [71, 72] ontology development methodologies.

To enhance the case retrieval and case adaptation, Maalel et al. [29] created domain ontology from which cases are instantiated in the case base and operational ontology in the form of decision rules to restrict the search space and guide case adaptation (see Fig. 10 [29]). The adaptation rules are not created in a standardized form suitable for ontology such as SWRL. In addition, the process of creating these rules is not straightforward. The COBRA was a domain-dependent model since it created ontologies for specific domain. JCOLIBRI (Cases and Ontology Libraries Integration for Building Reasoning Infrastructures) solved this problem and created a knowledge intensive and domain-independent architecture for CBR [64, 73, 74].

C. Domain independent ontological CBR framework

Díaz-Agudo et al. [75] created a domain independent architecture to help in the design of knowledge intensive CBR systems.

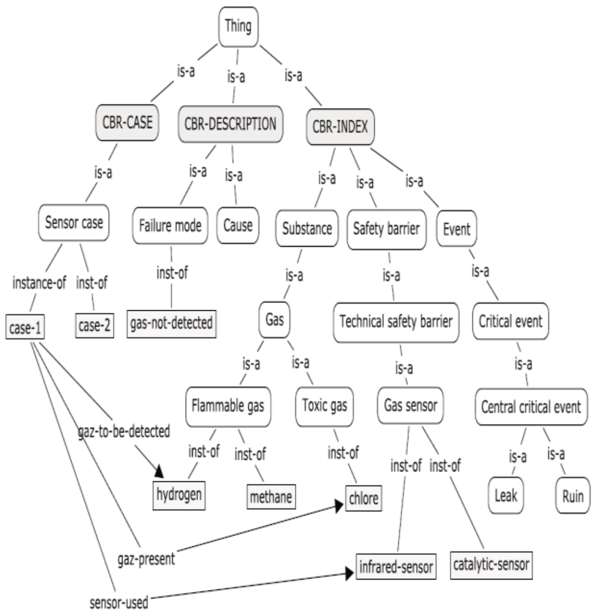


Fig. 9. Part of the case model

This model solves the knowledge-acquisition-bottleneck problem that faced all previous ontological methods. It is based on knowledge acquisition from a library of application-independent ontologies and the use of CBROnto, ontology with the common CBR terminology that guides case representation and allows the description of flexible, generic and reusable CBR Problem Solving Methods (PSM).

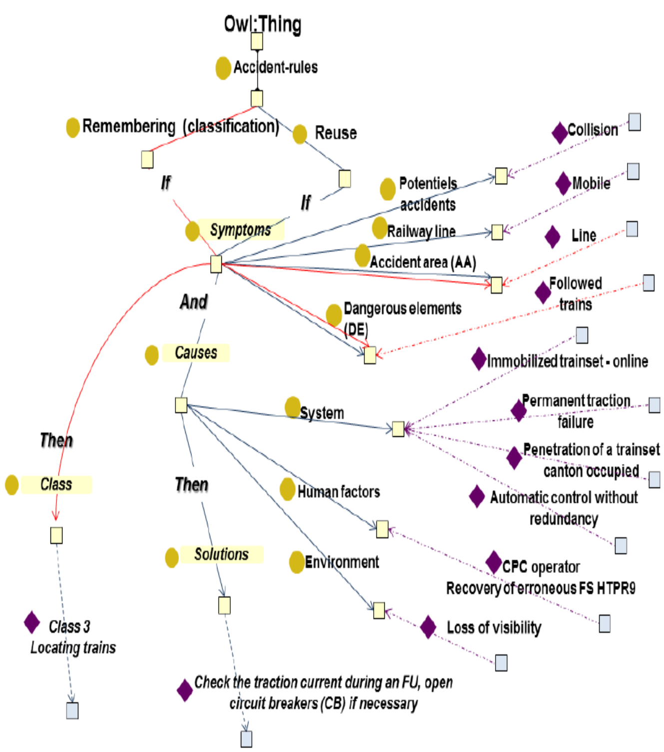


Fig. 10. Part of the operational ontology

PSM describe the reasoning process of a Knowledge Based System in an implementation- and domain-independent manner [76]. As shown in Fig. 11 [76], COLIBRI has two-layered architecture. The lower layer provides with domain specific knowledge while the top layer is used as a bridge between the domain knowledge and the generic PSMs. Ontology Server contains all reusable and formal domain specific ontological knowledge. This way, the specific domain model is interchangeable and the same knowledge could play different roles within different contexts of problem resolution. Moreover, COLIBRI integrates different knowledge sources, range from general domain knowledge to CBROnto knowledge about tasks and methods.

To take advantage of the domain knowledge acquired by reusing ontologies from Ontology Server, the knowledge needed by the CBR methods, or at least part of it, should be expressed in a similar way. CBROnto provides terminology about CBR that captures CBR semantically important terms and provides vocabulary for describing issues involved in the CBR methods. CBROnto includes CBR dependent but domain-independent terms that make possible different types of CBR [77]. These terms are used as the junction between the domain knowledge and the Problem Solving Methods that are defined using CBR terminology but with a domain-independent perspective (Fig. 11). CBROnto aims to unify case specific and general domain knowledge representational needs. All domain terms (concepts and relations) are classified according to the role they play in CBR methods. CBROnto terminology serves as the syntactic and semantic “glue” between the domain terminology and the reusable and generic PSMs.

That mechanism allows the CBR methods to be domain independent because they only refer to the CBROnto terms. CBROnto ontology includes general terminology (Fig. 12 [17]):

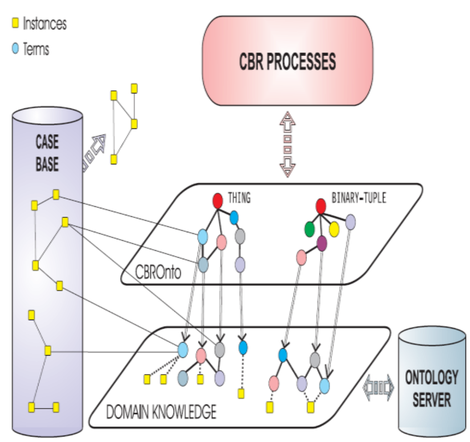


Fig. 11. COLIBRI Architecture

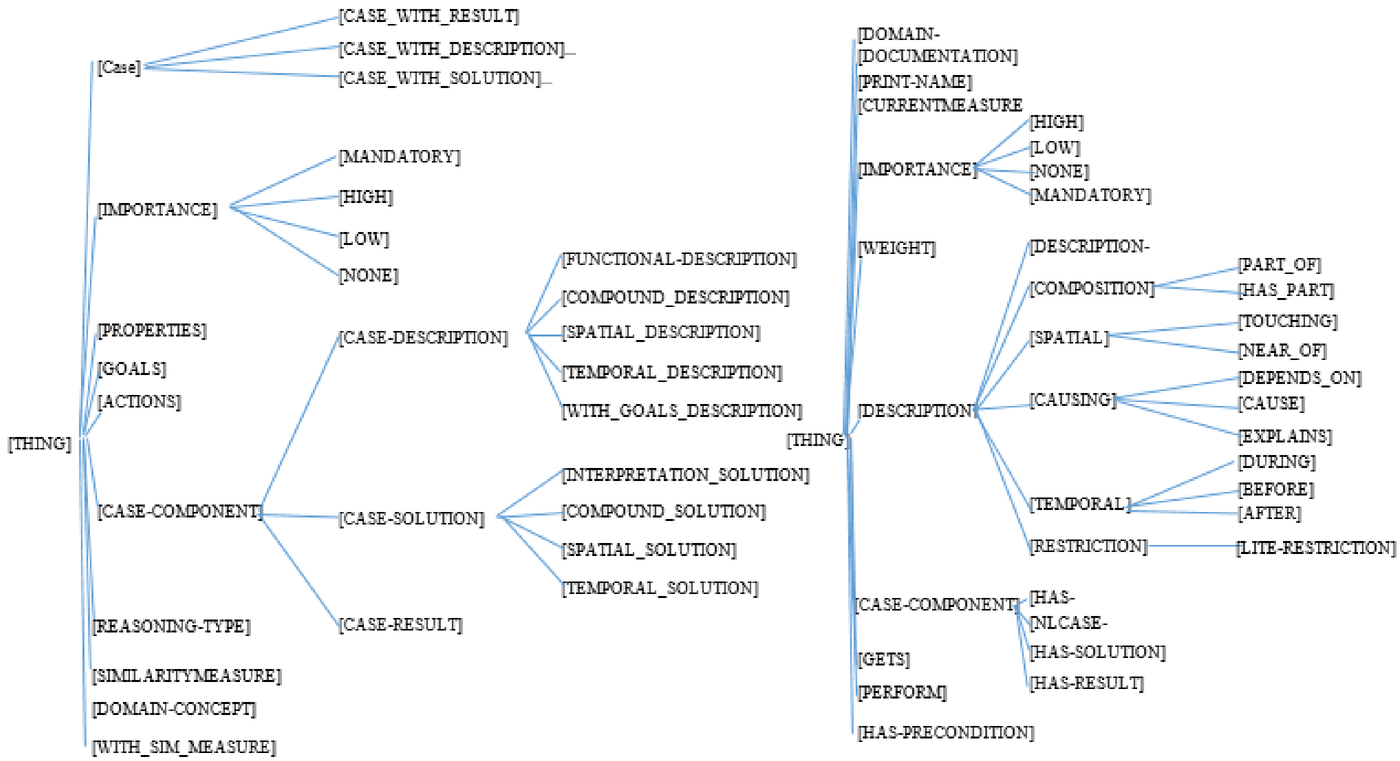


Fig. 12. Fragment of the CBROnto hierarchies

- Related with the tasks and methods hierarchies.
- Related with the definition of the case structures; and related with different knowledge roles used in the PSMs; and terms used to organize and classify the domain knowledge.

Note that, designer doesn't classify one by one every domain term because due to the inheritance mechanism only the top-level terms in the hierarchies should be classified.

The activities performed by the CBR application designer to model a domain, and to formalize it as a knowledge base include:

- The designer determines what domain is to be modelled, and selects from the library those ontologies that are potentially useful.
- The domain terminology from the ontologies has to be integrated as two term hierarchies: the concept hierarchy rooted by the CBROnto's Thing concept, and the relation hierarchy, rooted by the CBROnto's Binary-Tuple relation (see Fig. 12).

Each case is described by CBROnto and domain vocabulary. In this sense, the CBR processes are domain-independent but they are guided by the domain terminology organized below (in the subsumption hierarchies) the CBROnto terms. This model describes CBR processes using tasks and methods (PSM). These tasks and methods have global CBROnto task/method hierarchies. This model used the task decomposition of [11] for CBR processes. CBROnto includes the capabilities for describing a library of PSMs associated to the main CBR tasks. CBROnto describes CBR PSMs by relating them to terms and relations regarding tasks, requirements and domain characteristics. CBROnto includes terms of the method description language that are used to formalize PSMs.

D. XML-based case representation with ontology

Recently, several XML-inspired case representation languages such as CBML and OML have been introduced into the CBR community. They are devised to facilitate case interchanging in the web and could be viewed as structured representation languages that facilitating the encoding of case knowledge into web documents. The following issues should be taken into consideration:

- Some standard vocabularies for case description are needed, which ensure the success of case interchanging and distributed case-based reasoning.
- Some conveniences for integrating domain vocabularies should be provided.
- The web case language should be flexible to fulfil the needs of both unstructured and structured case representations.

Huajun et al. [78, 79] propose a web-oriented case representation RDF-based Case Markup Language CaseML for encoding case knowledge into web documents that allowed the usage of case base in the semantic web. To achieve the purpose of globally interpreted case base, the following issues are considered:

- Some standard vocabularies for case description are needed, which ensure the success of case interchanging and distributed case-based reasoning.
- Some conveniences for integrating domain vocabularies (ontologies) should be provided.

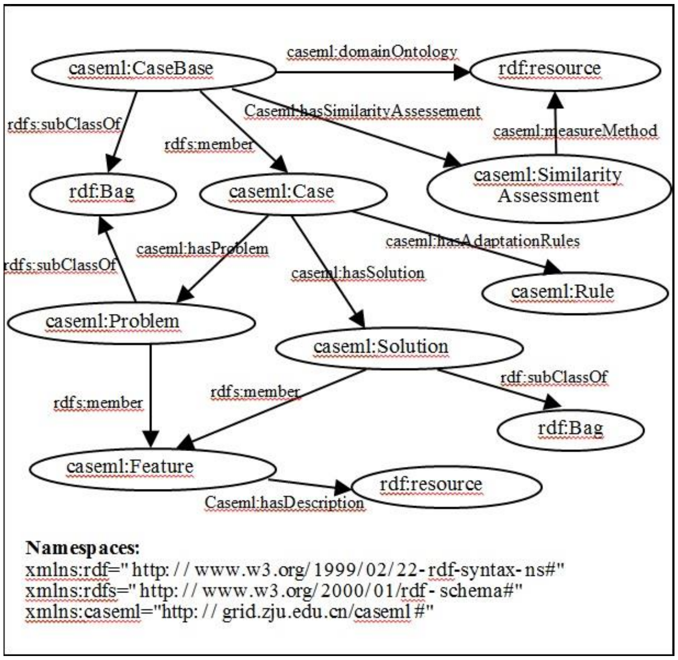


Fig. 13. The RDF graph for CaseML

The authors defined sets of standard classes (i.e. CaseBase, case, problem, solution, etc) and properties (i.e. domainOntology, hasProblem, hasSolution, etc) to define the structure of the case base ontology shown in Fig. 13 [78]. CaseML offers basic building blocks for publishing case knowledge onto the web and facilitates the sharing and interchanging of experience knowledge and building distributed CBR systems. Besides, it integrated domain ontologies with the case base ontology with enhanced CBR processes. What is more, [78] provided a generic architecture for the CBR (OpenDisCBR) in an open and distributed environment. This architecture emphasises on the integration of case knowledge with web ontologies. In this architecture, heterogeneous case bases cooperate with each other through some domain web ontologies.

E. OWL based and medical domain case representation methodology

Juarez et al [32] proposed case representation ontology for medical domains. This method depends on heavyweight ontologies, and temporal and context aspects are added (see Fig. 14 [32]). Case ontology is integrated with domain ontology for semantic case retrieval. The domain ontology contains domain specific concepts and standard concepts as ICD10. Context knowledge is collected from patient information as demographic information. Cases are instances of this ontology. This technique defines five kinds of cases: Complete, Valued, Solved, Contextualized, and Valid (and their opposites: Incomplete, Unvalued, Unsolved, Uncontextualized and Invalid). These kinds allow similarity at 3 levels of discrimination: (1) case representation ontology inference criteria, (2) top level description criteria, and (3) problem similarity criteria.

As a result, using ontology in case representation and in reasoning as a domain vocabulary enhances the CBR systems.

When case base is in the form of ontology, Ontology Description Logic Inference can be used to find relationships between cases. Reusability and sharing is also enhanced very much. What is more, the same ontology can be used in different systems and in different environments. Integration between ontological case base and domain ontology is enhanced and cases can contain textual, numerical and concept features. The semantic similarity and retrieval of cases is achieved, where users can express their request in a variety of terminology and the system understand user query by ontology terminology and ontological reasoning (i.e. description logic).

V. A Comparison Between Ontological CBR Methods

To the best of our knowledge, there are no previous researches for comparing case representation methods neither traditional nor ontological. The previous works as [9] have concentrated on the CBR systems as a whole and case representation has not mentioned. As a result, the paper depends on the existing ontological CBR systems as a whole for comparison. A comparative study between various systems that use ontology will be done. These systems are heterogeneous, and the paper depends on the self-explanatory features in each system and introduce ours. This strategy is followed by many researches [8]. The purpose of this comparison is to discover the weakness points and the challenges for the future enhancements. The main focus is on:

- Whether the system uses only simple traditional features as textual, numerical.
- Whether the system uses a traditional features and ontological features.
- Whether the system uses ontologies for case base.
- The form of integrated domain knowledge (not used, rules or ontology).
- Whether the case base ontology includes default knowledge, temporal knowledge and context knowledge.
- Whether the adaptation knowledge utilized ontology or rules.
- Depending on the application domain, whether the system used standard domain ontology as SNOMED CT in medical domain.
- The querying and retrieving capabilities of the system, whether it support semantic retrieval, the indexing structures used, and whether it support results explanation.
- Whether the ontology semantic is enhanced using rules. This metric is very important in medical domain

because medical ontology logic can be improved by domain expert and CPG rules.

- Whether the system support representation of cases with different internal structure.
- Whether case updating and maintenance is supported.

These are the most important metrics for comparing CBR systems. Moreover, the usage of ontology affects all these parameters as retrieval query, explanation, etc. Some other metrics such as the integration with other AI techniques, feature weighting methods, case mining, feature selection and/or extraction, integration with other reasoning systems, and others are discussed in other works as [8] and these aspects have little relation with case representation. One exception is the relationship between fuzzy logic and case representation. In the recent period, especially in medical domain, the representation of vague knowledge in case base has gotten a great attention. The introduction of fuzzy ontology and fuzzy semantic rules [80] will enhance the case representation. However, this aspect will be handled in future works. Because ontology-based CBR systems outweigh CBR systems that depend on traditional methods for case representation, the comparison will involve four KI-CBR systems and nineteen measures (see table 1). The paper checks whether a system support a feature or not. The most complete system is JCOLIBRI. Its completeness is 52.6% that is a low level. All of the methods suffer from shortages. CBR systems face great challenges that need solutions in future researches. The success of CBR system cannot be achieved without the cooperation of all these aspects ranging from query creation to case base maintenance.

VI. Semantic Retrieval Methods

The key to case-based reasoning is to retrieve the most similar case in a fast and accurate way [81]. Thus, the case similarity measurement is distinctly important which has a direct influence on the matching process. In some applications of CBR, it may be adequate to assess the similarity of the stored cases in terms of their surface features where similarity is computed by k nearest neighbour algorithm. In other applications, it may be necessary to use derived features obtained from a case's description by inference based on domain knowledge. In complex applications, cases are represented by complex structures (such as graphs) and structural similarity is required for retrieval. In case of surface features retrieval, a CBR system retrieves the k cases with maximum similarity to the target problem. However, sequentially processing all cases in memory has complexity O (n), where n is the number of cases. Optimization techniques are required such as parallel processing, indices or creating binary tree that organize cases in case base according to their similarity. Structural similarity is computationally expensive. One way for enhancement is to combine surface and structural similarity as in MAC/FAC model [82], Spread Activation Model [83], using generalized cases [84]. Improvement to the retrieval algorithms includes techniques for improving the speed of retrieval and for improving solution quality. Problems likely to affect solution quality include the use of inadequate similarity measures, noise, missing values in cases, unknown values in the description of the target problem, and the so-called heterogeneity problem that arises when different attributes are used to describe different cases. Each case representation method has a suitable retrieval algorithm. The paper concentrates on the ontological or semantic similarity measurement. The similarity computation of two ontology concepts or instances can be divided into two components: a concept-based similarity (or inter-class similarity) that depends on the location of the concepts in the ontology, and a slot-based similarity (or intra-class similarity) that depends on the fillers of the common attributes between the compared objects. Let q, q' be two instances of the ontology. The concept-based similarity, simcpt, is computed using the measure of Wu and Palmer [85] defined as follows (Eq. 1):

$$sim_{cpt}(q,q') = \frac{2 * prof(LCS(q,q'))}{prof(q) + prof(q')} \quad (1)$$

Dendani [86] adds the weights of the attributes wq to enhance similarity. Weight can be represented as attribute in ontology, and there are many methods to calculate it (Eq. 2).

$$\mathrm{Sim}_{\mathrm{cpt}} = Wq * \frac{2 * MAX(prof(lcs(q,q')))}{prof(q) + prof(q')} \quad (2)$$

Wang et al [87] defined this function as inverse of Eq. 1 (Eq. 3):

$$\mathrm{Sim}_{\mathrm{cpt}} = \frac{prof(q) + prof(q')}{2 * prof(lcs(q,q'))} \quad (3)$$

Garrido [88] defined a simple method for semantic similarity with low semantics as in Eq.4:

$$\mathrm{sim}(q,q') = \begin{cases} 1, & if\ q = q' \\ 3/4, & if\ q\ differs\ 1\ level\ of\ q' \\ 1/2, & if\ q\ differs\ 2\ levels\ of\ q' \\ 0, & if\ q\ differs\ more\ than 2\ level\ of\ q' \\ & \text{or if q is not in the ontology} \end{cases} \quad (4)$$

Where *prof* is the depth of a concept or an instance in the ontology hierarchy (only inheritance relations), and *LCS* is the Least Common Subsumer concept of two instances. In a special case, when *q* and *q'* represent the same instance in the ontology, then: *prof* (*LCS* (*q*, *q'*)) =*prof* (*q*), and thus: $sim_{cpt}(q, q') = 1$.

The more specific a concept that subsumes the concepts being compared, the more similar the concepts are. The above two approaches for concept similarity can be enhanced by adding the context knowledge in the ontology. This way the only concepts valid to comparison must be in the same context.

Another possibility for improvement is using the Similarity Region that is sub-hierarchy of the ontology where concepts and instances are comparable with each other [66]. In addition, other relations such as PART_OF, CAUSE need to be considered in similarity measurement in some ways.

The slot-based similarity measure, sim_{slt} is defined as follows (Eq. 5):

$$sim_{slt}(q,q') = \frac{\sum_{s \in CS} sim(q.s, q'.s)}{|CS|} \quad (5)$$

TABLE I. COMPARISON BETWEEN CBR METHODS, √ =SUPPORT, ×=NOT SUPPORT

Measure \ Method	Juan et al [56]	COBRA [57]	Juarez et al [89]	JCOLIBRI [16]
Simple features as numerical, text, and lists outside ontology	√	×	×	×
Ontological features	√	√	√	√
Ontology for case base	×	√	√	√
Domain ontology	√	√	√	√
Adaptation knowledge (rules/ontology)	√	√	×	×
Represent context knowledge	×	√	√	√
Represent default knowledge	×	×	×	√
Query's case enrichment	×	×	×	×
Case base update and maintenance	×	×	×	×
Cases with different structures	×	√	×	√
Temporal reasoning	×	×	√	√
Case indexing	×	√	×	√
Embed rules in case ontology	×	√	×	×
Explanation capability	×	×	×	×
Semantic retrieval or semantic query	√	√	√	√
Classical query	√	×	×	×
Incomplete or ambiguous input case	×	×	×	×
Use an abstract case base ontology	√	×	×	√
Integrate standardized ontologies with domain ontologies (as ICD)	×	×	√	×
Total = 19	**7**	**9**	**7**	**10**

Where *CS* (Common Slots) is the set of common simple attributes (properties) of q and q', $|CS|$ is its cardinality, $q.s$(or $q'.s$) represents the simple attribute s of q (or q'), and $sim(q.s, q'.s)$ is the similarity between the two simple attributes. To compute this similarity, they have defined two calculation modes that can be associated to attributes:

– ignore: for the properties that must not be taken into account in the similarity.

– exact: a strict mode that allows verifying the equality of property values. By considering these modes, the similarity *sim* ($q.s$, $q'.s$) is calculated as follows (Eq. 6):

$$sim(q.s, q'.s) = \begin{cases} 1 & if\ (w_{q.s} = exact) and\ (v_{q.s}, v_{q'.s}) \\ 0 & else \end{cases} \quad (6)$$

Where $w_{q.s}$is the calculation mode associated to the simple attribute $q.s$, and $v_{q.s}$ is the value of this attribute in q.

This method does not handle three points: (1) if the attribute is also a concept, it may be needed a loop of local-global similarity, (2) the method deals with exact quantitative attributes. When attributes have inexact values which need ontology-based fuzzy CBR or when the values are text, interval (time) or list,(3) when the two measured attributes are concepts which have different number of properties, and (4) the function speak about exact or not exact evaluation. It also does not take into account the largest and smallest values of the measured attribute. The global similarity measure of q and q' is given by the following formula (Eq. 7):

$$sim\ (q, q') = (1 - \alpha) \times sim_{cpt}(q, q') + \alpha \times sim_{slt}(q, q') \quad (7)$$

Where α is a parameter allowing controlling the importance of the slot-based similarity in the calculation.

The global similarity between the two cases C_1, C_2 can be calculated as follow (Eq. 8):

$$sim(C_1, C_2) = \frac{\sum_{i=1}^{n} w_i * sim(q_i, q'_i)}{\sum_{i=1}^{n} w_i} \quad (8)$$

Where C_1 (q_1, q_2 … q_n), C_2 (q'_1, q'_2 … q'_n) are the two compared classes. w_i is the weight of attribute q_i.

The above methodologies compare the query case with all cases in the case base ontology, but case base size is increasing exponentially when new cases are retained. Case base clustering, multi-way indexing, context knowledge, case classification ontology, and/or combination with RBR are critical to lower the search space and enhance case retrieval speed especially in time critical systems as ICU. Moreover, the semantic relationships between cases problem features can be inferred using DL inference beside similarity functions. The semantic relationships between cases solution features can be used to discover solution for unsolved cases. What's more, the Eq. 8 assumes that both query and retrieved cases have the same number and type of features. The number of features may not be the same between cases, the features may not be comparable because of its semantic, noise may be exists in the query case. Adding defaults and benefit from ontological reasoning can mitigate this problem. Similarity is also increasingly being combined with other criteria to guide the retrieval process, such as how effectively the solution space is covered by the retrieved cases; how easily their solutions can be adapted to solve the target problem; and how easily the proposed solution can be explained. What is more, query can be represented as small ontology. This way ontology matching between query ontology and case base ontology with support of domain ontology and DL reasoning can enhance semantic retrieval. These points require further research.

VII. CBR FUTURE CHALLENGES

Ontological CBR has many challenges to reach its full functionality. Challenges exist in all aspects and processes of the CBR system such as: case base creation, query building, case semantic retrieval, case adaptation, case retention, case base update and maintenance. Here, the paper will discuss some of these challenges.

1) Case solution adaptation has many techniques range from manual to generative (replays the method of deriving the retrieved solution on the new problem). Adaptation knowledge may be in the form of rules that are not fully compatible with ontological CBR. In order to determine which rules must be included in the system, and a deep analysis of the domain is required. Unfortunately, CBR is often applied to domains poorly understood or difficult to codify in the form of rules. So the leaders in the field have sometimes argued for postponing or avoiding the automatic adaptation. One challenge is how to auto learn the adaptation knowledge by discovering the semantic relationship between case description features (concepts) and formulate semantic rules in the same ontology to guide case adaptation (semi-) automatically. In addition, how to represent these rules in formats compatible with ontological case bases is another challenge. The best way to combine case adaptation rules in CBR system is by using ontology itself or using rule format designed for ontology as SWRL. SWRL is designed to add rule logic into OWL ontologies. Only some of the systems develop automatic adaptation strategies whereas the majority of the systems/projects provide for manual/conventional adaptation [89, 90, 91]. Ontology can provide more intelligence in case adaptation algorithms [73, 64].

2) Ontology engineering is critical in ontological CBR. Invention of a suitable ontology construction methodology for CBR case base and domain ontology in connection with patient medical record is a critical research area. It will enhance the integration of Case-Based and Ontology-Based Reasoning [92], and the discovered case base structure will require new indexing, semantic retrieval algorithm and similarity metrics. Until now, there is no ontology engineering methodology specific for CBR in the medical domain. This model will be different from existing ones because of the complexity and richness of medical domain: the existing standard terminologies as UMLS, standard ontologies as Disease Ontology, upper ontologies (i.e. Basic Formal Ontology (BFO), DOLCE, General Formal Ontology (GFO), and Unified Foundation Ontology (UFO)), vagueness in data, integration with EHR, etc.

3) Data pre-processing steps are critical to prepare medical data to form case bases because medical data are incomplete, inconsistent, vague, and detailed in most cases. It includes data aggregation, summarization, normalization, fuzzification, coding, integration, cleaning, etc. AI and data mining techniques help in this filed [93, 94]. The selection, mining and extraction of relevant features for case representation and weights for these features are open problems [95, 96]. The problem becoming complicated in the recent medical CBR systems due to a complex data format where the data are coming from sensors, images, time series or free-text format. The solution ranges from automatic one as genetic algorithms, or done manually by domain expert. The weights may be static for all situations or dynamic according to the context of execution. Adding default knowledge for describing classes is critical, and this allows reasoners to perform default reasoning with defaults added to class descriptions [17].

4) Cases are represented using simple or concept attributes. For medical domains, other multimedia attributes as images could add more semantic.

5) Reasoning with incomplete, inconsistent, vague and/or inaccurate data is expected in medical domain. Soft computing can enhance the functionality of CBR system [97]. For example, the use of fuzzy sets allows a flexible encoding of case characteristics as linguistic terms [98]. Cases are stored in fuzzy database or fuzzy ontology. During retrieval, the fuzzy similarity of a case can be calculated using a fuzzy membership function and weighted fuzzy pattern matching. This similarity can enhance the semantic similarity achieved by using ontology. All numeric parameters of the CBR system (e.g., feature weights, value of k in k-NN, shape of fuzzy similarity membership functions) can be maintained using a genetic algorithm and ANN. Inductive methods can be used to cluster case bases and find representative and redundant cases, which can be used to direct case base maintenance. Moreover, query creation connected with patient record that contains all patient medical data and connected with rule-base background knowledge will enhance new case creation or enrichment.

6) In medical domain, the domain ontology can benefit from and reuse existing standard ontologies as SNOMED, UMLS, ICD, etc. [89].These ontologies provide standardized terminology to represent findings, diseases, procedures, medications, sites, and organisms. Without these deep domain ontologies, CBR systems would not have been able to perform acceptable clinical assistance. However, coding of EHR data and extraction of reference set from these large ontologies is a big challenge. The open question is how to use these ontology to achieve semantic interoperability between EHR systems [89, 99], and ease case collection from distributed databases.

7) Temporal data representation in domain ontology, case base, and the query is critical especially in ICU and chronic diseases patients where temporal and continuous evaluation is essential. Temporal data is represented in case, and handled in case retrieval algorithm [100, 101, 102]. Time representation in case ontology is standardized in OWL ontology [103] and requires temporal similarity functions for effective retrieval [104, 105]. However, the application of temporal CBR requires more research [106]. Moreover, handling uncertainty in temporal data is critical especially for medical data [107, 108].

8) Distributed CBR on the web is critical to share, integrate and distribute knowledge. It will be advantageous to develop CBR systems as Web services, to receive patient input data from the Internet, securely, to process them against several CBR systems, combine with non-CBR systems, and give back a consolidated result from several sources.

9) In medical domain, there are two types of knowledge. (a) The general knowledge including domain ontologies, standardized terminologies (i.e. SNOMED CT, ICD and

UMLS) and CPG. Domain ontology provides ground service to specify the meaning of the terms used in case description. The challenges in this point is the encoding of EHR data using a selected ontology, the creation of suitable subset of this ontology for your domain, and the creation of efficient semantic case retrieval algorithm [109]. CPGs can be represented in the form of rules [110]. In [111] clinical pathways are represented in prototypical cases. (b) The experience knowledge that are represented in cases. CPGs can enhance the reasoning process of CBR because these rules can be represented in the form of ontology (using SWRL) and enrich knowledge in case bases and domain ontologies.

*10)*The number of initial cases in case base affects the efficiency of CBR system. The creation of ontology engineering methodology, to extract cases as ontology instances from EHR, is critical issue. In other words, the case-base ontology population by cases from EHR raw or prepared data. The cases must have a standard structure that may utilize HL7 RIM data model, and standard content that utilize standard terminologies. When CBR systems are able to take advantage of patients' representations in electronic health records, they will become applicable to a wide range of diseases.

*11)*Heterogeneous case base contains cases with different structures or with different number and types of attributes. This case requires enhanced case retrieval algorithms [68]. Ontology enhance the creation of dynamic structure case base very much [66].

*12)*Defining a Medical Context Ontology for the domain explicitly species a set of medical context, which are used for retrieving only cases highly relevant to the new case [112]. A context can be defined as a set of attributes relevant for a given retrieval that is a set of constraints on the patient clinical state.

*13)*No researches have been done in the establishment of semantic relations between case problem attributes, between case solution attributes and between the two. These relationships have benefits in query answering, complex case decomposition, case enrichment, case adaptation, etc.

*14)*Because of space restrictions, the paper will not discuss the challenges of soft CBR including the integration with fuzzy logic, statistics, neural networks, and data mining and how these technologies can enhance the functionality of CBR systems.

VIII. Conclusion

This paper has reviewed the CBR case representation formalisms. They can be divided into two categories, traditional and ontological methods. The traditional methods have many limitations such as the case features have no relations to each other and users have to express their queries for new cases exactly as represented in case base. The similarity and retrieval of cases is static and based on exact matching. There are no inference mechanisms in the case base. On the other hand, integrating ontologies as domain terminologies with traditional case representation methods can enhance the sharing and querying capabilities. The optimum solution is achieved by using ontologies in representation of cases and domain knowledge. This action creates what is named knowledge intensive CBR. Sharing, semantic retrieval, case representation issues are achieved. This paper has also conducted a comparison between ontological CBR methods, and it has concluded that JCOLIBRI is the best approach. The paper has discussed the semantic retrieval in case based reasoning and suggest the challenges for the future research in ontological CBR. As a result CBR could be a valid approach for building CDSS, but more investigations are needed. As future works, we will study the case retrieval algorithms, the soft-CBR techniques, the integration between CBR and EHR environment. We will study how the results of this paper can be extended for other new systems or new metrics.

References

[1] S. Begum, M. Ahmed, P. Funk, N. Xiong, B. Schéele, "A case based decision support system for individual stress diagnosis using fuzzy similarity matching," Computational Intelligence. Blackwell Publishing.; vol. 25, no. 3, pp. 180-195, 2009.

[2] J. Corchado, J. Bajo, A. Abraham, "GERAmI. Improving healthcare delivery in geriatric residences," IEEE Intelligent Systems, vol. 23, no. 2, pp. 19-25, 2008.

[3] C. Marling, P. Whitehouse, "Case-based reasoning in the care of Alzheimer's disease patients," 4th International Conference on Case-Based Reasoning (ICCBR 01), pp. 702-715, 2001.

[4] S. Montani, L. Portinale, G. Leonardi, R. Bellazzi," Case-based retrieval to support the treatment of end stage renal failure patients," Artificial Intelligence in Medicine, vol. 37, no. 1, pp. 31-42, 2006.

[5] I. Watson, "Applying Case-Based Reasoning, Techniques for Enterprise Systems," Morgan Kaufmann Publishers Inc. 1997.

[6] J. Kolodner, D. Leake, "A tutorial introduction to case-based reasoning. In Case-Based Reasoning: Experiences, Lessons and Future Directions," AAAI/MIT Press, pp. 31-65, 1996.

[7] I. Bichandaritz, S. Montani, Advances in case-based reasoning in the health sciences," Artificial Intelligence in Medicine, vol. 51, no. 2, pp. 75-146, 2011.

[8] S. Begum, M. Ahmed, P. Funk, N. Xiong, M. Folke, "Case-Based Reasoning Systems in the Health Sciences: A Survey of Recent Trends and Developments," IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 41, no. 4, pp. 421 – 434, 2011.

[9] I. Bichindaritz, C. Marling, "Case-based reasoning in the health sciences: what's next?" Artificial Intelligence in Medicine, vol. 36, no. 2, pp. 127–135, 2006.

[10] A. Mobyen, B. Shahina, F. Peter, "Case Studies on the Clinical Applications using Case-Based Reasoning," Proceedings of the IEEE Federated Conference on Computer Science and Information Systems, pp. 3–10, 2012.

[11] A. Aamodt, E. Plaza, "CBR: foundational issues, methodological variations and system approaches," AI Communications, vol. 7, no. 1, pp. 39-59, 1994.

[12] A. Cordier, B. Fuchs, J. Lieber, A. Mille, "Acquisition interactive des connaissances d'adaptationintégrée aux sessions de raisonnement à partir de cas - Principes, architecture IAKA et prototype KAYAK," In: Actes du 15ème atelier de Raisonnement à Partir de Cas (RàPC 2007), pp. 71–84, 2007.

[13] F. Gavin, S. Zhaohao, "R5 model for case-based reasoning. Knowledge-Based Systems," vol. 16, no. 1, pp. 59–65, 2003.

[14] M. Pantic, L. Rothkrantz, "Case-based reasoning for user-profiled recognition of emotions from face images," Proceeding of IEEE International Conference of Multimedia and Expo, vol. 1, pp. 391 – 394, 2004.

[15] T. Jirapond, B. Veera, P. Puntip, "A new similarity measure in formal concept analysis for case-based reasoning," Journal of Expert Systems with Applications, vol. 39, no. 1, pp. 967–972, 2012.

[16] BelénDíaz-Agudo, A. Pedro, González-Calero, "An Ontological Approach to Develop Knowledge Intensive CBR Systems," Ontologies Integrated Series in Information Systems. Springer, vol. 14, pp. 173-213, 2007.

[17] B. Díaz-Agudo, P. González-Calero, "An Architecture for Knowledge Intensive CBR Systems," Advances in Case-Based Reasoning (EWCBR'00); Springer, pp. 37-48, 2000.

[18] A. Aamodt, "Knowledge Intensive CBR in CREEK," In the proceeding of 7th European Conference on CBR (ECCBR), vol. 3155, pp. 1-15, 2004.

[19] M. Richer, "The knowledge contained in similarity measures," Invited Talk on the ICCBR, Sesimbra, Portugal, 1995.

[20] A. Bouhana, A. Fekih, et al. "An integrated case-based reasoning approach for personalized itinerary search in multimodal transportation systems," Transportation Research Part C, vol. 31, pp. 30–50, 2013.

[21] E. Armengol, A. Palaudàries, E. Plaza, "Individual Prognosis of Diabetes Long-term Risks: A CBR Approach," Methods Inf Med, vol. 40, no. 1, pp. 46-51, 2001.

[22] I. Watson, Applying Case-Based Reasoning: Techniques for Enterprise Systems. San Francisco, CA: Morgan Kaufmann Inc. 1997.

[23] S. Nitsuwat, W. Paoin, Development of ICD-10-TM Ontology for a Semi-automated Morbidity Coding System in Thailand," Methods Inf Med, vol. 51, no. 6, pp. 519-528, 2012.

[24] S. Simon, P. Sankar, "Foundations of Soft Case-Based Reasoning," 1st ed. Wiley-Interscience, 2004.

[25] B. Ralph, K. Janet, P. Enric, "Representation in case-based reasoning," The Knowledge Engineering Review. Cambridge University Press, vol. 00, no. 0, pp. 1–4, 2005.

[26] J. Kolodner, "Case-Based Reasoning," San Mateo, CA: Morgan Kaufmann Publishers; 1993.

[27] J. Hunt, "Evolutionary case based design, Progress in Case-based Reasoning," Lecture Notes in Computer Science; Springer Berlin Heidelberg, vol. 1020, pp. 17–31, 1995.

[28] B. Allen, Case-based reasoning: business applications. Communications of the ACM, vol. 37, no. 3, pp. 40–42, 1994.

[29] A. Maalel, L. Mejri, H. Hadj-Mabrouk, H. Ben Ghézela, "Towards a Case-Based Reasoning Approach Based on Ontologies Application to Railroad Accidents," Proceeding of Third International Conference of Data and Knowledge Engineering (ICDKE), pp. 48-55, 2012.

[30] M. Richter, "On the notion of similarity in case-based reasoning," Mathematical and Statistical Methods in Artificial Intelligence. Springer-Verlag, pp. 171-184, 1995.

[31] Z. Cheng., L. Chen, X. Jia, H. Zeng, "Case Representation and Retrieval in the Intelligent RCM Analysis System," Advances in Electronic Commerce. Web Application and Communication Advances in Intelligent and Soft Computing, vol. 149, pp. 21-28, 2012.

[32] J. Juarez, J. Salort, J. Palma, R. Marin, "Case representation ontology for case retrieval systems in medical domains," ACM Proceedings of the 25th IASTED International Multi-Conference: artificial intelligence and applications, pp. 168-173, 2007.

[33] C. Jian-guo, C. Yong-zhao, "Research on operation standard expert system based on ontology and CBR," IEEE 18Th International Conference on Industrial Engineering and Engineering Management (IE&EM) Part 3, pp. 1932 – 1936, 2011.

[34] Y. Guo, J. Hu, Y. Peng, "A CBR system for injection mould design based on ontology: A case study," Journal of Computer-Aided Design, vol. 44, no. 6, pp. 496-508, 2012.

[35] B. Morteza, S. Abbas, N. Rassoul, "A case-based reasoning system development for statistical process control: Case representation and retrieval," Journal of Computers & Industrial Engineering, vol. 63, no. 4, pp. 1107–1117, 2012.

[36] B. Porter, R. Bareiss, R. Holte, Concept learning and heuristic classification in weak-theory domains," Journal of Artificial Intelligence, vol. 45, no. 1–2, pp. 229–263, 1990.

[37] Z. Ming, C. Zhimin, H. Wenjing, C. Xu, "Representing and matching simulation cases: A case-based reasoning approach," Journal of Computers & Industrial Engineering, vol. 59, no. 1, pp. 115–125, 2010.

[38] M. Negnevitsky, "Artificial Intelligence: A Guide to Intelligent Systems," 2nd ed. Addison-Wesley; 2004.

[39] F. Baader, D. Calvanese, D. McGuinness, D. Nardi, P. Patel-Schneider, "The description logic handbook: theory, implementation, and applications," New York, NY, USA: Cambridge University Press; 2003.

[40] E. Plaza, "Cases as terms: A feature term approach to the structured representation of cases," Proceedings of the 1st International Conference on Case-Based Reasoning. Berlin: Springer, pp. 265–276, 1995.

[41] R. Bergmann, "Experience Management: Foundations, Development Methodology, and Internet-Based Applications," Springer, 2002.

[42] M. Manago, R. Bergmann, S. Wess, R. Traph¨oner, "CASUEL: A Common Case Representation Language," ESPRIT Project INRECA, No. 6322, Deliverable D1, Univ. of Kaiserslautern, Kaiserslautern Germany, 1994.

[43] B. Stahl B. Similarity measures for object-oriented case representations. Proceedings of the European Workshop on Case-Based Reasoning. EWCBR'98. 1998.

[44] R. Weber, K. Ashley, B. Stefanie, "Textual case-based reasoning," The Knowledge Engineering Review, vol. 20, no. 3, pp. 255–260, 2006.

[45] R. Burke, K. Hammond, V. Kulyukin, S. Lytinen, N. Tomuro, S. Schoenberg, "Question answering from frequently-asked questions files: experiences with the FAQ Finder system," AI Magazine, vol. 18, no. 1, pp. 57–66, 1997.

[46] D. Wilson, S. Bradshaw, "CBR textuality," Expert Update, vol. 3, no. 1, pp. 28–370, 2000.

[47] E. Rissland, J. Daniels, "The synergistic application of CBR to IR," Artificial Intelligence Review, vol. 10, no. 5–6, pp. 441–475, 1996.

[48] M. Lenz, H. Burkhard, "Case retrieval nets: Basic ideas and extensions," Advances in Artificial Intelligence, Berlin: Springer, pp. 227–239, 1996.

[49] K. Gupta, W. Aha, "Towards Acquiring Case Indexing Taxonomies from Text," In Proceeding of 6th International Florida Artificial Intelligence Research Society Conference, pp. 172-177, 2004.

[50] R. Weber, A. Martins, R. Barcia, "On legal texts and cases" Citation: Textual Case-Based Reasoning: Papers from the AAAI-98 Workshop, Technical Report WS-98-12. AAAI Press, pp. 40–50, 1998.

[51] S. Brüninghaus, K. Ashley, "Bootstrapping case base development with annotated case summaries," Case-Based Reasoning Research and Development, Lecture Notes in Artificial Intelligence, Springer Berlin Heidelberg, vol. 1650, pp. 59–73, 1999.

[52] S. Brüninghaus, K. Ashley, "The role of information extraction for textual CBR," Case-Based Reasoning Research and Development, Lecture Notes in Artificial Intelligence, Springer, vol. 2080, pp. 74–89, 2001.

[53] B. Stefanie, K. Ashley, "Reasoning with textual cases," Proceedings of the 6th international conference on Case-Based Reasoning Research and Development (ICCBR'05). Springer-Verlag Berlin, Heidelberg, pp. 137-151, 2005.

[54] K. Ashley, B. Stefanie, "Developing Mapping and Evaluation Techniques for Textual Case-Based Reasoning," AAAI Tech-nical Report; WS-98-12. 1998.

[55] R. Bergmann, W. Wilke, "On the role of abstraction in case-based reasoning," Proceedings of the 3rd European Workshop on Case-Based Reasoning. Berlin: Springer, pp. 28–43, 1996.

[56] I. Watsona, S. Perera, "A hierarchical case representation using context guided retrieval," Journal of Knowledge-Based Systems, vol. 11, pp. 285–292, 1998.

[57] M. Maher, D. Zhang, "CADSYN: Using case and decomposition knowledge for design synthesis," Artificial Intelligence in Design, Oxford, UK: Butterworth-Heinmann, 1991.

[58] A. Ram, A. Francis, "Multi-plan retrieval and adaptation in an experience-based agent," Case-Based Reasoning: Experiences, Lessons, and Future Directions. Menlo Park, CA, US: AAA1 Press/The MIT Press. 1996.

[59] R. Bergmann, W. Wilke, I. Vollrath, S. Wess, "Integrating General Knowledge with Object-Oriented Case Representation and Reasoning," In H. D. Burkhard and M. Lenz editors, 4th German workshop on CBR: HU Berlin, pp. 120-127, 1996.

[60] A. Aamodt, “A Knowledge-Intensive Integrated Approach to Problem Solving and Sustained Learning,” PhD. Dissertation. University of Trondheim, Department of Electrical Engineering and Computer Science, Trondheim, 1991.

[61] P. Skalle, A. Aamodt, “Knowledge-based decision support in oil well drilling; Combining general and case-specific knowledge for problem solving,” Proceedings of ICIIP-2004, International Conference on Intelligent Information Processing. Beijing, 2004.

[62] R. Studer, R. Benjamins, D. Fensel, “Knowledge engineering: Principles and methods,” Data & Knowledge Engineering, vol. 25, no. 1–2, pp. 161–198, 1998.

[63] G. Yuan, YinghongPeng, H. Jie, "Research on high creative application of case-based reasoning system on engineering design," Journal of Computers in Industry, vol. 64, pp. 90–103, 2013.

[64] J. Recio-Garía, B. Díaz-Agudo, “Ontology based CBR with jCOLIBRI,” Applications and Innovations in Intelligent Systems XIV, Springer, pp.149-162, 2007.

[65] R. Bergmann, M. Schaaf, “Structural case-based reasoning and ontology-based knowledge management: A perfect match,” Journal of Universal Computer Science, vol. 9, no. 7, pp. 608–626, 2003.

[66] A. Abou Assali, D. Lenne, B. Debray, “Heterogeneity in Ontological CBR Systems,” Successful Case-based Reasoning Applications I, Studies in Computational Intelligence, vol. 305, pp. 97-116, 2010.

[67] L. Lamontagne, G. Lapalme, “Raisonnement `a base de castextuels: Etat de l’art et perspectives,” Revue d’intelligenceartificielle, vol. 16, pp. 339–366, 2002.

[68] A. Abou Assali, D. Lenne, B. Debray, “Case Retrieval in Ontology-Based CBR Systems,” KI 2009: Advances in Artificial Intelligence, Lecture Notes in Computer Science, Springer Berlin Heidelberg, vol. 5803, pp. 564-571, 2009.

[69] B. Smyth, M. Keane, “Retrieving Adaptable Cases: The Role of Adaptation Knowledge in Case Retrieval,” Proceeding EWCBR '93 Selected papers from the First European Workshop on Topics in Case-Based Reasoning, Springer-Verlag London, vol. 837, pp. 209–220, 1994.

[70] A. Maalel, L. Mejri, H. Mabrouk, H. Ben Ghezela, “Toward a Knowledge Management Approach Based on an Ontology and Case-based Reasoning (CBR) Application to Railroad Accidents,” Sixth International Conference on Research Challenges in Information Science (RCIS), pp. 1 – 6, 2012.

[71] A. Assali, D. Lenne, B. Debray, “Ontology Development for Industrial Risk Analysis,” IEEE 3rd International Conference on Information and Communication Technologies: From Theory to Applications (ICTTA), Damascus, Syria. 2008.

[72] M. Fernandez, A. Gomez-Perez, N. Juristo, “Methontology: From Ontological Art towards Ontological Engineering.” Proceedings of the AAAI97. Spring Symposium Series on Ontological Engineering, pp. 33–40, 1997.

[73] B. Díaz-Agudo, P. González-Calero, J. Recio-García, A. Sánchez-Ruiz-Granados, “Building CBR systems with jcolibri,” Journal of Science of Computer Programming, vol. 69, no. 1–3, pp. 68–75, 2007.

[74] A. Essam, A. Salem, “A Breast Cancer Classifier based on a Combination of Case-Based Reasoning and Ontology Approach,” Proceedings of the IEEE International Multi-conference on Computer Science and Information Technology, pp. 3–10, 2010.

[75] B. Díaz-Agudo, P. González-Calero, “An ontological approach to develop knowledge intensive CBR systems,” Ontologies Integrated Series in Information Systems, US Springer, vol. 14, pp. 173-213, 2007.

[76] V. Benjamins, D. Fensel, “Special Issue on Problem-Solving Methods,” International Journal of Human-Computer Studies (IJHCS), vol. 49, no. 4, pp. 305–313, 1998.

[77] B. D´ıaz-Agudo, P. Gonz´alez-Calero, “CBROnto: a task/method ontology for CBR,” Proceeding of the 15th International FLAIRS’02 Conference (Special Track on CBR). AAAI Press, pp. 101–106, 2002.

[78] H. Chen, Z. Wu, “On Case-based Knowledge Sharing in Semantic Web,” Proceeding of the 15th IEEE International Conference on Tools with Artificial Intelligence, pp. 200 – 207, 2003.

[79] H. Chen, Z. Wu “CaseML: A RDF-based Case Markup Language for Case-based Reasoning in Semantic Web,” Proceedings of Workshop on: From structured cases to unstructured problem solving episodes for experience-based assistance, 5th International Conference on Case-based Reasoning, ICCBR, 2003.

[80] Z. Ma, F. Zhang, L. Yan, J. Cheng, “Fuzzy Knowledge Management for the Semantic Web,” Studies in Fuzziness and Soft Computing. Springer Berlin Heidelberg, vol. 306, 2014.

[81] L. Bobrowski, “Induction of Similarity Measures and Medical Diagnosis Support Rules through Separable, Linear Data Transformations,” Methods Inf Med, vol. 45, no. 2, pp. 200-203, 2006.

[82] K. Forbus, D. Gentner, K. Law, “MAC/FAC: A model of similarity-based retrieval,” Cognitive Science, vol. 19, no. 2, pp. 141-205, 1994.

[83] M. Brown, “An underlying memory model to support case retrieval,” InWess S., Althoff K.-D. & Richter M. (eds.), Topics in Case-Based Reasoning. Berlin: Springer, pp. 132-143, 1994.

[84] B. Mougouie, R. Bergmann, “Similarity assessment for generalized cases by optimization methods,” In Proceedings of the Sixth European Conference on Case-Based Reasoning. Berlin: Springer, pp. 249- 263, 2002.

[85] Z. Wu, M. Palmer, “Verbs Semantics and Lexical Selection,” Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics (ACL), pp. 133–138, 1994.

[86] N. Dendani, M. Khadir, S. Guessoum, B. Mokhtar, “Use a domain ontology to Develop Knowledge intensive CBR systems for fault diagnosis,” IEEE International Conference on Information Technology and e-Services, pp. 1-6, 2012.

[87] D. Wang, W. Xiang, G. Bo Zhang, “Research on Ontology-Based Case Indexing in CBR,” IEEE International Conference on Artificial Intelligence and Computational Intelligence (AICI), vol. 4, pp. 238 – 241, 2009.

[88] J. Garrido, M. Hurtado, M. Noguera, J. Zurita, “Using a CBR approach based on ontologies for recommendation and reuse of knowledge sharing in decision making,” IEEE Eighth International Conference on Hybrid Intelligent Systems HIS'08, pp. 837-842, 2008.

[89] I. Bichindaritz, “Memoire: A framework for semantic interoperability of case-based reasoning systems in biology and medicine,” Journal of Artificial Intelligence in Medicine, vol. 36, pp. 177—192, 2006.

[90] F. Paz, S. Rodriguez, J. Bajo, M. Corchao, “Case-based reasoning as a decision support system for cancer diagnosis: A case study,” Int. J. Hybrid Intell. Syst. (IJHIS), IOS press, pp. 97–110, 2008.

[91] M. D’Aquin, J. Lieber, A. Napoli, “Adaptation knowledge acquisition: a case study for case-based decision support in oncology,” Comput. Intell, vol. 22, no. 3–4, pp. 161–176, 2006.

[92] M. Kowalski, H. Klüpfel, S. Zelewski, D. Bergenrodt, A. Saur, “Integration of Case-Based and Ontology-Based Reasoning for the Intelligent Reuse of Project-Related Knowledge,” Lecture Notes of Efficiency and Logistics. Springer Berlin Heidelberg, pp. 289-299, 2013.

[93] X. Xie, L. Lin, S. Zhong, “Handling missing values and unmatched features in a CBR system for hydro-generator design,” Computer-Aided Design, vol. 45, pp. 963–976, 2013.

[94] K. Borges, I. Tronto, R. Lopes, J. Silva, “A Data Pre-Processing Method for Software Effort Estimation Using Case-Based Reasoning,” Clei electronic journal, vol. 16, no. 3, 2013.

[95] S. Guessoum, M. Laskri, J. Lieber, “RESPIDIAG: A Case-Based Reasoning System for the Diagnosis of Chronic Obstructive Pulmonary Disease,” Expert Systems with Applications, vol. 41, no. 2, 1, pp. 267–273, 2014.

[96] Y. Li, M. Xie, T. Goh, “A study of mutual information based feature selection for case based reasoning in software cost estimation,” Expert Systems with Applications, vol. 36, pp. 5921–5931, 2009.

[97] Y. Cheng, H. Melhem, “Monitoring bridge health using fuzzy case-based reasoning. Advanced Engineering Informatics, vol. 19, pp. 299–315, 2005.

[98] R. Kuo, Y. Kuo, K. Chen, “Developing a diagnostic system through integration of fuzzy case-based reasoning and fuzzy ant colony system,” Expert Systems with Applications, vol. 28, pp. 783–797, 2005.

[99] I. Bichindaritz, "Semantic interoperability of case bases in biology and medicine," Artif. Intell. Med., (Special issue on case-based reasoning in the health sciences), vol. 36, no. 2, pp. 177–192, 2006.

[100] Z. Zhengxing Huanga, J. Juarezb, H. Duana, H. Lia, "Length of stay prediction for clinical treatment process using temporal similarity," Expert Systems with Applications, vol. 40, no. 16, pp. 6330–6339, 2013.

[101] S. Montani, L. Portinale, "Case Based Representation and Retrieval with Time Dependent Features," Case-Based Reasoning Research and Development, Springer Berlin Heidelberg, vol. 3620, pp. 353-367, 2005.

[102] I. Bichindaritz, "MNAOMIA: Reasoning and Learning from Cases of Eating Disorders in Psychiatry," Proceeding of AMIA Annual Fall Symposium, vol. 965, 1996.

[103] W3C, Time Ontology in OWL. http://www.w3.org/TR/owl-time. Last seen 11 Feb. 2015.

[104] R. Schmidt, L. Gierl, "Temporal Abstractions and Case-Based Reasoning for Medical Course Data: Two Prognostic Applications," Machine Learning and Data Mining in Pattern Recognition, Springer Berlin Heidelberg, vol. 2123, pp. 23-34, 2001.

[105] M. Jære, A. Aamodt, P. Skalle, "Representing Temporal Knowledge for Case-Based Prediction," Advances in Case-Based Reasoning, Springer Berlin Heidelberg, vol. 2416, pp. 174-188, 2002.

[106] M. Sánchez-Marré, U. Cortés, M. Martínez, J. Comas, I. Rodríguez-Roda, "An Approach for Temporal Case-Based Reasoning: Episode-Based Reasoning," Case-Based Reasoning Research and Development, vol. 3620, pp. 465-476, 2005.

[107] S. Badaloni, M. Falda "Temporal-based medical diagnoses using a Fuzzy Temporal Reasoning System," J. Intelligent Manufacturing, vol. 21, no. 1, pp. 145-153, 2010.

[108] J. Juarez, F. Guil, J. Palma, R. Marin, "Temporal similarity by measuring possibilistic uncertainty in CBR," Journal of Fuzzy Sets and Systems, vol. 160, no. 2, pp. 214-230, 2009.

[109] G. Melton, S. Parsons, F. Morrison, A. Rothschild, M. Markatou, G. Hripcsak, "Inter-patient distance metrics using SNOMED CT defining relationships," Journal of Biomedical Informatics, vol. 39, no. 6, pp. 697–705, 2006.

[110] U.S. Department of Health and Human Services, National Guideline Clearinghouse. A national clinical guideline, http://www.guideline.gov/index.aspx, last seen 22 Jan. 2015.

[111] I. Bichindaritz, Prototypical Cases for Knowledge Maintenance in Biomedical CBR, Case-Based Reasoning Research and Development, Springer Berlin Heidelberg, vol. 4626, pp. 492-506, 2007.

[112] S. Montani, "How to use contextual knowledge in medical case-based reasoning systems: A survey on very recent trends," Artificial Intelligence in Medicine, vol. 51, pp. 125–131, 2011.

Arabic Alphabet and Numbers Sign Language Recognition

Mahmoud Zaki Abdo
Electronic, communication, and computer department
Faculty of engineering, Helwan University
Cairo, Egypt

Sameh Abd El-Rahman Salem
Electronic, communication, and computer department
Faculty of engineering, Helwan University
Cairo, Egypt

Alaa Mahmoud Hamdy
Electronic, communication, and computer department
Faculty of engineering, Helwan University
Cairo, Egypt

Elsayed Mostafa Saad
Electronic, communication, and computer department
Faculty of engineering, Helwan University
Cairo, Egypt

***Abstract*—This paper introduces an Arabic Alphabet and Numbers Sign Language Recognition (ArANSLR). It facilitates the communication between the deaf and normal people by recognizing the alphabet and numbers signs of Arabic sign language to text or speech. To achieve this target, the system able to visually recognize gestures from hand image input. The proposed algorithm uses hand geometry and the different shape of a hand in each sign for classifying letters shape by using Hidden Markov Model (HMM). Experiments on real-world datasets showed that the proposed algorithm for Arabic alphabet and numbers sign language recognition is suitability and reliability compared with other competitive algorithms. The experiment results show that the increasing of the gesture recognition rate depends on the increasing of the number of zones by dividing the rectangle surrounding the hand.**

Keywords*—hand gestures; hand geometry; Sign language recognition; image analysis; and HMM

I. Introduction

The main problem in the language of deaf people makes it difficult to translate thoughts and feelings into words and phrases understandable and aware. The normal people translate ideas into words audible, but the deaf people translate ideas into visual signs through the hands movement.

Over the years used the deaf and dumb signs among themselves. It became the different sign language of each community in the world. These signs are only one thing for the deaf and dumb to communicate with each other and the outside world to them [1]. There has been a growing interest in the recognition of human hand movements. Normally, there is no problem when deaf persons communicate with each other by using their common sign language. The problem appears when a Deaf people want to communicate with a non-deaf people. Usually, both will be disgruntled in a very short time [2].

A Sign language is different from country to another country; the researchers attempt to unify the sign language in each country separately have been carried out such as Jordan, Saudi Arabia, and Egypt to help persons of the deaf and dumb for each country. The researchers are working on hand gestures in different sign languages as the Australian Sign Language (Auslan) [3], the Chinese Sign Language (CSL) [4], the American Sign Language (ASL) [5], and the Dutch Sign Language. The Arabic Sign Language has not received attention in researchers [7].

All Previous researches on sign languages depend on glove or vision based methods [6]. In the glove based method, the user wears special devices, like special gloves or markers, the system related with data on the hand shape and motion. While in the vision method, the system recognizes the gestures by using image processing techniques without putting any limitation on the user [7].

The work [2] created an automatic translation system for gestures of manual alphabets in the Arabic sign language recognition. It does not rely on using any visual markings or gloves. The extracted features depend on two stages, feature-vector-creation stage and edge detection stage. It used multilayer perceptron (MLP) classifier and minimum distance classifier (MDC) to detect 15 characters only of 28 characters.

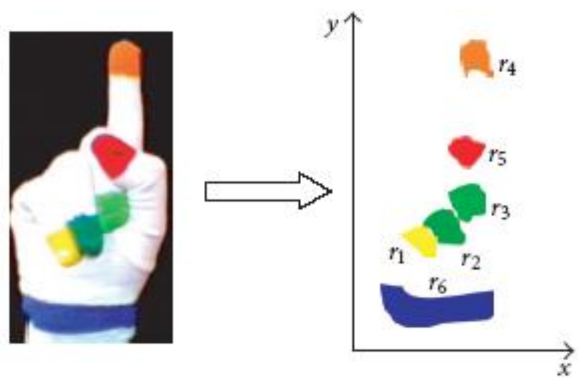
The research work in [7] investigated appearance-based features for the deaf person- vision-based on sign language recognition. It does not depend on a segmentation of the input images, and he used the image as a feature. The system used a combination of features including PCA, hand trajectory, hand position, and hand velocity. The grey scale image with a reduced frame size 195x165 pixels and downscaling to 32x32 pixels used **rwth-boston-104** database.

A system of the recognition and translation of the numbers was designed [8]. The system is consist of four main phases; pre-processing phase, feature extraction phase, interpolation phase and the classification phase. The extracted features are scale invariant and make the system more flexible. The experimental result revealed that the system was able to recognize a representing numbers from one to nine based on the minimum Euclidean distance between the numbers.

The research work in [9] introduced new two features for American Sign Language recognition: those are kurtosis position and principal component analysis PCA. Principal component analysis is used as a descriptor to provide a measure for hand orientation and hand configuration. PCA has been used before in sign language as a dimensionality reduction. As a local feature for measuring edges and reflecting the position, Kurtosis position is used. It used motion chain code that represents the movement of hand as a feature. The system input is a sign from RWTH-BOSTON-50 database, and the recognition error rate of the output is 10.90%.

A system for the recognition and translation of the Arabic letters was designed [10] . The system depends on the inner circle position on the hand contour and divides the rectangle surrounding by the hand shape into 16 zones. The extracted features are scale invariant. Experiments revealed that the system was able to recognize Arabic letters based on the hand geometry. The experiment results shown that the different signs gesture recognition rate of Arabic alphabet for were 81.6 %.

The research work in [11] used Adaptive Neuro-Fuzzy Inference system (ANFIS). The system used 30 Arabic sign language alphabets visually. The recognition rate of the system was 93.55%. The research work in [12] built an ArSL system and measures the performance of ArSL data collected. The system based on Polynomial Classifiers. It collected a 30 letter of ArSL. It collected the data by using gloves marked with six different colours at different regions as shown in Fig. 1 [12]. The recognition rate is 93.41 %



(a) (b)

Fig. 1. (a) Colored gloves and (b) output image segmentation

This paper is organizes as follows. Section two explains HMM classifier. Section three presents the proposed system. Section four shows the experimental data. Section five explains the experimental results. Section six presents the conclusions.

II. HMM Classifier

HMM is used as a classifier for speech [13] and used in sign language recognition systems. In HMM-based approaches, the information of each sign is modelled by a different HMM. The model that gives the highest likelihood is selected as the best model and the test sign is classified as the sign of that model [14]. It consists of a set of N states where the transition from each state to another state. It is denoted by Eq. 1:

$$\lambda = (A, B, \pi) \tag{1}$$

- The state transition probability distribution $A = \{a_{ij}\}$ where its elements represent the transition probability from each state to another state. State transition coefficients having the properties Eq. 2 and Eq. 3.

$$a_{ij} \geq 0 \tag{2}$$

$$i \geq 1 \ , j \leq N$$

$$\sum_{j=1}^{N} a_{ij} = 1 \tag{3}$$

- The observation symbol probability distribution in state j, $B = \{b_j(k)\}$ where its elements represent the probability of certain observation to occur at a particular state$\{1 \leq j \leq N \ , 1 \leq k \leq M\}$, where M is a number of observation sequence O1 O2. . . OM
- The initial state distribution $\pi = \{\pi_i\}$, $1 \leq j \leq N$

III. Proposed System

The proposed system, as shown in Fig. 2, consists of five phases, skin detection, removing background, face and hands isolating, Observation detection, and Hidden Markov Model HMM classifier. The Maximum recognition probability Wi, i = 1. . .N, where N is a number of letters, is corresponding to letter detection.

The system components described in the following subsections: Sub section (A) presents skin detection and removing background. . Sub section (B) presents face and hands isolating. Sub section (C) presents observation detection and HMM.

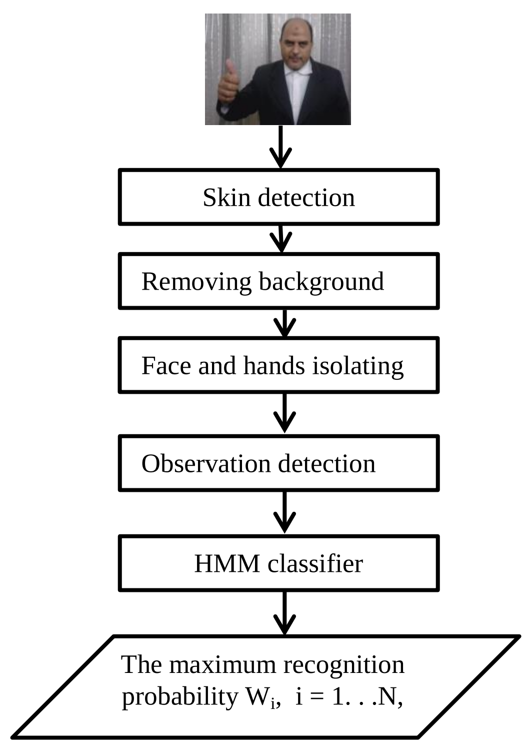


Fig. 2. Proposed system architecture

A. *Skin Detection and Background Removal*

The algorithm uses skin detection [15]. The algorithm adopts skin colour detection as the first step. Due to YCbCr color space transform, YCbCr is faster than other approaches

[16, 17]. The algorithm calculates the average luminance Y_{avg} of the input image as given in Eq.4. $Y_{avg} = \sum y_{i,j}$ (4)

Where $y_{i,j} = 0.3\,R + 0.6\,G + 0.1\,B$ is normalized to the range {0 to 255}, where i, j are the indices of the pixel in the image.

According to Y_{avg}, the algorithm can calculate the compensated image $C_{i,j}$ by the following equations Eq.5 and Eq.6 [15]:

$$R'_{i,j} = \left(R_{i,j}\right)^{\tau}$$

$$G'_{i,j} = \left(G_{i,j}\right)^{\tau} \quad (5)$$

$$C_{i,j} = \left\{R'_{i,j}, G'_{i,j}, B_{i,j}\right\}$$

$$\text{Where} \quad \tau = \begin{cases} 1.4, & Y_{avg} < 64 \\ 0.6, & Y_{avg} > 192 \\ 1, & \text{otherwise.} \end{cases} \quad (6)$$

It should be noted that the algorithm compensates the colour of R and G to reduce computation. Due to chrominance (Cr) which can well represent human skin, the algorithm only consider Cr factor for colour space transform to reduce the computation. Cr is defined as follows Eq. 7 [17]: $Cr = 0.5R' - 0.419G' - 0.081B$ (7)

Accordingly, the human skin binary matrix can be obtained as follows:

$$S_{ij} = \begin{cases} 0, & 10 < Cr < 45 \\ 1, & \text{otherwise} \end{cases} \quad (8)$$

Where '0' is the white point and '1' is the black point. The algorithm implements a filtration by a 5 × 5 mask.

First, the algorithm segments Sij into 5×5 blocks, and calculate show many white points in a block. Then, every point of a 5 × 5 block is set to white point when the number of white points is greater than half the number of total points. Otherwise, if the number of black points is more than a half, this 5 × 5 block is modified to a complete black block, as shown in Fig. 3 [16].

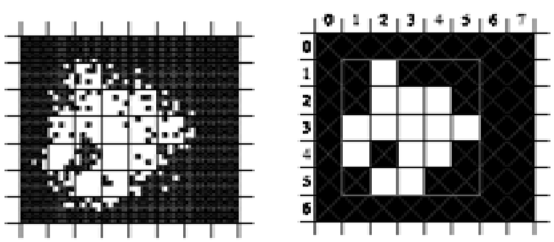


Fig. 3. (a) An example of Sij (b) Noise removal by the 5×5 filter

Fig. 4 shows the resultant image shapes after skin detection and removing the background [18] of image.

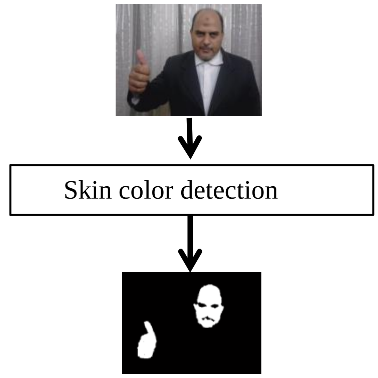


Fig. 4. Skin colour detection and removing background

B. Face and Hand Isolating

The algorithm tracks the objects in each image. The algorithm neglected the small objects, and then detects the largest objects as hands and the face. The algorithm isolates the hand and face as in Fig. 4. After detecting the skin colour and removing background the position of the face and hands can be isolate and detected as Fig.5.

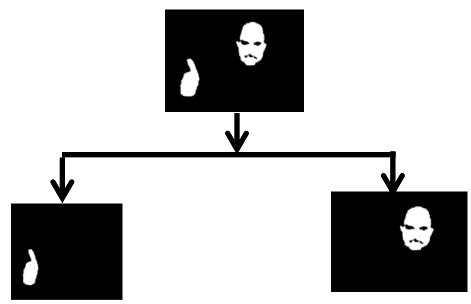


Fig. 5. Isolating the face and hands

Figure 4 shows the detected skin with background removal. The image contains a right hand and a face. The algorithm detects the hand and a face by the position and shape of each. Fig.5 shows isolating the face and hands, then isolate the right hand to detect the letter.

C. Observation Detection and HMM

The proposed algorithm divides the rectangle surrounding by the hand shape in Fig. 5 into 16 zones as in Fig. 6.

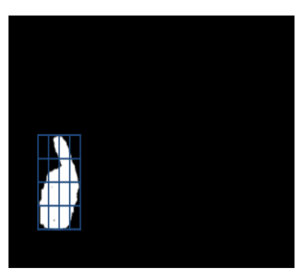


Fig. 6. Hand contour detecting

Col 4	Col 3	Col 2	Col 1	
4	3	2	1	Row1
8	7	6	5	Row 2
12	11	10	9	Row 3
16	15	14	13	Row 4

Fig. 7. 16 Zones hand shape

Fig. 8 summarize a creation of observation sequence and compute the HMM $\lambda = (A, B, \pi)$ for each letter and test a letter to get maximum $P(O|\lambda)$.

Step 1: Divide the rectangle surrounding by the hand into 16 zones as Fig. 6 and Fig. 7.

Step 2: Count the number of white pixel in each zone.

Step 3: Sort the zone numbers in ascending order depends into a number of white pixels in each zone.

Step 4: The observation vector of the letter is a vector of a sorted zone numbers, numbers between 1 and 16.

Step 5: Train the HMM for each letter $\lambda = (A, B, \pi)$ to maximize $P(O|\lambda)$

Step 6: To test a letter: Given the observation sequence $O=O_1O_2 \ldots O_{16}$ and a model $\lambda = (A, B, \pi)$ for this letter, then compute $P(O|\lambda)$ for each letter. The target letter is the maximum $P(O|\lambda)$.

Fig. 8. The proposed algorithm of calculating the observation vector and using HMM to train and test the letters

IV. EXPERIMENTAL DATA

To tune and test the proposed system, Arabic Alphabet Sign Language Recognition ArASLR database (ArASLRDB) is generated as follows.

The ArASLRDB corpus consists of 29 alphabet Arabic letters and numbers from 0 to nine as shown table 1. Clothes of signers are differently and the brightness of their clothes is completely different from person to other.

The image of the database is 640 x 480 pixels saved in jpg file format. The implementation is using the following as table 1:

Number of signs: 38.

Number of images: 357.

Number of training images: 253.

Number of testing images: 104.

Average images per sign: 9.4.

Average training images per sign: 6.7.

An average test image per sign: 2.7.

Percentage of training images per sign: 70.87%.

Average testing images per sign: 29.13%.

The programe is implemented using a Windows based MATLAB (R2013a).

TABLE I. A`RABIC ALPHABET DATABASE

	Arabic Sign name	English Sign Name	Number of images		Arabic Sign name	English Sign Name	Number of images
1	أ	Alef	11	20	ف	fa3	7
2	ب	Ba3	8	21	ق	Kaaf	11
3	ت	Ta3	10	22	ك	Kaf	9
4	ث	Tha3	7	23	ل	Laam	11
5	ج	Geem	12	24	م	Meem	10
6	ح	Ha3	10	25	ن	Noon	14
7	خ	Kha3	7	26	هـ	ha3	6
8	د	Dal	8	27	و	Waw	10
9	ذ	Thal	10	28	ى	Ya3	11
10	ر	Ra3	8	29	لا	Laa	10
11	ز	Zay	10	30	1	1	11
12	س	Seen	10	31	2	2	10
13	ش	Sheen	6	32	3	3	10
14	ص	Sad	8	33	4	4	6
15	ض	Dad	10	34	5	5	14
16	ط	Da3	9	35	6	6	9
17	ظ	Thaa3	8	36	7	7	11
18	ع	Aien	6	37	8	8	8
19	غ	Ghain	10	38	9	9	11

V. Experimental Result

In this research, the HMM is applied on the ArASLRDB with 29 Arabic alphabet sign language. The recognition system is tested when dividing the rectangle surrounding by the hand shape in Fig. 5 into 4, 9, 16, and 25 zones.

- At 16 zones: the recognition rate changes with changing the number of states until arrive to 100 % at 19 states, as shown in Fig.9.
- At 4 and 9 zones: the recognition rate cannot arrive to 100%, as shown in Fig. 9.
  - At 4 zones: the recognition rate is very poor and cannot exceed 40%.
  - At 9 zones: the recognition rate cannot exceed 97% whatever increases the number of states.
- At 25 zones: the recognition rate changes with changing the number of states until arriving 100 % at 18 states, as shown in Fig.9.
- The average time to execute the proposed algorithm to arrive of 100% recognition rate is shown in table 2.

TABLE II. Result of the Proposed System for Different Number of Zones

Zone number	State number	Time (Sec)	Recognition Ratio
4 zones	30	0.0332	37.5 %
9 zones	30	0.0708	96.93 %
16 zones	19	0.0902	100 %
25 zones	18	0.1065	100 %
36 zones	13	0.1358	100 %
49 zones	12	0.1531	100 %
64 zones	11	0.1872	100 %

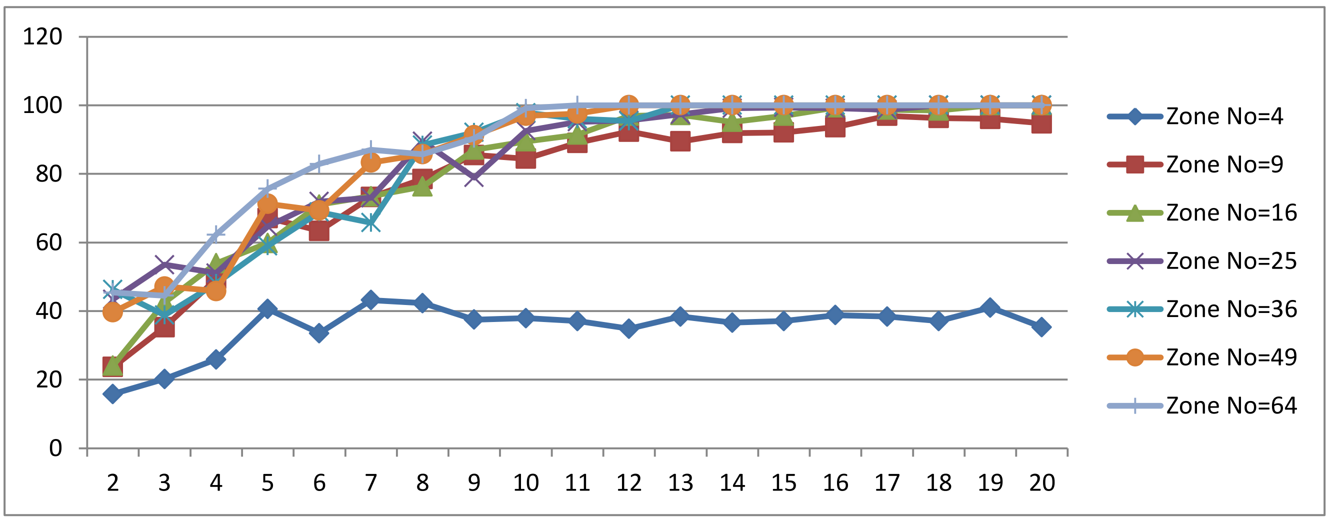


Fig. 9. The recognition rate verse us the change of the number of states for 4, 9, 16, and 25 zones

Finally, the best number of zones=16 with 19 states to recognition Arabic alphabet of sign language. The algorithm can achieve to 100% recognition rate with increase the zone number more than or equal 16 zones but more time is required.

As shown in table 3, Reference [2] used minimum distance classifier (MDC) and also used multilayer perceptron (MLP) classifier to detect 15 characters only of letters with recognition rate 91.7 % and 83.7 % respectively. Reference [11] recognized Arabic letters based on the hand geometry and the recognition rate of Arabic alphabet for different signs was 81.6 %. This system can reach a 100 % recognition rate with increasing number of zones and number of states. Reference [13] used Gloves marked with six different colour, the system used polynomial classifiers to recognize 30 letters with recognition rate of 93.41 %. Reference [12] did not use gloves and used ANFIS to recognize 30 letters by recognition rate of 93.55 %.

TABLE III. COMPARISON WITH ARSL ALPHABET RECOGNITION

	Instruments used	Number of Letters	Classifier	Recognition Rate
ArASLR	**None: Free Hands**	**29 Letters and 9 Numbers**	**HMM**	**100 % on more than 16 zones**
EL-Bendary etl. [2]	None: Free Hands	15 Letter	MDC MLP	91.3% 83.7 %
Jarrah, etl [12]	None: Free Hands	30 Letters	ANFIS	93.55 %
Assaleh, etl [13]	Gloves marked with six different colour	30 Letters	polynomial classifiers	93.41%
ArSLAT [11]	None: Free Hands	29 Letter	Outer of the inner circle zones	83.16

VI. CONCLUSIONS

In this paper, a new feature is used to recognize the Arabic Alphabet sign language via HMM. The proposed system is demonstrated experimentally. The phases of the proposed algorithm consists of skin detection, background exclusion, face and hands extraction, feature extraction, and also classification using Hidden Markov Model (HMM). The proposed algorithm isolates the hand from the image to recognize the letter. The proposed algorithm divides the rectangle surrounding by the hand shape into zones. The best number of zones is 16 zones. The observation of HMM is created by sorting zone numbers in ascending order depending on the number of white pixels in each zone. Experimental results show that the proposed algorithm achieves 100% recognition rate with minimum execution time at 16 zones with 19 states.

REFERENCES

[1] El-Bendary, Nashwa, et al. "ArSLAT: Arabic Sign Language Alphabets Translator." 2010 International Conference on Computer Information Systems and Industrial Management Applications (CISIM), 2010.

[2] A. A.Youssif, Amal Elsayed Aboutabl, Heba Hamdy Ali,"Arabic Sign Language (ArSL) recognition system using HMM ", IJACSA, vol. 2, no. 11, 2011.

[3] E.J. Holden, G. Lee and R. Owens, Automatic recognition of colloquial Australian sign language, IEEE Workshop on Motion and Video Computing 2, pp.183–188, Dec. 2005.

[4] F. Chen, C. Fu and C.,"Hand gesture recognition using a real time tracking method and hidden Markov models", Huang, Image and Vision Computing 21, pp. 745–758, Mar. 2003.

[5] Starner, "Visual Recognition of American Sign Language Using Hidden Markov Models", Doctoral dissertation, Massachusetts Institute of Technology, 1995.

[6] M. Al-Rousan, O. Al-Jarrah, and M. Al-Hammouri, "Recognition of dynamic gestures in Arabic sign language using two stages hierarchical scheme", The International Journal of Intelligent and Knowledge Based Engineering Systems, vol. 14, no. 3, 2010.

[7] Rybach, D., "Appearance-based features for automatic continuous sign language recognition", Diploma Thesis, RWTH Aachen University, Germany, 2006.

[8] Mahmoud Zaki Abdo, Alaa Mahmoud Hamdy, Sameh Abd El-rahman Salem, El-sayed Mostafa Saad. "An interpolation based technique for sign language recognition." Radio Science Conference (NRSC), 30th National. IEEE, 2013.

[9] Zaki M. M., Shaheen S. I., "Sign language recognition using a combination of new vision based features, Pattern Recognition Letters", vol. 32, Issue 4, pp. 572-577, 2011.

[10] Mahmoud Zaki Abdo, Alaa Mahmoud Hamdy, Sameh Abd El-rahman Salem, El-sayed Mostafa Saad," Arabic sign language recognition", International Journal of Computer Applications, pp. 0975 – 8887, vol. 89, no. 20, Mar. 2014.

[11] Al-Jarrah, Omar, and Alaa Halawani. "Recognition of gestures in Arabic sign language using neuro-fuzzy systems." Artificial Intelligence 133.1-2, pp. 117-138, 2001.

[12] Assaleh, Khaled, and M. Al-Rousan., "Recognition of Arabic sign language alphabet using polynomial classifiers", Journal on Applied Signal Processing , pp. 2136-2145, 2005

[13] Rabiner, Lawrence R., and Biing-Hwang Juang. "An introduction to hidden Markov models." ASSP Magazine, IEEE 3.1, pp. 4-16, 1986.

[14] Oya Aran, "Vision based sign language recognition: modelling and recognizing isolated signs with manual and non-manual components", 2008, Phd thesis.

[15] Pai, Y. T., Ruan, S. J., Shie, M. C., & Liu, Y. C. A simple and accurate color face detection algorithm in complex background. In Multimedia and Expo, IEEE International Conference on IEEE, pp. 1545-1548, 2006.

[16] S. Gundimada, Li Tao, and v. Asari, "Face detection technique based on intensity and skin color distribution", in 2004 International Conference on Image Processing, vol. 2, pp. 1413–1416, Oct. 2004.

[17] K. P. Seng, A. Suwandy, and L.-M. Ang, "Improved automatic face detection technique in color images", TENCON 2004, vol. 1, pp. 459–462, Nov. 2004.

[18] Khaled, H., Sayed, S. G., Saad, E. S. M., & Ali, H. "Hand gesture recognition using modified 1$ and background subtraction algorithms". Mathematical Problems in Engineering, 2015.

Comparative Study Between METEOR and BLEU Methods of MT: Arabic into English Translation as a Case Study

Laith S. Hadla
Department of Translation, Faculty of Arts, Zarqa University
Zarqa - Jordan

Taghreed M. Hailat
Faculty of IT and CS
Yarmouk University
Irbid, Jordan

Mohammed N. Al-Kabi
Computer Science Department
Zarqa University
P.O. Box 132222
Zarqa 13132, Jordan

***Abstract*—The Internet provides its users with a variety of services, and these services include free online machine translators, which translate free of charge between many of the world's languages such as Arabic, English, Chinese, German, Spanish, French, Russian, etc. Machine translators facilitate the transfer of information between different languages, thus eliminating the language barrier, since the amount of information and knowledge available varies from one language to another, Arabic content on the internet, for example, accounts 1% of the total internet content, while Arabs constitute 5% of the population of the earth, which means that the intellectual productivity of the Arabs is low because within internet use Internet's Arabic content represents 20% of their natural proportion, which in turn encouraged some Arab parties to improve Arabic content within the internet. So, many of those interested specialists rely on machine translators to bridge the knowledge gap between the information available in the Arabic language and those in other living languages such as English.**

This empirical study aims to identify the best Arabic to English Machine translation system, in order to help the developers of these systems to enhance the effectiveness of these systems. Furthermore, such studies help the users to choose the best. This study involves the construction of a system for Automatic Machine Translation Evaluation System of the Arabic language into language. This study includes assessing the accuracy of the translation by the two known machine translators, Google Translate, and the second, which bears the name of Babylon machine translation from Arabic into English. BLEU and METEOR methods are used the MT quality, and to identify the closer method to human judgments. The authors conclude that BLEU is closer to human judgments METEOR method.

Keywords—component; Machine Translation; Arabic-English Corpus; Google Translator; Babylon Translator; METEOR; BLEU

I. Introduction

The term Machine Translation (MT) dates back to the 1950's., and it is one of the earliest areas of research within Natural Language Processing (NLP) field. Until this moment, the accuracy of machine translation is lower than that of professional translators. There are different methods to translate from one natural language into another, and these methods are adopted by Online Machine Translation Systems such as Statistical Machine Translation (SMT), Hybrid Machine Translation (HMT), Rule-based, Knowledge-based, Interlingua, Direct, Transfer, and Example-based. The first two approaches (SMT and HMT) are the most widely used approaches nowadays.

The professional human translators are the best to evaluate the translation quality of Machine Translation Systems, but this way costs time, money, and effort consuming as the human translation. Therefore, many new methods are proposed by researchers to automatically evaluate the quality of the output of Machine Translation Systems. The utilization of these methods is not constrained to the automatic evaluation of MT systems, but it can be used in Software Development Life Cycle (SDLC) of MT systems, to enhance the efficiency of software under construction, analyze errors, and MT system benchmark. All these automatic MT evaluation methods depend on a core idea of making a comparison of the corresponding candidate translations and reference translations. We have to consider the fact that the correct human translation is not unique, and the list of valid reference translations is not limited. Therefore, this type of evaluation is considered a subjective, since it is highly correlated to human judgments (reference translations), and this leads to the difficulty. Manual (human) evaluation of MT is characterized by direct interpretation and accuracy relative to automatic evaluation of MT, but it costs money and time relative to automatic evaluation of MT.

Furthermore, the disadvantages of manual evaluation are non-reusability and subjectivity. On the other hand, automatic evaluation of MT is characterized by reusability, speed and free of charge, and it has a list of cons presented in the literature. The first generation of automatic MT evaluation methods depends on lexical similarity (n-gram -based) measures to compute their scores that represent the lexical matching between corresponding candidate Translations and reference translations [1].

Two widely used methods to automatically evaluate MT systems are used in this study. BLEU (Bilingual Evaluation Understudy) method is one of the earlier methods cast in this field, and it is used in this study. Therefore, as noted before, that the earlier methods of automatic MT evaluation depend on lexical similarity (n-gram -based) measures to compute their scores. BLEU score value is between 0 and 1, where 1

This research is funded by the Deanship of Research at Zarqa University / Jordan.

indicates the candidate translation and reference translation are fully matched, and 0 indicates the candidate translation and reference translation are completely different. BLEU values close to 1 indicates the similarity of the two translations is high, and BLEU values close to 0 indicate the similarity of the two translations is low. Those who use BLEU can benefit from its language independence and high correlation with human judgment. Furthermore, as other similar methods it has its pitfalls [1], [2]. The core idea of this widely used method based on the use of modified ngram precision, and so it needs to compute the number of common n-grams in the corresponding candidate and reference translations regardless of the position of matched n-grams. Then, the number of common words is divided by the total number of words in the candidate translation.

The second method used in this study is called METEOR 1.5 (Metric for Evaluation of Translation with Explicit ORdering) is an automatic evaluation metric for the machine translation output. Lavie, Kenji and Jayaraman study [3] proposes and casts METEOR metric for the first time in 2004, and aimed to improve correlation with human judgments of MT quality at the segment level. METEOR scores machine translation hypotheses by aligning them to one or more reference translations. Alignments are based on exact, stem, synonym, and paraphrase matches between words and phrases. METEOR has several features that are not found in BLEU, such as stemming and synonym matching, along with the standard exact word matching. On the other hand, BLEU and NIST metrics are based on precision alone, but the METEOR metric mentioned before uses precision and recall. Researchers proved that using precision and recall by METEOR leads to a higher correlation with human judgment at the sentence or segment level relative to metrics like BLEU and NIST [4]. Furthermore, METEOR score includes a fragmentation penalty that considers how well ordered the matched unigrams of the candidate translation are with respect to the reference. Carnegie Mellon University releases five versions of METEOR (Version 1.0, Version 1.2, Version 1.3, Version 1.4, and Version 1.5) on its Web page (http://www.cs.cmu.edu/~alavie/METEOR/).

Semitic languages include the following list of languages sorted according to native speakers: Arabic, Amharic, Hebrew, Aramaic, etc. The number of Arabic native speakers is widely varied from 220 to 400million people, besides Muslims who use it during the practice of their faith [5]. The Arabic language is one of the official languages used by all member states of the Arab league, and it is one of the UN official languages. Nowadays, two types of Arabic language are used, the first type is called Modern Standard Arabic (MSA) and it used mainly in official correspondence, books, journals, newspapers, etc., the second type includes a number of varying Arabic dialects that used in homes, markets, cafes, chatting, etc. Therefore, the spoken vernaculars are varied from country to country, and sometimes from village to a near village. The Arabic language is different from the English language since it has 28 Arabic letters, written from right to left in cursive style. The shape of the letter inside a word depends on its position (initial, medial, or final). The Arabic language lacks to the capitalization found in the English language [6].

This research is funded by the Deanship of Research at Zarqa University / Jordan.

Many studies exhibit different methods to improve MT of Arabic into other languages [7], [8], [9], [10], and [11].

In this study, researchers set out to conduct experiments that were used to benchmark the two methods (BLEU and METEOR 1.5) used for MT evaluation, and the two free online MT systems (Google Translate (https://translate.google.com) and Babylon system (http://translation.babylon.com/)) used to translate the 1033 Arabic sentences into English. Therefore, this study includes building an Automatic Machine Translation Evaluation System from Arabic into English using METEOR 1.5 and BLEU methods. The data set is divided to the sentence types (imperative, declarative, exclamatory, and interrogative). It is to be mentioned that these sentences were used in a previous study [12].

This paper is structured as follows: Section II reviews the related studies on automatic MT evaluation, and specifically those related to METEOR and BLEU methods. In Section III, we describe the framework and methodology of this study. In Section IV, we present our experimental results of the system designed and implemented by the second author and the results of two free online machine translation systems using a small data set consisting of 1033 Arabic sentences. Section V presents the conclusion from this paper. Finally, section VI presents plans to extend and improve this paper.

II. Related Work

Kirchhoff, Capurro, and Turner study [13] categorize the evaluation of machine translation (MT) into three main categories: human evaluation category, automatic evaluation category, and embedded application evaluation category.

This section starts with presenting studies related to Metric for Evaluation of Translation with Explicit Ordering (METEOR) method to automatically evaluate machine translation. Afterward, papers related to (BLEU) method are presented. Last and not least this section presents eleven studies related to the automatic evaluation of MT that includes Arabic.

Lavie, Kenji and Jayaraman in their study [3] cast a new metric in 2004 called METEOR to automatically evaluate MT systems. Some of the deficiencies of the BLEU score attempted to be addressed by METEOR metric. METEOR is based on a generalized concept of unigram matching between the Candidate Translation and Reference translation. METEOR flexibly matches unigrams using stemming and WordNet synonyms that does not require exact matching of words between the Candidate Translation and the Reference translation. This first metric METEOR attempts to determine all generalized unigram matches between the candidate translation and reference translation, then it starts computing METEOR score using a combination of unigram-precision, unigram-recall, and a measure of fragmentation to measure how well-ordered the matched words in the candidate translation are in relation to the reference. Banerjee and Lavie study [4] tests METEOR using LDC TIDES 2003 Arabic-into-English and Chinese-into-English data sets, to prove it yields better results than its counterparts (BLEU, NIST, Precision, Recall, F_1, and F_{mean}). In other words, they prove that

METEOR is closer to human judgments relative other metrics using the Arabic and Chinese data sets. Lavie and Agarwal study [14] presents an improvement to METEOR presented in [3] and [4]. This improved version supports additional languages like Spanish, French, and German, in addition to English, Arabic, and Chinese [14]. METEOR metric accounts for reordering to enhance the correlation with human judgments of translation accuracy [15].

Denkowski and Lavie [16] present in their study an improvement to METEOR metric. The new improved version is called METEOR-NEXT, and it includes METEOR's original features, besides paraphrases; and more sophisticated metrics that use deeper linguistic information.

Another study by Denkowski and Lavie [17] presents an improvement to METEOR metric. The new METEOR 1.3 includes improved text normalization, higher-precision paraphrase matching, and discrimination between content and function words. Furthermore, this improved version of METEOR metric includes ranking and adequacy versions and overcome some weaknesses of previous versions of METEOR such as noise in the paraphrase matching, lack of punctuation handling and discrimination between word types.

The second part of this section presents studies related to (BLEU) as a second method used to automatically evaluate MT quality.

It is mentioned before in this study that the number of valid reference translations of a certain source text is not limited to one, two, three, etc. valid reference translations. Based on this fact in 2002 Papineni et al. [2] cast BLEU to automatically evaluate the accuracy of the output of MT system using one, two, or more reference translations beside the corresponding candidate translation. BLEU is an n-gram based metric, where n ranges from 1 to 4. BLEU scores are highly affected by the number of reference translations, and that means the more reference translations per candidate translation there are, the higher BLEU score is. Therefore, BLEU requires multiple good reference translations.

Modified n-gram precision is an improved version of n-gram precision that aims to identify and avoid rewarding false positives outputted by MT systems. Brevity penalty is a correcting factor used to prevent short candidate translations relative to their reference counterparts from receiving a high a score. Small variations in candidate and reference translation lengths have a small impact on overall BLEU score. BLEU score is a product of multiplying modified n-gram precision at a sentence level by brevity penalty factor.

Although, many previous studies propose an enhanced BLEU method, only three studies are presented in this section due to space limitation.

Babych and Hartley [18] presents a modified version of BLEU method which uses Weighted n-gram Model that depends on a modified version of tf -idf to compute the weights of different words according to their importance, and S-score weights are used for translation Adequacy and Fluency. The S-score helps to weigh Content words differently from common words. DARPA-94 MT French-English evaluation corpus, which has 100 news text is used in their study to evaluate the effectiveness of the enhanced BLEU model. In their experiments, they used five MT systems to translate each of the 100 French news texts into English, and four of these candidate translations are evaluated by professional human translators. The corpus used by them has 2 English reference translations for each of the 100 French news texts. The results of their experiments reveal that their enhanced method scores for fluency are consistent with the base-line BLEU scores for fluency, but their enhanced method scores for adequacy outperform the base-line BLEU scores for adequacy. This modified version of BLEU can use only one reference translation, and yields a reliable result.

Another proposed extended BLEU method is presented in a study conducted by Yang, Zhu, Li, Wang, Qi, Li and Daxin [19]. They proposed assigning different weights to different part-of-speech (POS) and different lengths of n-gram. The information related to POS and lengths of n-gram are introduced to a linear regression model within the classical BLEU framework. This extension to BLEU does not affect the language independence of the original BLEU. Experimental results of the extended BLEU method show it is more effective than the baseline BLEU method.

An extended version of BLEU called AMBER is presented in Chen and Kuhn [20] study. The extended version of BLEU includes several new penalties instead of the brevity penalty used in the original BLEU. Furthermore, the computation of their metric includes text processing operations and the use of F-measure instead of precision and, therefore, they have to compute recall and precision before computing the F-measure. AMBER test results show it is more effective than the original BLEU metric. There is relatively little number of studies in the literature concerned with the evaluation of Arabic MT systems.

Therefore, Guessoum and Zantout [21] decided to evaluate four English-Arabic commercial MT systems (ATA, Arabtrans, Ajeeb, and Al-Nakel) using their new proposed to evaluate MT systems. The evaluation results show poor performance generally, except the lexical coverage of the domain of the Internet and Arabization.

Al-Haj and Lavie [22] study refers to the challenges facing statistical machine translation (SMT) such as Google Translate to translate from or into a morphologically rich language, and this challenge is magnified when translating into a morphologically rich language like Arabic. They addressed this challenge in the framework of a detailed description English-to-Arabic phrase-based statistical machine translation (PBSMT). Morphological segmentation and tokenization decisions have a great impact of the effectiveness of English-to-Arabic PBSMT outputs. Al-Haj and Lavie [22] present BLEU scores of different morphological segmentation schemes. Therefore, they deduce that a proper choice of segmentation has a significant effect on the performance of the SMT.

All the studies that use BLEU and METEOR methods use reference translations, in addition to candidate translations, but an interesting study conducted by Palmer [23] to automatically evaluate candidate translations depends on user-centered method and do not reply on reference translations. Palmer's method compares the outputs of MT systems and then ranking them, according to their quality. A number of professional users who have the necessary linguistics skills are used to rank candidate translations. The tests of Palmer's method [23] include seven MT systems, four Arabic-into-English MT systems, and three Mandarin (simplified Chinese)-into-English MT systems. Palmer study [23] is based on spoken language transcripts, and not on a textual data set.

Arabic dialects are the real languages used by most of the people in the Arab world to communicate with each other at homes, markets, restaurants, hospitals, etc. Arabs use many dialects that vary from a place to another, and Iraqi Arabic is one of these dialects that used in Iraq as the name indicates, and it is close to dialects used in the gulf region. Condon et al. study [24] presents an automatic method to evaluate the quality of Iraqi Arabic-English speech translation dialogues. They show that normalization has a positive effect on making the candidate translations closer to human judgments.

Adly and Al-Ansary [10] study presents an evaluation of an MT system that based on the interlingua approach, and Multilanguage MT system called Universal Network Language (UNL) system.

They address in their study [10] the evaluation of English-Arabic MT using three metrics BLEU, F_1 and F_{mean}, and conclude that UNL MT accuracy outperforms other MT systems. Alansary, Nagi, and Adly [25], and Al-Ansary [26] studied the effect of UNL MT system on translation from/into the Arabic language.

Carpuat, Marton, and Habash's [7] study overlaps with our study since it is concerned with translation from Arabic into English. They addressed in their study three main challenges: reordering, subject detection, and Arabic verb in Statistical Machine Translation. Furthermore, Carpuat, Marton, and Habash's [7] proposed a reordering of Verb Subject (VS) construction into Subject Verb (SV) construction for alignment only to minimize ambiguities. The results of their proposal show an improvement in BLEU and TER scores.

Alqudsi, Omar, and Shaker [27] conducted a good survey about available MT techniques, and exhibited some of the linguistic characteristics of the Arabic language with an emphasis on linguistic characteristics that have negative effects on MT. The study [27] presents a summary of the strengths and weaknesses of the main methods used in MT from Arabic into English.

A preliminary study is conducted by Hailat, Al-Kabi, Alsmadi, and Shawakfa [28] to evaluate the effectiveness of translation from English into Arabic of two free online MT systems (Google Translate and Babylon machine translation systems). They used a small data set that consists of 200 English sentences. BLEU was used to automatically evaluate the accuracy of each system. The evaluation results indicate Google Translate system is more effective than its counterpart.

The authors of the previous study [28] decide to improve their study using a larger data set of English sentences relative to the data set used in their previous study. They used in their new study [29] the same two online MT systems (Google Translate (https://translate.google.com) & Babylon (http://translation.babylon.com/) they used before, and they use the same method (BLEU) used before to automatically evaluate MT systems. They conclude that Google Translate is generally more accurate than its counterpart.

ATEC is another metric usually used to automatically evaluate the outputs of MT systems. The effectiveness of each of the 2 free Online Machine Translation systems "IMTranslator" and "Google Translate MT system" is explored by Al-Deek, Al-Sukhni, Al-Kabi, and Haidar [30] to conclude the Google Translate is more accurate than its counterpart.

A closely related study to our is conducted by Hadla, Hailat, and Al-Kabi study [12], to identify the best of two online machine translation systems (Google Translate and Babylon MT systems) to translate from Arabic into English. BLEU method is used by these authors to evaluate translation effectiveness of the above two online MT systems under consideration. They used more than 1000 Arabic sentences in their study to conduct their benchmark, where each Arabic sentence is accompanied by 2 reference English translations. The Arabic sentences they used are classified into four classes, where each class represents one of the four basic sentence functions (declarative, exclamatory, interrogative, and imperative). Hadla, Hailat, and Al-Kabi study [12] study concludes that Google Translate system is more accurate than Babylon MT system in terms of translation from Arabic into English.

III. The Methodology

This study is based on a data set constructed by Hadla, Hailat, and Al-Kabi [12], that consists of 1033 Arabic sentences with two reference translations of each Arabic sentence in the data set. This is an open access data set that can be downloaded from the following URL:https://docs.google.com/spreadsheets/d/1bqknBcdQ7cXOKtYLhVP7YHbvrlyJlsQggL60pnLpZfA/edit?usp=sharing

The 1033 Arabic sentences of the above data set is distributed among four basic sentence functions (declarative: 250 Arabic sentences, interrogative: 281 Arabic sentences, exclamatory: 252 Arabic sentences, and imperative: 250 Arabic sentences).

Figure 1 shows the main steps of the methodology, and how to extract n-grams from the Arabic, English Candidate, English Reference sentences, to be used to compute METEOR 1.5 and BLEU scores for Google Translate and Babylon machine translation system. Afterward, the closest score to the human judgment is determined.

METEOR method is used to automatically evaluate machine translation systems, and it uses word matching in target and reference translations to evaluate the accuracy of the machine translation. METEOR score is based mainly on word-to-word matches between target and reference translations. When more than one reference translation is available, the METEOR score is computed independently for each reference translation and the best METEOR score is adopted. METEOR consists of two main components, the first is a flexible monolingual word aligner component, and the second component is a scorer [32]. METEOR creates a word alignment between the two target and reference translations in the comparison process. Word alignment means the mapping between words in candidate and reference translations so that every word in each translation maps to at most one word in the other translation. Word-mapping modules are used to produce incremental alignment. These modules include modules such as the "exact" module that maps two words if they are fully matched. METEOR second module called "porter stem", and this module uses Porter Stemmer to yield stems that are mapped if they are fully matched. METEOR third module called "WN synonymy", and this module uses "synset" in WordNet to yield synonyms that mapped if they are fully matched [33]. The "porter stem" and "WN synonymy" modules do not support the Arabic language; therefore, Arabic is partially supported by METEOR 1.5 word-mapping modules [32].

The BLEU-score formula is a product of multiplying Brevity Penalty (BP) by geometric average of the modified n-gram precisions. Therefore, we have to start computing the geometric average of the modified n-gram precisions. Afterward, the length of the candidate translation (c), and the length of the effective reference corpus (r) have to be computed, in order to be able to compute the Brevity Penalty (BP). Formula 1 [2] shows how to make Brevity Penalty Reduced exponentially in (r / c).

$$BP = \begin{cases} 1 & \text{if } c > r \\ e^{\left(1-\frac{r}{c}\right)} & \text{if } c \le r \end{cases} \tag{1}$$

Formula 2 exhibits how to compute the final BLEU score.

$$BLEU = BP \times \exp\left(\sum_{n=1}^{N} w_n \log p_n\right) \tag{2}$$

Where $N = 4$ and uniform weights $w_n = (1/N)$ [2].

METEOR score is computed for each corresponding candidate and reference translations, and it proposed by (Banerjee, and Lavie, 2005) [4] to automatically evaluate MT. The range of METEOR score is between 0 and 1. Unlike base-line BLEU score that depends on precision only, METEOR's score uses recall in addition to the precision, with more emphasize on Recall. Furthermore, METEOR incorporates stemming and if English is the target WordNet is used to yield English synonyms.

The computation of METEOR score needs computing unigram precision (*P*), and unigram recall (*R*) first in order to be able to compute *F*-mean as shown in the following formula (3) [4]:

$$Fmean = \frac{10PR}{R+9P} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots..(3)$$

Afterward, METEOR method computes a penalty for a given alignment as shown in the following formula (4) [4]:

$$Penalty = 0.5\left(\frac{\#chunks}{\#unigrams\_matched}\right) \ldots\ldots\ldots\ldots(4)$$

The formula of computing the final METEOR score is shown in the following formula (5) [4]:

$$METEOR\,Score = Fmean\left(1 - Penalty\right) \quad \ldots\ldots\ldots\ldots..(5)$$

The higher score whether it represents BLEU or METEOR means that the candidate is closer to reference translation. Therefore, the higher BLEU or METEOR score means it is closer to human judgment. METEOR assigns a score in the range of 0 to 1 to every candidate translation [31].

The values of the BLEU metric range from 0 to 1 [2]; where the value of 1 means that the candidate translation fully matched reference translation, and the value of 0 means that the candidate translation and the corresponding reference translation are completely different.

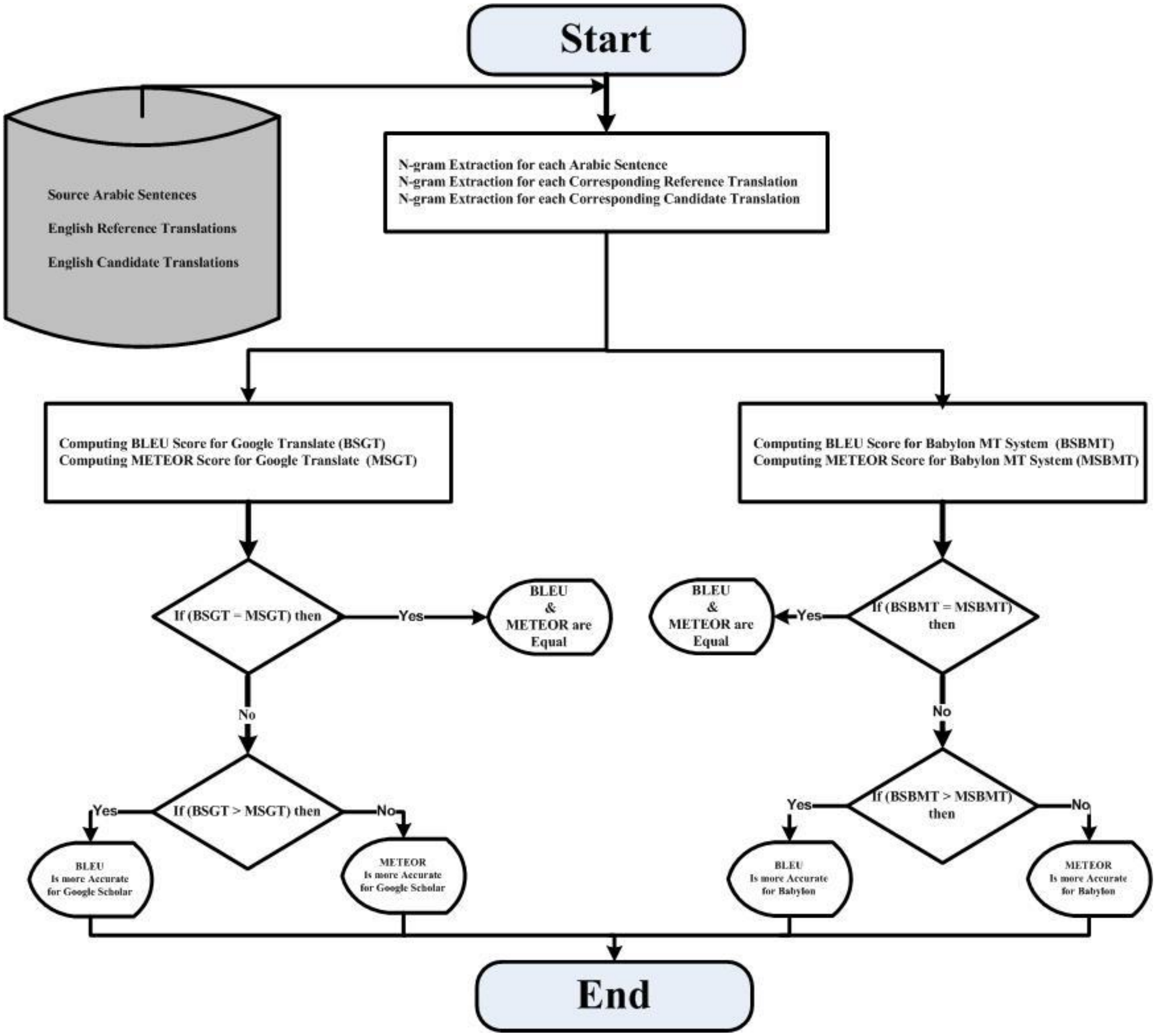


Fig. 1. Flowchart of BLEU and METEOR Evaluation Methodology

IV. The Evaluation

This study uses two well-known automatic evaluation methods (METEOR & BLEU). METEOR method measures precision and recall of unigrams when comparing a hypothesis translation against a reference one, and is characterized by its flexibility to match words (semantic matching) using stemming, paraphrase and WordNet synonyms [4].

When conducting tests on Google Translate and Babylon MT systems, the following was noted:

1) The conducted tests reveal within the outputs of Babylon MT system some Arabic words that indicate the incapability of Babylon MT system to translate these words into English, but, further investigation that these Arabic words with other inputted Arabic sentences are translated correctly into English.

Babylon MT system could not translate the words that contain related pronouns "الضمائر المتصله", for example in case of the source sentence in Arabic (I heard a word that made me laugh) " سمعتكلمةاضحكتني", was translated by Babylon MT as: "I heard the word اضحكتنى" and this note is already mentioned by the authors of [12].

TABLE I. BLEU AVERAGE PRECISION FOR EACH TYPE OF SENTENCES

Type / Translator	Declarative Sentence	Exclamation Sentence	Imperative Sentence	Interrogative Sentence	Average
Babylon MT System	0.3475	0.3686	0.5189	0.3588	0.39845
Google Translate System	0.4486	0.3378	0.5453	0.4668	0.449625

TABLE II. METEOR 1.5AVERAGE PRECISION FOR EACH TYPE OF SENTENCES

Respect to / Type	Ref1 + Google	Ref2 + Google	Ref1 + Babylon	Ref2 + Babylon
Declarative Sentence	0.39	0.32	0.33	0.30
Exclamation Sentence	0.36	0.28	0.20	0.19
Imperative Sentence	0.56	0.41	0.46	0.36
Interrogative Sentence	0.35	0.30	0.38	0.34
Average	0.415	0.328	0.343	0.298

2) Babylon machine translation system could not translate multiple Arabic sentences at one time while Google Translate has the feature of translating a set of Arabic sentences at one time.

The use of BLEU to evaluate and test these two online MT systems reveals that for some sentences the BLEU scores are equal to Google Translate & Babylon MT systems. However, the effectiveness of Google Translate system is generally better than the effectiveness of its counterpart as shown in Table 1.

Table 1 shows the evaluations of BLUE scores for each type of the 1033 Arabic sentences (imperative, declarative, exclamatory, and interrogative), and BLEU average precision values are presented for Babylon MT System and Google Translate System.

BLEU average precision values are used to identify the best MT system. Table 1 shows that Google Translate system is generally better than its counterpart, since it has a higher BLEU average precision. Furthermore, Table 1 shows Babylon MT System is better than Google Translate system in translating Arabic exclamation sentences into English. We have to note that the values of Table 1 are fully matched with those presented by [12] since we use their data set and their method (BLEU).

Table 2 shows the METEOR 1.5 average precision for the 4 types of sentences. Overall, the translations precision is below 50% except in imperative sentences that are translated by Google MT system (56%).

The values of BLEU and METEOR 1.5 methods in Table 1 and Table 2 generally show that Google Translate is more accurate than Babylon MT. Furthermore, these values do not imply that METEOR 1.5 method is more accurate than BLEU method. This deterioration in the scores of METEOR 1.5 relative to BLEU scores due to the Arabic language is not fully supported by METEOR 1.5, and therefore stems and synonyms are not used by METEOR 1.5 system.

V. CONCLUSION

In this study, two well-known Automatic MT Evaluation methods (METEOR 1.5 & BLEU) are used to identify the evaluation method that is closer to human judgment. Furthermore, this study includes tests to the effectiveness of two online MT systems (Google Translate & Babylon MT) systems to translate the 1033 Arabic sentences into English.

Most of the methods used to automatically evaluate the accuracy of the translation of MT system are based on a comparison between candidate and reference translations. This type of studies need a standard corpus, but, unfortunately, no standard corpus accepted by all researchers was found, except for an Arabic-English data set constructed and released by Hadla, Hailat, and Al-Kabi [12] to be used by the researchers was found. Therefore, the present study used this data set.

The second author developed an Arabic BLEU System to compute the BLUE score. We have found out in Table 1 that the overall translation precision for Google Translate system is 0.449625, and the overall translation precision for the Babylon MT system is0.39845using BLEU method. On the other hand, when METEOR 1.5 [35] is used we found in Table 2, that the overall translation precision for Google Translate system is 0.3715 ((0.415 + 0.328)/2), and the overall translation precision for the Babylon MT system is 0.3205 ((0.343+0.298)/2).

The second author developed an Arabic BLEU System to compute the BLUE score. Blue method results are presented in Table 1, and it clearly shows that the overall translation precision for Google Translate MT system is approximately 0.45, and the overall translation precision for the Babylon MT system is approximately 0.4. Therefore, it is deduced that the translation accuracy from Arabic into English of Google Translate MT system is more accurate than its counterpart. On the other hand, when METEOR 1.5 [34] is used, it is found, in Table 2, that the overall translation precision for Google Translate system is 0.3715 ((0.415 + 0.328)/2), and the overall translation precision for the Babylon MT system is 0.3205 ((0.343+0.298)/2). Once again, it is deduced that the translation accuracy from Arabic into English of Google Translate MT system is more accurate than its counterpart.

Babylon MT system proves to be more effective in translating exclamatory Arabic sentences to English.

Furthermore, it is concluded that BLEU method is closer to human evaluation than METEOR 1.5 for the translation from Arabic into English. This unexpected result is due to the fact that version METEOR 1.5, which does not fully support the Arabic language, is used in this study.

VI. Future Work

As future work, we would like to extend the scope of the study, by using a larger data set of sentences, use more automatic evaluation MT methods like ROUGE, NIST and RED, and use more languages.

Furthermore, we plan to study the effect of using an Arabic Stemmer like Khoja Stemmer [35] on the results of METEOR method.

References

[1] J. Giménez, and L. Màrquez, “Linguistic measures for automatic machine translation evaluation,”*Machine Translation* vol.24, no. 3-4, pp. 209-240, 2010.

[2] K. Papineni, S. Roukos, T. Ward, and W.J. Zhu, “*BLEU: a method for automatic evaluation of machine translation*,” in Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02). Stroudsburg, PA, USA, pp. 311-318, 2002.

[3] A. Lavie, K. Sagae, and S. Jayaraman, "*The Significance of Recall in Automatic Metrics for MT Evaluation*," In Proceedings of the 6th Conference of the Association for Machine Translation in the Americas (AMTA-2004), pp. 134–143, Washington, DC, September, 2004.

[4] S. Banerjee, and A. Lavie, "*METEOR: An Automatic Metric for MT Evaluation with Improved Correlation with Human Judgments*" in Proceedings of Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization at the 43rd Annual Meeting of the Association of Computational Linguistics (ACL-2005), Ann Arbor, Michigan, June 2005.

[5] Arabic language - *Wikipedia*, the free encyclopedia. [cited 06 July 2015]. Available from:http://en.wikipedia.org/wiki/Arabic_language

[6] K. C. Ryding, “*A Reference Grammar of Modern Standard Arabic*,” Cambridge: Cambridge University Press, 2005.

[7] M. Carpuat, Y. Marton, and N. Habash, “*Improving Arabic-to-English Statistical Machine Translation by Reordering Post-verbal Subjects for Alignment*,” in Proceedings of the ACL 2010 Conference Short Papers, pp. 178–183, Uppsala, Sweden, 2010.

[8] R. Al Dam, and A. Guessoum, “*Building a neural network-based English-to-Arabic transfer module from an unrestricted domain*," In Proceedings of IEEE International Conference on Machine and Web Intelligence (ICMWI), pp.94-101, 2010.

[9] J. Riesa, B. Mohit, K. Knight, and D. Marcu, “*Building an English-Iraqi Arabic Machine Translation System for Spoken Utterances with Limited Resources*", In the Proceedings of INTERSPEECH, Pittsburgh, USA, 2006.

[10] Adly, N. and Alansary, S. 2009, “*Evaluation of Arabic Machine Translation System based on the Universal Networking Language*,” in Proceedings of the 14th International Conference on Applications of Natural Language to Information Systems “NLDB 2009”, pp. 243-257, 2009.

[11] Y. Salem, A. Hensman, and B. Nolan, “Towards Arabic to English Machine Translation,”*ITB Journal*, Issue 17, pp. 20-31, 2008.

[12] L. S. Hadla, T. M. Hailat, and M. N. Al-Kabi, "Evaluating Arabic to English Machine Translation," *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 5, No. 11, pp. 68 - 73, 2014.

[13] K. Kirchhoff, D. Capurro,and A. M. Turner,“A conjoint analysis framework for evaluating user preferences in machine translation,”*Machine Translation*, vol.28, no. 1,pp. 1-17, 2014.

[14] A. Lavie, and A. Agarwal, "*Meteor: An Automatic Metric for MT Evaluation with High Levels of Correlation with Human Judgments*," in Proceedings of the Second Workshop on Statistical Machine Translation, Prague, pp. 228–231, June 2007.

[15] A. Lavie, M. J. Denkowski, "The Meteor metric for automatic evaluation of machine translation," *Machine Translation*, September 2009, Volume 23, Issue 2-3, pp. 105-115, 2009.

[16] M. Denkowski, and A. Lavie, "*Extending the METEOR Machine Translation Evaluation Metric to the Phrase Level*," Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, California, pp. 250-253, 2010.

[17] M. Denkowski and A. Lavie, “*Meteor 1.3: automatic metric for reliable optimization and evaluation of machine translation systems*,” in Proceedings of the 6th Workshop on Statistical Machine translation of the Association for Computational Linguistics (ACL-WMT '11), pp. 85–91, ACL Press, 2011.

[18] B. Babych, and A. Hartley, “*Extending the BLEU MT evaluation method with frequency weightings*,” In Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL '04). Association for Computational Linguistics, Stroudsburg, PA, USA, Article 621, 2004.

[19] M. Yang, J. Zhu, J. Li., L. Wang, H. Qi, S. Li., and L. Daxin, “*Extending BLEU Evaluation Method with Linguistic Weight*,” 2008. ICYCS 2008. The 9th International Conference for Young Computer Scientists, pp.1683-1688, 2008.

[20] B. Chen, and R. Kuhn, “*AMBER : A modified BLEU, enhanced ranking metric*,”in Proceedings of the 6th Workshop on Statistical Machine Translation, Edinburgh, UK, pp. 71-77, 2011.

[21] A. Guessoum, and R. Zantout, “A Methodology for Evaluating Arabic Machine Translation Systems,”*Machine Translation*, issue 18, pp. 299-335, 2005.

[22] H. Al-Haj, and A. Lavie A., “The Impact of Arabic Morphological Segmentation on Broad-coverage English-to-Arabic Statistical Machine Translation,”*Machine Translation*, vol. 26, no. 1-2, pp. 3-24, 2012.

[23] D. D. Palmer, “*User-centered evaluation for machine translation of spoken language*," in Proceedings. (ICASSP '05). IEEE International Conference on Acoustics, Speech, and Signal Processing, vol.5, pp. v/1013- v/1016, 2005.

[24] S. Condon, M. Arehart, D. Parvaz, G. Sanders, C. Doran, and J. Aberdeen, “Evaluation of 2-way Iraqi Arabic–English speech translation systems using automated metrics", *Machine Translation*, vol. 26, Nos. 1-2, pp. 159-176, 2012.

[25] S. Alansary, M. Nagi, and N. Adly, “*The Universal Networking Language in Action in English-Arabic Machine Translation*,” In Proceedings of 9th Egyptian Society of Language Engineering Conference on Language Engineering, (ESOLEC 2009), Cairo, Egypt, 2009.

[26] S. Alansary, “*Interlingua-based Machine Translation Systems: UNL versus Other Interlinguas*", in Proceedings of the 11th International Conference on Language Engineering, Cairo, Egypt, 2011.

[27] A. Alqudsi, N. Omar, and K. Shaker, "Arabic Machine Translation: a Survey", *Artificial Intelligence Review*, pp.1-24, 2012.

[28] T. Hailat,, M. N. AL-Kabi, I. M. Alsmadi, E. Shawakfa, “*Evaluating English To Arabic Machine Translators*,” 2013 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT 2013) - IT Applications & Systems, Amman, Jordan, 2013.

[29] M. N. Al-Kabi, T. M. Hailat, E. M. Al-Shawakfa, and I. M. Alsmadi, “Evaluating English to Arabic Machine Translation Using BLEU,”*International Journal of Advanced Computer Science and Applications (IJACSA)*, vol. 4, no.1, pp. 66-73, 2013.

[30] H. Al-Deek, E. Al-Sukhni, M. Al-Kabi, M. Haidar, “*Automatic Evaluation for Google Translate and IMTranslator Translators: An Empirical English-Arabic Translation*,” The 4th International

Conference on Information and Communication Systems (ICICS 2013). ACM, Irbid, Jordan, 2013.

[31] L. S. Hadla, T. M. Hailat, M. N. Al-Kabi, "Evaluating Arabic to English Machine Translation," *International Journal of Advanced Computer Science and Applications (IJACSA)*, SAI Publisher, 5(11), pp. 68-73, 2014.

[32] A. Agarwal and A. Lavie, "*Meteor, m-bleu and m-ter: Evaluation Metrics for High-Correlation with Human Rankings of Machine Translation Output*," in Proceedings of the Third Workshop on Statistical Machine Translation, Columbus, Ohio, pp. 115–118, 2008.

[33] *Meteor 1.5: Automatic Machine Translation Evaluation System, Code* by Michael Denkowski WebsiteGithub. [cited August 31, 2015]. Available from: http://www.cs.cmu.edu/~alavie/METEOR/README.html.

[34] Alon Lavie and Abhaya Agarwal. 2007. "*Meteor: an automatic metric for MT evaluation with high levels of correlation with human judgments*". In Proceedings of the Second Workshop on Statistical Machine Translation (StatMT '07). Association for Computational Linguistics, Stroudsburg, PA, USA, 228-231.

[35] M. Denkowski, A. Lavie, "*Meteor: Automatic Machine Translation Evaluation System*." [cited August 31, 2015]. Available from:http://www.cs.cmu.edu/~alavie/METEOR/

[36] Shereen Khoja - *Research*. [cited August 31, 2015]. Available from: http://zeus.cs.pacificu.edu/ shereen/research.htm

AUTHORS PROFILE

Laith Salman Hassan Hadla, born in Baghdad/Iraq in 1970. He obtained his PhD in Machine Translation from Al-Mustansiriya University in (2006), his masters' degree was in stylistic translation from Al-Mustansiriya University in (1995), and his bachelor degree is in Translation from Al-Mustansiriya University in (1992). Laith S. Hadla is an assistant professor at the Faculty of Arts, at Zarqa University. Before joining Zarqa University, he worked since 1993 in many Iraqi and Arab universities. The main research areas of interest for Laith S. Hadla are machine translation, translation in general, and linguistics. His teaching interests fall into translation and linguistics

Taghreed M. Hailat, born in Irbid/jordan in 1986. She obtained her MSc. degree in Computer Science from Yarmouk University (2012), and her bachelor degree in Computer Science from Yarmouk University (2008). Currently, she is working at Irbid chamber of Commerce as a Computer Administrator and previously as a trainer of many computer courses at Irbid Training Center

Mohammed Al-Kabi, born in Baghdad/Iraq in 1959. He obtained his Ph.D. degree in Mathematics from the University of Lodz/Poland (2001), his master's degree in Computer Science from the University of Baghdad/Iraq (1989), and his bachelor degree in statistics from the University of Baghdad/Iraq(1981). Mohammed Naji Al-Kabi is an assistant Professor in the Faculty of IT, at Zarqa University. Prior to joining Zarqa University, he worked many years at Yarmouk University in Jordan, Nahrain University and Mustanserya University in Iraq. He also worked as a lecturer in PSUT and Jordan University of Science and Technology (JUST). Al-Kabi's research interests include Information Retrieval, Sentiment analysis and Opinion Mining, Big Data, Web search engines, Machine Translation, Data Mining, & Natural Language Processing. He is the author of several publications on these topics. His teaching interests focus on information retrieval, Big Data, Web programming, data mining, DBMS (ORACLE & MS Access).

Analysis of Medical Domain Using CMARM: Confabulation Mapreduce Association Rule Mining Algorithm for Frequent and Rare Itemsets

Dr. Jyoti Gautam
HOD CSE and Associate Professor (CSE)
JSSATE
Noida, India

Neha Srivastava
Post Graduate of Computer Science & Engineering
JSSATE
Noida, India

***Abstract*—In Human Life span, disease is a major cause of illness and death in the modern society. There are various factors that are responsible for diseases like work environment, living and working conditions, agriculture and food production, housing, unemployment, individual life style etc. The early diagnosis of any disease that frequently and rarely occurs with the growing age can be helpful in curing the disease completely or to some extent. The long-term prognosis of patient records might be useful to find out the causes that are responsible for particular diseases. Therefore, human being can take early preventive measures to minimize the risk of diseases that may supervene with the growing age and hence increase the life expectancy chances. In this paper, a new CMARM: Confabulation-MapReduce based association rule mining algorithm is proposed for the analysis of medical data repository for both rare and frequent itemsets using an iterative MapReduce based framework inspired by cogency. Cogency is the probability of the assumed facts being true if the conclusion is true, means it is based on pairwise item conditional probability, so the proposed algorithm mine association rules by only one pass through the file. The proposed algorithm is also valuable for dealing with infrequent items due to its cogency inspired approach**.

Keywords—association rule mining; cogency; confabulation theory; medical data mining

I. Introduction

Data mining is one of the most vital and motivating area of research with the objective of finding meaningful information from huge data sets. In present scenario, Data mining is well accepted in healthcare field because there is a need of efficient analytical methodology for detecting unknown and valuable information in health data.

It also provides several benefits such as detection of the fraud in health insurance, availability of medical solution to the patients at lower cost, detection of causes of diseases and identification of medical treatment methods. It also helps the healthcare researchers for making efficient healthcare policies, constructing drug recommendation system, developing health profiles of individuals etc. Therefore, it can be concluded that Medical data mining has a great potential to discover the hidden relationship in the data sets of medical domain. This will help in understanding the prevailing situations in healthcare industry with respect to patients, their medical conditions and treatments. The most important thing in medical field is prognosis, i.e. an opinion, based on medical experience, of the likely course of a medical condition or the early diagnosis of any disease which helps in curing the disease at early stage and increase the life expectancy chances.

The data generated by the health organizations is very complex, heterogeneous in nature, and voluminous due to that it is difficult to analyze the data in order to make any important decision regarding patient health. So, there is a need to generate a robust, simple and computationally efficient tool for analyzing and extracting important information from such complex data sets. The development and understanding of such tools is the core business of data mining. Therefore, it is considered as an important area of study to explore the extent of association among such datasets. This is the reason that make association rule mining a crucial task.

Studies [11], [14] suggest that Confabulation association rule mining (CARM) approach, based on cognitive learning for generating association rules lead to generation of more interesting rules. CARM mine association rules by only one pass through the file as it is based on pairwise item conditional probability. Hence, it is more efficient for dealing with infrequent items. However, the confabulation based network is manually configured for selecting active items and interesting rules. So, Building such networks might require expert knowledge. Also, the value of parameters for finding items and for mining association rules must be re-configured when applied to different datasets. This limitation makes the Confabulation based approaches inflexible in handling different datasets. To address this limitation, an algorithm named CMARM: Confabulation MapReduce based association rule mining algorithm is proposed.

In the proposed algorithm, the concept of MapReduce framework of distributed computing is use for implementing the confabulation based association rule mining algorithm for both frequent and rare itemsets. MapReduce framework is regarded as the most successful computing platform for analyzing voluminous data as it adopts a data centric approach of distributed computing with the thought of “moving computation/processing closer to data”. Also it provides higher level of abstraction which keeps many system level details hidden from the programmers and allows them to focus more on the problem explicit computational logic.

Hadoop is use here for implementing the proposed algorithm which is an open source implementation of MapReduce framework written in Java. Also, Java is use to write map and reduce function in the discovery and the mining phase. To improve the performance and execution time of MapReduce job, data is compress while writing them in HDFS. The main focus of this work is to make an automatic procedure that learns the structure of a confabulation network from the input datasets and mine the association rules for frequent and rare itemsets by strengthening the knowledge link between itemsets.

The remainder of this paper is organized as follows. Section II outline the related works in finding frequent and rare itemsets and different algorithm used for association rules mining. The proposed CMARM algorithm is introduced in section III and here, also discuss the performance evaluation which will be done by comparing the proposed algorithm with those of classical methods by using a graph in terms of memory usage, execution time and number of produced rules. Lastly, section IV concludes this paper.

II. RELATED WORKS

Commonly, Association rule mining (ARM) approaches use the concept of support-confidence to measure the potency of association rules. This concept may work splendidly for frequent itemsets mining but it is not necessarily suitable when rare itemsets are required. It can be explain as if the value of minimum support is considered high, then only few interesting rules are generated; and if it is considered low, then many uninteresting rules may be generated [11]. Even for low minimum support, some interesting rules may be lost if their corresponding items are rare. One more problem of this concept is its computational complexity as it required multiple scanning of the data sets for finding the support of itemsets.

In recent years, the main focus is given to the nature inspired algorithms for solving many today's challenging problems, including different perspective to ARM. For instance, kuo and shih proposed ant-based algorithm in 2007 and also show that the proposed method is able to provide more condensed rules than Apriori method and also, computational time is reduced. Later, in 2010, Suneetha and Krishnamoorti proposed an Organized Transaction Selection Approach, in which the rules are generated by selecting transactions according to the highest order first basis and hence avoiding unnecessary patterns generation. Therefore, the major advantage of this approach is the reduction of database scans and hence overcome relatively higher time complexity of Apriori algorithm.

To improve the quality of the generated rules, some multi-objective algorithms have been developed with more measures considered, other than only confidence factor or predictive accuracy. Ghosh and Nath, in 2004, introduces comprehensibility, interestingness, and confidence factor as three interestingness measures of rules with the objective to model the association rule mining problem as a multi-objective problem [2]. The proposed algorithm used these three measures to strengthen the quality of generated rules but sampling decreases the accuracy. In 2014, A. Gupta, R. Arora, R. Sikarwar and N. Saxena proposed another algorithm for web usage mining using improved frequent pattern tree algorithm [1], in which the system operates in three phases namely; Preprocessing module, Apriori or FP growth algorithm module- comprises: frequent itemset generation followed by rules derivation. The main drawback of Apriori algorithm is that the candidate set creation is costly, especially if a large number of patterns exist. The main drawback of FP-growth algorithm is lacks of good candidate generation method. Future research can combine FP-Tree with Apriori candidate generation method to solve the disadvantages of both Apriori and FP-growth [1].

Later in 2014, M.Parmar, M. Sutaria and M. Joshi proposed an approach for finding out frequent item set using comparison based technique, in which the author used a MINHEAP tree algorithm in place of FP-Growth tree to find most frequent item which is already sorted and present in root of heap tree [12]. In the same year, 2014, K. Mutakabbir, S.Mahin and M. Hasan, find the frequent pattern within a genetic sequence using unique pattern indexing and mapping techniques and showed that the proposed algorithm required multiple scanning of the database [6]. Kuo proposed a particle swarm optimization – based approach for association rule mining and showed that it is more efficient and has better runtime than genetic algorithm (GA) as in GA for each fitness evaluation needs to calculate support and confidence for corresponding rules, which is time consuming. In comparison to the population– based algorithm, neural network is an exemplar that promises time efficiency and high performance. Logically, this can be proving by observing human's remarkable ability to efficiently process voluminous data. These networks are based on the concept of human nervous system, which is considered as one of the most intelligent creature on earth. The neural network is nonlinear model that is easy to use and understand compared to statistical methods and also non-parametric model while most of statistical methods are parametric model that need higher background of statistic. So far, there are only a few researches that have applied neural network to ARM technique. Self-organizing map (SOM) is type of Artificial Neural Network (ANN), which is widely used in many data mining tasks, such as clustering and dimensionality reduction. The reduction in dimensionality allows people to visualize and interpret what would otherwise be, for all intents and purposes, scrawled data. Nohuddin discovered frequent trends in social network and use SOM to cluster these trends for better understanding of the nature of the trends. Furthermore, Yeh proposed a new approach called association reasoning neural network (ARNN) based on multilayer perceptron- network, where the number of hidden neurons act as the support threshold that control the generation of rules with low support value.

In 2014, A. Soltani and M. R. Akbarzadeh proposed Confabulation inspired association rule mining (CARM) approach [11], based on cogency inspired measure for generating rules. Cogency inspiration can lead to more intuitive rules. Also, cogency-related computations need pairwise item co-occurrences; hence findings rules requires only one pass through the file. Since, file access, particularly for large files can be significantly time consuming, therefore it can be concluded that the proposed algorithm is superior to the

Apriori algorithm due to one-time file access. In this paper, the parameter for selecting active items and for selecting interesting rules are predefined by the user. It would be desirable to set these parameters automatically using data set statistics. Moreover, the use of matrices in the implementation of CARM seems to basically make this a problem of statistical analysis.

Recently, in March 2015, Mansurul A.Bhuiyan and Mohammad Al Hasan proposed a frequent subgraph mining algorithm called FSM-H [10]. Here, author used an iterative MapReduce based framework for FSM-H and showed that it is complete as it returns all the frequent subgraphs for a given user-defined support, and also efficient as it applies all the optimizations that the latest FSM algorithms adopt. This algorithm is limited to the size of a graph which is equal to the number of edges it contains.

In June 2015, Qiuwen Chen, Qing Wu, Morgan Bishop, Richard Linderman and QiuQiu proposed an algorithm for self-structured confabulation network for fast anomaly detection and reasoning [14]. In this paper, the author proposed an automatic procedure that learns the structure of a confabulation network from the incoming dataset. The constructed model consists of well-defined nodes that capture both spatial and temporal relations among the features of the dataset. This work will be further extended by improving the workload distribution among the network nodes so that testing instance can be dynamically assigned to multiple heterogeneous devices.

III. Methodology

The proposed algorithm CMARM is designed as an iterative MapReduce process using Confabulation approach to find out the association rules for frequent and rare itemsets. Confabulation theory is an information processing model of human cognition introduced by Hecht-Nielsen [19]. It is based on the concept of cognitive learning. Cognitive learning concerned with the acquisition of problem-solving abilities, intelligence and conscious thought. It uses existing knowledge and generates new knowledge; therefore, lead us to the generation of more intuitive rules.

A. MapReduce Framework Implemenation

MapReduce is a linearly scalable programming model that enables distributed computation over voluminous data. The model provides two abstract functions: map and reduce each of which defines a mapping from one set of key-value pairs to another. The map function corresponds to the "map" operation in functional programming, whereas the reduce function corresponds to the "fold" operation in functional programming. The input to a MapReduce job starts as data stored on the underlying distributed file system. The Mapper is applied to every input key-value pair to generate an arbitrary number of intermediate key-value pairs. The Reducer is applied to all values associated with the same intermediate key to generate output key-value pairs. The files (input and output) of MapReduce are managed by a distributed file system. Hadoop is an open source implementation of MapReduce framework written in Java languages.

B. CMARM Algorithm

In the proposed CMARM algorithm, an additional parameter along with support and confidence threshold values is used to select active items and interesting rules in association rule mining. The value of this additional parameter is computed using cogency approach that leads us to formulation of more intuitive rules.

The proposed algorithm is divided into two main modules:

1) Frequent and Rare itemsets discovery,

2) Generate Strong Association Rule using confabulation theory and cogency measure.

In the first module, MapReduce framework is used to represent the knowledge links with weak strength using the threshold value of support count. This phase results in generation of frequent and rare itemsets such that whose support count is greater than the minimum specified support count value. In the second module, the strong association rules are generated based on confabulation theory, which states;

$$Fr = \{x \in I \,|supp(x) \geq S_\theta\}$$

It considers S_1= Fr where S_1 is 1-itemsets. After finding all frequent and rare n-itemsets, the algorithm generates all rules using their support, confidence and interestingness parameter for selecting active items and interesting rules that strengthen the knowledge link between the itemsets. Here, the value of interestingness parameter is estimated by mapper in terms of changes for both user-feature and item-feature pair. Then, the Reducer calculate the sum of changes and apply the calculated sums to each item rating by updating appropriate feature and use this output value of reduce function for next iteration.

C. Design Model

The design model of proposed CMARM algorithm is shown below:

1) Initially, parallel scan first divide the dataset information horizontally into 'P' node subsets and distribute it to 'Q' nodes supersets.

2) The various 'P' nodes are then processed again using mapping function.

3) Then once the method is completed, every node scans its own information sets then generates set of Candidate item set Fr.

4) Initially, the support count of every Candidate itemsets is about to one. This Candidate itemsets Fr is split into R partitions and sent to 'R' nodes with their support value count. Here, Min_Support is also defining which describes value of minimum support count. The Candidate itemsets Fr is discarded whose support count is less than Min_Support.

5) Once the algorithm has collected the set of all frequent and rare itemsets of all sizes that survived the support threshold. The next step is to extract strong associated rules from frequent and rare itemsets.

6) CMARM uses MapReduce framework to extract significant rules from all frequent and rare itemsets. If all frequent and rare itemsets can fit in computer memory and if the processing time is not that big then Hash table data *structure can be used to hold the data thrown from the map function. In this case, the key will be the left-part and the value will be set of (right-part: support) entries for frequent and rare itemsets. In the distributed implementation of this step, data are thrown to distributed file system and the Map-Reduce middleware is responsible to sort the entries, to fetch them, groups and sent to the reduce functions. Here, Mapper also calculates the changes for both user-feature and item-feature pair. These changes are used as a parameter for selecting active items and interesting rules from the frequent and rare itemsets. Then, Reducer calculates the sum of changes and applies sums to each item rating by updating appropriate feature and sends this to mapper function again for next iteration.*

D. Implementation

The CMARM algorithm will implement using Hadoop which is open source of MapReduce framework written in Java language. The Map and Reduce function in the discovering and the mining phase of the Association Rule Mining will be written in Java Language. The choice to use Java language is motivated on several grounds. First, the construction of a runnable jar file is the easiest way to run a distributed program with Hadoop. Second, a Java implementation allows future integration with the Apache Mahout parallel machine learning library, a widely used package among machine learning end-users and researchers alike. To improve the performance and execution time of MapReduce job, the data is compress while writing them in HDFS. Also, global counter is use provided by Hadoop to track the stopping point of the iterative mining task.

Following are the Steps to implement proposed CMARM algorithm with Map Reduce framework:-

Step 1: Maps the input dataset to N partitions, where N = number of slave machines.

Step 2: Reduce phase would take the immediate key-value pairs emitted in the map phase.

Step 3: Send them altogether to the master node for further collecting the number for count per item.

3.1 To generate frequent and rare itemsets in form of key-value pair. This describes the number of occurrences of individual itemsets.

3.2 To generate the candidate sets from the source data file. It first discard those items that occur minimum than the support threshold value by looking up the global Hash map list and then recursively call mapping() function.

Step 4: Mapper calculates changes for both user-feature and item-feature pair, then using this value along with support and confidence strong association rule is generated.

Step 5: Reducer calculates the sum of changes and apply sums to each item rating by updating appropriate feature and use output of this reduce for next iteration.

E. Performance Evaluation

The relative performance of CMARM algorithm will analyze by comparing against two mining algorithm, one is CARM: Confabulation association rule mining algorithm and other is FIN algorithm, which is variant of FP tree algorithm. The Performance analysis will be shown by using a graph in terms of memory usage, execution time and number of produced rules. To verify the CMARM algorithm, both synthetic and real data sets are used, which is obtained from the health data repository.

IV. Conclusion

The foremost objective of this paper is to analysis of medical data repository to find out rare and frequent occurring diseases that may occur with growing age. So that, human being can take early preventive measures to minimize the risk of diseases that may supervene with the growing age and hence increase the life expectancy chances. Since, the data generated by medical domain is very vast and heterogeneous in nature; therefore, a new ARM algorithm is proposed called CMARM to mine the association rules. CMARM algorithm uses an iterative MapReduce based framework inspired by cogency. This algorithm uses the concept of cognitive learning for making clinical decisions which are often made based on doctor's intuition and experience rather than on the knowledge rich data hidden in the database. The proposed work can be further enhanced and expanded for the automation of long-term prognosis of diseases

References

[1] A. Gupta, R. Arora, R. Sikarwar, and N. Saxena, "Web usage mining using improved frequent pattern tree algorithm," IEEE conference on communication system and network technologies, pp. 573-578, 2014.

[2] A. Ghosh, and B. Nath, "Multi-objective rule mining using genetic algorithm," Information science, vol. 163, pp. 123-133, June 2004.

[3] D. Agnihotri, K. Verma, and P. Tripathi, "Pattern and cluster mining on text data," IEEE conference on communication system and network technologies, pp. 428-432, 2014.

[4] Faraz Rasheed, and Reda Alhaji, "A framework for periodic outlier pattern detection in time-series sequences," IEEE transactions on cybernetics, vol. 44, no.5, pp. 569-582, May 2014.

[5] G. Liu, H. Zhang, and L. Wong, " A Flexible approach to finding representative pattern sets," IEEE transaction on knowledge and data engineering , vol. 26 , no. 7, pp. 1562-1574, July 2014.

[6] K. Mutakabbir, S. Mahin, and M. Hasan, " Mining frequent pattern within a genetic sequence using unique pattern indexing and mapping techniques," IEEE conference on informatics electronics & Vision, 2014.

[7] K. Srinivas, B. Kavihta Rani, and Dr. A. Govrdhan, "Applications of data mining techniques in healthcare and prediction of heart attacks," International journal on computer science and engineering, vol. 02, no. 02, pp. 250-255, 2010.

[8] L. Cagliero, and P. Garza, "Infrequent weighted itemset mining using frequent pattern growth," IEEE transactions on knowledge and data engineering, vol. 26, no. 4, pp. 903-915, April 2014.

[9] L. Wang, David, Wai-Lok, and R. Cheng, "Efficient mining of frequent item sets on large uncertain databases," IEEE transactions on knowledge and data engineering, vol. 24, no. 24, pp. 2170-2183, December 2012.

[10] M. A. Bhuiyan and M. A. Hasan, "An iterative mapreduce based frequent subgraph mining algorithm," IEEE transactions on knowledge and data engineering, vol. 27, no. 3, pp. 608-620, March 2015.

[11] M. R Akbarzadeh-T, and A. Soltani, "Confabulation inspired association rule mining for rare and frequent itemsets," IEEE transactions on neural networks and learning systems," vol. 25, no. 11, pp. 2053-2064, November 2014.

[12] M. Parmar, M. Sutaria, and M. Joshi, "An approach for finding frequent itemset done by comparison based technique," International journal of computer science and mobile computing," vol.3, issue. 4, pp. 996-1001, 2014.

[13] M. Saravanan, and R. RJ, "Performance study of association rule mining algorithms for dyeing processing system," Innovative systems design and engineering, vol. 2, no. 5, pp. 34-43, 2011.

[14] Qiuwen Chen, Qing Wu†, Morgan Bishop†, R. Linderman† and Qinru Qiu, "Self-structured confabulation network for fast anomaly detection and reasoning," Air force research laboratory, Information Directorate, 2015.

[15] R. Agrawal, R. Srikant and others, "Fast algorithms for mining association rules," IEEE conference on informatics electronics & vision, vol. 1215, pp. 487-499, 1994.

[16] R. D. Canlas Jr., "Data mining in healthcare: Current applications and issues," International Journal on computer science and engineering," vol. 26, 2009.

[17] R. Jin, "An efficient implementation of apriori association web mining," Proc. workshop on high performance data web mining, April 2011.

[18] S. Bai and S. Bai, "The maximal frequent pattern mining of DNA sequence," IEEE conference on informatics electronics and vision, pp. 23-26, 2009.

[19] Srikant R, Agrawal R., "Mining sequential patterns: generalizations and performance improvements," Proceeding of 5th international conference on extending database technology (EDBT'96), pp. 3–17, Mar 1996.

[20] S. Q. Wang, Y. B. Yang, Y. Gao, G. P. Chen, and Y. Zhang, "Mapreduce based closed frequent itemset mining with efficient redundancy filtering," Proc. IEEE 12th International conference data mining workshops, pp. 49–453, 2012.

[21] Yang Xiang, Philip R.O. Payne, and Kun Huang, "Transactional database transformation and its application in prioritizing human disease genes," IEEE transactions on computational biology and bioinformatics, vol. 9, no. 1, pp. 294-304, January 2012.

[22] Z. Zhao, Da. Yan and W. Ng, "Mining Probabilistically Frequent Sequential Patterns in Large Uncertain Databases, IEEE Transactions on Knowledge and Data Engineering , vol. 26 , no. 5, pp.1171-1184, May 2014.

On Attack-Relevant Ranking of Network Features

Adel Ammar, Khaled Al-Shalfan
College of Computer and Information Sciences
Al-Imam Mohammad Ibn Saud Islamic University (IMSIU)
Riyadh, KSA

Abstract—**An Intrusion Detection System (IDS) is an important component of the defense-in-depth security mechanism in any computer network system. For assuring timely detection of intrusions from millions of connection records, it is important to reduce the number of connection features examined by the IDS, using feature selection or feature reduction techniques. In this scope, this paper presents the first application of a distinctive feature selection method based on neural networks to the problem of intrusion detection, in order to determine the most relevant network features, which is an important step towards constructing a lightweight anomaly-based intrusion detection system. The same procedure is used for feature selection and for attack detection, which gives more consistency to the method. We apply this method to a case study, on KDD dataset and show its advantages compared to some existing feature selection approaches. We then measure its dependence to the network architecture and the learning database.**

Keywords—Intrusion detection; network security; feature selection; KDD dataset; neural networks

I. Introduction

A network intrusion is any attempt or action aiming at compromising the confidentiality, integrity or availability of a computer or network. Intrusion detection systems (IDSs) are software or hardware systems that automate the process of monitoring the events occurring in a computer system or a network, in order to prevent intrusions, and detect threatening breaches in information security. Depending on the type of analysis performed, IDSs can be classified into two groups; a) signature-based, which rely on a regularly updated database of known attack signatures, and b) anomaly-based, which build a statistical model of the system's normal behavior, using machine learning or data mining techniques, and considers any excessive deviation from this behavioral profile to be an intrusion. During the last decade, anomaly detection has attracted the attention of many researchers to overcome the weakness of signature-based IDSs in detecting novel attacks. Nevertheless, due to the rapidly increasing network traffic, it becomes of significant interest for an anomaly-based IDS to rank the importance of input features, since the elimination of irrelevant or useless inputs leads to a simplification of the problem and may allow faster and more accurate detection. This is especially critical for the construction of an efficient real-time IDS able to comply with the constraints of high speed networks. In fact, some of the connection features may be irrelevant or redundant which results in lengthy detection process and degrades the performance of the IDS. For this aim, this article presents a feature selection method based on Neural Networks (NN), applied on the problem of classifying traffic features according to their relative contribution to attack detection.

Section II introduces the method and describes its theoretical basis. Section III details the results of a case study for both a single and multiple output classification NN and reviews the advantages and limitations of the method. Finally, section IV draws a conclusion for the present work and mentions some open issues for future works.

II. Theoretical Basis

Feature selection is the process of removing features from the original data set that are redundant or not very relevant with respect to the task that is to be performed, such as noisy features. Feature selection can be seen as a search problem on the power set of the set of available features, which is a combinatorial problem. The method proposed here for selecting connection features is a heuristic based on feed-forward neural networks. It has been applied in another application by [1] and theoretically formulated by [2] who called it HVS (Heuristic for Variable Selection). Nevertheless, it has not yet been applied to intrusion detection, to the best of our knowledge.

We introduce the features that need to be ranked as inputs of a feed-forward neural network (with a single hidden layer) used as a classifier that distinguishes attacks from normal traffic. After the training process on a representative learning database, we assess the relative contribution of each feature as follows. We expect the contribution C_{js} of a neuron j of the hidden layer to the output s according to the formula:

$$C_{js} = \frac{|W_{js}|}{\sum_{k=1}^{Nh} |W_{ks}|} \tag{1}$$

Where W_{ks} is the weight of the connection between a hidden neuron k and the output s and N_h is the number of hidden neurons. Then, we obtain the contribution of an input neuron i to the output according to the formula:

$$C_{is} = \sum_{j=1}^{N_h} C_{js} \cdot \frac{|W_{ij}|}{\sum_{k=i}^{N_i} |W_{kj}|} \tag{2}$$

```
Contribution(neuron_i, layer_j)
    // neuron_i belongs to layer_j
    If layer_j=number_of_layers then
return 1;  // Output layer
    C=0
    For k=1 to number_neurons(layer_i+1)
    // Layers are numbered ascendingly
from input to output
     C = C + weight(neuron_i, neuron_k) /
     sum_weights(layer_i, neuron_k)
     *Contribution(neuron_k, layer_i+1)
End
Return C
End
```

Algorithm 1. Computation of a neuron's contribution

Where W_{ij} is the weight of the connection between the input neuron i and a hidden neuron j and N_i, is the number of inputs. The sum of input contributions is, therefore, equal to 1. Note that (1) and (2) can be generalized to multiple outputs and multiple hidden layers and reduced to a single recursive formula if we define the contribution of output neurons as being equal to 1, according to Algorithm 1.

III. Case Study on KDD Database

A. Distinction between normal and abnormal traffic: single output NN

1) Calculation of features' contribution

We have applied the HVS method described above, in a case study, to the KDD 99 intrusion detection benchmark [3]. This database originated from the 1998 DARPA Intrusion Detection Evaluation Program that was prepared and managed by MIT Lincoln Labs. The objective was to assess and evaluate research in intrusion detection [4]. Lincoln Laboratories created an environment to acquire TCPdump data in a local area network (LAN) simulating a typical U.S. Air Force network hit by multiple attacks over nine weeks. The training data set collected during the first 7 weeks occupied reached about five million connection records, presenting 23 different attack types, whereas the test data obtained during the last two weeks accounted for around two million connection records, with 18 additional attack types. The dataset was then summarized into network connections with 41 features per connection (Table 1). In order to measure the relevance of these features, we constructed a NN with a single output that distinguishes between normal traffic and attacks. The learning database that we used to train the NN consists of a 1% random extraction (4,940 samples) from the original KDD learning set (containing 494,021 connection records). A learning database with such a size is sufficient to achieve an accuracy rate of 92% on the KDD test set (composed of 311029 independent connection records).

Fig. 1 depicts the obtained results, after applying the HVS method following (1) and (2). Features # 20 and 21 take a null contribution because they are constant in the whole KDD learning set. The same can be noticed for features # 9 and 15, which are almost constant. In fact, more than 99.999% of the KDD learning set connection records contain a null value for these two features. Features 7, 11 and 18 could also be excluded from the learning database since their contribution is remarkably little; while the most significant features are # 10, 22, 23, 34, 36, 39.

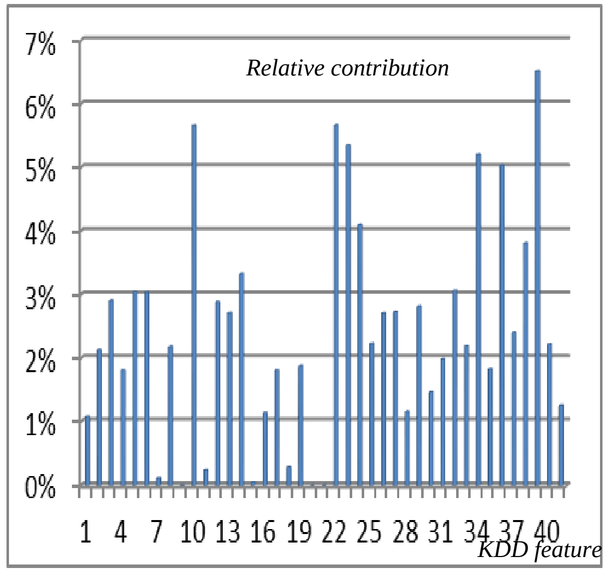


Fig. 1. Relative contribution of each of the KDD 41 features to the detection of attacks (distinction between normal traffic and attacks of various types)

2) Checking the consistence of the method

In order to verify the consistence of the results, we selected a set of most significant features (calculated as in the section above) to be set as inputs of the classification NN, and compared the results with those obtained with the full set of inputs. Figure 2 shows these results after applying the networks to the testing databases. We note that we can keep only the most influential 12 features (out of 41), without significantly deteriorating neither the overall accuracy rate (Figure 2) nor the false positive and false negative rates (FPR and FNR, Figure 3). With only 9 features, we still obtain a relatively small FNR (2%), while the FPR reaches 15%.

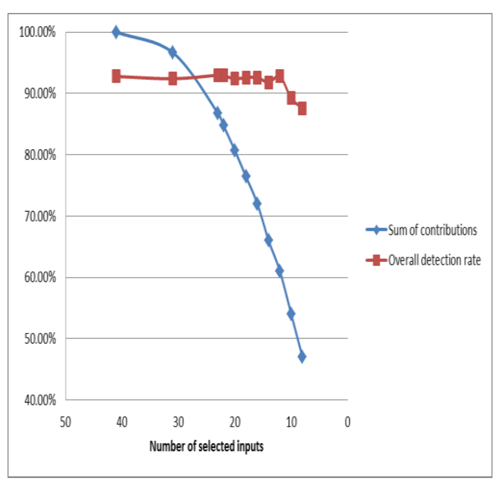


Fig. 2. Evolution of the overall accuracy rate according to the number of selected inputs

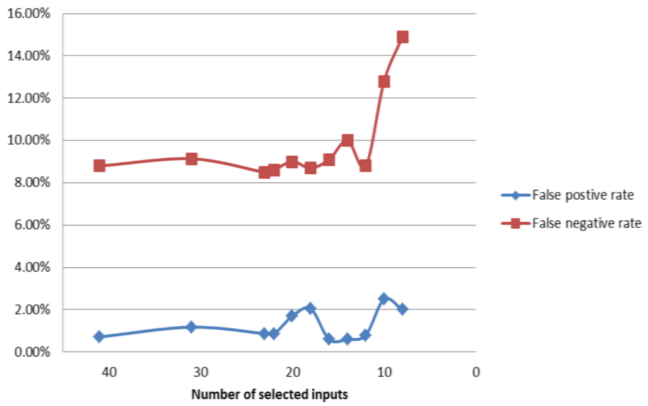
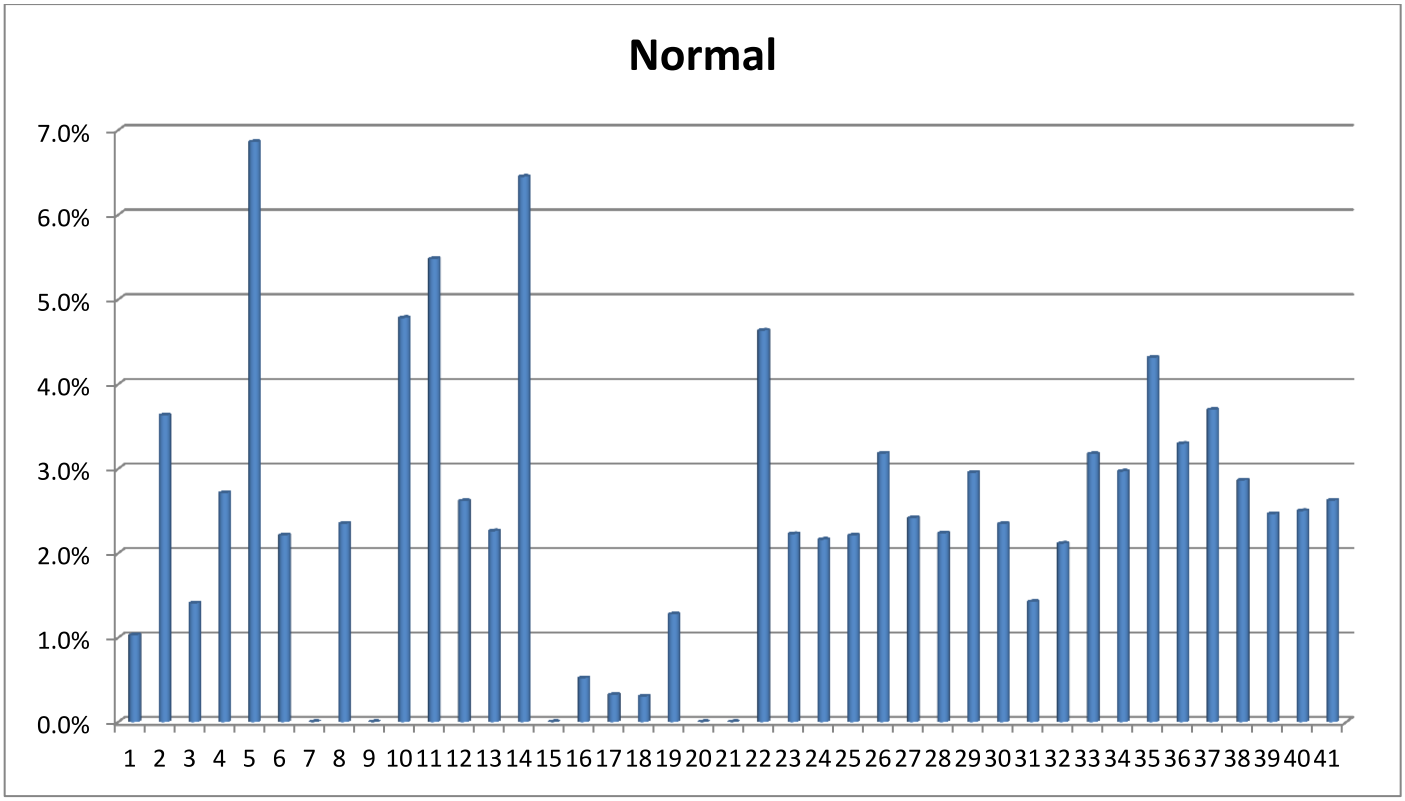


Fig. 3. Evolution of the false postive rate according to the number of selected inputs

B. *Distinction between normal and different types of attacks: multiple outputs NN*

We have also tested the HVS approach on a multiple output NN (that distinguishes the different attack classes) so that the features can also be ranked according to their contribution to identify each attack class. The KDD dataset divides attacks into 4 types:



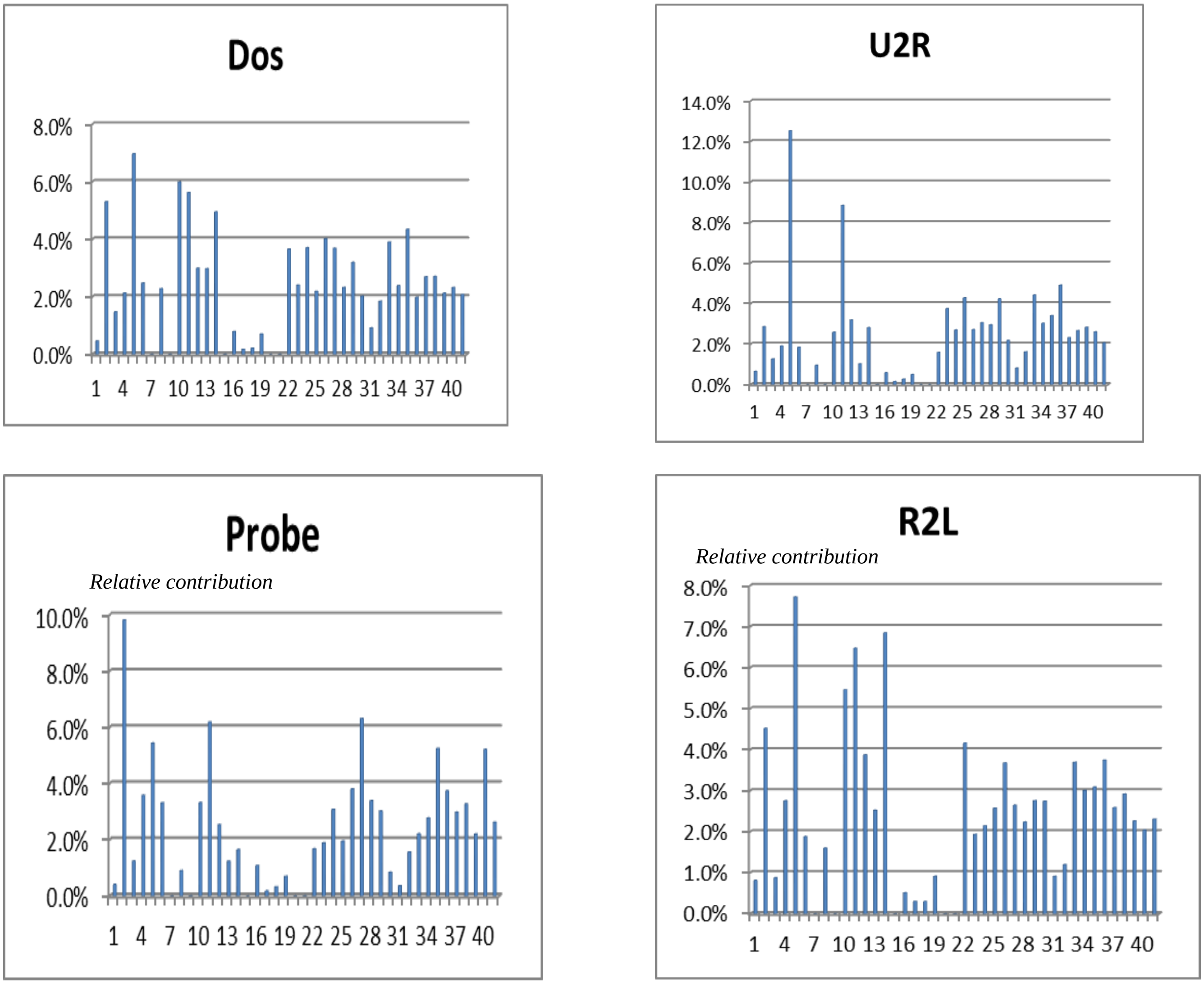


Fig. 4. Relative contribution of each of the KDD 41 features to the detection of normal traffic and different classes of attacks

1) Probe: any attempt to gather information about a network of computers for the purpose of circumventing its security controls.

2) DOS (Denial Of Service): causing computing or memory resources to be too busy to handle legitimate requests.

3) U2R (User to Root): unauthorized access to local superuser (root) privileges, by exploiting some vulnerability.

4) R2L (Remote to Local): unauthorized access from a remote machine, to gain local access as a user of that machine.

We constructed a classification NN with 5 outputs, each one corresponding to a class of traffic (normal traffic plus the four above mentioned types of attacks). Figure 4 illustrate the relative contribution of each of the KDD features to the detection of each of the five traffic classes. Some features, such as 'source bytes' (# 5) are important for all classes, while others are specific to one class (#14 for normal traffic, #2 for Probe, #10 for DOS, #11 for U2R, #14 for R2L). Unsurprisingly, constants features (#9, 15, 20, 21) take a null contribution for all classes. We see that Figure 4 (normal traffic) shows some differences compared to Fig. 1 due to the fact that in one case we have a 2-class classification problem, and in the other, we have a 5-class problem.

C. Advantages of the method

The results shown above are consistent with those obtained by [5] and [6]. The latter used a totally different method which consists in deleting one of the features and measuring its impact on the result, using either a Neural Network or an SVM classifier. Such a technique is known as a wrapper model [7]. Compared to this approach, the method we have presented above shows several advantages:

- The deletion-based method needs to run as many trainings as the number of features, each time deleting one of the features while the HVS method ranks all the features after a unique training, and does not imply any complicated computation, which makes it more suitable for a lightweight IDS.

- The HVS method tends to be more accurate in selecting relevant features than the method used by [6] as explained in section III.A.2.
- The HVS method distinguishes well between features than the SVM based feature ranking used by [6] which yields remarkably close accuracy results for most of the features, with so slight variations that they could be of random origin.
- The HVS method reveals to be more precise in detecting irrelevant features than the method presented in [6]. For example, while features 20 and 21 are constant in the whole KDD learning dataset (as previously noticed by [5]), and features 9 and 15 almost constant and they were not detected as the least important features in [6].

On the other hand, in term of consistence of HVS method, we note that we can keep only the most important 12 features (out of 41), without significantly deteriorating neither the overall accuracy rate (Figure 2) nor the false positive and false negative rates (Figure 3). This number of features is close to the one retained by [10] (11 features) using rough sets and genetic algorithms. [6] conducted a similar test but showed a significant deterioration when selecting the most important 34 features (the overall accuracy rate decreased from 87% to 81% and the false positive rate increased from 6.7% to 18%). This tends to prove that the selection feature method proposed here is considerably more accurate than other cited methods. It should be also noticed that these latter results shown by [6] are not consistent with the Figures they obtained during the feature ranking since the deletion of only one feature (#10 or #35) decreased the accuracy of their network to less than 55%. They did not precise on which database they tested their result. Intuitively, the results they gave for the SVM classification suggests that they tested on only a part of the KDD training dataset (so with a very close distribution to that of the learning database) while we tested on the independent KDD testing dataset (which an entirely different distribution of attacks, and containing new attack types), which is more realistic. Obviously, testing on the training data set yields an artificially high performance.

Furthermore, the contributions of the inputs, calculated using the HVS method, are largely independent of the network architecture, as shown in Figure 5. This Figure depicts the result of use of the HVS method to five networks with different internal architectures. The five tests show very close results. Nevertheless, this stands only if the number of hidden neurons is sufficient to resolve the classification problem.

D. Limitations

While the HVS method is lowly dependent on the network's architecture, it depends more on the composition and the size of the learning dataset. We have tested it on different learning databases randomly extracted (with a fixed size) from the original KDD learning set. Figure 6 shows the result of this test. The dependence is significantly more marked than in Figure 5, but the maximum standard deviation remains under 0.04. The features that present the larger variation are #11 and #17. This is because they are nearly constant over the KDD learning database (only 63 nonzero values for the first and 22 for the second over 494,021 connection records). Consequently, some random extractions may exclusively contain null values for these features (thus yielding null relative contributions) while other random extractions may over represent these features.

Furthermore, this method does not deal with correlation between input features. Consequently, in order to yield more relevant results, it should ideally be applied after a de-correlation preprocessing step, such as a principal component analysis (PCA). Nevertheless, we skipped this step in the present paper, in order to compare our results to existing methods that have been applied directly to the KDD features as they are.

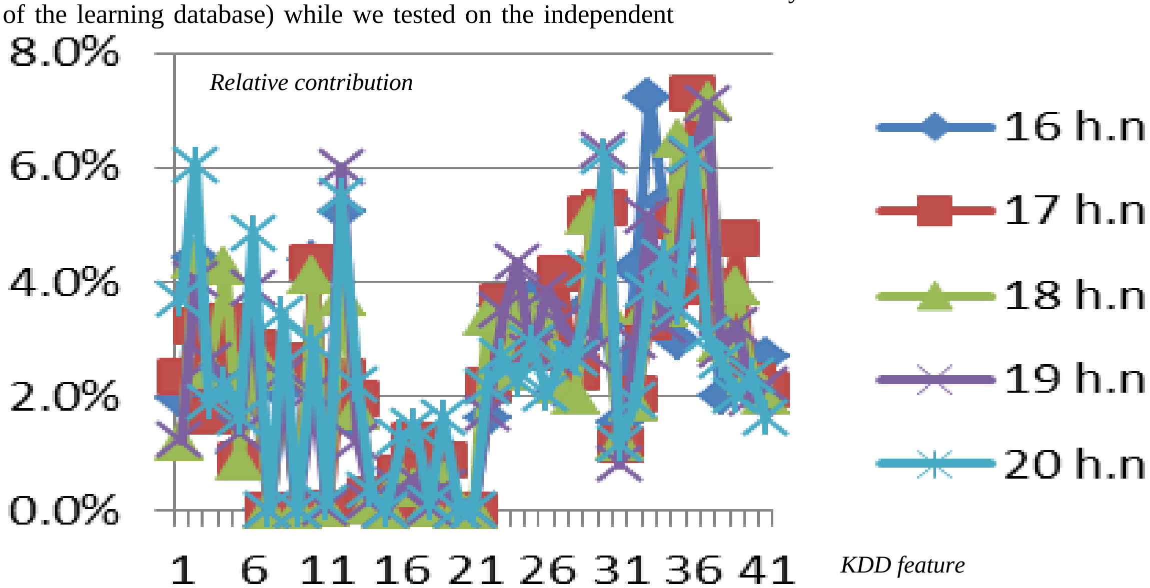


Fig. 5. Relative contribution of each of the KDD 41 features to the detection of normal traffic, calculated for five different networks (with a number of hidden neuron varying from 16 to 20)

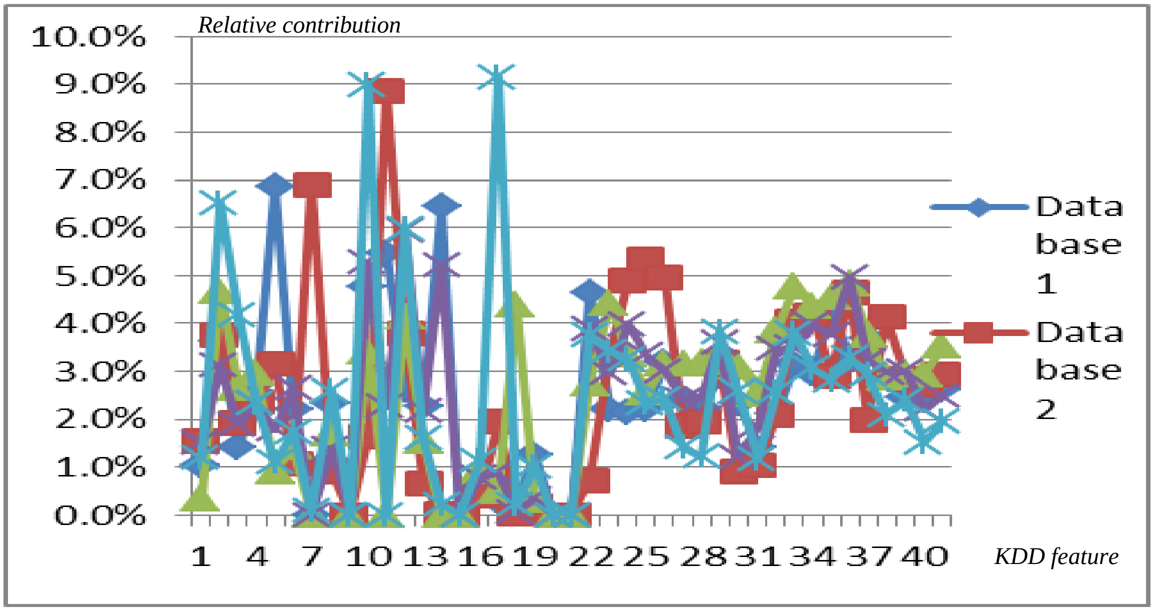


Fig. 6. Relative contribution of each of the KDD 41 features to the detection of normal traffic, calculated for five different randomly extracted learning databases of the same size

TABLE I. LIST OF KDD FEATURES WITH THEIR DATA TYPES

Feature number	Feature	Type
1	Duration	continuous
2	Protocol type	symbolic
3	Service	symbolic
4	Flag	symbolic
5	Source bytes	continuous
6	Destination bytes	continuous
7	Land	symbolic
8	Wrong fragment	continous
9	Urgent	continous
10	Hot	continous
11	Number of failed logins	continous
12	Logged in	Symbolic
13	Number of "compromised" conditions	continous
14	Root shell	continuous
15	"su root" command attempted	continuous
16	Number of "root" accesses	continuous
17	Number of file creations	continuous
18	Number of shells prompts	continuous
19	Number of operations on access files	continuous
20	Number of outbound commands	continuous
21	Is host login	symbolic
22	Is guest login	symbolic
23	Count	continuous
24	Service count	continuous

25	Syn error rate	continuous
26	Service Syn error rate	continuous
27	Rej error rate	continuous
28	Service Rej error rate	continuous
29	Same service rate	continuous
30	Different service rate	continuous
31	Service different host rate	continuous
32	Same destination host count	continuous
33	Same destination host and service count	continuous
34	Same destination host and service rate	continuous
35	Different services on current host	continuous
36	Connec. to current host with same src port	continuous
37	Connec. to same service from diff. hosts	continuous
38	Connec. to current host with an S0 error	continuous
39	Connec. to current host and specified service that have an S0 error	continuous
40	Connec. to the current host with RST error	continous
41	Connec. to the current host and specified service with RST error	continous

IV. Related Works

There exists other feature selection methods also based on neural networks, theoretically described in [8], which we should consider and compare in future works, in the context of intrusion detection. The one we used is the simplest to calculate. In fact, feature selection techniques that use complex calculations are inefficient for large scale data.

Besides, several recent papers presented various feature selection techniques applied to the KDD features. Reference [9] proposed a hybrid approach combining the information gain ratio (IGR) and the k-means classifier. Reference [10] proposed a feature selection method based on Rough Sets, improved Genetic Algorithms and clustering. Then they used the SVM classifier for performance evaluation on the KDD database. Reference [11] proposed a clustering-based classifier selection method. The method selects the best classifier on similar clusters, compares it with the best classifier on the nearest cluster and chooses the better one to make the system decision. It showed better results than the Clustering and Selection (CS) method. Reference [12] constructed binary classifiers at local sensors to distinguish each class from the rest. The authors used both a synthetic and the KDD99 datasets to confirm the improved performance of the pairwise feature subset selection algorithm for multiclass classification problems. This approach could be also applied to the method we used in this article by constructing five single-output neural networks for each of the traffic data five classes. Reference [12], however, did not include U2R and R2L attacks in the experiments, due to their extremely small amount in the database. This is the most challenging aspect of the KDD contest, and it should not be discarded. Reference [13] applied Artificial Bee Colony algorithm (ABC) to determine free parameters of support vector machine (SVM) and to achieve the optimum feature selection for IDSs from KDD Cup 99 data set. Reference [14] used Logistic Regression which is similar to SVM, and found it to be superior to traditional feature selection techniques such as Discriminant Analysis and PCA, after an evaluation on the KDD dataset. Reference [15] also evaluated the performance of standard feature selection methods; CFS (Correlation-based Feature Selection), IG (Information Gain) and GR (Gain Ratio), but on the enhanced NSL-KDD dataset [16] which has been suggested to solve some of the inherent problems of the KDD99 data set. Although, the NSL-KDD dataset still suffers from some of the problems pointed out by [17]. Finally, [18] compared various techniques of feature selection and feature reduction on the Kyoto 2006+ dataset, which is more recent than KDD but which contains a reduced set of features.

The objective of the present paper was to prove the efficiency of the HVS feature selection technique when applied to the intrusion detection problem, which has not been evaluated before. We should however compare this method to the various techniques cited above in a future work. Nevertheless, most of the cited works tested their methods on an extraction from the KDD learning database. They did not test them on the KDD database originally dedicated to testing and containing new attacks as we did in this paper. This demonstrates the potential of the method to detect new attacks and gives more realistic results than the results produced by testing on a part of the KDD learning database.

V. Conclusion and Future Work

We have shown that the HVS method we presented in this work can be directly and efficiently applied to the problem of intrusion detection, in order to assess the most important features that contribute to attack detection. We could then select a set of most relevant features to accelerate the detection process, and construct an IDS with higher performance in terms of accuracy and execution time. The attribute-selection algorithm can therefore be applied as a pre-processing technique for classification. An important advantage of the approach, compared to existing methods (like [10] or [13] for example), is that the same technique (feed-forward neural networks) can be used for both feature selection and attack detection, which gives more consistency to the method. Furthermore, the method is almost independent of the used networks' architecture. Further rigorous tests should be conducted to measure accurately the dependence of the HVS method to the learning database, with databases of different sizes. This dependence should not be an obstacle, however, since, in most applications, the learning database is set once for all.

Acknowledgment

This work is a partial result of a research project supported by grant number INF 36-8-08 from King Abdul Aziz City for Sciences and Technology (KACST), Riyadh, Kingdom of Saudi Arabia.

References

[1] P.M. Wong, T.D. Gedeon, and I.J. Taggart, "An Improved Technique in Porosity Prediction: A Neural Network Approach," IEEE Transactions on Geoscience and Remote Sensing, vol. 33(4), 971-980, 1995.

[2] M. Yacoub and Y. Bennani, "HVS: A Heuristic for Variable Selection in Multilayer Artificial Neural Network Classifier", International Conference on Artificial Neural Networks and Intelligent Engineering, ANNIE '97, Missouri, USA, 1997, pp. 527-532.

[3] The 1998 intrusion detection off-line evaluation plan. MIT Lincoln Lab., Information Systems Technology Group. http://www.11.mit.edu/IST/ideval/docs/1998/id98-eval-11.txt, 25 March 1998.

[4] Knowledge discovery in databases DARPA archive. Task Description. http://www.kdd.ics.uci.edu/databases/kddcup99/task.html

[5] H. G. Kayacik, A. N. Zincir-Heywood, and M. I. Heywood, "Selecting Features for Intrusion Detection: A Feature Relevance Analysis on KDD 99 Intrusion Detection Datasets," Proceedings of the Third Annual Conference on Privacy, Security and Trust, St. Andrews, Canada, October 2005.

[6] S. Mukkamala and A. H. Sung, "Identifying Significant Features for Network Forensic Analysis Using Artificial Intelligence Techniques". In the International Journal on Digital Evidence, vol. 1(4), 2003.

[7] H. Liu, and Y. Lei, "Toward integrating feature selection algorithms for classification and clustering," Knowledge and Data Engineering, IEEE Transactions on 17.4 (2005): 491-502.

[8] P. Leray and P. Gallinari, "Feature selection with neural networks," Behaviormetrika, vol. 26(1), pp. 145–166, 1999.

[9] Araújo, N.; de Oliveira, R.; Ferreira, E.-W.; Shinoda, A.A.; Bhargava, B.; , "Identifying important characteristics in the KDD99 intrusion detection dataset by feature selection using a hybrid approach," Telecommunications (ICT), 2010 IEEE 17th International Conference on, vol., no., pp.552-558, 4-7 April 2010.

[10] Y. Guo; B. Wang; X. Zhao; X. Xie; L. Lin, and Q. Zhou; , "Feature selection based on Rough set and modified genetic algorithm for intrusion detection," Computer Science and Education (ICCSE), 2010 5th International Conference on, vol., no., pp.1441-1446, 24-27 Aug. 2010.

[11] A. Mi and L. Hai, "A clustering-based classifier selection method for network intrusion detection," Computer Science and Education (ICCSE), 2010 5th International Conference on, vol., no., pp.1001-1004, 24-27 Aug. 2010.

[12] H. D. Nguyen and Q. Cheng, "An efficient feature selection method for distributed cyber attack detection and classification," Information Sciences and Systems (CISS), 2011 45th Annual Conference on, vol., no., pp.1-6, 23-25 March 2011.

[13] J. Wang, T. Li, and R. Ren, "A real time IDSs based on artificial Bee Colony-support vector machine algorithm," Advanced Computational Intelligence (IWACI), 2010 Third International Workshop on, vol., no., pp.91-96, 25-27 Aug. 2010.

[14] M. F. Wu, "Protocol-based classification for intrusion detection", WSEAS International Conference. Proceedings. Mathematics and Computers in Science and Engineering. Eds. Qing Li, S. Y. Chen, and Anping Xu. No. 7. World Scientific and Engineering Academy and Society, 2008.

[15] H. S Chae, B. O. Jo, S. H. Choi, and T. K. Park, "Feature Selection for Intrusion Detection using NSL-KDD," Recent Advances in Computer Science, 2013, 184-187.

[16] M. Tavallaee, E. Bagheri, W. Lu, and A. Ghorbani, A detailed analysis of the KDD CUP 99 data set, 2009 IEEE Int. Conf. Comput. Intell. Security Defense Appl., 2009, pp. 53–58.

[17] J. McHugh, "Testing intrusion detection systems: a critique of the 1998 and 1999 darpa intrusion detection system evaluations as performed by lincoln laboratory," ACM Transactions on Information and System Security, vol. 3, no. 4, 2000, pp. 262–294.

[18] A. Ammar, "Comparison of Feature Reduction Techniques for the Binominal Classification of Network Traffic", Journal of Data Analysis and Information Processing, vol. 3, no. 2, pp. 1-10.

Generating Representative Sets and Summaries for Large Collection of Images Using Image Cropping Techniques and Result Comparison

Abdullah Al-Mamun
Department of Computer Science
University of Trento
Trento, Italy

Dhaval Gandhi
Department of Computer Science
University of Trento
Trento, Italy

Sheak Rashed Haider Noori
Department of Compute Science & Engineering
Daffodil International University
Dhaka, Bangladesh

***Abstract*—The collection of photos hosted on photo archives and social networking sites has been increasing exponentially. It is really hard to get the summary of a large image set without browsing through the entire collection. In this paper two different techniques of image cropping (random windows technique and sequential windows technique) have been proposed to generate effective representative sets. A ranking mechanism has been also proposed for finding the best representative set.**

***Keywords*—summarization; representative set; image collection; diversity; coverage**

I. Introduction

Due to technological growth in electronic gadgets and digital media such as mobiles, tablets, digital cameras, memory cards and many more, there is a big increase in large image collection in our hard drives and in the web storage. The main purpose of capturing photos is to keep and refresh memories about our life events. The new coming trend is to share photos with family and friends using social networking web-sites. Nevertheless, the growth of images raises challenges such as the difficulty in browsing large image set while avoiding large number of duplicates and similar images.

The goal of this paper is to find the diverse representative set and summary of the collection of images. To deal with this problem, firstly, focus has been put on a cropped image that is called window. This means the concentration is placed on some portion instead of whole image to generate the representative set. This approach is initiated from the idea of n-gram model. An n-gram is a subsequence of n items from a given sequence. It is a model [1] based on text and is widely used in statistical natural language processing. The items can be phonemes, syllables, letters, words or base pairs according to the application. The idea is to find out whether this model can be applied to process images and what will be the outcomes. Therefore, two algorithms are proposed for windows cropping, namely random and sequential. That is, the cropping points are generated randomly and sequentially respectively.

On these cropped images image recognition is performed. Because image features have many properties that make them suitable for matching and differing images of an object or scene.

Some of the features are invariant to image scaling and rotation [2]. Scale-invariant feature transform (or SIFT) algorithm is used for finding and computing descriptors of each images. SIFT features are extracted from a set of candidate images and stored in a database. By applying K-means clustering algorithm those descriptors for image features go into clusters respectively. After that the centroid image of all clusters are fetched which means that the centroid images are the representative images of each respective cluster.

As it is intended to have highly representative images, namely it is needed a small set of images that are highly dissimilar; a ranking mechanism is developed to select more representative images from the given image set created above. First, image matching is performed by individual comparison of each features based on Euclidean distance of their feature vectors. Then the ratio of number of matching points to summation number of the detected key points between images is computed. A higher ratio indicates a larger possibility of similarity of two images. At last the ratio is sorted by ascending order, so that we are free by top *n* images from the sorted array. The overall process is depicted in figure 1. The evaluation of the performances of these two methods is performed by subjective evaluation.

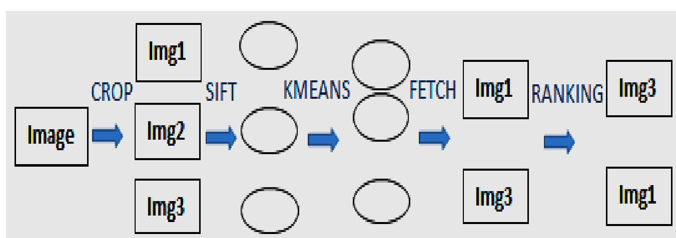


Fig. 1. proposed process of image processing

Some of the important related works are reported in Section 2. The problem statement more specifically is described in Section 3. In Section 4, a brief introduction of the followed approaches, used to generate representative sets is discussed. Section 5 gives the detailed windows cropping mechanism. Section 6 provides the clustering details and in Section 7 the ranking mechanism is provided. The evaluation of the experiment is presented in Section 8.

II. RELATED WORK

The problem for generating representative set and summarization has been done by many researchers. Some of the excellent and most related work are reported here. Scene Summarization for Online Image Collections by Simon [5], examines the distribution of images in the collection to select a set of canonical views to form the scene summary, using clustering techniques on visual features using SIFT. The author summarized all images based on likelihood, coverage and orthogonality. The proposed approach followed in this paper is similar to the clustering technique and selection categories. But, the key difference between [5] and the proposed approach is generating windows before clustering phase. In [6], the author summarizes images based on spatial patterns in photo sets, as well as textual-topical patterns and user (photographer) identity cues. The key difference between [6] and the proposed approach is that the geo-referenced images are not used in the experiments of the later case. In this paper strong focus is placed on low level features of the images. In [7], the author generates diverse and representative image search results for landmarks based on context- and content-based tools. To do that the author used location and other metadata as well as tags associated with images, and the images' visual features. The work [7] is somewhat different with Simon because the author starts from the tags that represent landmarks while the author used also SIFT for the visual features comparison. The difference with the approach proposed in this paper is that metadata associated with images as well as image tags are not used. The only concentration is put on visual features of images.

In [8], the problem is of selecting iconic images to summarize general visual categories. The author defined conic images as high-quality representatives of a large group of images consistent both in appearance and semantics. The approach to find such groups was to perform joint clustering in the space of global image descriptors and latent topic vectors of tags associated with the images. The author has also used a ranking mechanism. The key difference is that in [8], it has assumed that there is one iconic view of the scene rather than a diverse set of representative views as are shown in this work.

In the absence of location metadata, temporal metadata was also considered in the past for the purpose of photo collection summarization. In [11], Graham et al. describe an algorithm to heuristically select representative photos for a given time period in a personal collection, utilizing patterns in human photo-taking habits. Additional time-based work aims to detect events in personal collections [10], which could be the basis for collection summarization. However, again, all these projects considered single-photographer collections only. Several projects [9, 10] use geographic data to organize photo collections in novel ways, for example, by detecting significant events and locations in a photo collection. Such structures could indeed be the basis for collection summarization. However, geographic data of the images used in the experiment presented in this paper are not available. In medical science the image processing is also being conducted to quantify how body cells move, divide and die over time. Traxtile [12] has been developed with Python program as a tool for directly measuring rates of cell division and death, by observing events under various conditions using time-lapse imaging, where the cell of interest is centered in an image tile that is cropped and zoomed from the original, and corresponding tiles from preceding and following frames are displayed in a montage.

In general the work presented in this paper is different in terms of windows generation technique in initial stage and ranking mechanism.

III. PROBLEM STATEMENT

It is initiated by defining some terminologies. Throughout the report, the term photo is used interchangeably with image, all of which refers to an ordinary 2D image. Collection is defined as a set of photos and windows as cropped images. The representative set is loosely defined as a subset that captures representativeness, relevance and breadth in the original collection.

Two different notations for the random windows and sequential windows are used. For the random window and for the sequential window, IN-C and SN-C are used respectively, where *N* is number of windows of each image and C is coverage need to be covered.

For example, with number of random or sequence windows that are 3 and with coverage 85%, it is represented as I3-85 if random images and S3-85 for sequential images respectively. A summary is a set of photos ordered by applying ranking mechanism and selecting any arbitrary number of images from the given representative set.

Given a set of photos P, the goal is to compute a representative set RS $\subseteq$ P and then summary S $\subseteq$ RS such that S represents highly diverse representative images of the photo set P.

IV. APPROACH FOR GENERATING REPRESENTATIVE SET AND SUMMARY

In this section the selection criteria for the representative set and overview of the proposed solution are introduced.

A. Selection Criteria

As there is no accurate formal model which constitutes a "good" representative set and summary of a collection of images, some simple heuristics are followed that try to model and capture human attention. These heuristics are as follows:

- Images are taken at a location that provides views of some important objects or landmarks.
- Image is more relevant and should be included in representative set, if it matches with many other images of the collection. The representative set and summary should contain highly distinct or diverse images.

B. Overview of Proposed Solution

The proposed solution has three different phases. In the first phase "Pre-cluster" there are three different techniques, followed by second phase "Clustering" phase which generates the representative set and the last phase is "Ranking mechanism" through which the summary is generated. The overview can be seen in the figure 2.

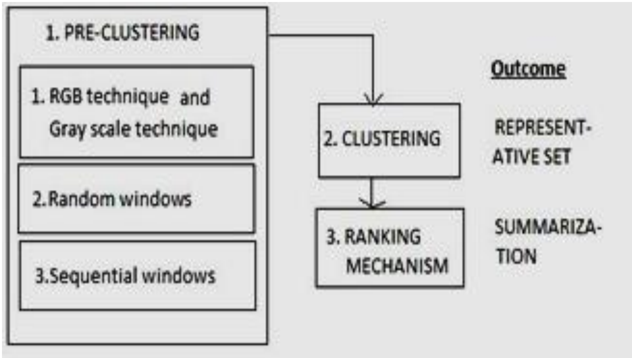


Fig. 2. Overall scenario of proposed solution

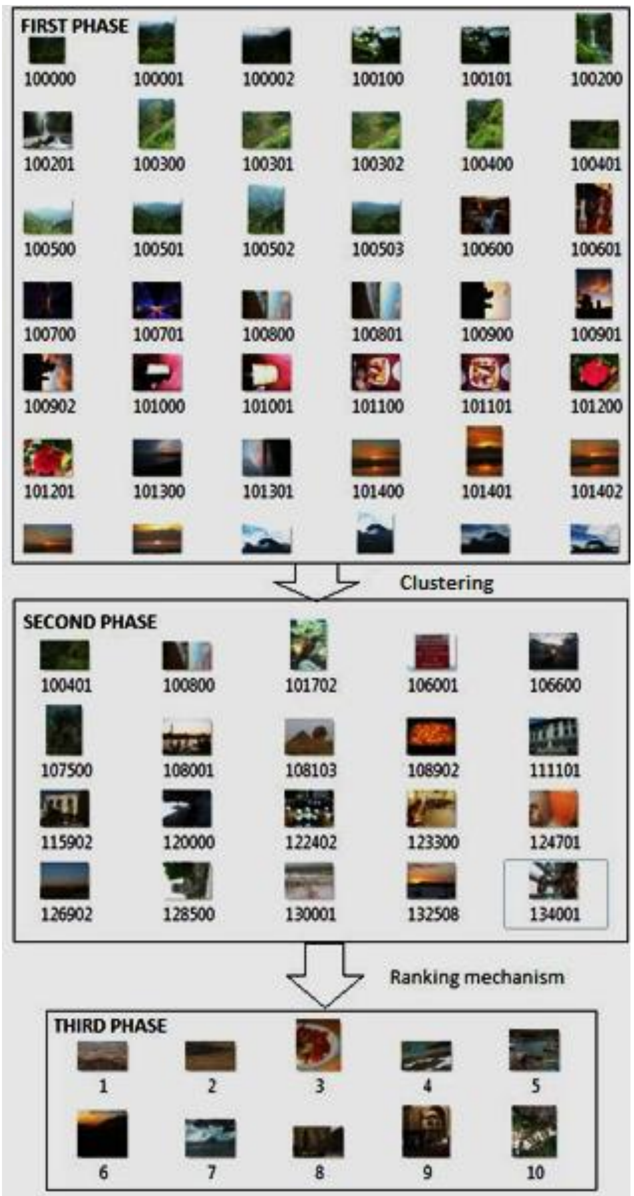


Fig. 3. Overall output scenario of proposed solution

In particular:

- Different pre-clustering techniques namely random and sequence are applied in order to generate windows. The input will be an original large image set.
- Clustering was applied to group appearance-wise similar images. The input will be windows generated from the pre-cluster phase. Then centroids of all clusters are fetched to generate a representative set.
- Ranking mechanism is applied on the representative set and a summary is generated by selecting any arbitrary number of top ranked images. So the input will be the representative set and the output will be a summary.

By applying these three phase, it will get the final outcome "summary". One can see the graphical overview of the scenario in figure 3 for summary of 10 images.

In the following subsections the algorithm is explained. First, different pre-clustering techniques are described. Then the clustering algorithm for fetching centroid is provided. Finally, how to rank each image of the representative set in order to generate summary is presented.

V. Pre-Cluster

In pre-cluster three different techniques are used which are as following.

A. *RGB and Gray scale Techniques*

Using these two pre-cluster techniques, images are stored in RGB form and clustering algorithm (k-means) is applied. For gray scale technique, all the images are converted in gray scale and are stored in vectors and again the cluster algorithm is applied in second phase. From the result of clusters it is found that the pictures with same appearance with different pixel values do not go to the same cluster. It means taking a picture of any particular landscape with normal mode and taking the same picture with zoom, would not go in the same clusters although both have same visual features. Because of this big drawback it is not possible to stand on finding a representative set and summarization which holds the selection criteria. Therefore, experiments with this pre-cluster stage are stopped and focus is set only on the next two techniques.

B. *Random Windows*

As it is mentioned earlier that the idea of cropping images initiated from n-gram model, however random windows are not suitable for the n-gram model. It is initiated with taking n-windows of same size with certain fix coverage of each images from random pixel point of an image. More formally, for each image of the dataset *P* and for the fix coverage *C* (i.e. window should cover some fixed portion of image like 85%, 75% and 66%), generate *N* random windows from (r_x, r_y) pixel point (i.e. 3 windows of 75% coverage of each images) where r_x and r_y is random point from where the image with fix coverage will be cropped.

One can see from the figure 4 that the black dots are windows initial points which are totally random. After having the N windows set of original data set, sift algorithm is applied to generate and store descriptors of each random windows for the next stages. Descriptors are K-by-128 matrix, where each row gives an invariant descriptor for one of the *K* key points. The procedure is mentioned in algorithm 1 and the output can be seen in figure 5 with respect to the original image.

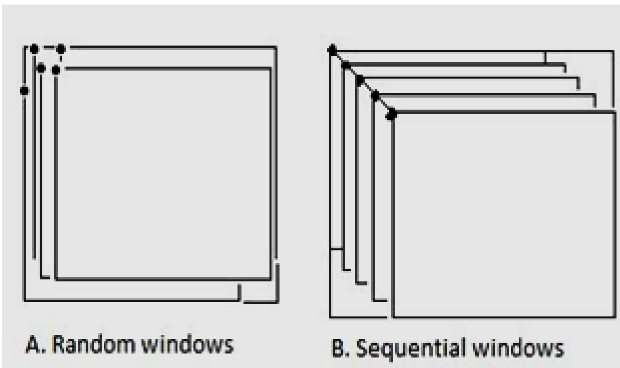


Fig. 4. Random and sequential windows

Algorithm 1: Random windows generation

Input: original image set
Output: *N* random windows of each images with certain coverage *C*.
1: set the coverage *C* and the number of windows *N* to generate random windows
2: **for** each image img **do**
get the size of img [rows columns] = size(img) manipulate window width and height
according to coverage *C*:
w_width = *C* * column;
w_height= *C* * rows;
3: **for** each image img, to generate *N* random windows **do**
generate random pixel point (r_x,r_y);
crop the image img : cropped_img = imcrop(img, [r_x, r_y, w_width, w_height])
save cropped_img (window) at the output directory
4: **end for**
5: **end for**

C. Sequential Windows

In this technique, N sequential windows are generated in order to implement n-gram model. So, the coverage C remains same as the previous technique but the window point is now fixed. More formally, for the each image of the dataset P and for the fix coverage C, N sequential windows are generated from (s_x,s_y) pixel point where s_x and s_y is sequential pixel point from where the image with fixed coverage will be cropped. Here sequence of windows are kept diagonally. One can see from the figure 4, windows initial points which are totally diagonal.

The idea is to capture certain coverage without losing any sort of information. In other words for any N windows, we must have all images between (0,0) and (X,Y) where X and Y are height and width of an original image, diagonally.

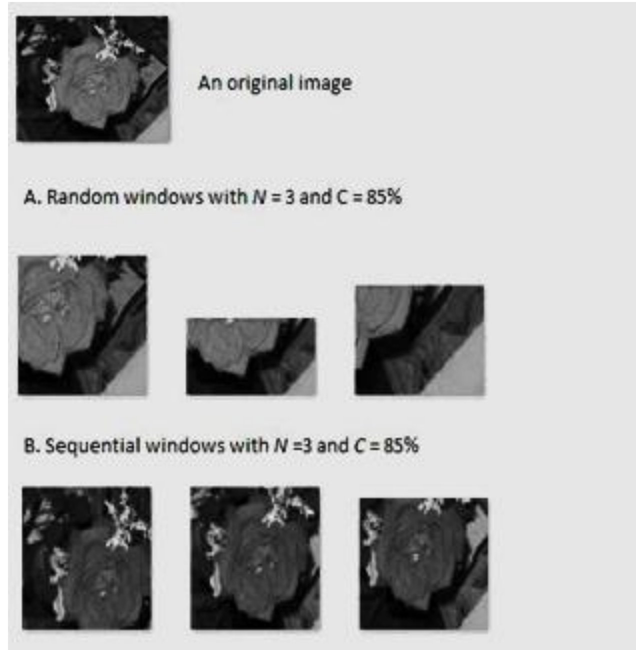


Fig. 5. Output of random and sequential windows

After having the n sequential windows set of original data set, we apply sift algorithm to generate and store descriptors of each random window for the next stages. The procedure is mentioned in algorithm 2 and the output can be seen in figure 5 with respect to the original image.

Algorithm 2: Sequence windows generation
Input: original image set
Output: *N* sequential windows of each image with certain coverage *C*.
1: set the coverage *C* and the number of windows *N* to generate
2: **for** each images img **do**
3: get the size of img [rows columns] = size(img)
4: manipulate window width and height according to coverage *C*:
w_width = *C* * column;
w_height= *C* * rows;
5: generate initial window from the initial pixel point(s_x0, s_y0) of the image img s_x0=0 and s_y0=0;
cropped_img=imcrop(img, [r_x, r_y, w_width, w_height]);
save cropped_img (window) in output directory
6: **for** each image img, to generate *N* sequence windows **do**
(for i=*N*;I>1;i--)
generate sequential pixel point (s_x,s_y) s_x= (rows – w_height)/*N*-1;
s_y= (columns – w_width)/*N*-1; crop the image img cropped_img=imcrop(img, [s_x, s_y, w_width, w_height]
save cropped_img (window) at the output directory
7: **end for**
8: **end for**

VI. CLUSTERING

As it is previously discussed descriptors of each image of the collection are needed in order to cluster. Therefore, first, it is started with the generation of descriptor of each image and then those are saved into a single array in order to cluster. The input will be image collection of the pre-cluster phase. Technically, the focus is set on the highly dissimilar representative images. For that local features of images are used. The local approach represents each image by a set of local featured descriptors computed at some interesting points inside the image [3]. For finding and computing descriptors of each images SIFT algorithm is used.

Now, K-means algorithm is applied on an array of descriptors of images. In statistics and data mining, K-means clustering is a method of cluster analysis which aims to partition n observations into *k* clusters in which each observation belongs to the cluster with the nearest mean [7]. Since the dataset is a large set (minimum 1000 images and maximum 5000 for *N*=5 windows) and the goal is to generate precise, minimal redundant and diverse informative overview of the image collection, it is decided to apply K- means twice. So, first, K-means with the *k* value is applied and a small subset of image set is generated then K-means is applied again in order to get reduced set of small set, which is more diverse .When K- means is applied, cluster results, centroids, sum and distances are achieved.

Now centroid image of all clusters are fetched which are the representative image of each clusters. The concept is, to find for each cluster the least distance image to the centroid. The distances are found from K-means output and after calculating the distances; we sort the nearest image of the cluster centroid for each cluster. That will be the representative set of the image collection.

After having 1st k-means subset, again K-means is applied second time on representative set which becomes precise and small representative set of the large image data set. So, from this phase we generate representative set and it is also useful for the next phase namely Ranking Mechanism. One can see the procedure for generating representative set in algorithm 3.

Algorithm 3: Clustering and generating representative set

Input: resultant image set of pre-cluster
Output: the representative set
1: **for** each images img **do**
  get descriptor or key points by calling sift function
  [image, descriptor] =sift (img)
  save each image descriptor in an array
  descriptor_images[img] = descriptor
2: **end for**
3: set number of clusters k and apply k-means on the descriptor_images array
  [Id, *C*, D] = kmeans(descriptor_images, k)
    where Id is image identification number, *C* is the assigned cluster number and the D is distance from the assigned cluster and other clusters as well.
4: find centroid image of each clusters:
  for each images i and j of cluster *C*
  if distance_image_i < distance_image_j
    //store the least distance image
    centroid = image_i
  end if
  save centroid at the output directory of the representative set
5: **end for**

VII. RANKING MECHANISM

Given the output from previous clustering step, it is desired to eliminate possible similar images in order to get more representative ones out of these inputs which compose the summarization. One other motivation behind this mechanism is to generate best set of images of the representative set and it helps the human based evaluation. In other words it would be hard to evaluate all images by participant. Instead of the full set we select most desirable set of images for example 10 images out of 20 images in order for the evaluation. Technically It goes like, on top of the function *Match.m* [2] some changes are made to fulfill the needs of the experiments of this paper and the number of matching points are then calculated. Besides, the number of key points of each image are achieved through *Sift.m*[2]. Then the ratio of number of matching points to summation number of detected key points among images is computed. A higher ratio indicates a larger possibility of similarity. Function *sum ()* is chosen here because the image sets used in this paper have the following features:

- They are supposed to be very dissimilar, which means the number of matching points, namely the numerator could be very small
- The number of key point's detected can varies a lot. for example, one is 100, the other one is 3000.

Based on these, the values are more like 10/(100+3000), (or 10/max(100,3000) if *max()* is used), where function *sum()* can represent the dissimilarity well. However, the function *max()* can be chosen since it also can do a good approximation. By sorting the ratios of all images, one is free to choose arbitrary number of images and these images are highly representative. The Ranking mechanism is presented in algorithm 4.

Algorithm 4: Ranking mechanism

Input: representative set
Output: ranking of each images of representative set
1: **for** each image i **do**
  for each image j from i+1 to n **do**
   // function Keypoint return the number of detected keypoints of an image using SIFT
   //function Matchingpoints return the number of matched points between to images calculate similarity ratio r[i j]=
  Matchingpoints(i,j)/Sum(Keypoint(i)+Keypoint(j));
2: **end for**
3: **end for**
4: **for** i from 0 to n **do**

```
    for j from i+1 to n
do
      compute the maximum similarity value mr[i]=Max[i j];
   end for
 end for
 sort mr[i]; // in ascending order.
select arbitrary number of images from array s which will be the summary S;
```

VIII. Evaluation

A. Initial Experimental Setup

First the data set is defined. The data set is of 1000 holiday images (ordinary 2D high quality images) taken from INRIA holidays data-set[1]. The dataset only contains images without any tags. The dataset includes a very large variety of scene types (natural, man-made, water and fire effects, etc) and images are in high resolution. Using windows cropping techniques we increase image sets from 1000 to *N**1000 images. For example, if 3 windows are found per image then there will be 3000 images for the overall experiments. For the evaluation, 6 data set have been selected on the basis of overall time, coverage and number of windows and also for the overall comparison only the sift output is used. The values for coverage *C*, number of windows *N* for the experiments are set. The idea is to set coverage higher in order to lose less information. It is considered to have only *N*=3 and 5 windows and *C* =66%, 75% and 85%. Two result image sets (one for random and the other for sequential windows) are selected of each coverage *C* for the evaluation while taking care of different *N* values.

For the clustering phase, the input image set or windows set is obtained from the first phase. After many initial experiments k=90 is set for the first clustering. As it is mentioned earlier that k-means is applied two times in order to get reduced set and highly dissimilar images, after having 90 clusters again k-means is applied with the value of k=20. The idea is to keep same k, so that windows of the original image mostly go to the same cluster because of high coverage *C*. The outcome of this phase is the representative set which have 20 images of the original data set.

For the ranking mechanism, the input of 20 images are taken from the second clustering phase and then ranking mechanism is applied on it and it gives rank of all images. Now, for taking into consideration of human based evaluation, 3 different sets of images top 10, 15 and 20 images are selected and these becomes summarization of the original big set. By doing this, it is intended to see whether changing the number of images in the summary set does give different results.

A questionnaire [figure 6] is devised for human evaluation of the experiment outcomes from different methods and is spread to 24 interviewees. It is intended to see the rating of the users or human evaluator for each sets. It is also intended to see, changing the number of images inside the summary affects the result. For the experiment, 10, 15 and 20 images for the summary are chosen for the each set. So, human evaluator has to check total 18 sets (6 generated set * 3 different summary of each image set). Before starting questionnaire, the original image collection is shown and the research perspective of this paper is explained to the interviewees.

As 3 sets of each image set are obtained which contains 10, 15 and 20 images respectively, the fixed pattern for each set is followed which can be seen in figure 6.

Image set 1: (I3-75)

A. **10 Images:** (1)Very Low (2) Low (3) good (4) Very good (5)Excellent

B. **15 Images:**

Do you think it's more or less representative than 10 images?
- Yes
- No

If "Yes" : 1)Very Low (2) Low (3) good (4) Very good (5)Excellent

C. **20 Images:**

Do you think it's more or less representative than 10 and 15 images?
- Yes
- No

If "Yes" : 1)Very Low (2) Low (3) good (4) Very good (5)Excellent

Fig. 6. Questionnaire pattern for each result sets

To understand deeply, an example is explained here. Suppose one of the representative set called "I3-75" is under evaluation. As it is mentioned earlier that 20 images are considered in a representative set, first interviewees are requested to rate 10 top ranked images generated from ranking mechanism and then 15 top ranked images are shown and participants are asked whether they think that these 15 images are more or less representative than 10 images. If they say "yes", then they are requested for the rating again otherwise directly whole 20 images summary is shown and the same question is asked in perspective of 10 and 15 images summary. Again if they say "yes", then again they are asked to rate for the 20 images summary. The pattern of the questionnaire will be same for rest of the five result sets.

B. Initial Time Based Evaluation

From figure-7 it can be observed, there is nominal difference in the time taken by windows generating algorithm (blue bar) for *N*=3 and *N*=5. The time taken by clustering algorithm is higher for both *N*=5, *C*=85% random and sequence windows. The reason of this is to fetch sift descriptors of all the windows and then process for the clustering. While for the set *N*=3, *C*=66% sequence took very less time. There is no much difference in time taken by the ranking mechanism for all sets. The set *N*=5, *C*=75% sequence windows took long time for the ranking mechanism in compare to others. Maximum total time is consumed by *N*=5, *C*=85% random windows set. The time taken by clustering and fetching centroids increases with the increase of number of windows *N* and the coverage *C*.

[1] INRIA project web-link: http://lear.inrialpes.fr/~jegou/data.php.

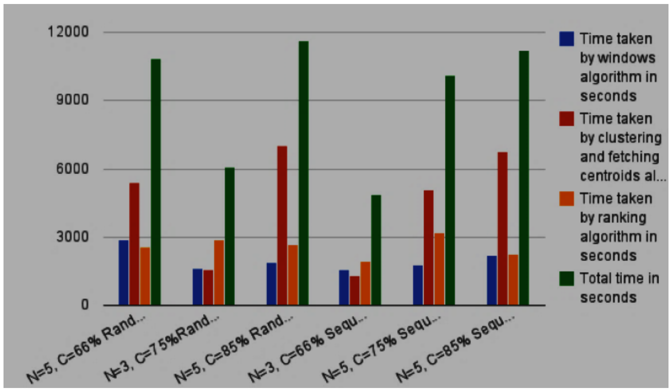


Fig. 7. Time consumed by different cropped window

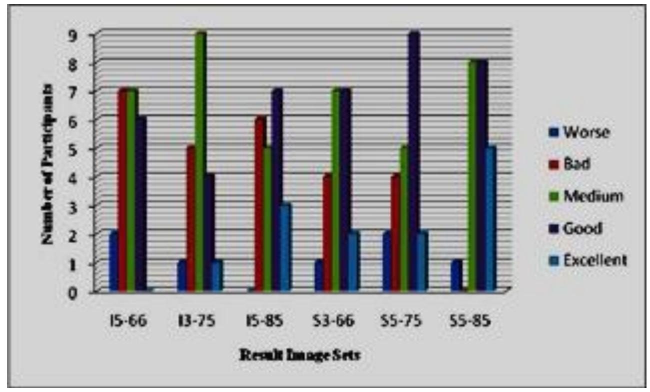


Fig. 9. Participant ratings for summary of 15-images

C. Human Based Evaluation

In this section the human based evaluation results are presented with the statistical evaluation. It is started with analyzing the participants rating and then the number of images inside the summary. First, all different summaries individually and then the common observation are evaluated. Once again, it can be recalled that six different summaries with different number of images inside are evaluated, where, three random window results namely I5-66, I3-75 and I5-85 as well as three sequential window results namely S3-66, S5-75 and S5-85 are used.

Figure 8 shows the participants rating summary of 10 images. One can observe that the worst result with the highest votes is I5-66. Similarly for the bad, medium and good results with the highest votes are I5-85, I3-75 and S3-66 respectively. The most excellent result is S5-85.

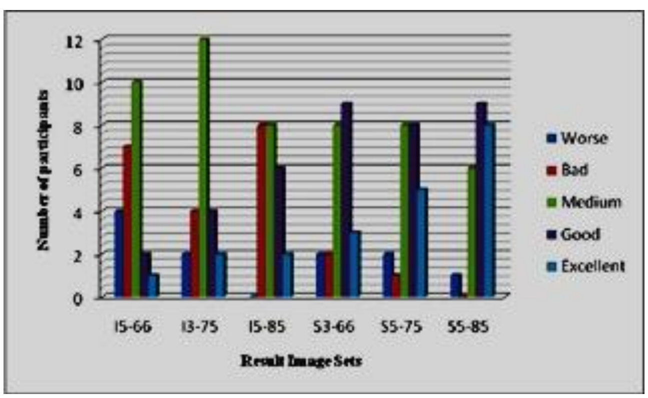


Fig. 8. Participant ratings for summary of 10-images

Figure 9 shows the participants rating summary of 15 images. One can observe that the worst result with the highest votes is I5-66. Similarly for the bad, medium and good results with the highest votes are I5-66, I3-75 and S5-75 respectively. The most excellent result is again S5-85 for the summary of 15 images.

Figure 10 shows the participants rating summary of 20 images. One can observe that the worst result with the highest votes is again I5-66 which similar to 10 and 15 images results. Similarly for the bad, I3-75 and S5-75 have the highest and same votes. The medium results with the highest and same votes are I5-66, I3-75 and S3-66. The good and excellent with the highest votes is S5-85.

General observation about the summary of 10, 15 and 20 images are as follows:

- Random window result sets give negative rating with the high portion of votes as worse, bad and medium while sequential window result sets give positive rating with the high portion of the votes in medium, good and excellent.
- If random windows and sequential windows are focused differently, one can observe a common thing. When the coverage increases, the result sets with higher coverage secures good and positive ratings.

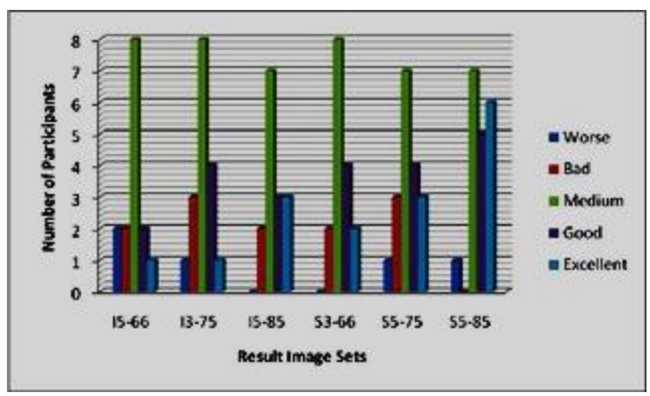


Fig. 10. Participant ratings for summary of 20-images

To understand these two general observations, scores for each summary is calculated. So, the calculation is based on the evaluators rating. For the worse, bad, medium, good and excellent rating, an integer number 1,2,3,4 and 5 is assigned respectively. Now, a formula is devised for calculating score of each result sets. The formula is defined as follows:

Total score = [((NE_worse *1) + (NE_bad *2)+ (NE_medium *3)+ (NE_good*4)+ (NE_excellent*5)) * 100] / Total NE

Here, NE: Number of Evaluators who voted.

Total score for each result set is a calculation for number of evaluators votes for each category multiply the assigned integer number and again multiply 100 and the value is divided by total number of evaluators voted. The reason of multiplying with 100 and division by total number of evaluators voted for the particular set is to achieve unique scoring pattern. As users are asked to look summaries with 15 or 20, they might be satisfied with summaries 10 or 15 images. So, the total number of evaluators votes of summaries 15 and 20 images could be less than the summary of 10 images.

Figure 11 shows the score of each result sets with different number of image. The points made in general observation can be easily observed with this graph. The total scores for each results set with different number of images (i.e. 10, 15 and 20) are more or less similar.

Another evaluation is based on the number of images inside the summary. Here, it is intended to know the evaluator's perception to see the summary with different number of images. As per all the completed experiments for each image set, initially, it is started with 10 images, then 15 images and finally, 20 images of the same representative set are shown.

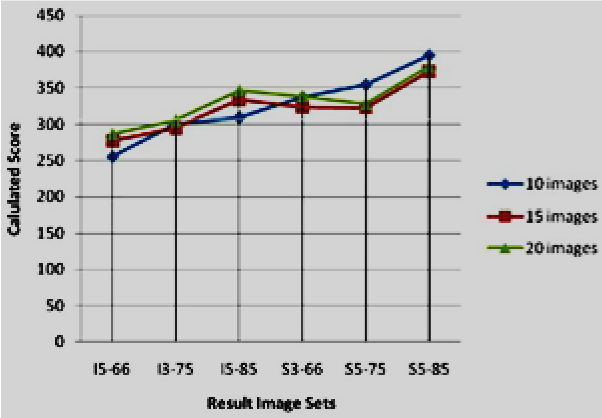


Fig. 11. Overall evaluation

The evaluation states that, after seeing summaries of 10 images, 89% of the evaluators were interested to rate the summary of 15 images. Figure 12 shows that the interest ratio of evaluators for the 15 images set. That means they would like to change the rating for the new 15 images summary of the same representative set. It can be either positive or negative. The result shows four results sets secure higher or equal total score than the 10 images, while two sets namely S5-75 and S5-85 give less score compared to 10 images.

Same for the next set of 20 images, there were 69% evaluator who would like to rate 20 images summaries. Figure 13 shows that the interest ratio of evaluators for the 20 images set. Evaluation shows that the summary with 20 images secures almost same result as 15 images. So, there is no much difference in the total score between the summary of 15 images and 20 images.

The interest level to see summary of 10 to 15 images and 15 to 20 images shows that the number of images inside the summary affects the result.

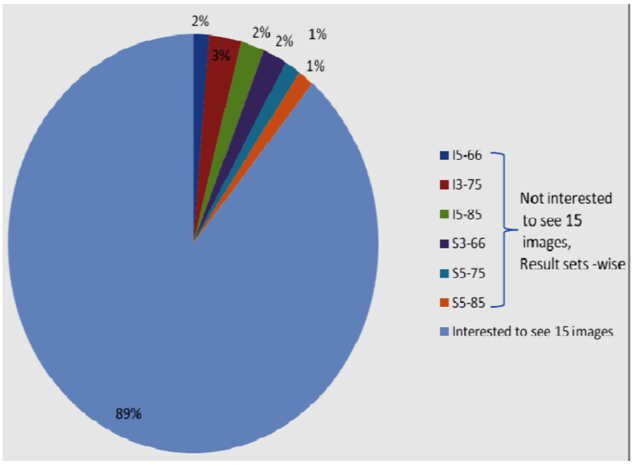


Fig. 12. Distribution of interest of participant to look summary of 15 images

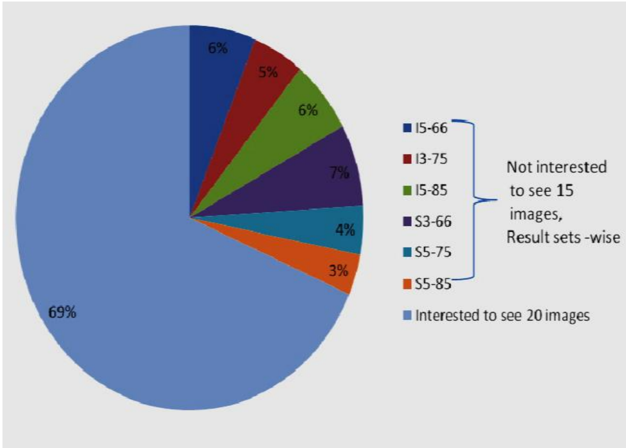


Fig. 13. Distribution of interest of participant to look summary of 20 images

IX. Conclusion and Future Work

The focus of this work is to generate the best representative set and summary of the large dataset by cropping images randomly and sequentially with different coverage. Though, the algorithm takes too much time for the overall computation, a good human based evaluation is found for sequential datasets rather than random datasets. It is observed that the higher coverage gives the best result regardless of sequential or random windows. In addition, it is also found that the number of images inside a summary varies the results.

For the future work, a further analysis on more values of number of windows *N* and coverage *C* can be done. The work can also be extended for finding out the faster algorithm to deal with the larger dataset. It is also planned to see the number of images inside the summary aspect in order to have a more comprehensive conclusion.

References

[1] Cavnar, William B., and John M. Trenkle. "N-gram-based text categorization." Ann Arbor MI 48113.2 (1994): 161-175..

[2] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision,60(2):91-110,2004.

[3] Lowe, David G. (1999). "Object recognition from local scale-invariant features". Proceedings of the International Conference on Computer Vision. 2. pp. 1150–1157.doi:10.1109/ICCV.1999.790410.

[4] Mor Naaman, Yee Jiun Song, Andreas Paepcke, and Hector Garcia-Molina. Automatic organization for digital photographs with geographic coordinates. In Proceedings of the Fourth ACM/IEEE-CS Joint Conference on Digital Libraries, 2004.

[5] I. Simon, N. Snavely, and S. M. Seitz. Scene summarization for online image collections. In ICCV '07: Proceedings of The 11th IEEE international Conference on Computer Vision. IEEE, 2007.

[6] A. Jaffe, M. Naaman, T. Tassa, and M. Davis. Generating summaries for large collections of geo-referenced photographs. Proc. Int. Conf. on World Wide Web, pages 853–854, 2006.

[7] Lyndon S. Kennedy and Mor Naaman. Generating Diverse and Representative Image Search Results for Landmarks. Proceedings of the 17th international conference on World Wide Web, NY,USA, 2008.

[8] T. L. Berg and D. A. Forsyth. Automatic ranking of iconic images. Technical report, U.C. Berkeley, January 2007.

[9] Pigeau and M. Gelgon. Organizing a personal image collection with statistical model-based ICL clustering on spatio-temporal camera phone meta-data. Journal of Visual Communication and Image Representation, 15(3):425–445, September 2004.

[10] Matthew Cooper, Jonathan Foote, Andreas Girgensohn, and Lynn Wilcox. Temporal event clustering for digital photo collections. In Proce of the eleventh ACM international conference on Multimedia, pages 364– 373. ACM Press, 2003.

[11] Adrian Graham, Hector Garcia-Molina, Andreas Paepcke, and Terry Winograd. Time as essence for photo browsing through personal digital libraries. In Proceedings of the Second ACM/IEEE-CS Joint Conference on Digital Libraries, 2002.

[12] Benjamin S. Braun. BioTechniques, Vol 59, No. 2, August 2015, pp. 82-86.

Smart City Architecture: Vision and Challenges

Narmeen Zakaria Bawany
Systems Research Laboratory,
FAST-National University of Computer and Emerging Sciences
Karachi, Pakistan

Jawwad A. Shamsi
Systems Research Laboratory,
FAST-National University of Computer and Emerging Sciences
Karachi, Pakistan

***Abstract*—The concept of smart city was born to provide improved quality of life to citizens. The key idea is to integrate information system services of each domain, such as health, education, transportation, power grid etc., of the city to provide public services to citizens efficiently and ubiquitously. These expectations induce massive challenges and requirements. This research is aimed to highlight key ICT (Information and Communication Technology) challenges related to adaptation of smart city. Realizing the significance of effective data collection, storage, retrieval, and efficient network resource provisioning, the research proposes a high level architecture for smart city. The proposed framework is based on a hierarchical model of data storage and defines how different stakeholders will be communicating and offering services to citizens. The architecture facilitates step by step implementation towards a smart city, integrating services, as they are developed in a timely manner.**

Keywords—Smart city; Data management; urban technology; socio-technical systems; smart city architecture

I. Introduction

Smart city brings enormous opportunities and exciting challenges. In general, a metropolitan area can be considered as smart when city operations and services such as healthcare, education, transport, parking, and electricity grid are supported through ICT infrastructure in order to facilitate efficiency and ease of operation. Some valid examples of such services would be looking for a job, applying for a driver's license, buying of car and property, change of the address, request for a passport, start of a new business, reporting of a crime, declaration of income taxes, seeking health services, and so on. All such services require execution of several services under an orchestrated coordination. The smart city design must be citizen-centric. Despite the complexity of the city's systems, the architecture must bring benefits to the people regardless of their ICT abilities.

The primary goals of the smart city include, offering digital means for supporting social needs in all daily transactions, to adapt the citizens to the notion of the information society and to collect information from the public departments and citizens in order to support sustainable growth of the city. Above all, development of smart city will give way to implementation of a citizen-centered public administration, where corruption phenomena and time-consuming bureaucratic procedures are eliminated [1],[2].

Smart city has been actively studied and researchers have come up with different definitions, frameworks, and implementations of smart city [3],[4],[5],[6],[7]. The key objective of almost all the research is to present a strategy to mitigate the problems generated by the urban population growth by using information and communication technology. Cities and megacities generate new kinds of problems. Difficulty in waste management, scarcity of resources, air pollution, human health concerns, traffic congestions, electricity generation, distribution, and billing, and inadequate, deteriorating, and aging infrastructures are among the more basic technical, physical, and material problems [8][9][10]. Many cities have taken the initiative in this direction and more work is in progress. [11],[12],[3],[13],[14]. In this above context, a smart city system can be considered as a massive information system comprising of several smaller but efficient subsystems that may be connected with each other. Each of these subsystems require efficient and enhanced capabilities to handle components for data storage, information retrieval, networking, and communication.

There are numerous challenges in realization of a city that monitors and integrates all of the city infrastructure and services to leverage the collective intelligence. From technical infrastructure, that needs to be put in place, to adaption of system by citizens various technical and non-technical obstacles are to be dealt with. For example, smart transport management system has immense potential to improve road safety and reduce traffic congestions. However, the system may require new types of networks such as Vehicular AdHoc network (VANETs) to facilitate the communication between vehicles. Huge infrastructure will be needed across the city to make intelligent traffic management possible. Similarly, there are stringent requirements of security and privacy on data, which will be collected and disseminated.

The purpose of this paper is to highlight key challenges related to information system management in smart cities and to propose a model to address these challenges. For instance, for a smart city, networking infrastructure should support extensive monitoring and comprehensive data collection. Further, efficient means of data storage and retrieval are also needed. A critical requirement for smart city is to make the relevant data available for applications in order to materialize the idea of smart city. This induces additional requirements of availability, scalability, security, and privacy. Considering the massive size of cities and increasing trends of urbanization, severity of these challenges and requirements are extensive. Considering these challenges, this paper presents a high-level architecture for smart city. Our proposed model focuses on efficient storage and distribution of data in order to facilitate different operations of smart city.

The model consists of hierarchical storage with distributed zones of data, which allows rapid and autonomous processing of data and permits instant decision making. The hierarchal model is also beneficial in providing data redundancy and reduces network latency. The hierarchal model also compliments service-oriented architecture, which restricts access to data and enhances and privacy. The zonal model provides access to data for development of third party applications which may subsequently lead to smart city app store.

The main contributions of this paper are follows:

- Accentuate key challenges in smart city implementation
- Propose a zone-based architecture for data storage and management in order to address key challenges for smart city information system management.
- Emphasize an open data model for smart city that gives way to third party application development.

II. Related Work

Mahmoud et al. [15] proposes centralized operational platform for smart city. He proposes pyramid like conceptual development pyramid with smart infrastructure at the bottom, followed by smart database resources, smart building management, smart interface and smart city at the top. The primary background of this research is to provide a single management system for the collective processing and management across multiple sub-systems, applications and controllers. The GIS automation platform is used for administering and maintaining the assets. Cisco, IBM, HP, and other global information and communications technology (ICT) players have launched their own smart planet and smart city initiatives. The smart city market is estimated to be hundreds of billions of dollars by 2020, with annual spending reaching nearly $16 billion [16]. Harrison et al's[5] , proposes a three layer architecture for smart city, an instrumented layer, interconnected layer and intelligent layer. Instrumentation enables the capture and integration of live real-world data through the use of sensors. The activities found at this level can measure water quality, collect electrical meter readings for a grid, or provide building measurements to determine its energy usage. Interconnected layer also called middle layer integrates the data collected at instrumented layer and maps various inputs into events of interest. This layer incorporates the data into an enterprise computing platform and the communication of such information among the various city services. The intelligent layer processes relevant city data in a broader context to identify city-relevant events that need to be analyzed or acted upon. A service-oriented architecture (SOA)-based model, along with existing applications and management systems, is used to transform data and perform analysis. Motivated by the limitations of current internet architecture highlighted by Future Internet Assembly Piro et al.[17] proposes the smart city platform based on the emerging Named Data Networking (NDN) network paradigm. The service platform proposed in this work is built on two levels: the service layer and the technology layer, both interacting through a NDN interface. For service provisioning, three consecutive phases are conceived (1) the Discovering phase, (2) the Security Initialization phase, and (3) the Service Usage phase. Dirks and Keeling [19] suggested that cities are based on six core systems comprising different infrastructures related to their key functions that is people, business, transport, communication, water and energy. Peoples system includes public safety, health and education while businesssystem refers to the environment that businesses face in terms of policy and regulation. All six systems together forms systems of systems. Anthopoulos and Tsoukalas[18] proposes a multi-tier generic architecture which can describe all the types of attributes needed to support the smart city context. This generic architecture contains the four layers. User Layer that consists of all e-service end-users and stakeholders of a smart city, Service Layer incorporates all the particular e-services being offered by the smart city, Infrastructure Layer contains network, information systems and other facilities, which contribute to e-service deployment and Information Layer presents all the information, which is required, produced and collected in the smart city.Doran et al.[19] has developed a smart city model describing the three main components that the smart city system aims to Integrate economic component, environment component and social component. Economic component includes public administration and economic actors. It covers governance models, urban regeneration, open data, big data, bandwidth, mobility, cloud computing, security, business intelligence, etc. Environmental component includes resources and managerial infrastructures. It covers water, air, energy and waste management, public and alternative transportation, geographical information, green buildings, green spaces, smart growth, climate change measurement, etc. Social component includes citizens. It covers community life, urban mediation, participatory democracy, social innovation, human-scale cities, civic participation, proximity services, etc.

III. Key Challenges

Smart city being a gigantic information system, comprising sub information systems, bring enormous challenges. Smart city initiatives are fundamentally based on ICT. The latest developments in cloud computing, Internet of Things, open data, semantic web, and future internettechnologies will be leading technologies to enable smart city development. These technologies can altogether provide infrastructure, applications and turn-key solutions for offering various services. ICT has immense potential to enhance the management and functioning of a city which in turn will benefit citizens. However, all these technologies have their inherent challenges and limitations. Combining all these technologies to form a complex system, like smart city, will magnify the amplitude of many long-standing challenges.

A. *IT Infrastructure*

The development of ICT infrastructure, from communication channels to sensors and actuators in physical space remains a huge barrier in taking a smart city initiative. Lack of infrastructure is a significant barrier in achieving smart city objectives [1][20][21][22][23][24]. Reliable, scalable and high speed network connectivity and infrastructure is a key foundation for integrating information systems across city. This infrastructure has to be in place before smart city services are offered to stake holders. Consequently, the adequate reliable IT infrastructure which tends to be scalable is a critical challenge for the implementation of smart city.

B. *Security and privacy*

As the trend move towards smart city, and different gadgets becomes integrated within everyday lives, personal data collection creating privacy issue is intrinsic. For example, an intelligent traffic management app that updates user about traffic congestion will require that location of the user is collected. To meet the security and privacy requirements is a fundamental challenge for smart city system where huge amount of sensitive data processing is involved [1][20][22][25][26][27]. Threats from hackers, intruders, viruses, worms, Trojans etc has immense potential to disrupt the services and bring down the whole system resulting in enormous losses. Extensive security approaches are required to secure sensitive data at levels including collecting, processing, storing and disseminating. Security and privacy are significant not only in making data and services available, but they are also essential in building citizens trust and confidence in using these systems.

C. *Big data Management*

Understandably, huge amount of data will be generated from all smart city systems. To handle various types of data, with varying velocity an efficient big data management system is required. This system has to be reliable and scalable with no downtime. Continuous generation, collection, processing and storage of massive heterogeneous data from countless smart city sensors has its own intrinsic challenges. Nevertheless, big data collected across city is extremely useful and is vital for achieving objective of smart city. For example, GPS sensors installed on vehicles may give valuable information about transport flow but it will generate huge amount of high velocity data.

D. *Cost*

Smart city evidently would require acquiring enormous IT infrastructure. Huge financial investment is to be obtained to put the system in place. Millions of sensors, thousands and thousands of networking equipment and computing devices will be needed to get end to end connectivity. Similarly, requirement of IT professional and consultancies will make up a considerable amount of expense. This investment is not limited to one time only, operational and maintenance cost of such a huge real time system will be much higher. To meet stringent reliability and efficiency requirements more resources will be required leading to higher overheads. For example, in case of smart traffic management system each car has to fitted with a sensors and thousands of road side units must be installed. Such system cannot afford a downtime and must be highly efficient and reliable. In a typical urban city, cost of implementing such a project may be around millions of dollars.

E. *Heterogeneous environment and Interoperability*

Smart city architecture is characterized by Heterogeneity of networks, applications, devices, platforms etc. For instance, Vehicular networks may require adhoc networks like VANETs while close range wireless devices may work on zigbee. The legacy applications and their integration with emerging technologies will also be significant challenge.

F. *Efficiency, Availability and Scalability*

Critical systems cannot afford downtime and requires very high availability. Challenge of meeting the tough requirements of availability is directly proportional to the size and complexity of the system. Smart city, not only have a huge infrastructure but its size and complexity will be increasing leaps and bounds as it goes operational. Massive amount of data generated across the city will make availability, scalability and efficiency a critical challenge.

To maintain the efficiency of such a gigantic system is obligatory. Performance optimization of each system, proficient planning, efficient utilization of resources and immediate response to queries are few of the key benefits of smart city. For example, it is expected that in the time of crisis different civil departments may communicate almost instantaneously with each other resulting in immediate access to emergency services. Consequently, resources requirements for cities can be calculated and projects can be initiated well before the demand crosses supply.

For example, energy requirements can be calculated via smart grid and government can plan to meet the energy requirements for upcoming years before it goes critical. Smart grid, if put in place, can assist in maximizing the productivity of current resources. Architectural scalability must be an intrinsic characteristic of such a system where data processing, management, analytics requirements may increase by leaps and bounds. Response time and throughput must not be compromised.

G. *Social Adaption*

The smart city seems to be an ideal solution to overcome existing and emerging urban population problems. Although, researchers have also identified challenges with reference to inequality, digital divide and changing cultural habits [1]. Social adaption of such a system requires changing social habits of citizens generally and city management people specifically.

H. *App development*

Faster development of new and innovative applications will required so that citizens can take maximum advantage of data that is being collected. If application development is limited to city management it is very likely that people will be disappointed by slow application development. For example, one of the key reasons behind Android’s success and wide adaption is its play store, its huge app base where countless applications are uploaded every day.

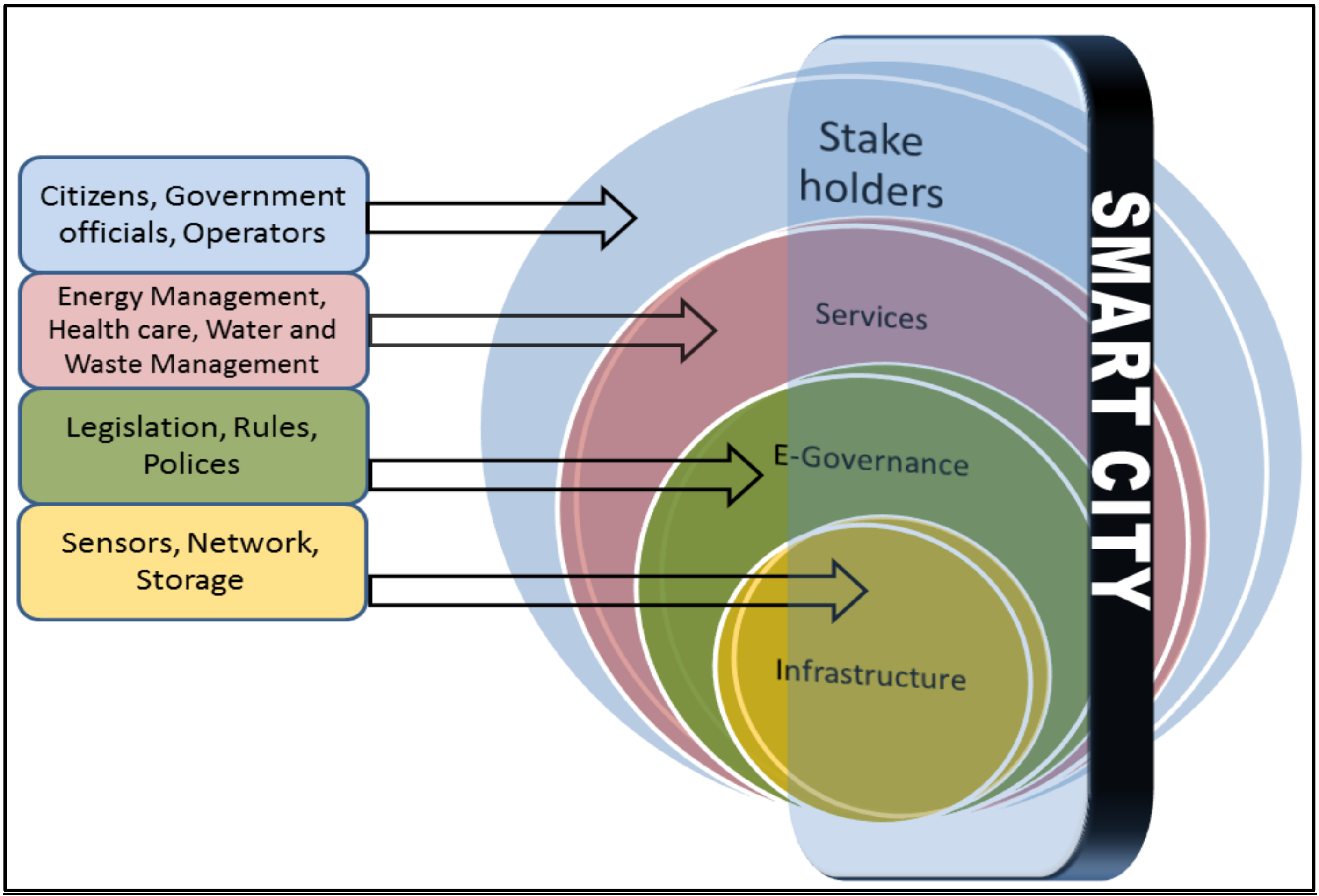


Fig. 1. Smart City Layers

IV. Proposed Architecture for Smart City

Figure 1 shows the key layers of a smart city. ICT infrastructure forms the foundationof a smart city. It is the fundamental layer on which all other components rely. ICT infrastructure comprises high speed wired and wireless network connectivity, high end data centers, physical space enrichment with smart devices, sensors, actuators and much more. E-governance layer facilitates the development of strategic connections between various departments of public sector organization. This layer formulates the policies, rules and legislations to improve the performance of government organizations and offer potential benefits to citizens. Depending upon the infrastructure and policies formulated at e-governance layer, various public services will be offered to citizens and other stake holders, in an efficient and effective manner, ubiquitously.

Traditionally, the city departments have been functioning independently, sharing limited information with other departments in overall city. ICT technologies and infrastructure that are in place in departments only focuses on its on operation. This setup creates lot of chaos and delays the process in implementing or executing a service. Citizens face hardships where information from multiple departments is required. Similarly, in case of disaster, reaction time of departments where mutual communication and coordination is required gets way slower. Resource planning on the basis of real time data is not possible.

This architecture focuses on sharing information not only across departments but also offer data services to other interested parties via open data model. We argue that in smart city, citizens should be provided with cross application services. Using real time data from any number of domains can speed up the reaction/completion time of process improving the overall efficiency of city services. By sharing information, which is the first source of value, the other domains and the city can gain insight and identify impending problems with the intent of responding before these problems occur or escalate. In a smarter city, responses to information and events can be almost instantaneous. Similarly, availability of real time data will lead to efficient resource planning which helps in optimizing resource utilization both at departmental level and overall city level. In a smarter city, city-wide operational processes using data from any number of domains can continuously predict and react to events and trends that are affecting the city.

A. *Holistic View*

Figure 2 shows the holistic view of smart city architecture. The approach focuses on managing the city as a system of sub-systems. Each autonomous sub-system is connected to

Central Data Management System (CDMS), that is fully integrated and interconnected with all sub-systems. All systems share their data with CDMS which in turn can provide cross domain services to citizens. CDMS acts as an integration point for information coming from sub systems. CDMS can use the information and data at its disposal to make better decisions in real time.

B. *Zone level architecture*

Each sub system collects data from different zones across the city as shown in Figure 3. Each zone maintains its own data center cloud at site level. For example Water and waste management system will have numerous sites across a city each maintaining its own zone level data. Similarly, Electricity supply and management, Gas management and other public services systems will also have various zonal sites across the city.

Traditionally, city is divided in zones for administration and each zone has its own public service offices. The proposed architecture fits well in such a scenario where each public service and or utility has a zonal representation. Keeping in view, social and technical challenges, it is assumed that smarter city objective will be achieved in an incremental manner. By introducing the concept of zone level service, this architecture supports step by step movement towards a smart city.

Each utility zonal site has its own autonomous system that comprises a local data center, wireless sensor and network infrastructure and technology and facility-related components. A zonal site is an environment that provides processing, storage, networking, management of data within a zone. These zonal sites interact with each other via web services to provide cross zone services to users.

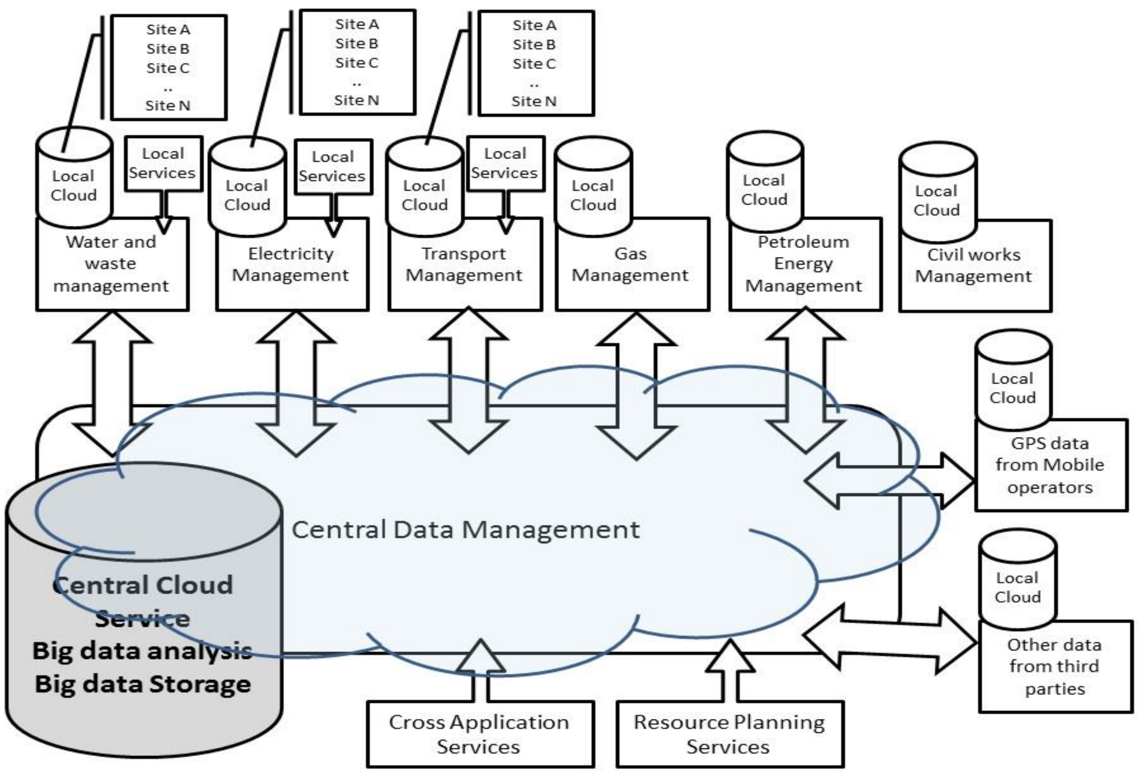


Fig. 2. Holistic view of proposed architecture

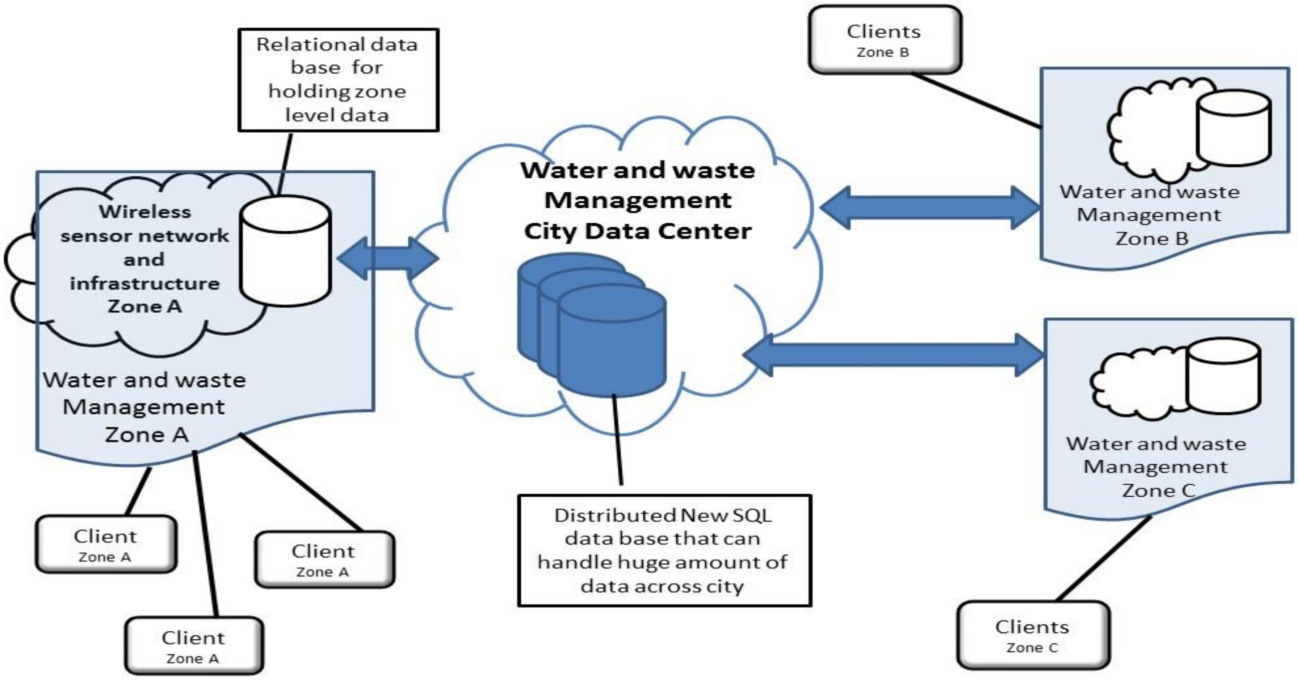


Fig. 3. Zone wise view of proposed architecture

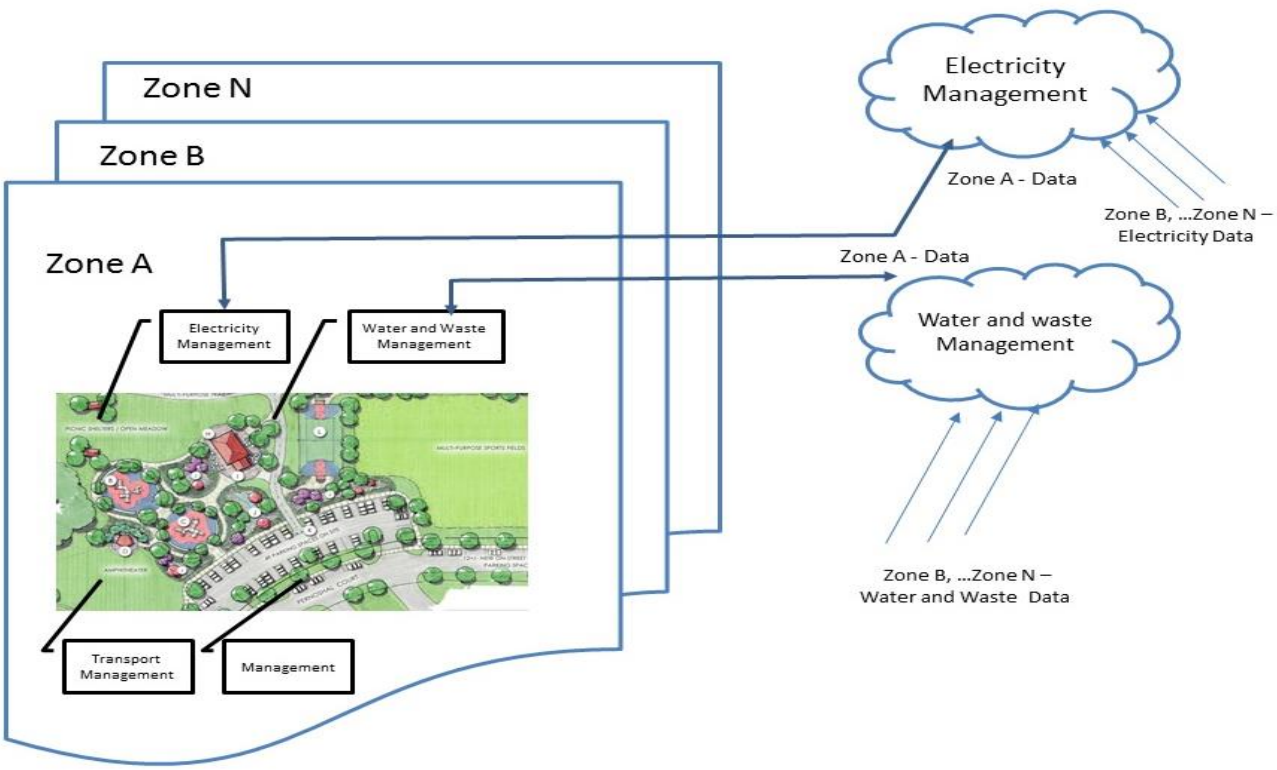


Fig. 4. Data Management at Zonal Sites of Water and Waste Management System. Zone A, Zone B and Zone C are connected to Center of Water and Waste Management System

C. *Data Management*

Figure 4 depicts water and waste management system for a city. Zone A maintains a local relation database to store data from sensors and other devices in its own region. Clients from Zone A will be receiving services from Zone A datacenter. Relational database is chosen because of its proven ACID(atomic, consistent, isolated, durable) properties. Query optimization may also help in fast processing of user query. All Zonal (Zone A, Zone B, Zone C…Zone N) utility data centers will be connected to main utility data center over reliable high bandwidth connection. Aggregation of data collected from all zonal data centers is stored in distributed New SQL databases like Google Spanner, Nuodb, ClustrixDB. It is assumed that huge data will be generated from all datacenters and its aggregation and center will require big data technologies for management. In order to offer fast and efficient retrieval of data to provide cross zone services New SQL databases are preferred over traditional relational databases and No SQL databases

D. *Service-oriented architecture*

All components of a smart city will be integrated using service oriented architecture. Smart city architecture is essentially a large-scale distributed system which is inherently complex and decentralized. Different platforms, heterogeneous environment and variety of sensors networks will lead to interoperability issues. Service Oriented Architecture with its open standards like XML, WSDL, SOAP, and UDDI not only provides interoperability among diverse platforms but also supports modular design, software reuse, interoperation and application integration. Services offered by one utility let say Water and waste management system can be utilized by Disaster Management System which may be entirely different platform.

E. *Open data Model*

It is assumed that huge data will be generated at each utility city center. The proposed architecture introduces a novel idea of open data model. Some of the data collected by utility center is made available to researchers and developers via API. Freelance developers, Professional Application developers and third parties will also have opportunities to develop new analytics tools, new services, etc. Revenue can be generated by charging fee or by advertising. Variety of application will be available for citizens. For example, Traffic management system for city will be collecting huge amount of data about vehicle movements and traffic jams. Applications that can predict peak traffic load resulting in traffic jams, or updates about VIP movements that results and road closures, etc. will be handy utility for citizens. Though implementing such a model has stringent privacy and security requirements.

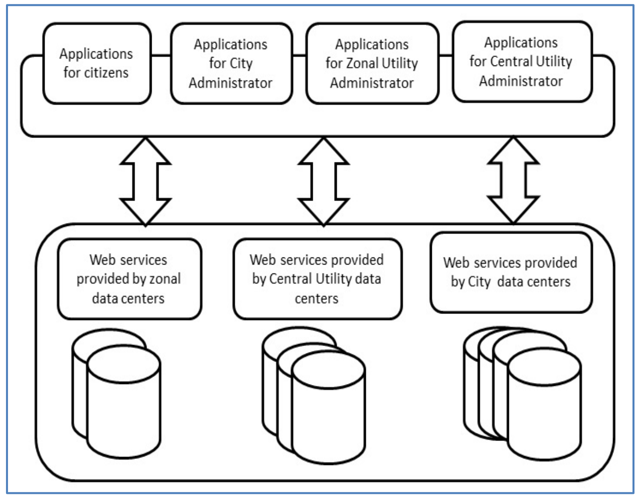


Fig. 5. Service Oriented Architecture

V. PROPOSED MODEL AND CHALLENGES

Table 1 lists the key challenges for smart city architecture along with the proposed solution. IT infrastructure and cost challenges include acquiring and laying down enormous network infrastructure (Wired and wireless, bandwidth, connectivity), Smart devices, sensors, kiosks, Wi-Fi hotspots and much more. The proposed model enables Zone wise implementation of each public service which does not require entire infrastructure at once. Nevertheless, this does not bring down the requirement of IT infrastructure but facilitates in moving steadily towards a smarter city. Smart city architecture is characterized by Heterogeneity of networks, applications, devices, platforms etc. The proposed model is based on Service Oriented Architecture. Exposing data services as web services can make information accessible to a wide variety of clients[28]. Architectural scalability and availability must be an intrinsic characteristic of smart city system where data processing, management and analytics requirements may increase by leaps and bounds. These critical systems cannot afford downtime and requires very high availability. The proposed model supports both horizontal and vertical scalability. Horizontal scalability means that more and more public services can be plugged in easily. Zone wise implementation of each public service provides vertical scalability ensuring Quality of Service(QoS) at zone level. Security and privacy are inherent challenges; the proposed model recommends exposing data as web service.WS-Security is particularly useful because it provides encryption-based, message-level security data[29][30]. In the proposed model, each public service is running its own data center at zone level.

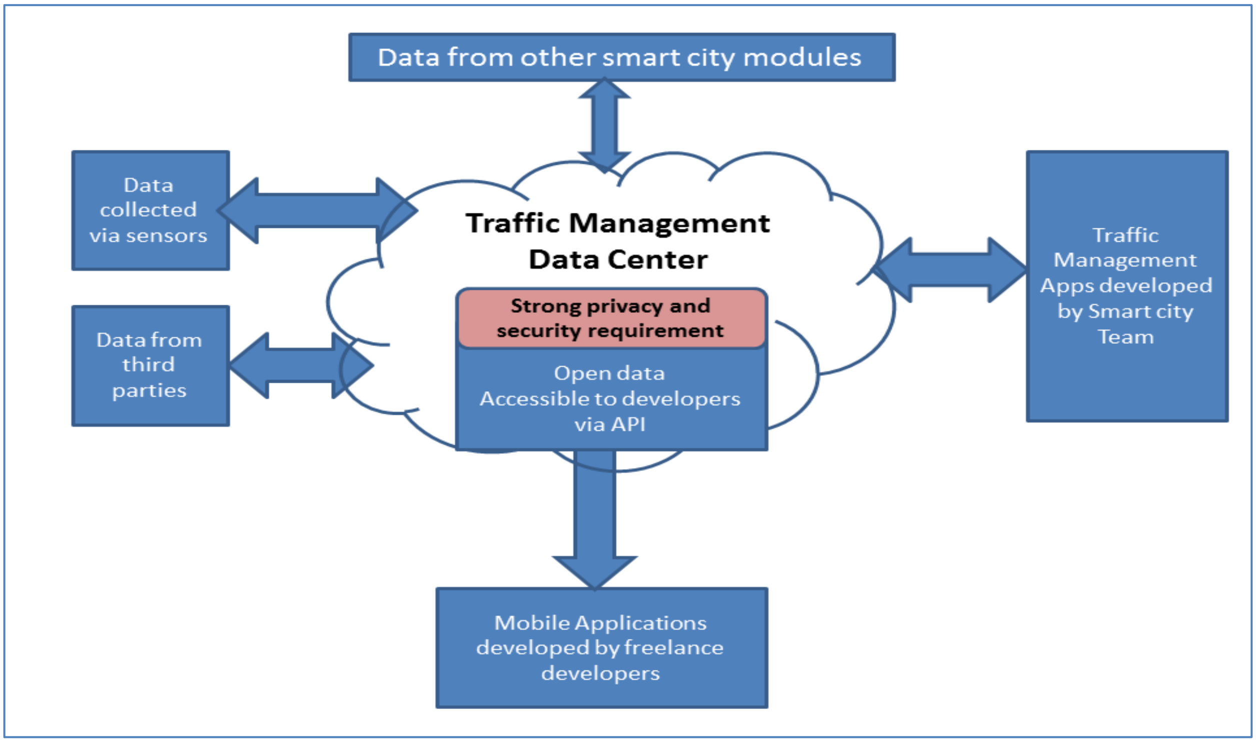


Fig. 6. Exposing data via API

However, they exchange information in real-time but the system itself is not controlled centrally. Attacks on one utility's zonal data center cannot bring down the whole system. Besides, redundancy for replication and efficiency is inherent in cloud based systems. This model also proposes that the anonymized data is exposed via web services, with the consent of user to ensure privacy. Nevertheless, personal data collected will be for greater good. According to McKinsey[31] , applying personal location data has the potential to provide more than $800 billion in economic value to individual consumers and organizations over the next decade.Efficiency is critical for the success and adaption of smart city system. Citizens will expect to get their problems resolved more efficiently. For e.g. more immediate access to emergency services, cities become more resilient in times of crisis. Similarly, efficient planning and utilization of resources is presumed. The proposed model include cross domain services that can process queries from multiple domains in real time. Various departments can co-ordinate and cooperate instantly and share their data via web services. Resource planning at central data management system (CDMS), and at each utility center results in efficient planning and utilization of resources as real data is available for analysis. Furthermore, hierarchical cloud model assists in efficiently handling big data that is continuously generated from heterogeneous sources.

TABLE I. KEY CHALLENGES AND PROPOSED MODEL

	Challenge	Proposed model
1	**IT Infrastructure**	**Zonal Sites of each public utility** Zone wise implementation of each public service does not require all the infrastructure at once. The proposed model streamlines step-wise implementation of smart city. Nevertheless, this does not brings down the requirement of IT infrastructure but facilitates in moving steadily towards a smarter city.
2	**Cost**	**Zonal Sites of each public utility** Zone wise implementation of each service may not require huge investment at once. In a long run, use of the ICT to deliver public services will reform the speed and effectiveness of public service delivery and administration, in turn, providing improved service delivery, Reduced consumer costs and Social benefits [32].
3	**Heterogeneous environment /Interoperability**	**Service Oriented Architecture.** The proposed model is based on Service Oriented Architecture. Exposing data services as web services can make data service information accessible to a wide variety of client. Web services makes it possible for systems to integrate with each other independent of underlying platform.
4	**Availability and Scalability**	**Zonal Sites of each public utility** The proposed model supports both horizontal and vertical scalability. Horizontal scalability means that more and more public services can be plugged in easily. Zone wise implementation of each public service provides vertical scalability ensuring Quality of Service(QoS) at zone level. Hierarchical cloud model, basically a cloud model, is characterized by high availability and scalability inherently. Furthermore, each zone can enhance its resources, as and when required, to ensure 24/7 availability to its users.
5	**Security**	**Service Oriented Architecture – Using Web Services** The proposed model recommends exposing data as web service. WS-Security is particularly useful because it provides encryption-based, message-level security data
6	**Privacy**	**Service Oriented Architecture – Using Web Services** The proposed model defines user as the owner of data. However, policies and rules can be defined by the e-governance layer. Data is only provided to known an authentic party. It is recommended that anonymized data is exposed via web services after the consent of user.
7	**Efficiency**	**Hierarchical Cloud Model** As data zone level queries will be entertained by zonal data center, the proposed model is expected to perform better. The cross domain services can process queries from multiple domains in real time. Various departments can co-ordinate and cooperate instantly and share their data via web services. Resource planning at central data management system (CDMS), and at each utility center results in efficient planning and utilization of resources as real data is available for analysis.
8	**Big data Management**	**Hierarchical Cloud Model** **Zonal Sites of each public utility** Zone wise distribution of data helps in storing and analyzing data efficiently. For example, city transport system will have zonal data centers across city. The data generated from each zone will be collected, stored and analyzed at zone level. Nonetheless, aggregated data of each zone will be sent to main city transport system. This division, helps in efficient management of big data.
9	**Social Adaption and app development**	**Open data model – Smart city App store**

VI. CONCLUSION

The paper presents a four layer view of a smart city. Requirements and challenges in realizing a smart city are presented. Hierarchical data management architecture is proposed that facilitates distributed data management at zone level enhancing efficiency, availability and scalability of services. Service oriented nature of the proposed architecture conforms to the heterogeneous environment of smart city. The Open data model for smart city will give way to researchers and free-lance programmers to contribute to smart city app development leading to smart city app store. E-governance layer is also one of the fundamental layers of smart city but this research deals mostly with the technicalities of smart city architecture. The socio-techno challenges like accessibility, digital divide, Adaptability etc. are not addressed by the proposed model.

REFERENCES

[1] H. Chourabi, T. Nam, S. Walker, J. R. Gil-Garcia, S. Mellouli, K. Nahon, T. a. Pardo, and H. J. Scholl, “Understanding Smart Cities: An Integrative Framework,” 2012 45th Hawaii Int. Conf. Syst. Sci., pp. 2289–2297, Jan. 2012.

[2] R. Giffinger and N. Pichler-Milanović, “Smart cities: Ranking of European medium-sized cities,” 2007.

[3] R. Giffinger, R., Fertner, C., Kramar, H., Kalasek and E. Pichler-Milanović, N., & Meijers, “Smart cities Ranking of European medium-sized cities,” Vienna, Austria Cent. Reg. Sci. (SRF), Vienna Univ. Technol., 2007.

[4] Bowerman, B., et al. "The vision of a smart city." 2nd International Life Extension Technology Workshop, Paris. 2000.

[5] Harrison, Colin, et al. "Foundations for smarter cities." IBM Journal of Research and Development 54.4 (2010): 1-16..

[6] Moss Kanter, Rosabeth, and Stanley S. Litow. "Informed and interconnected: A manifesto for smarter cities." Harvard Business School General Management Unit Working Paper 09-141 (2009).

[7] Alawadhi, Suha, et al. "Building understanding of smart city initiatives." Electronic Government. Springer Berlin Heidelberg, 2012. 40-53.

[8] Marceau, Jane. "Introduction: Innovation in the city and innovative cities." Innovation: Management, Policy & Practice 10.2-3 (2008): 136-145.

[9] D. Washburn, U. Sindhu, S. Balaouras, R. A. Dines, N. M. Hayes, and L. E. Nelson, "Helping CIOs Understand ' Smart City ' Initiatives," 2010.

[10] R. Hollands, "Will the real smart city please stand up? Intelligent, progressive or entrepreneurial?," City, 2008.

[11] Bakıcı, Tuba, Esteve Almirall, and Jonathan Wareham. "A smart city initiative: the case of Barcelona." Journal of the Knowledge Economy 4.2 (2013): 135-148.

[12] Caragliu, Andrea, Chiara Del Bo, and Peter Nijkamp. "Smart cities in Europe." Journal of urban technology 18.2 (2011): 65-82.

[13] L. Hopkins and P. K. Isom, "Front cover Smarter Cities Series : Understanding the IBM Approach to Efficient Buildings."

[14] Kehoe, Mike, et al. "Smarter cities series: A foundation for understanding IBM smarter cities." Redguides for Business Leaders, IBM (2011).

[15] Al-Hader, Mahmoud, et al. "Smart city components architicture." Computational Intelligence, Modelling and Simulation, 2009. CSSim'09. International Conference on. IEEE, 2009.

[16] P. Research, "Smart Cities. Intelligent Information and Communications Technology Infrastructure in the Government, Buildings, Transport, and Utility Domains," 2011. [Online]. Available: http://www.navigantresearch.com/.

[17] G. Piro, I. Cianci, L. A. Grieco, G. Boggia, and P. Camarda, "Information Centric Services in Smart Cities."

[18] Anthopoulos, Leo G., and Ioannis A. Tsoukalas. "The implementation model of a Digital City. The case study of the Digital City of Trikala, Greece: e-Trikala." Journal of e-Government 2.2 (2006): 91-109.

[19] Daniel, Sylvie, and Marie-Andree Doran. "geoSmartCity: geomatics contribution to the smart city." Proceedings of the 14th Annual International Conference on Digital Government Research. ACM, 2013..

[20] P. Suresh, "Understanding Challenges in e-Governance," Better Account. with e-governance, pp. 61–63, 2011.

[21] Dillon, J., and W. Pelgrin. "E-government/commerce in New York State." Office of Technology, New York, NY (2002)..

[22] Alghamdi, Ibrahim A., Robert Goodwin, and Giselle Rampersad. "E-government readiness assessment for government organizations in developing countries." Computer and Information Science 4.3 (2011): p3..

[23] Giffinger, Rudolf, et al. Smart cities-Ranking of European medium-sized cities. Vienna University of Technology, 2007.

[24] Vasseur, Jean-Philippe, and Adam Dunkels. Interconnecting smart objects with ip: The next internet. Morgan Kaufmann, 2010..

[25] Wan, Jiafu, et al. "From machine-to-machine communications towards cyber-physical systems." Computer Science and Information Systems 10.3 (2013): 1105-1128.

[26] Sanchez, Aurora, et al. "The relationship between IT for communication and e-government barriers." AMCIS 2003 Proceedings (2003): 104.

[27] Laplante, Phillip A. "Who's Afraid of Big Data?." IT Professional 15.5 (2013): 6-7..

[28] Hamad, Hatem, Motaz Saad, and Ramzi Abed. "Performance Evaluation of RESTful Web Services for Mobile Devices." Int. Arab J. e-Technol. 1.3 (2010): 72-78.

[29] Yuan, Eric, and Jin Tong. "Attributed based access control (ABAC) for web services." Web Services, 2005. ICWS 2005. Proceedings. 2005 IEEE International Conference on. IEEE, 2005.

[30] Atkinson, Bob, et al. "Web services security (WS-Security)." Specification, Microsoft Corporation (2002).

[31] Manyika, James, et al. "Big data: The next frontier for innovation, competition, and productivity." (2011).

[32] G. Bonham, J. Seifert, and S. Thorson, "The transformational potential of e-government: the role of political leadership," 4th Pan Eur. Int. Relations Conf. Univ. Kent, 2001.

[33] Hannola, Lea, and Päivi Ovaska. "CHALLENGING FRONT-END-OF-INNOVATION IN INFORMATION SYSTEMS." Journal of Computer Information Systems 52.1 (2011).

[34] Barba, Carolina Tripp, et al. "Smart city for VANETs using warning messages, traffic statistics and intelligent traffic lights." Intelligent Vehicles Symposium (IV), 2012 IEEE. IEEE, 2012.

[35] Khazanchi, Deepak. "Information technology (IT) appropriateness: the contingency theory of 'fit'and IT implementation in small and medium enterprises." Journal of Computer Information Systems 45.3 (2005): 88-95.

[36] Chen, Kuanchin, and Alan I. Rea. "Protecting personal information online: A survey of user privacy concerns and control techniques." Journal of Computer Information Systems 44.4 (2004): 85-92.

[37] Shelton, Taylor, Matthew Zook, and Alan Wiig. "The 'actually existing smart city'." Cambridge Journal of Regions, Economy and Society 8.1 (2015): 13-25.

[38] Ricciardi, Francesca, and Stefano Za. "Smart City Research as an Interdisciplinary Crossroads: A Challenge for Management and Organization Studies." From Information to Smart Society. Springer International Publishing, 2015. 163-171.

[39] Kitchin, Rob. "The real-time city? Big data and smart urbanism." GeoJournal79.1 (2014): 1-14.

[40] Sanchez, Luis, et al. "SmartSantander: IoT experimentation over a smart city testbed." Computer Networks 61 (2014): 217-238.

Video Summarization: Survey on Event Detection and Summarization in Soccer Videos

Yasmin S. Khan
ME student, Department of Computer Engineering
D. Y. Patil College of Engineering, Akurdi
Savitribai Phule Pune University
Pune, India

Soudamini Pawar
Assistant Professor, Department of Computer Engineering
D. Y. Patil College of Engineering, Akurdi
Savitribai Phule Pune University
Pune, India

***Abstract*—In today's world, the rapid development of digital video and editing technology has led to fast growing of video data, creating the need for effective and advanced techniques for analysis and video retrieval, as multimedia repositories have made browsing, delivery of contents (video) and video retrieval very slow. Hence, video summarization proposes various ways for faster browsing among a large amount of data and also for content indexing. Many people spend their free time to watch or play different sports like soccer, cricket, etc. but it is not possible to watch each and every game due to the longer timing of the game. In such cases, the users may just want to view the summary of the video that is just an abstract of the original video, instead of watching the whole video that provides more information about the occurrence of various incidents in the video. It is preferable to watch just highlights of the game or just review/trailer of a movie. Apparently, summarizing a video is an important process. In this paper, video summarization approaches are discussed, that can generate static or dynamic summaries. We present different techniques for each mode in literature. We have discussed some features used for generating video summaries. As soccer is the world's most famous game played and watched, it is taken as a case study. Research done in this domain is discussed. We conclude that there is a broad perspective for further research in this field.**

Keywords—Summarization; Sports Summarization; Soccer

I. INTRODUCTION

The rapid development of digital video capture and editing technology has led to fast growing of video data that creates the need for effective and advanced techniques for video retrieval. In this era, life is becoming too busy no has free time. The users do not have enough time to watch the entire video. In such cases, the user may just want to view the summary of the video that is just an abstract of the original video, instead of watching the whole video that provides more information about the occurrence of various incidents in the video. Many people spend their free time to watch or play different sports like soccer, cricket, basketball, etc. but it is not possible to watch each and every game due to lengthy time of the game. So many people prefer to watch highlights of the game. So as per user's requirement there is a growing need for summarizing videos. Video summarization is a process that facilitates fast browsing among large video collections. It also allows more efficient content indexing. Video summarization refers to creating a summary of a video that addresses three main points. (1) The video summary should contain scenes and events not only as short as possible from the video but also the most important one. For example, in a soccer game, the summary must contain goals, fouls, shot boundaries, goal attempt, and some other important scenes. (2) The video summary should maintain a continuous connection amongst scenes. It means that the video summary should not contain video segments connected in a blind way. (3) The summarized video should not contain any redundancy. That is, the video summary should have a free repetition that is very difficult to achieve. It is necessary to detect various events in the game to generate a summary. There are two types of video summarization: Static summarization and Dynamic Summarization. As we all know soccer is the world's most famous game played and watched, the survey is discussed on soccer.

We have divided this paper into different sections. Section II describes Video Summarization, types, techniques, methods. Section III sports video summarization that takes soccer game as a case study. Section IV describes Literature Review.

II. VIDEO SUMMARIZATION

Video summarization is considered as one of the most important feature that makes the search easier and useful than before. To develop efficient indexing and search techniques to manage the huge amount of video data, new technologies need to be researched. Using this, people can use it to get the actual idea and the important events as well as scenes without watching the full original and long videos of several hours. The developed techniques in video summarization can be used for various domains, such as surveillance videos, consumer videos, movies, sports, news, etc. The summary produced can be static or dynamic i.e. it can be either KeyFrames or Video Skims. Video summarization is a tool for generating a short summary of a video, as the name implies, can either be a sequence of stationary images called key frames or moving images called video skims.

A. *Static Video Summarization*

These are also called representative frames also called R-frames, still-image abstracts or static storyboard. This type of video summarization can be classified in three different ways. These are as Classification based on sampling, classification based on shot segmentation, classification based on scene segmentation.

In [22], a similar method is discussed that consists of extracting the keyframes. Keyframes are extracted by pre-sampling uniformly or randomly the original video sequence. Keyframe extraction is a fundamental process in video content management that involves selecting one or multiple frames that will represent the content of the video and used for generating video summaries.

B. Dynamic Video Summarization

The idea of video skimming or dynamic summarization is generating a short video composed of informative / important scenes from the original video. The user receives an abstract view of the video story. The story is in video format [17]. For dynamic summarization also known as skimming, most techniques extract and segment video clips from the original video. Some of the techniques/ mechanisms for dynamic video summarization include applying SVD (Singular Value Decomposition), motion model [18]. In [20] and [19], method based on semantic analysis technique is applied for skimming. Compared with static summarization, there are relatively few works being addressed for dynamic video skimming. Most techniques are based mainly on visual information. Some other approaches make use of audio and linguistic information.

In [20], a dynamic video abstraction scheme for movie videos is presented. The proposed method is based on the progress of stories. The proposed approach attempts to comprehend video contents from the progress of the overall story and human semantic understanding. Here, the properties of two-dimensional histogram entropy of image pixels are adopted first, to segment a video into shots. Then, semantical meaning scenarios are obtained. It is done according to the spatio-temporal correlation among detected shots. Lastly, general rules of special scenario and common techniques of movie production are exploited. It is done to achieve the progress of a story in terms of the degree of progress between scenarios to the overall story.

C. Summarization Based On Clustering Techniques

The basic idea is clustering together similar frames/shots and then extracting some frames per cluster as key frames. These methods are different in features as color histogram, luminance, and motion vector and clustering algorithms as k-means, hierarchical [15].

A fuzzy c-means clustering algorithm is used in [23], where the original video is segmented into frames and these frames are considered as basic elements. Then color features are extracted in HSV color space. Then for grouping the frames Fuzzy c-means clustering algorithm is used. Then one frame (KeyFrame) per cluster is selected. The clustering output is a membership matrix that represents most represntable frame from each cluster. Paper [16] presents a new approach called VGRAPH. It uses keyframes extraction process as a shot-based method. For that it requires video segmentation by detecting the shot boundaries. First, the original video is pre-sampled. So that it can reduce the number of frames to be processed. Second, the pre-sampled video is segmented into shots using the color features. These features are extracted using the color histogram computed from the HSV. Next, noise frames are eliminated. And then second frame is selected as a shot representative. At last, the keyframes are extracted using nearest neighbor graph. This graph is built from the texture features extracted from the shots R-frames using Discrete Haar Wavelet Transforms. In [25], cluster based techniques are further divided into four further classes as, techniques based on similar activity based, k-means based, partitioning based and spectral based. A disadvantage of the most of the methods that relay on clustering algorithms is making them computationally very complex for real time applications.

In [25], M. Ajmal et al have categorized video summarization techniques in six different categories based on mechanism used and overall processing. The hierarchical classification of these techniques is described by author. The six major categories discussed are based on features, events, shot selection, cluster, trajectory and mosaic. Feature based techniques have wide scope of research as it is further divided on the basis of color, motion, gesture and object based and more.

III. Summarizing Sports Videos

Sports videos are mainly contents of some fascinating events that capture attention of the user. Many people prefer for summarized version of sports video rather than to watch full lengthy videos. Full version of the video may contains many non-significant events like advertisement, unnecessary playbacks, replays etc. Even if a generic sports video summarization system is efficient and useful, the summarization technique in a domain-specific way, like soccer videos, may present much more conveniences to users. Many sports broadcasters and web sites use editing effects such as super-imposed text captions and slow-motion replay scenes to discriminate the key events. For that reason, high level semantics can be perceived by using these editing effects.

Main part of the event detection in sports video is shot boundary detection. Various methods are already purposed for this like temporal video segmentation, frame based segmentation and event detection [1][2][3]. Shot view classification contains detection of various views like Long view, close up view, medium view and out of field view. There are various techniques proposed [1][4], some of the techniques use dominant color of frames for view classification. For close up view dominant color can be skin color of the player, for long view we can say that dominant color will be green as the background of the soccer field is green. Replays are mainly played into slow motion; most of the broadcasters play replays between graphical logos [5]. Replays can be detected by identifying such logos.

IV. Literature Review (Soccer Game as Case Study)

Various applications of video summarization touch different domains such as Consumer video applications, Personal videos, Image-Video databases management, Sports videos and surveillance videos and news videos. Apparently, Media organizations and TV broadcasting companies have shown considerable interest in these applications. Hence, Sports video summarizing is a vast domain to study. In this section various techniques/approaches are discussed in the literature of the same field.

A Dynamic summarization method is presented in [14] for the automatic extraction of summaries in soccer videos. It is based on shot detection, shot classification and Finite State Machines. Four stages discussed in the same are: playfield segmentation, shot detection using the Discrete Cosine Transform Multi-Resolution (DCT-MR) and finally, soccer video word extraction and finding out the appropriate sub-words. These sub words present summaries using the FSM and domain knowledge, where a set of rules are defined to present the semantic states in soccer game. It also explores the interesting relations between syntactic structure and the semantics of the video. Playfield segmentation is a preprocessing step. Shot detection is a step where different types of shot transition are extracted, shot classification is done in three major classes; long shot, medium shot and close-up shot using statistical method.

The proposed system in [24] is capable of detecting seven events in soccer games such as goal, foul, non-highlights, card, goal attempt, corner, offside, etc. It uses Chow-Liu Tree for structure estimation of Bayesian Network. For better pattern recognition it gives good approximation results. It has been proven that it provides a better or at least as good approximation for a discrete multivariate probability distribution.

Pattern recognition-based techniques generally extract some audio-visual features ie. Mid-level features, low level features etc. and then by using a classifier, the events or high level semantics (high level features) are detected. In [7], an automatic method is proposed that utilizes a subspace-based data mining method for feature extraction. That method is generic such that it does not use any prior knowledge in the detection process and can be considered as a domain-free method. It uses a C4.5 decision tree classifier. In [8], another method is proposed that uses a specific dimension reduction method, called mixture modality projection (MMP), to obtain high level features from low and mid-level features. Some alternative pattern recognition techniques include the use of a dynamic Bayesian network (DBN) for capturing the temporal pattern of extracted features during soccer events. For sports video highlight detection, a hybrid approach that integrates some audio-visual statistics into logical rule-based models is reported in [9]. It utilizes the play-break sequence as a semantic unit of sports videos. The method has been applied to different sports; including soccer, basketball, and Australian football.

J. Liu et al [10] had proposed a programmed player area, unsupervised naming and capable player following framework ordered broadcast soccer features. The discovery module joins background displaying and boosting acknowledgment. Naming is proficient through unsupervised player appearance learning. The outcome can be utilized for group procedures and player activity investigation, high-light distinguishing proof, et cetera. The framework can in like manner be joined with diverse applications, for instance, vision-based human-machine interaction effort. While most of players can be perceived and followed by their framework, a couple of cases, for case in point, long obstructions, feature smear, sudden camera movement and player tangle, may direct to disappointment. They wanted to plan more gainful MCMC proposals, and improve the naming and following execution by playfield enrollment and directions incitement in future works.

J. Shen et al, in [13], exploit a subspace collection method to achieve rapid and accurate video event classification employ a subspace grit method. The technology is prepared for saving the intra-modal geometry of specimens inside an matching class and disentanglement individual classes. With the structure, feature vectors bring in miscellaneous sort of multi data can be efficiently predictable from distinct modalities and identities onto a unified subspace, on that recognition technology can be performed. In addition, the training phase is finished one time and they had a combined alteration matrix to expand miscellaneous modalities.

V. Conclusion

Video summarization has attracted researchers considerably and as a result, various algorithms, mechanisms and techniques have been proposed. In this paper, a review of the research in two forms of video summarization: static summary and the video skim is carried out. Regardless of the methods used, that are static or dynamic forms, the evaluation process showed that the techniques proposed produces video summaries of high visual quality, and some approaches are suitable for real-time video processing. However, a valid evaluation method can support the field to another level. There is not any best technique for abstracting a video sequence, as video abstraction is still in the research phase largely. Also, practical applications are still limited. So, there is a scope of research in many fields such as personalized videos, consumer videos, and movie videos as well.

References

[1] A. Ekin, A. M. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarization", IEEE Transactions in Image Processing, vol. 12, no. 7, pp. 796–807, Jul. 2003.

[2] C. L. Huang, H. C. Shih, and C. Y. Chao, "Semantic analysis of soccer video using dynamic Bayesian network", IEEE Transactions Multimedia, vol. 8, no. 4, pp. 749–760, Aug. 2006.

[3] D. W. Tjondronegoro et al, "Knowledge-discounted event detection in sports video", IEEE Transactions vol. 40, no. 5, pp. 1009–1024, Sep. 2010.

[4] D. A. Sadlier, N. E. Oconnor, "Event detection in field sports video using audio-visual features and a support vector machine", IEEE Transactions. Circuits System Video Technology vol. 15, no. 10, pp. 1225–1233, Oct. 2005.

[5] A. Nagasaka, Y. Tanaka, "Automatic video indexing and full-video search for object appearances", in Visual Database Systems II, E. Knuth and L. M.Wegner, Eds. Amsterdam, The Netherlands: North-Holland, 1992, pp. 113–127.

[6] L. Xie, S. F. Chang, A. Divakaran, and H. Sun, "Structure Analysis of Soccer Video with Hidden Markov Models",ICASSP New York, USA: Columbia University, Apr. 2002

[7] M. Shyu, Z. G. Xie et al "Video semantic event/concept detection using a subspace-based multimedia data mining framework", IEEE Transactions Multimedia, vol. 10, Feb. 2008.

[8] J. Shen, D. Tao, X. Li, "Modality mixture projections for semantic video event detection", IEEE Transactions Circuits System. Video Technology., vol. 18, no. 11, pp. 1587–1596, Nov. 2008.

[9] D. W. Tjondronegoro, and Y. P. Chen, "Knowledge-discounted event detection in sports video," IEEE Trans. Syst., Man, Cybern. Part A: Syst. Humans, vol. 40, no. 5, Sep. 2010

[10] J. Liu, X. Tong, W. Li, et-al "Automatic player detection, labeling and tracking in broadcast soccer video", Pattern Recognition, vol. 30, no. 2, pp. 103–113, Jan., 2009.

[11] G. Zhu, M. Yang, et al, "Detecting video events based on action recognition in complex scenes using spatio-temporal descriptor", in Proc. ACM MM, 2009, pp. 165–174.

[12] A. M. Tekalp. "Digital Video Processing", Englewood Cliffs, NJ: Prentice- Hall, 1995.

[13] J. Shen, D. Tao, and X. Li, "Modality mixture projections for semantic video event detection", IEEE Transactions. Circuits System. Video Technology., vol. 18, no. 11, pp. 1587–1596, Nov. 2008.

[14] Y. Tabii et al "A new method for soccer video summarization based on shot detection, classification and finite state machine", in SETIT 2009 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 22-26 March 2009.

[15] E. Asadi and N. M. Charkari, "Video Summarization Using Fuzzy CMeans Clustering", 20th Iranian Conference on Electrical Engineering, (ICEE2012). IEEE, 15-17 May 2012.

[16] M. K.M., et-al "VGRAPH: An Effective Approach for Generating Static Video Summaries", in IEEE International Conference on Computer Vision Workshops, IEEE, 2013.

[17] M. S. Lew, et-al "Content-Based Multimedia Information Retrieval: State of the Art and Challenges", ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), vol. 2, February 2006.

[18] Y.-F. Ma and H.-J. Zhang, "A model of motion attention for video skimming", Proceedings International Conference on Image Processing, vol. 1. IEEE, 2002, pp. 129–132.

[19] Y. Tabii and R. O. H. Thami, "A new method for soccer video summarization based on shot detection, classification and finite state machine", SETIT 2009 5th International Conference: Sciences of Electronic, Technologies of Information and Telecommunications, 2009.

[20] S. Zhu, Z. Liang, and Y. Liu, "Automatic Video Abstraction via the Progress of Story", Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2010, vol. 6297, pp. 308–318.

[21] Zaynab El khattabi et-al, "Video Summarization: Techniques and Application", International Scholarly and Scientific Research & Innovation, 2015.

[22] E. Mendi, H. B. Clemente, and C. Bayrak, "Sports video summarization based on motion analysis", Computers and Electrical Engineering, vol. 39, pp. 790–796, April 2013.

[23] E. Asadi and N. M. Charkari, "Video Summarization Using Fuzzy CMeans Clustering", in 20th Iranian Conference on Electrical Engineering, (ICEE2012). IEEE, 15-17 May 2012.

[24] Mostafa Tavassolipour, Mahmood Karimian, and Shohreh Kasaei, "Event Detection and Summarization in Soccer Videos Using Bayesian Network and Copula", IEEE Transactions on circuits and systems for video technology, vol. 24, no. 2, February 2014.

[25] Muhammad Ajmal, Muhammad Ashraf et-al, "Video Summarization: Techniques and Classification", Computer Vision and Graphics, Springer Verlag, 2012.

Approximation Algorithms for Scheduling with Rejection on Two Unrelated Parallel Machines

Feng Lin, Xianzhao Zhang*, Zengxia Cai

College of Science
Linyi University
Linyi, Shandong, China 276005

***Abstract*—In this paper, we study the scheduling problem with rejection on two unrelated parallel machines. We may choose to reject some jobs, thus incurring the corresponding penalty. The goal is to minimize the makespan plus the sum of the penalties of the rejected jobs. We first formulate this scheduling problem into an integer program, then relax it into a linear program. From the optimal solution to the linear program, we obtain the two algorithms using the technique of linear programming rounding. In conclusion, we present a deterministic 3-approximation algorithm and a randomized 3-approximation algorithm for this problem.**

Keywords—Scheduling; Rejection; Approximation algorithm; Linear programming; Rounding

I. INTRODUCTION

The unrelated parallel machine scheduling problem to minimize makespan, $R\|C_{\max}$ following the notation of Graham et al. [1], is one of the classic NP- hard problems in combinatorial optimization. This problem is mentioned in many works concerning approximation algorithms [2 ,3], and it has received much attention in the past few decades with many extant approximation algorithms, among which the currently best 2-approximation algorithm is due to [4], who also show that the problem does not admit an algorithm with approximation ratio smaller than $3/2$, unless P = NP. One special case for this problem-Each job can only be assigned to a subset of the machine set with the same processing time-which is also known as the restricted assignment problem. Does there exist an approximation algorithm with approximation ratio better than 2? This problem is regarded as one of the ten open problems in combinatorial optimization [3].

The scheduling problem with rejection arises in make-to-order production systems with limited production capacity and tight delivery requirements, where simultaneous job rejection and scheduling decisions have to be made for maximizing the total revenue. In such systems, accepting orders without considering their impact on the whole system may delay some of the orders beyond their due dates. To be able to preserve the high quality of service to customers of accepted orders, the manufacturer has to determine which orders to accept and how to schedule them to maximize total revenue.

Many important results concerning the parallel machine scheduling problems with rejection appear in recent years. In reference [5] Engels et al. develop some techniques to design approximation algorithm for the general problem with rejection. Their main technique is to reduce a problem with rejection to a scheduling problem without rejection based on the linear programming rounding method. Hoogeveen, Skutella and Woeginger [6] consider the preemptive scheduling with rejection, and the goal is to optimize the preemptive makespan on the m parallel machines plus the sum of the penalties of the rejected jobs. They provide a complete classification of these scheduling problems on complexity and approximability. On the variant with an arbitrary number of unrelated machines, which is APX-hard, they propose a 1.58-approximation algorithm for it. Moreover, their results for unrelated machines may be carried over to the corresponding preemptive open shop scheduling problem with rejection. Li and Yuan [7] consider several parallel machine scheduling problems with deteriorating jobs and rejection. The objective is to minimize the scheduling cost of the accepted jobs plus The total penalty of the rejected jobs. They propose two fully polynomial time approximation schemes for the problems under consideration. In reference [8] Gerstl and Mosheiov study scheduling problems with rejection and general position-dependent processing times on identical parallel machines, and they introduce efficient algorithms for the problems, which run in $O(n^{m+3})$ time (which is polynomial for a given number of machines). There are many other important results ([9-13]).

In this paper, we study the following problem: We have two machines $\{M_1, M_2\}$, together with a job set $\{J_1, \cdots, J_n\}$. The processing time of job J_j on machine M_i is $p_{ij}(i=1,2)$. We may choose not to process job J_j, thus incurring a penalty q_j. The objective is to optimize the makespan for the processed jobs plus the sum of the penalties of the rejected jobs. If we denote by J_A the set of processed jobs, and J_R the set of rejected jobs, the problem may be expressed as $R_2|rej|C_{\max}(J_A)+\sum_{j\in J_R} q_j$.

We organize this paper as follows: In section Two we give a deterministic 3-approximation algorithm for the problem under consideration. In section Three a randomized 3-approximation algorithm is presented for this problem. In section Four we give some concluding remarks.

II. A 3-Approximation Algorithm for $R_2|rej|C_{\max}(J_A)+\sum_{j\in J_R} q_j$

We introduce a decision variable z_j for job $J_j\,(j=1,2,\cdots,n)$, with the following meaning:

$$z_j=\begin{cases}1, if-job-J_j-is-rejected\\0, else\end{cases}$$

For each machine-job pair $(i,j)(i=1,2;\ j=1,2,\cdots,n)$, we introduce a decision variable x_{ij} with the following meaning:

$$x_{ij}=\begin{cases}1, if-job-J_j-is-processed-on-M_i\\0, otherwise\end{cases}$$

Based on the above notations we formulate the scheduling problem into an integer program.

$$\begin{aligned}\min\ & T+\sum_{j=1}^n q_j z_j\\ s.t.\ & x_{1j}+x_{2j}+z_j=1,\ \ j=1,2,\cdots,n\\ & \sum_{j=1}^n p_{1j}x_{1j}\le T,\\ & \sum_{j=1}^n p_{2j}x_{2j}\le T,\\ & x_{ij}\in\{0,1\},\ \ i=1,2;\ j=1,2,\cdots,n\\ & z_j\in\{0,1\},\ \ j=1,2,\cdots,n.\end{aligned}\tag{1}$$

We first relax integer program (1) into the following linear program.

$$\begin{aligned}\min\ & T+\sum_{j=1}^n q_j z_j\\ s.t.\ & x_{1j}+x_{2j}+z_j=1,\ \ j=1,2,\cdots,n\\ & \sum_{j=1}^n p_{1j}x_{1j}\le T,\\ & \sum_{j=1}^n p_{2j}x_{2j}\le T,\\ & x_{ij}\ge 0,\ \ i=1,2;\ j=1,2,\cdots,n\\ & z_j\ge 0,\ \ j=1,2,\cdots,n.\end{aligned}\tag{2}$$

We denote by $(x_{1j}^*,x_{2j}^*,z_j^*)$ the optimal solution to linear program (2). Obviously we have $T^*+\sum_{j=1}^n q_j z_j^*\le OPT$.

Here $T^*=\max\left\{\sum_{j=1}^n p_{1j}x_{1j}^*,\sum_{j=1}^n p_{2j}x_{2j}^*\right\}$, OPT stands for the optimal objective value for the scheduling problem.

We apply the following rounding procedure to $(x_{1j}^*,x_{2j}^*,z_j^*)$.

Rounding Procedure R_1

Step 1: Whenever $z_j^*\ge\alpha$ $(0<\alpha<1)$, *set* $\bar z_j=1$ *(job* J_j *is rejected).*

Step 2: Otherwise set $\bar z_j=0$ *(job* J_j *is processed).*

Step 3: Whenever $x_{1j}^*\ge x_{2j}^*$, *set* $\bar x_{1j}=1$. *(job* J_j *is processed on machine* M_1*).*

Step 4: Else set $\bar x_{2j}=1$. *(job* J_j *is processed on machine* M_2*).*

We now analyze the quality of the solution obtained by Rounding Procedure R_1.

We first take a look at the makespan $\bar T$, here $\bar T=\max\{\bar L_1,\bar L_2\}$.

$$\bar L_1=\sum_{j:\bar x_{1j}=1}p_{1j}=\sum_{j:x_{1j}^*>\frac{1-\alpha}{2}}p_{1j}<\frac{2}{1-\alpha}\sum_{j:x_{1j}^*>\frac{1-\alpha}{2}}p_{1j}x_{1j}^*$$

$$\le\frac{2}{1-\alpha}\sum_{j=1}^n p_{1j}x_{1j}^*\le\frac{2}{1-\alpha}T^*.$$

Similarly we have $\bar L_2\le\frac{2}{1-\alpha}T^*$. Therefore we get $\bar T\le\frac{2}{1-\alpha}T^*$.

As for the total penalty for the rejected jobs, we have

$$\sum_{j:\bar z_j=1}q_j=\sum_{j:z_j^*\ge\alpha}q_j\le\frac{1}{\alpha}\sum_{j:z_j^*\ge\alpha}q_j z_j^*\le\frac{1}{\alpha}\sum_{j=1}^n q_j z_j^*$$

So we have

$$\bar T+\sum_{j:\bar z_j=1}q_j\le\frac{2}{1-\alpha}T^*+\frac{1}{\alpha}\sum_{j=1}^n q_j z_j^*$$

$$\le f(\alpha)\left(T^*+\sum_{j=1}^n q_j z_j^*\right)\le f(\alpha)OPT.$$

Here $f(\alpha)=\max\left\{\frac{2}{1-\alpha},\frac{1}{\alpha}\right\}$ is the approximation ratio. The value for $f(\alpha)$ is minimized if and only if $\frac{2}{1-\alpha}=\frac{1}{\alpha}$, i.e., $\alpha=\frac{1}{3}$, the minimum value for $f(\alpha)$ is 3.

Based on the discussions above we propose a 3-approximation algorithm for $R_2|rej|C_{\max}(J_A)+\sum_{j\in J_R} q_j$.

Algorithm 1

Step 1: Formulate the scheduling problem into an integer program (1).

Step 2: Relax integer program (1) to linear program (2).

Step 3: Solve linear program (2) and obtain an optimal solution $\left(x_{1j}^*, x_{2j}^*, z_j^*\right)$.

Step 4: Set $J_R=\left\{j: z_j^* \geq \frac{1}{3}\right\}$, $J_A=\left\{j: z_j^* < \frac{1}{3}\right\}$.

Step 5: Regulate $J_{M_1}=\left\{j: x_{1j}^* \geq x_{2j}^*, z_j^* < \frac{1}{3}\right\}$, $J_{M_2}=\left\{j: x_{2j}^* \geq x_{1j}^*, z_j^* < \frac{1}{3}\right\}$.

Step 6: Process the jobs in $J_{M_i}(i=1,2)$ *continuously on machine* M_i *in an arbitrary order.*

Theorem 1. Algorithm 1 is a 3-approximation algorithm for $R_2|rej|C_{\max}(J_A)+\sum_{j\in J_R} q_j$.

III. A Randomized 3-Approximation Algorithm for $R_2|rej|C_{\max}(J_A)+\sum_{j\in J_R} q_j$

In section Two we have a deterministic principle that determines the set J_R, J_{M_1} and J_{M_2}. In this section the processed jobs are assigned to machine M_1, M_2 with some probability.

We still use program (2) as a linear program relaxation for the scheduling problem. For the optimal solution $\left(x_{1j}^*, x_{2j}^*, z_j^*\right)$ to linear program (2), we apply the following rounding procedure.

Rounding Procedure R_2

Step 1: Whenever $z_j^* \geq \alpha$ $(0<\alpha<1)$, *set* $\bar{z}_j=1$ *(job* J_j *is rejected)*

Step 2: Otherwise set $\bar{z}_j=0$ *(job* J_j *is processed). Job* J_j *is assigned to machine* $M_i(i=1,2)$ *with probability* $\tilde{x}_{ij}=\frac{x_{ij}^*}{x_{1j}^*+x_{2j}^*}$.

In rounding procedure R_2, the set of rejected jobs J_R is a deterministic set. While the schedule formed by the jobs in J_A is a randomized one. We denote by $L_i(i=1,2)$ the load on machine M_i in the randomized schedule, and by T the makespan for the schedule. Obviously we have $T=\max\{L_1, L_2\}$.

For processed job J_j, obviously we have $z_j^*<\alpha$, $x_{1j}^*+x_{2j}^*>1-\alpha$, thereby we get $\tilde{x}_{ij}=\frac{x_{ij}^*}{x_{1j}^*+x_{2j}^*}<\frac{x_{ij}^*}{1-\alpha}$.

$$E[L_1]=\sum_{j:z_j^*<\alpha} p_{1j}\tilde{x}_{1j} < \frac{1}{1-\alpha}\sum_{j:z_j^*<\alpha} p_{1j}x_{1j}^*$$

$$\leq \frac{1}{1-\alpha}\sum_{j=1}^{n} p_{1j}x_{1j}^* \leq \frac{1}{1-\alpha}T^*$$

Similarly we have $E[L_2]\leq \frac{1}{1-\alpha}T^*$.

Lemma 1. $E[T]\leq E[L_1]+E[L_2]$.

Proof: $\forall J_{M_1}\subseteq J_A$.

The notation $\left(J_{M_1}, J_A \setminus J_{M_1}\right)$ stands for the random event that jobs in J_{M_1} are assigned to machine M_1, while the jobs in $J_A \setminus J_{M_1}$ are assigned to machine M_2.

Obviously we have

$$P_r\left(J_{M_1}, J_A \setminus J_{M_1}\right)=\prod_{j:j\in J_{M_1}}\tilde{x}_{1j}\prod_{j:j\in J_A\setminus J_{M_1}}\tilde{x}_{2j}$$

$$L_1\left(J_{M_1}, J_A \setminus J_{M_1}\right)=\sum_{j:j\in J_{M_1}} p_{1j}$$

$$L_2\left(J_{M_1}, J_A \setminus J_{M_1}\right) = \sum_{j: j\in J_A\setminus J_{M_1}} p_{2j}$$

$$E[L_1] = \sum_{J_{M_1}\subseteq J_A}\left(\sum_{j: j\in J_{M_1}} p_{1j}\right) P_r\left(J_{M_1}, J_A \setminus J_{M_1}\right)$$

$$E[L_2] = \sum_{J_{M_1}\subseteq J_A}\left(\sum_{j: j\in J_A\setminus J_{M_1}} p_{2j}\right) P_r\left(J_{M_1}, J_A \setminus J_{M_1}\right)$$

$$E[T] =$$

$$\sum_{J_{M_1}\subseteq J_A} \max\left\{\sum_{j: j\in J_{M_1}} p_{1j}, \sum_{j: j\in J_A\setminus J_{M_1}} p_{2j}\right\} P_r\left(J_{M_1}, J_A \setminus J_{M_1}\right)$$

Obviously we have $E[T] \le E[L_1] + E[L_2]$. The proof is completed.

From the discussions above we have $E[T] \le \dfrac{2}{1-\alpha} T^*$.

As mentioned in section 2, we have an upper bound for the total penalty of the rejected jobs

$$\sum_{j:\bar{z}_j=1} q_j = \sum_{j: z_j^* \ge \alpha} q_j \le \frac{1}{\alpha}\sum_{j: z_j^* \ge \alpha} q_j z_j^* \le \frac{1}{\alpha}\sum_{j=1}^{n} q_j z_j^*$$

Hence we have

$$E[T] + \sum_{j:\bar{z}_j=1} q_j \le \frac{2}{1-\alpha} T^* + \frac{1}{\alpha}\sum_{j=1}^{n} q_j z_j^*$$

$$\le \max\left\{\frac{2}{1-\alpha}, \frac{1}{\alpha}\right\} OPT$$

The approximation ratio $\max\left\{\dfrac{2}{1-\alpha}, \dfrac{1}{\alpha}\right\}$ is minimized when $\alpha = \dfrac{1}{3}$, and the minimum value is 3.

We give a randomized 3-approximation algorithm for $R_2|rej|C_{\max}(J_A) + \sum_{j\in J_R} q_j$ in the following way.

Algorithm 2

Step 1: Formulate the scheduling problem into an integer program (1).

Step 2: Relax integer program (1) to linear program (2).

Step 3: Solve linear program (2) and obtain an optimal solution $\left(x_{1j}^*, x_{2j}^*, z_j^*\right)$.

Step 4: Set $J_R = \left\{j : z_j^* \ge \dfrac{1}{3}\right\}$, $J_A = \left\{j : z_j^* < \dfrac{1}{3}\right\}$.

Step 5: For job j *satisfying* $z_j^* < \dfrac{1}{3}$, *assign job* j *to machine* $M_i\,(i=1,2)$ *with probability* $\tilde{x}_{ij} = \dfrac{x_{ij}^*}{x_{1j}^* + x_{2j}^*}$.

Step 6: Process the jobs in $J_{M_i}\,(i=1,2)$ *continuously on machine* M_i *in an arbitrary order.*

Theorem 2. Algorithm 2 is a randomized 3-approximation algorithm for $R_2|rej|C_{\max}(J_A) + \sum_{j\in J_R} q_j$.

IV. Conclusion

In this paper, we study the scheduling problem with rejection on two unrelated parallel machines. We may choose not to process some jobs, thus incurring the corresponding penalty. The goal is to minimize the makespan plus the sum of the penalties of the rejected jobs. We present a deterministic 3-approximation algorithm and a randomized 3-approximation algorithm for this problem. We obtain the two algorithms using the technique of linear programming rounding.

Acknowledgment

The authors are thankful to the reviewers for their valuable suggestions. This work was supported by the Chinese Society of Logistics and the China Federation of Logistics and Purchasing Project (2015CSLKT3-199), the Logistics Teaching and Research Reformation Projects for Chinese Universities (JZW2014048, JZW2014049), the Applied Mathematics Enhancement Program of Linyi University, and the national college students' innovation and entrepreneurship training program (201410452004).

References

[1] R. L. Graham, E. L. Lawler, J. K. Lenstra and A. H. G. Rinnooy Kan, “Optimization and approximation in deterministic sequencing and scheduling: A survey”, Annals of Discrete Mathematics, vol. 5, pp. 287-326, 1979.

[2] V. V. Vazirani, Approximation Algorithms, Springer, Berlin, 2003.

[3] D. P. Williamson, D. B. Shmoys, The Design of Approximation Algorithms, Cambridge University Press, London, 2011.

[4] J. K. Lenstra, D. B. Shmoys and E. Tardos, “Approximation algorithms for scheduling unrelated parallel machines”, Mathematical Programming, vol. 46, pp. 259-271, 1990.

[5] D. W. Engels, D. R. Karger, S. G. Kolliopoulos, S. Sengupta, R. N. Uma and J. Wein, “Techniques for scheduling with rejection”, Journal of Algorithms, vol. 49, pp. 175-191, 2003.

[6] H. Hoogeveen, M. Skutella and G. J. Woeginger, “Preemptive scheduling with rejection”, Mathematical Programming Ser B, vol. 94, pp. 361-374, 2003.

[7] S. S. Li and J. J. Yuan, “Parallel-machine scheduling with deteriorating jobs and rejection”, Theoretical Computer Science, vol. 411, pp. 3642-3650, 2010.

[8] E. Gerstl and G. Mosheiov, “Scheduling on parallel identical machines with job-rejection and position-dependent processing times”, Information Processing Letters, vol. 112, pp. 743-747, 2012.

[9] D. Shabtay, N. Gaspar and M. Kaspi, “A survey on offline scheduling with rejection”, Journal of Scheduling, vol. 16, pp. 3-28, 2013.

[10] L. Epstein and H. Z. Haider, “Online scheduling with rejection and withdrawal”, Theoretical Computer Science, vol. 412, pp. 6666-6674, 2011.

[11] L. Epstein and H. Z. Haider, “Preemptive online scheduling with rejection of unit jobs on two uniformly related machines”, Journal of Scheduling, vol. 17, pp. 87-93, 2014.

[12] L. F. Lu, C. T. Ng and L. Q. Zhang, “Optimal algorithms for single-machine scheduling with rejection to minimize the makespan”, International Journal of Production Economics, vol. 130, pp. 153-158, 2011.

[13] X. Min, Y. Q. Wang, J. Liu and M. Jiang, “Semi-online scheduling on two identical machines with rejection”, Journal of Combinatorial Optimization, vol. 26, pp. 472-479, 2013.

Wireless Sensor Networks for Road Traffic Monitoring

Kahtan Aziz
College of Engineering Computing.
Al Ghurair University
Dubai, United Arab Emirates

Saed Tarapiah
Telecommunication Engineering Dept.
An-Najah National University Nablus,
Palestine

Mohanad Alsaedi
Amarah Gas Power Plant Project
Amarah, Iraq

Salah Haj Ismail
Aleppo University
Aleppo, Syria

Shadi Atalla
Lavoro Autonomo (LA)
Torino, Italy

***Abstract*—Wireless Sensor Networks (WSNs) consist of large number of sensor nodes. Each node is empowered by a communication interface that is mainly characterized by low power, short transmission distance and minimal data rate such as the maximum data rate in ZigBee technology is 256 kbps, while approximately the physical transmission range between 10 to 20 meters. Currently, WSN Technology is being distributed over a large roadway of areas, in order to monitor traffic and environmental data. This approach allows several Intelligent Transport Systems (ITSs) applications to exploit the primary collected data in order to generate intelligent decisions based on earlier valuable selected information. Therefore, in this work we present a MAC protocol that is suitable for WSN where its nodes are assigned to a linear topology. The investigated protocol is realized by adapting an already existing Jennic MAC protocol. We demonstrate the validity of the MAC by building a complete end-to-end road traffic monitoring system using 4 Jennic nodes deployed in an indoor environment with the aim to prove the MAC potential in meeting the expectations of ITS applications. It is appropriate to mention that the proposed implementation just considers stationary WSN nodes.**

Keywords*—*Wireless Sensor Networks(WSN); Linear Topology; Road Monitoring; Jennic MAC

I. Introduction

Intelligent Transportation Systems (ITSs) starts to receive much attention recently by research institutes, industrial factories and standardization entities as they affect widely the life of people. The scope of ITSs is to provide fundamental services and applications that will improve transportation and mobility safety and enhance the available resources and time which influences on driving speed and affects crash rates. ITS services and applications depend on deploying advanced technologies and distributing them on the intelligent infrastructure systems and vehicles systems. These advanced technologies consists of physical world perceive technologies that are able to provide real world measurements and convert them into the digital world, storage capabilities to work on the digital measurements which will save and analyse these data and the communication technologies of both wired and wireless technologies to exchange the collected data with the vehicles and their infrastructure and vice versa [1] [2].

A system that monitors and reports the physical condition of roads such as slipperiness factor, humidity and road works ... etc; and estimates traffic on different road segments would provide very useful information for further analysis and decision that would be taken or generated by the system. Information generated from this system could be integrated with SMS based services that alert users about congestion, automatic traffic light timers, geographic information systems that suggest less congested paths or roads which are less damaged, systems that trigger road maintenance work and analysis tools that help to manage traffic and plan extensions to the road network,which will enhance the efficiency of traffic circulation , minimize risks and time to take corrective actions both on design and management levels. There are several challenges in building such a system; Such challenges fall in the areas of sensing, signal processing, communication links, protocol design, information storage and retrieval. Traffic on the road or condition of the road can only be determined through some dedicated and in purpose designed sensors.

Road traffic scenarios change dynamically, where many previous studies have been conducted in order to characterize the dynamic road traffic scenario [3] and the response to congestion must be swift. Therefore, communication protocol for such a system must be near real time to provide meaningful and useful data.

Traffic monitoring systems generate huge amount of data where the systems by its turn must process these useful information, especially those systems that need historical information to correctly estimate current state of traffic.

Traffic monitoring systems generally try to count, classify or estimate speed of vehicles moving on the road.

It is clear that a road condition and traffic monitoring system must be economically feasible so some tunning to the system can be applied such as reducing the number of system components and reusing and utilizing existing infrastructure.

The system should be feasible and able to avoid the need of digging up roads or creating additional infrastructure in form of laying wires or making overhead structures. It should be built using cheap and easily available commercial. The aim of this work is to build a road monitoring system that is able to better quantify a road anomaly. To achieve such goal we propose an approach to improve road monitoring through wireless sensor networks (using Jennic jn-5148 wireless sensors by NXP) to build a wireless sensor network where several sensors distributed along a road in forced linear topology.

II. Related Work

The early works, [4] and [5], can be considered of earliest works that sought to provide a traffic monitoring application based WSN concepts and technology rather than relying on traditional methods. However, the authors of these two works showed a sophisticated MAC protocol which is divided into two parts. The first part is suited for sensor nodes deployed under surface of pavements along sides of the roads. The second part is designed for access points. Furthermore, the overall resulted network protocol supports bidirectional communication links. The first link is to transfer collected and sensory readings from the sensor nodes to the access point via multi-hop fashion. While the later link is used to carry data packets and synchronization commands from the reverse direction (from the access point to sensor nodes) in single hop manor. Such approaches leads to asymmetric communication between sensor nodes and access points in order to cope with the difference in their communication range such property inconsistent with our assumption that all the nodes have similar communication capabilities and these two works are suitable for ITS applications that is using WSN deployed in star topology not for Linear topology.

The authors of the earlier paper [6] presents a brief survey of the important aspects for LWSN namely its applications as well as issues. Mainly, the paper classifies LWSN topologies into three main categories thin, thick and very thick. The thin structure applies where all sensor nodes are physically deployed in a line.

This section discusses the primary MAC implementations that are supported by the Jennic sensor nodes. Which is the most suitable to be considered and integrated in the proposed approach (Namely IEEE 802.15.4, ZigBee and Jennic proprietary MAC). At the end of this section, a general discussion of related research work are provided.

Supported MAC Implementations are : I) IEEE 802.15.4 is a standard protocol that builds the physical layer (PHY) and Media Access Control layer (MAC). This standard is suitable for low power, short range and low bit rate in particular the Wireless Personal Area Networks (WPAN). The JenNet and ZigBee Pro (will be explained briefly bellow) are built on top of IEEE 802.15.4 implementation while keeping the ability to develop a novel application through interfacing it directly with the IEEE 802.15.4 API (discarding the JenNet and ZigBee Pro implementations). II) JenNet is a proprietary network protocol implemented by Jennic Network [7]. JenNet aims to simplify the development of new application on top of Jennic wireless devices through providing the implementation of full network layer - the third layer on the OSI model-. Actually, JenNet runs on top of IEEE 802.15.4 standard implementation. JenNet is supplied with an API, known as the Jenie API, to facilitate the interaction between the application and the JenNet stack. III) ZigBee Pro is an industry-standard protocol where the two protocols are implemented on top of the IEEE 802.15.4 protocol. Zigbee Pro provides the network stack implementation of the OSI reference model in order to facilitate the implementation of new wireless network application developments. Zigbee Pro and JenNet Operating System (OS) API are used to realize new application for Jennic devices.

Feature	IEEE 80.2.15.4	JenNet	ZigBee Pro
Supported Topologies	Star Point-to-Point	Tree Star Line	Mesh
Maximum Number of nodes	50 nodes	500 nodes	50 nodes
Third Party Interoperability	no provisioning	no provisioning	Interoperability through ZigBee public profiles and compliance/certification
Licensing Cost	Free	Free	ZigBee Alliance membership and product certification

TABLE I: supported MAC implementations feature Comparison

We concisely present the supported MAC implementations in the properties, which are most relevant to this work comparison in terms of, Application field, Technology, Topology, MAC, Sensor, and Energy. We emphasize that JenNet and ZigBee Pro are using IEEE 802.15.4 as the foundation to build end-to-end WSN applications. We include IEEE 802.15.4 here to form a baseline.

A survey on the different mac layer protocols for linear wireless sensor networks is further discussed in [8]

III. Linear Topology MAC Protocol Design

The implementation of the proposed protocol just considers a network which is a linear topology for wireless sensor network (LWSN). Such network consists of low power nodes connected via wireless links arranged on a single line. These nodes are located along the side of the road segments, as depicted in Figure 1. Usually, linear topologies exist on pipelines, rivers and railways. Many other possible candidate applications could be also considered as in [9], [10] and [11] not only for speed monitoring or temperature measurement, but also for GIS data collection to enhance the health of important structures as tunnels, bridges, and kinetic traffic structures to avoid disasters and improve environmental quality by measuring co2 emissions inside tunnels and ventilate it properly in times of congestion, vibration of bridges and avoid collapse caused by resonance,surface humidity and protect vehicles from slipping away by forcing proper speed ... etc. For such topology, ZigBee cluster-tree can be implemented as well to exchange data between nodes, where this method suffers from limitations in LWSN: the limited number of children, and as well as the maximum number of children routers, and the maximum tree depth as stated and discussed by [12]. Therefore, the goal of this work is to design a new MAC protocol that would take advantage of the linear topology nature and characteristics in order to minimize the energy consumption needed to exchange the collected date among the sensor nodes.

Fig. 1: wireless sensors placed on bridge [13]

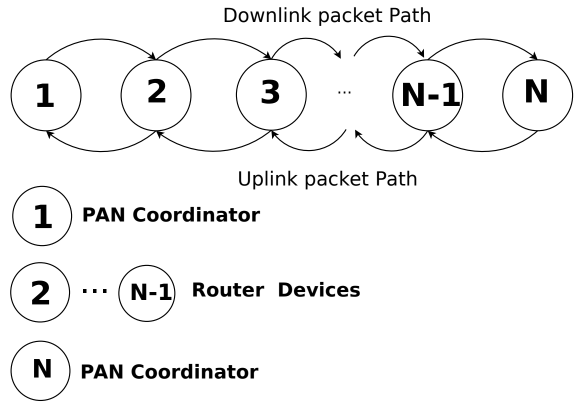


Fig. 2: LWSN Topology and type of network devices

Thus, several requirements have been taken into consideration.

A. LWSN Node Types

The nodes of LWSN based network are of the following general types, which classified based on their roles in the network:

- I Personal Area Network (PAN) Coordinator: There must be one and only one PAN Coordinator. Where it performs the following network functions: a) Assigning a PAN ID (a unique 16-bit value) to the network. b) Selecting radio frequency to be assigned for network operation. c) Identify and assign a short address (a unique 16 bit value) to itself. d) Handling requests from other devices to join the network e) Relaying messages from one node to another.
- II Router Node: A LWSN network can have one or more routers. Each of these routers serves its own children and its roles include:
  - a Handling requests from other devices to join the network.
  - b Relaying messages from one node to another.
- III End Device: This is a node which has a sensor and/or actuator functions but no coordinating functionality. The term “End Device” is not used in the IEEE 802.15.4 standard, but is commonly used in the field (wireless sensor network).

For better understanding, we consider LWSN as depicted in Figure 2. Which consists of one node acts as the network PAN coordinator, one End Device while other nodes act as router nodes. the nodes of LWSN are arranged in a logical line.The start of the line is the PAN Coordinator and the end of the line is the End Device. The line sequence passes by order starting from one router node to the next router nodes until it reaches the End Device.

Therefore, to send a data packet from the Edge device to the network Coordinator, the message will be sent via router nodes, which relay the message to the next router node until it arrives the Coordinator node. This linear topology uses two type of packets (Downlink and Uplink), the Downlink packet is a request from gateway node (coordinator) to all nodes asking them to answer with Uplink packet with their current reading as shown in Figure 2.

B. Nodes Addressing

The Simple way to refer to a node in a network in the LWSN under study is by a numeric address. In JenNet, the MAC address which is a unique and permanent 64-bit value assigned by the manufacturer to a device is used which is also called an extended address. JetNet uses 64 bits address to identify a node in the network address.

C. Network Protocol Stack

The LWSN node software stack is illustrated in Figure 3. At *the application level*, the user applications that make use of the node provided services. The application that is defined by user works together with the network mainly via the Jenie API. Network Protocol Level is the JenNet network layer that controls network routing and addressing by calling functions in the IEEE 802.15.4 MAC layer.Network Protocol Level responsibilities are:

1) Initialize and setup the communication in the network.
2) controls joing and leaving the nodes of the network by allowing them to join or removing it.
3) Route messages to their planned routes.
4) grant secure communication link for the routed messages between the nodes of the network.

The Physical/Data Link Level. This Link Level is offered by the IEEE 802.15.4 standard and defined by two divided layers:

1) Data Link layer: This Link layer is offered by the IEEE 802.15.4 MAC layer. Where, the main MAC layer services are to provide reliable communication link at the MAC layer as well to provide fragmentation and reassemble of data. Moreover, MAC layer provides addressing by assigning Physical MAC Address to the data frames.
2) Physical layer: This layer is offered by the IEEE 802.15.4 PHY layer. It is working with the physical transmission medium interface, exchange data bits with this medium, and exchange data bits with the above layer (the Data Link layer).

The Jenie API offers the main strategy for the user application to work with the JenNet software stack. The API includes

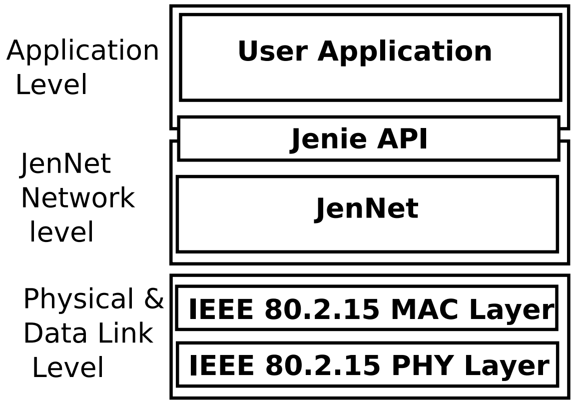


Fig. 3: Detailed Software and Network Stack for LWSN node

C functions and related resources (data types, enumerations, etc.), and provide a an easy-to-use and simple interface that is built to streamline the development of wireless networks applications. Finally, *the JenNet API* may be used in combination with the Jenie API to access the underlying JenNet stack layer features. The JenNet API includes C functions which provide extra rules where nodes can join a network, inter-network communication and the JenNet stack operation.

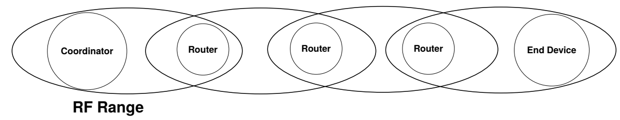


Fig. 4: Detailed Software and Network Stack for LWSN node

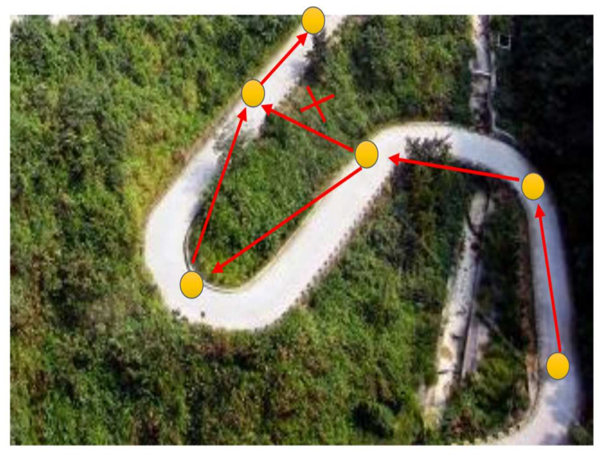


Fig. 5: WSN in Practical situation where RF range of nodes overlap randomly

IV. End-to-End Application Implementation

To boost the implementation time and cost, the work uses JenNet application template as the base foundation for the Linear Topology MAC protocol. This template is provided by Jennic [7] and the details is in the Application Note JenNet Application Template (JN-AN-1061) [14].

Actually, the JenNet Template presents a good foundation for building a new JenNet applications. As an individual basic code is specified for each node type: Router, Coordinator and End Device. The developers of the embedded System can change the provided code to adapt it to their own application needs. The provided application templates consider the following:

1) the considered network topology is A Tree only.
2) Just one device is selected to be the PAN Coordinator.
3) one other device At least (to act as an End Device Or Router).
4) The Network Application ID and the PAN ID have a pre-identified values.

In fact the last 3 assumed points by the Jennic Template fits perfectly with the works engineering requirements. While, the first point (a Tree topology) needs to be redesigned and to be implemented in order to fit with the desired application and linear topology. Thus, paper main contribution is to provide the design guideline and implementation methods to generate a valid linear topology.

The main methods can be applied to generate a linear topology lay on the following ways:

1) Physical Distance to force a Linear topology in LWSN: In the Jennic template the linear topology is part of tree topology but the constrain is the distance i.e. if an engineer wants to have a linear topology he/she must arrange the network nodes in linear way and assures that just one node resides in the Radio Frequency (RF) range of another node figure 4. This method suffers from major problem when the nodes are near each other (this is common practical situation). In this case the resulted topology will change to another topology may be tree topology or any other topology. For further details, Figure 5 illustrates practical case of the problem.
2) Higher Network Layer to force a Linear topology in LWSN: This method resolves the later described using distance based method problem. The proposed method is to modify the network stack code to make sure that the resulted network topology is a linear one, this method works always even if nodes are in same RF range. The following steps are used to force Linear Topology: re-implementing the Authority function [14]: to accept or reject any node based on configuration file deployed in each node. The configuration file in a node includes list of the nodes' addresses to be connected to the node hosting the file, any unknown node (not in the addresses list) will be rejected. This method ensures always that the resulted topology is linear one as requested by the work engineering requirements.

V. Experements and Measurements

This section details preliminary experiments performed to validate the proposed methods and it includes the collected results to show the network performance. In these experiments all the 3 types of network nodes are included, PAN coordinator node, router nodes and End Device. All the experiments are

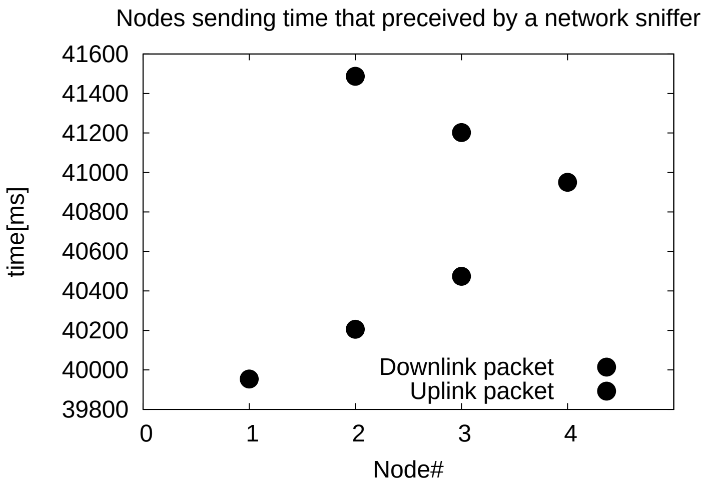


Fig. 6: Performance Evaluation

carried out with Jennic JN5148-EK010 Evaluation KIT [15]. The End-to- End application scenario represents a road monitoring system, where the road administrative authorities monitor the road environment temperature and the network nodes battery levels in real time manor. Such information could be visualized and controlled using a web application remotely, and furthermore, have the ability to control the lights appliances that are installed in both sides of the road.

A. Experiments with LWSN of 4 Nodes

The LWSN built with 4-nodes, after the network starts. The coordinator node sends *Downlink* request with address $0xFFFFFFFF$ and specify the required sensor value,in this case the temperature readings data as well as the battery voltage data sensors. Measuring the LWSN Round Trip Time (RTT): RTT = The time of Coordinator start sending *Downlink* packet until receiving the corresponding *Uplink* packet from the End Device. For this the 4 nodes setup where distance spacing between the nodes are identical and is equal 1 meter, the average RTT values is 1.533 seconds.

Figure 6 shows, the measured time in millisecond, observed by a network sniffer deployed in the radio frequency range of all the running network nodes, the measured time increases linearly with distance from node to coordinator node. The packet flow starts by initiating a packet request, DownLink packet, from the coordinator, where each packet received and processed by all the router nodes until it arrives to the End device. Which in turn, start a new uplink packet in response. The response packet will include the End Device temperature and voltage readings and forwarded to the nearest router. The router then append its reading to the end of the uplink packet and forward it to the next router, this process done by a router will be repeated by every router in the network until the packet arrives its destination that is the coordinator node in these experiments.

VI. Conclusions and Future Work

This paper studies a MAC protocol for wireless sensor networks (WSNs) that focuses on road traffic monitoring applications, which can be implemented on monitoring serious issues like bridges vibration, pollution and gases emissions in tunnels and any factor could affect traffic and drivers health, and for GIS data collection for better practices targeted by Intelligent Transportation Systems(ITSs). The *Jennic JN5148* evaluation kit attached with some dedicated sensors is the main used hardware component for building a road monitoring application, that utilizes the proposed *MAC* method protocol. Furthermore, the proposed protocol is designed to support WSN where its nodes are arbitrary deployed and distributed to form in a linear topology (LWSN). This work demonstrates the validity of the MAC protocol by building a complete end-to-end road traffic monitoring system by using 4 nodes deployed in an indoor environment. Regrading the ingoing work, we intend to increase the scalability of the system in order to support more nodes in the wireless sensor network compared to current implementation. Whereas, our target is to extend the proposed MAC scheme to support 20 nodes. In order to achieve this goal we propose to modify the current addressing method; Instead of using the typical 8 bytes for node address, we could just use two bytes for node address. The proposed MAC overcomes the typical cluster tree algorithm [12] by providing reliable communication links and avoiding well-known limitation of the earlier mentioned algorithm. Furthermore, the main features of the proposed *MAC* are to overcome some of limitations of the Jennic propriety MAC implementation by decreasing the protocol overhead, increasing maximum data packet size, providing a bidirectional communication links and supporting the linear network topology.

References

[1] S. Tarapiah, S. Atalla, and R. AbuHania, "Smart on-board transportation management system using gps/gsm/gprs technologies to reduce traffic violation in developing countries," *International Journal of Digital Information and Wireless Communications (IJDIWC)*, vol. 3, no. 4, pp. 96–105, 2013.

[2] S. Tarapiah and S. Atalla, "Public transportation management system based on gps/wifi and open street maps," *International Journal of Advanced Computer Science and Applications (IJACSA).*, vol. 6, no. 1, pp. 189–194, 2015.

[3] B. M. Velichkovsky, S. M. Dornhoefer, M. Kopf, J. Helmert, and M. Joos, "Change detection and occlusion modes in road-traffic scenarios," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 5, no. 2, pp. 99–109, 2002.

[4] S. Coleri, S. Y. Cheung, and P. Varaiya, "Sensor networks for monitoring traffic," in *Allerton conference on communication, control and computing*, 2004, pp. 32–40.

[5] S. Y. Cheung, S. C. Ergen, and P. Varaiya, "Traffic surveillance with wireless magnetic sensors," in *Proceedings of the 12th ITS world congress*, vol. 1917, 2005, p. 173181.

[6] Z. Wang, X. Zha, and X. Qian, "The application and issuse of linear wireless sensor networks," in *System Science, Engineering Design and Manufacturing Informatization (ICSEM), 2011 International Conference on*, vol. 2. IEEE, 2011, pp. 9–12.

[7] Jennic, "Jennic Website," http://www.jennic.com/, 2010, [Online; accessed 01-April-2015].

[8] R. SOKULLU and E. DEMİR, "Mac layer protocols for linear wireless sensor networks: A survey."

[9] S. Tarapiah, S. Atalla, N. Muala, and S. Tarabeh, "Offline public transportation management system based on gps/wifi and open street maps," in *Computational Intelligence, Communication Systems and Networks (CICSyN), 2014 Sixth International Conference on*. IEEE, 2014, pp. 182–185.

[10] S. Tarapiah, S. Atalla, and B. Alsayid, “Smart on-board transportation management system geo-casting featured,” in *Computer Applications and Information Systems (WCCAIS), 2014 World Congress on*. IEEE, 2014, pp. 1–6.

[11] S. Tarapiah, R. AbuHania, I. Hindi, and D. Jamal, “Applying web based gps/gprs ticketing and tracking mechanism to reduce traffic violation in developing countries,” in *The International Conference on Digital Information Processing, E-Business and Cloud Computing (DIPECC2013)*. SDIWC - Digital Library, 2013, pp. 102–106.

[12] M. D. Sarr, F. Delobel, M. Misson, and I. Niang, “Automatic discovery of topologies and addressing for linear wireless sensors networks,” in *Wireless Days (WD), 2012 IFIP*. IEEE, 2012, pp. 1–7.

[13] libelium, “Smart Roards,” http://www.libelium.com/smart_roads_wsn_smart_infrastructures/, [Online; accessed 01-April-2015].

[14] N. Semiconductors, “Application Note JenNet Application Template (JN-AN-1061),” http://www.jennic.com/files/support_documentation/JN-AN-1061-JenNet-ApplicationTemplate.pdf, 2010, [Online; accessed 01-April-2015].

[15] Jennic, “Jennic JN5148-EK010 Evaluation KIT,” http://www.jennic.com/files/support_files/JN-UG-3062-JN5148-EK010-User-Guide.pdf, 2010, [Online; accessed 01-April-2015].

AL-S^{2}m: Soft road traffic Signs map for vehicular systems

Ammar LAHLOUHI
Department of Computer Science,
University of Batna 2,
05 000 Batna, Algeria

***Abstract*—In this paper, we describe AL-S^{2}m, a roadmap with traffic signs to be used in vehicular systems. AL-S^{2}m is part of a more general system of traffic signs (TSs) management, called AL-S^2, which includes two sides: central map server and client vehicular system. The server allows establishing, maintaining and disseminating AL-S^{2}m. The client localizes the vehicle in AL-S^{2}m and detects TSs. In this paper, we focus on the AL-S^{2}m establishment. AL-S^{2}m can handle variable TSs and its update is easy, which keeps it coherent with the reality. Also, it improves the map-matching algorithm. We implemented AL-S^{2}m easily using an Android device.**

Keywords*—Road Traffic Signs, Roadmap, Map-Matching, Driver Assistance Systems, Autonomous vehicles

I. Introduction and Motivation

Advanced Driver Assistance Systems (ADAS) have the capacity to increase driver safety by observing the driver and its environment and providing information, warnings, or even taking actions. Integrating ADAS into adaptive cruise control system can reduce the cognitive load of the driver and supports safe driving. Completely automated tasks of the driver lead to autonomous vehicles AV. Both ADAS and AV need to be aware of local traffic laws, such as the right of way and speed limits, as well as information such as city limits to comply with traffic rules. The transport authority specifies such rules in traffic signs (TSs).

Several important works tried to make vehicular systems aware of TSs. The well known of such works are those based on TSs detection and recognition (DRS) (see [8] for instance) using a camera onboard of the vehicle. Such systems are more consistent with the reality since they report the state of the reality as it is when the vehicle moves by detecting, particularly, variable and temporary TSs.

High precision real-time DRS is a hard task [11]. In spite of several works on DRS, such as [1], [3], [5], [6], [12], [14], current DRSs are lengthy in recognition time, sometimes, deduce an incorrect detection, and require a special powerful processing devices and a material for Nocturne vision during darkness time. Also, bad weather conditions (snowing, raining and fog) and direct sunlight or sunlight reflections by TSs can affect camera sign detection. Even with the best design and under perfect weather conditions, cameras are still far away from performing as accurately as the drivers vision. The limited field of view and fixed related position on the vehicle may cause the camera system to miss a TS, specifically in sharp curves. In some occasions, other vehicles may hide TSs. Even with a very good image processing system, it may not be easy to determine that the TS does not belong to the road in which the vehicle travels.

Several works use digital maps for navigation purposes that the vehicular system can exploit, such as the ones proposed in [9], [10], [18]. Firstly, we must build and store a detailed roadmap including TSs. Next, the estimator of the vehicular system filters online measurements and gives the position of the vehicle in the roadmap. Then, one can make TSs detection independently of atmospheric disturbances and traffic situations. An interesting approach [7], applied to speed limit TSs only, combines a camera based with a map-based system to provide information that is more precise. The first detects variable and temporary speed limits while the second detects implicit ones, such as when entering from a country road into a town when there is no TS indicating a speed limit.

Two difficulties remain not issued with the important system proposed in [7]. Firstly, the TSs detection with a camera-based system remains problematic. Secondly, with the static nature of current roadmaps where their design aims updating them for long periods, they cannot satisfactorily handle temporary and variable TSs.

Given the previous weaknesses, a map-based solution can be more significant. However, current roadmaps have particularly two difficulties: (1) They are usually incoherent with the reality since they are often outdated, and (2) they cannot handle variable TSs (vTSs). Firstly, the roadmaps incoherence is due to the difficulty of their updating, which includes a lot of information at once. A solution resides in specialized (simplified) roadmaps that one can update easily. Secondly, the problem of vTSs is due to static nature of the current maps. A solution for vTSs resides in dynamic maps.

In this paper, we propose an entirely map based system, called AL-S^2 (Augmented Lighted – Soft Signs). AL-S^2 takes into account the two previous solutions. Its achievement includes two parts: central map server and client vehicular system. The server establishes, maintains and disseminates a specific roadmap (AL-S^{2}m). The client localizes the vehicle in AL-S^{2}m and detects TSs. AL-S^2 tracks then the movement of the vehicle and when it reaches the visibility field of a TS, it displays it. In this paper, we focus on the specific roadmap AL-S^{2}m only.

In AL-S^{2}m, we represent the road network as a simple topology of roads with their associated TSs. Updating AL-

S^{2}m consists in simple operations (adding, deleting and modifying) on the representations of the roads and TSs. Such operations can be broadcasted timely and wirelessly by the central map server to AL-S^{2}m. Broadcasting such information is not problematic with current generations of communication systems. Timely broadcasting such operations allows handling temporary TSs.

The remainder of the paper describes AL-S^{2}m representation (virtualizing roads and TSs). We organized it in four sections, in addition to a conclusion. Section II describes the roadmap representation. It focuses mainly on representing the elements of AL-S^{2}m, its achievement, and the map-matching method. Section III focuses on traffic signs. It consist of TSs representation (including the variable TSs), their insertion in AL-S^{2}m, and their detection during the vehicle movement. Section IV describes an implementation of AL-S^{2}m using an Android device. Section V presents an experimentation of AL-S^{2}m.

II. ROADMAP

Localizing a vehicle on a roadmap is made by using a map-matching algorithm (see [15], for a survey). The general purpose of such algorithm is to identify the correct segment of the road on which the vehicle is traveling at a given time and to determine the position of the vehicle on such segment [16], [17]. In spite of their differences, the map-matching algorithms consider a linear representation of the road regardless of its width (see Fig. 1, for an illustration). Such representation leads to a confusion in determining the position of the vehicle among several roads, especially when it comes to parallel or near roads, such as the problem at Y-junctions [15]. Matching the road direction (see Fig. 2, for such direction) with that of the vehicle has improved the situation, but the problem of parallel roads close to the same direction remains unsolved.

AL-S^{2}m uses a special map-matching algorithm, which is based on representing the road as a polygon (see Table 1, for an illustration). The polygon representation is more coherent with the reality of the road than the linear one. Such representation leads to a more precise map matching. This section describes such representation and the different configurations, adopted in AL-S^{2}m. To simplify the discussion, we illustrate and summarize the design elements in equations.

AL-S^{2}m design uses the following scenario. We consider the vehicle as a mobile point V; see (1). Initially, AL-S^2 determines the coordinates of V, and locates V on a given road cR (Current road) in AL-S^{2}m using a map-matching algorithm. Then, AL-S^{2}m tracks the motion of V until it leaves cR. While tracking V, AL-S^2 searches if there is a TS on cR, which is in the visibility-field vf of the driver, to identify it. By leaving cR, V either leaves AL-S^{2}m entirely or enters a new cR to follow the same scenario.

$$V = \langle x, y, z \rangle \tag{1}$$

The rest of this section explains the representation of AL-S^{2}m elements (Subsection II-A) and its achievement (Subsection II-C), and then describes the used map-matching method (Subsection II-B).

Fig. 1: Highway 401 within Ontario, Canada. In spite of its width, the road is represented in the map by a single line.

A. Representation of Roadmap Elements

AL-S^{2}m (see (2)) combines the topological localization with the geometric accuracy of metric methods. We consider AL-S^{2}m as a directed graph (N, E) where N is a set of nodes and E is a set of arcs. A node can be a road, an intersection or a roundabout. An arc from a node A to a node B specifies that when the vehicle leaves A, it can enter B. E will be distributed on the nodes of the graph such that each node A will have its own exit nodes E. For this, we associate B to A if there is an arc from A to B. Consequently, E will not appear in (2) but rather it appears in the road elements, see (3). AL-S^{2}m includes also the Vehicle (V) and the current Road (cR) on which V moves.

$$ALS2m = \langle N, V, cR \rangle \tag{2}$$

In the following subsections, we describe the different road configurations, and their virtualization in AL-S^{2}m: Single One-Way, Single Two-way, Multiple Channels and Interchange.

1) Single One-way Road: We represent a single one-way road R as a triplet; see (3), where:

- P: a polygon delimiting R,
- G: a rectangular cuboid (see explanation above),
- E: a set of roads, which are in exit relation with R.

$$R = \langle P, G, E \rangle \tag{3}$$

To take into account the fact that the road is not planar but rather a three-dimensional shape, we represent it as a skew polygon, referred below simply by polygon. A skew polygon is a polygon whose vertices are not all coplanar, see [19] for more details, and Fig. 3 for an illustration of a simple skew polygon.

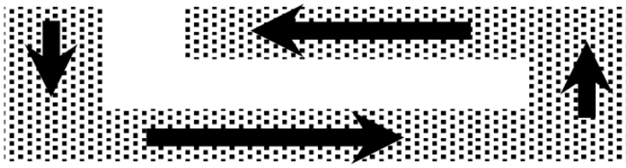


Fig. 2: A road with several directions that can be opposite. The arrows indicate the roads directions.

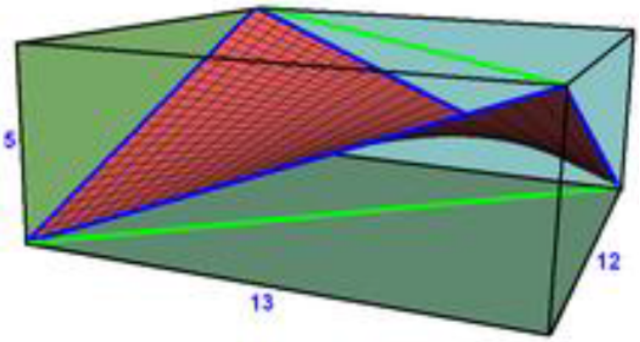


Fig. 3: An illustration of a skew polygon and its minimum bounding box.

In a polygon representation, we made the map matching simply by a point-in-the-box which determines if a given point is inside the box (rectangular cuboid) bounding the polygon (see Subsection II-B, for our map-matching method). For this reason, we add a rectangular cuboid (G) to (3). We compute G using an algorithm of a minimum bounding box. Such an algorithm proposed firstly in [19] and then improved in [13], and [4] describes a recent interesting heuristic. The description of such algorithms is outside this paper. Note that its performance is not important since we use it once at the map creation only.

For a road R with various directions, we consider it as a sequence of roads (four roads, in the example of Fig. 2), one road by direction. Each of them has its own direction. We made simply the transition between such roads by their exit relations E of (3).

For a curved road R (as illustrated in Fig. 4-a), we approximate it by a polygon, see Fig. 4.b. However, if such a form contains several directions we consider each portion of a given direction as a separate road and the road will consist of a sequence of roads, as shown in Fig. 4-c. As handled, we modeled easily parallel curved roads, as given in Fig. 4-d.

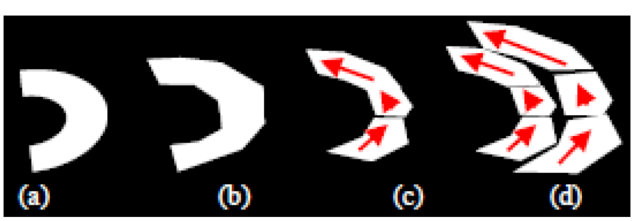
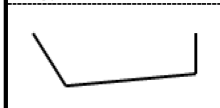
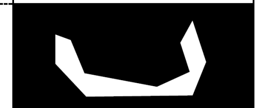
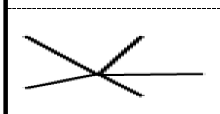
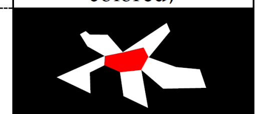
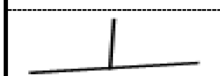
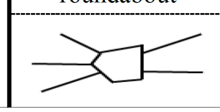
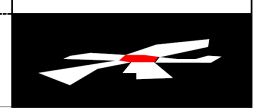


Fig. 4: Curved roads and their representation in AL-S^{2}m.

c

TABLE I: LINEAR (SPS) vs POLYGONAL (AL-S^{2}M) REPRESENTATIONS OF ROADS.

Road configuration	SPS	AL-S²m
Sample road	Three road segments as Lines	A polygon
		
Intersection	Five road segments	Six polygons including one for the intersection part (red colored)
		
T – junction	Three road segments	Four polygons including one for the intersection
		
Y – junction	Three road segments	Four polygons including one for the intersection
		
Roundabout	Ten road segments including five segments (constituting visually a polygon) constituting the roundabout	Five polygons (one for each road leaving the roundabout) and a sixth for the roundabout itself
		

2) Single Two-way Channel and U-turn Detection Problem: We consider the two-way road as two roads, R and its inverse R^{-1}. When the vehicle is in the direction of R and makes a U-turn, it leaves R and enters R^{-1}. The exit relation of R^{-1} will then include R and vice-versa, see (4).

$$
\begin{array}{l} R = \langle P, G, \theta_R, E \cup \{R^{-1}\} \rangle \\ R^{-1} = \langle P, G, \theta_R + \pi, E \cup \{R\} \rangle \end{array} \tag{4}
$$

The definition of a vehicle V, given in (3), does not allow the vehicle to be aware of having made a U-turn. Then, we add a direction angle θ_R to the definition of the road R, as given in (4). Such angle for R^{-1} will be $\theta_R + \pi$. When V (with θ_V as its direction) is in R and the expression U ($U = (|\theta_V - \theta_R| > (\pi/2))$) is true, V has made a U-turn. Since U-turn uses the direction of the vehicle, we add θ_V to the definition

Fig. 5: A five-level interchange (The High Five Interchange in Dallas, Texas, United States).

of the vehicle of (1), as given in (5).

$$V = \langle x, y, z, \theta_V \rangle \tag{5}$$

When the vehicle makes a U-turn, it leaves R to enter R^{-1}. We search for R^{-1} in the exit roads of R ($E \cup \{R^{-1}\}$) as any other exit road of R. Note that the exit roads of R^1 (noted E^{-1}) are different from those of R; they are those of its opposite direction.

$$\begin{aligned} C1 &= \langle P_{C1}, G_{C1}, \theta_{C1}, E_{C1} \cup \{C2\} \rangle \\ C2 &= \langle P_{C2}, G_{C2}, \theta_{C2}, E_{C2} \cup \{C1, C3\} \rangle \\ C3 &= \langle P_{C3}, G_{C3}, \theta_{C3}, E_{C3} \cup \{C2\} \rangle \end{aligned} \tag{6}$$

3) Multiple Channels Road: We consider a multiple channels road essentially as a sequence of roads (channels). We made the transition between channels as that between other roads, by using the exit relation E of (4), i.e., E will include the side channels. For example, in (6), we give a representation of a road with three channels C1, C2 and C3, where C2 is in the middle.

4) Interchange: An interchange (see Fig. 5, for an illustration) is a road junction that typically uses grade separation, and one or more ramps, to permit traffic on at least one highway to pass through the junction without directly crossing any other traffic stream. In the case of a representation by a polygon, the vehicle V belongs to a given ramp RP before it will be above other ramps. Consequently, in AL-S^{2}m, we just check if V still belongs to RP without considering other ramps. This is not possible with the linear representation (especially in two-dimensions) where the map-matching is done by comparing the distances between the vehicle and the segments representing the ramps.

B. Roadmap Achievement

Several professional technics, methods, tools and devices are currently useful in determining the elements of the road, given in (4), such as LIDAR (Laser Interferometry Detection and Ranging) and roads aerial images. Such operations are not problematic tasks for road professionals. Their description is outside this paper. However, we describe in the experimentation (see Section V) a simplistic method that we used to determine polygons of roads. In addition to shaping the polygon delimiting an element of the road, we specify an approximation of the road direction (see (4)).

C. Map-Matching

Recall that in AL-S^{2}m:

- a vehicle (see (5)) is represented by a point (x_V, y_V, z_V) and a direction θ_V,
- a road element R (see (4)) is represented by: (1) a skew polygon (P), (2) surrounded by a rectangular cuboid (G), (3) its direction θ_R, and (4) roads E in exit relation with R.

A map matching consists in determining the road element of AL-S^{2}m to which belongs the point (x_V, y_V, z_V) representing the vehicle. Initially, we must seek the road element to which the vehicle belongs among all the road elements of AL-S^{2}m. However, when the vehicle leaves a particular road element R, it can enter another road element among those of its exit relation E. The map-matching algorithm will seek then such element among those of E only.

The problem of map matching consists in verifying if (x_V, y_V, z_V) belongs to a rectangular cuboid (G) of a given road, with the condition NU is true (NU=($\mid \theta_V - \theta_R \mid \leq \pi/2$)); i.e., the vehicle is in the road direction, not in its opposite. In the following paragraphs, we explain the map-matching method used in AL-S^{2}m in order to determine the rectangular cuboid to which the point (x_V, y_V, z_V) belongs; such method is an exercise of elementary geometry. We briefly describe it in the following paragraphs.

A plane P (see (7)) divides the space into three parts. The first comprises the points q with e_P (x,y,z)>0 and the second includes the points r with e_P (x,y,z)<0, while the third consists of the points p of the plane P itself with e_P (x,y,z)=0. In order for the point (x_V, y_V, z_V) to be located between two plans P1 and P2, it must verify the inequality given in (8).

$$e_P(x, y, z) = a_P \times x + b_P \times y + c_P \times z + d_P = 0 \tag{7}$$

$$(e_{P1}(x_V, y_V, z_V) \times e_{P2}(x_V, y_V, z_V)) < 0 \tag{8}$$

A rectangular cuboid is defined by six rectangular planes Pi of equations e_Pi (x,y,z), i $\in$ [1, 6]. A point (x_V, y_V, z_V)

belongs to a rectangular cuboid if it is located between the six planes Pi, i ∈ [1, 6]. To determine whether the point (x_V, y_V, z_V) belongs to a given rectangular cuboid, we must determine that such point is between the parallel planes Pi, i ∈ [1, 6]. Assuming that the parallel planes of the rectangular cuboid are arranged according to the following pairs: (P1, P2), (P3, P4) and (P5, P6). In this case, the point (x_V, y_V, z_V) belongs to such a road if (In(x_V, y_V, z_V)=true), see (9). Such map-matching method is very easy to implement, and consumes fewer resources, i.e.; it does not require a powerful device for its execution and responds timely.

$$In(x_V, y_V, z_V) = \left(\begin{array}{c} (e_{P1}(x_V, y_V, z_V) \times e_{P2}(x_V, y_V, z_V)) < 0 \\ \bigwedge(e_{P3}(x_V, y_V, z_V) \times e_{P4}(x_V, y_V, z_V)) < 0 \\ \bigwedge(e_{P5}(x_V, y_V, z_V) \times e_{P6}(x_V, y_V, z_V)) < 0 \end{array}\right) \quad (9)$$

III. TRAFFIC SIGNS

In this section, we describe a representation of traffic signs (Subsection III-A), their placement on the roadmap AL-S^{2}m (Subsection III-B), and the method used for their detection (Subsection III-C).

A. *Traffic Signs Representation*

A hard (material) traffic sign (HTS) has an image Im, it is fixed on a specific place with coordinates $(x_{HTS}, y_{HTS}, z_{HTS})$ on a given road R, and its effect starts from some straight segment that passes by HTS coordinates (see Fig. 7). HTS will be seeable by the driver at some distance d, from which its visibility-field (vf) starts. In the following, we describe the particular representation of TSs in AL-S^{2}m and then we show how to handle variable TSs. However, before considering TSs representation, we modify the road equation (see (4)) so it includes a list S of TSs that belongs to R, see (10).

$$R = \langle P, G, \theta_R, S, E \rangle \quad (10)$$

1) Traffic Sign Representation in AL-S^{2}m: In a similar way to HTSs, we can represent a soft TS by its coordinates H=$(x_{HTS}, y_{HTS}, z_{HTS})$ and its image Im. In such representation, different drivers cannot detect the TSs with the same distance from their effect area. For instance, in Fig. 7, only the vehicle V3 can detect the TS whereas V1 and V2 cannot in spite of their same distance from the beginning of the effect field of the TS. Therefore, the reaction of the drivers will not be the same at the same distance from the TS effect. Such imprecision is already a weakness of HTSs.

To overcome such detection problem, we represent the TS by straight segment ts, in two dimensions space, as illustrated in Fig. 7. The ts segment passes by the point H=$(x_{HTS}, y_{HTS}, z_{HTS})$.

In AL-S^{2}m, the point (x_V, y_V, z_V) representing the vehicle moves inside a rectangular cuboid G, representing the road, see Fig. 6. For a better coherence, we consider a TS as a plan, termed pTS, rather than a simple straight line. The pTS equation will be determined by considering the two following facts: (1) pTS contains the point H=$(x_{HTS}, y_{HTS}, z_{HTS})$ and (2) pTS is parallel to a given plane of the rectangular cuboid G, say P1 of equation e_{P1}(x,y,z)=0. This leads to the pTS equation, see (11). Finally, a soft TS will be represented by: (1) its image Im, and (2) the coefficients of pTS $co_{TS} = (a_{TS}, b_{TS}, c_{TS}, d_{TS})$, see (12), where we discarded the constant k.

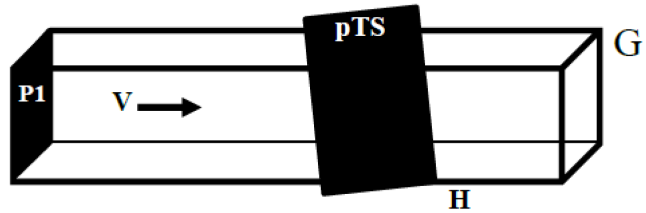


Fig. 6: The TS represented as a plan (pTS) in the rectangular cuboid G. The vehicle (point V) moves inside G (containing the road polygon) in a given direction from the plan P1.

$$\begin{array}{r} e_{TS}(x, y, z) = k \times a_{P1} \times (x - x_{HTS}) + k \times b_{P1} \\ \times(y - y_{HTS}) + k \times c_{P1} \times (z - z_{HTS}) = 0 \end{array} \quad (11)$$

$$\begin{array}{l} TS = \langle Im, co_{TS} \rangle = \langle Im, a_{TS}, b_{TS}, c_{TS}, d_{TS} \rangle \\ \quad =< Im, a_{P1}, b_{P1}, c_{P1}, \\ \quad -(a_{P1} \times x_{HTS} + b_{P1} \times y_{HTS} + c_{P1} \times z_{HTS}) > \end{array} \quad (12)$$

2) Traffic Signs Out of Their Roads: Some TSs must be seeable when the vehicle travels on a road element R while they belong to an element of its exit relation E, such as those signaling a roundabout. In AL-S^{2}m, we assign such TSs to the all roads where they must be visible, including the road where we made effectively the HTS.

3) Variable Traffic Signs: A variable TS (vTS) is a TS that changes according to some conditions (traffic, weather, time). vTSs can be seen as several conditioned simple TSs (sTSs) as follows: $(C_1 : sTS_1, \cdots, C_n : sTS_n)$, see (13). In the latter, each sTS$_i$ is represented by its image Im$_i$ while the TSs coordinates co$_{TS}$ are common to the all sTSs.

$$TS = \langle [C_1 : Im_1, \cdots, C_n : Im_n], co_{TS} \rangle \quad (13)$$

To handle the conditions, we associate to a vTS a behavior that verifies such conditions to display the image of a sTS$_i$ with the verified condition. We ranked vTS conditions into three categories:

1) Timed (such as tricolor lights, daytime, nighttime): The conditions are time intervals. AL-S^{2}m verifies the time (as a condition) and identifies the associated sTS,
2) Surrounding physical environment (such as traffic density, weather): It depends on the sensors available on the vehicle and the communication protocol between the vehicular system and AL-S^{2}m,
3) Remote (modifiable by transport authority): It depends on the communication protocol between the transport authority and AL-S^{2}m.

To take into consideration vTSs, we replace (12) of TS by (13). To illustrate the use of (13) in modelling the TSs, we

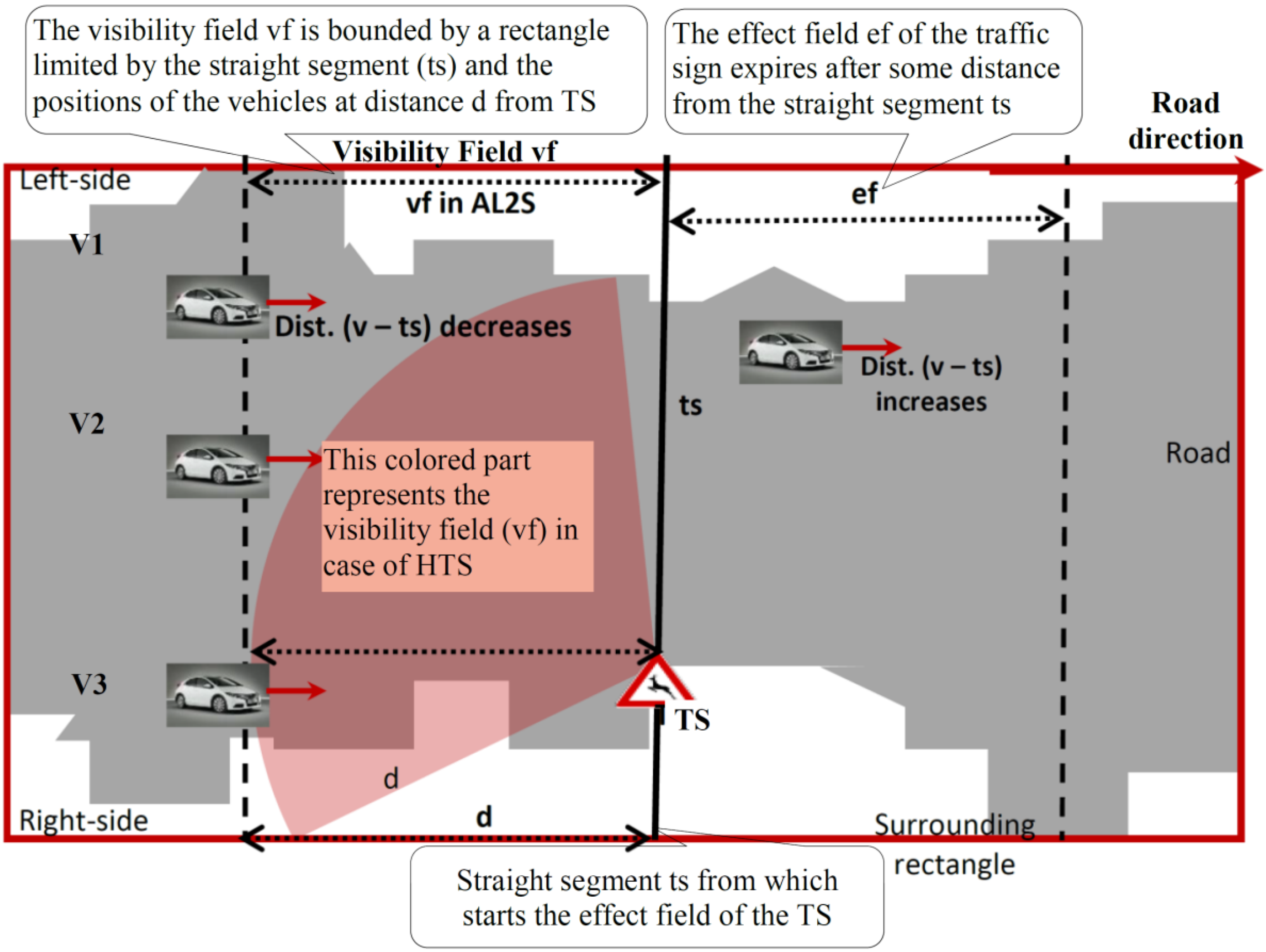


Fig. 7: TS as a plan. The road and a TS projected into a two dimensions space. The road is colored in grey, the vf of the HTS is colored in orange and vf of the soft TS (in AL-S^{2}m) is indicated by a discontinuous arrow.

give two cases: simple TS (sTS) and a Tricolor Lights (TL). For simple TS, see (14), it is sufficient to put with the TS image the true value as its condition.

$$sTS = \langle [true : Im], co_{TS} \rangle \tag{14}$$

The tricolor lights TL modeling needs some explanation. The TL sign changes regularly over the time. For this aim, we take into consideration the starting time tg (hour, minute and second) of the green light. Let us consider the transition time between colors as follows: gy (green-yellow), yr (yellow-red) and rg (red-green). The AL-S^{2}m system determines the TL color (TS image) at time t according to the conditions given in the equation of TL; see (15).

$$TL = \left\langle \begin{bmatrix} (0 \leq (t - t_g) mod(gy + yr + rg) \\ \quad < gy) : green, \\ (gy \leq (t - t_g) mod(gy + yr + rg) \\ \quad < gy + yr) : yellow, \\ (gy + yr \leq (t - t_g) mod(gy + yr \\ \quad + rg) < gy + yr + rg) : red \end{bmatrix}, co_{TS} \right\rangle \tag{15}$$

B. Placing Traffic Signs on the Roadmap AL-S^{2}m

In AL-S^{2}m, we can place a TS by two ways (see Subsection V-B, for more information). In the first (made on the ground), a traffic agent equipped with AL-S^{2}m system (hardware and software) moves to the TS location and requests AL-S^{2}m system to put such TS. AL-S^{2}m system proceeds then as follows: (1) it determines the position of the traffic agent (the TS coordinates) using a positioning system, (2) it determines the road R to which the TS belongs using the map-matching algorithm, and (3) it inserts the TS in AL-S^{2}m.

To insert a TS, AL-S^{2}m system: (1) requests TS information from the traffic agent (TS image, time of green light if any, $\cdots$), (2) computes the coefficients of the plan equation associated to the TS (co$_{TS}$, see (12)), and, finally, (3) adds TS information to AL-S^{2}m.

In the second way (made directly on AL-S^{2}m), an agent specifies a given road R and the information allowing to determine the coordinates of TS, such as a distance from a given point in R. Then, AL-S^{2}m determines the coordinates where it must place the TS. Finally, it inserts such TS in AL-S^{2}m as described in the previous paragraph.

C. Detecting Traffic Signs

AL-S^{2}m system detects a given TS when the vehicle is located in the visibility-field vf of such TS. vf is specified by the distance d (see Fig. 7) between the vehicle (point (x_V, y_V, z_V)) and the plan TS. We continue considering the detected TS while d decreases (the vehicle approaching TS). When d starts increasing (the vehicle moves away from TS), then the detected TS can be forsaken.

TSs detection involves two parts: tracking the vehicle and displaying the TS. For the first part, the vehicle, equipped with AL-S^{2}m, starts from a point SVP and moves on the road. AL-S^{2}m system determines the SVP coordinates using a positioning system. Then, AL-S^{2}m follows the movement of the vehicle and determines its successive positions following a dead reckoning process (see [2], for dead reckoning). However, such process has the disadvantage of errors accumulation that requires correction of the position for its use over long distances. For this aim, we used sensors fusion to correct the position of the vehicle. In this paper, we focus on the roadmap. The correction of the position is then outside this paper.

To detect a TS, AL-S^{2}m system determines firstly the road R on which the vehicle is located (to which belongs the starting point SVP). Each road R contains a list of TSs (rTSs) ordered according to their distances from a given side, say P1, of the rectangular cuboid G of the road R. P1, see Fig. 6, is the side situated at the opposite of the road direction θ_R.

AL-S^{2}m system determines then the first TS in the visibility field of the vehicle, in order to display it. We must search such first TS, as explained in the next paragraph. When a given TS is displayed, AL-S^{2}m system takes the next TS from rTSs to do the same as the previous TS and so on.

To seek the first TS, we use the following process. When the vehicle enters a road element (road, intersection or roundabout) coming from another element, we consider the first TS of rTSs and check if it is in the visibility field of the vehicle. We continue the verification until such TS becomes in the visibility field where we display it. It is not necessary to check other TSs since the vehicle must enter firstly the visibility field of the first TS of the list rTS. However, when the vehicle enters a road element from the outside of AL-S^{2}m, we must seek the nearest TS, which can be not the first element of rTSs since the vehicle can enter between TSs. To search such nearest TS, we follow the process given in Fig. 8. In the next paragraphs, we give some clarifications on such process.

At time intervals dt, the vehicle crosses different successive points (positions). We termed two successive points by old position (OP) and new position (NP). For readability reason, we shortened by TS the plan containing the TS (pTS).

For a given TS, if the distance d(NP, TS) between NP and TS is less than d(OP, TS) between OP and TS, i.e., the distance between the vehicle and such TS decreases, the vehicle approaches the TS. In this situation, the current TS is ahead of the vehicle and the searched TS is found.

In contrast, if the distance between the vehicle and TS increases, the vehicle moves away from TS. In such situation, the current TS is behind the vehicle and we must search for another TS.

A special case is when the distance between the vehicle and the TS remains unchanged. This happens when the vehicle is stopped or moves in parallel to TS. In such situation, we cannot decide. We will then get a new position of the vehicle after dt and then reiterate the same process.

To compute the distances d(OP, TS) and d(NP, TS), we use the formula of elementary geometry of a distance between a given point and a plan. For d(OP, TS), we obtain the equation (16), in which we can substitute aTS, bTS, cTS and dTS by their associated expressions, as given in (12), containing the coefficients of pTS. We can obtain the d(NP, TS) expression by substituting OP by NP.

IV. Implementation

To implement AL-S^{2}m, we used an Android device, endowed with accelerometers, gyro-meters and a GPS sensor. Android is a Java based language. The SensorManager class of API of Android sensors allows registered applications to receive data from sensors. After registration, changed data of the sensor are sent to SensorEventListener as a SensorEvent containing information generated from the given sensor (for Android programming and sensors use, see its official web site http://developer.android.com/ (Accessed 2015 11 26)).

For AL-S^{2}m, we identified four main object classes (AL-S^{2}m, R, V and TS). We already described their attributes in the associated equations. We recall them in (17).

Here, we describe some significant aspects of implementing such classes according to the following AL-S^2 implementation scenario. AL-S^{2}m identifies the current road cR on which the vehicle runs initially (see Subsection II-B) and then passes the control to cR by invoking the method cR.in. The latter tracks the vehicle in cR by verifying if it belongs to the cuboid rectangle G of cR. While the vehicle is in cR, cR.in checks whether the vehicle is on vf of some TS to detect it (see Subsection III-C). After leaving cR by the vehicle, cR.in returns the control to AL-S^{2}m. This latter seeks another road cR and iterates such process until the vehicle leaves completely AL-S^{2}m.

We implemented the vehicle by a thread. Its run() method continually updates the coordinates of the vehicle, by using the accelerometers of the Android device, and its direction, by using its gyro-meters.

Both AL-S^{2}m and cR objects use the coordinates of the vehicle and its direction. Several methods of TS and R can also access the vehicle attributes. Then, we declared the getters of

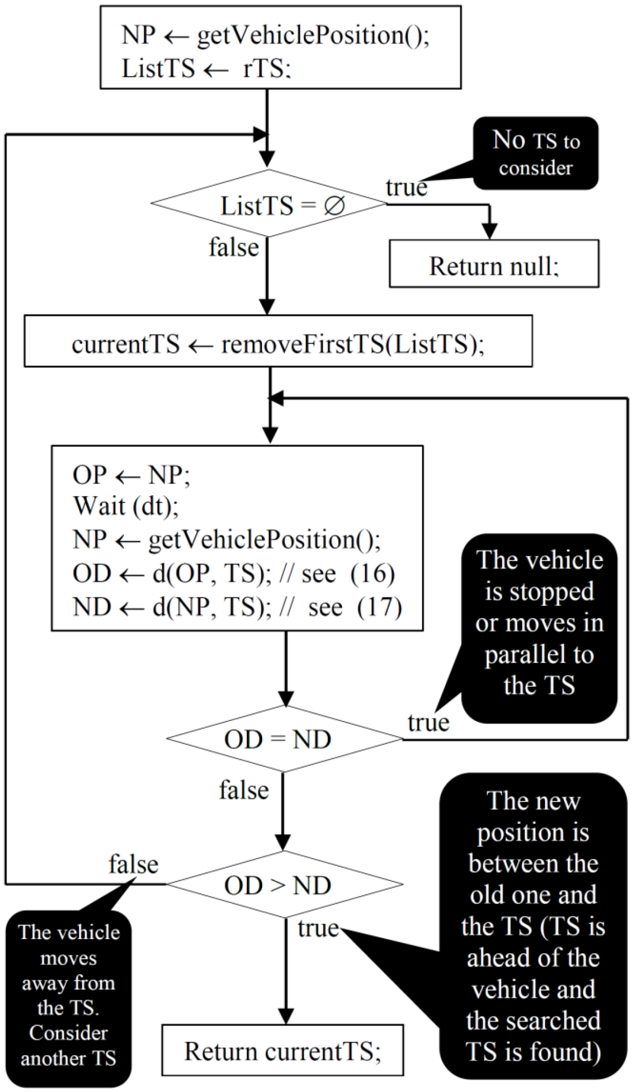


Fig. 8: Seeking the first TS. The symbols used in the diagram are well known (rectangle: processing; lozenge: testing; arrow: control flow). However, we added a bullet symbol for comments.

the vehicle (getx(), gety(), getz(), get-teta(), getvx(), getvy(), getvz()) as synchronize.

We also implemented TS as a Java thread. Its run() method allows displaying and dropping the TS image.

$$d(OP,TS) = \frac{|x_{OP} \times a_{TS} + y_{OP} \times b_{TS} + z_{OP} \times c_{TS} + d_{TS}|}{\sqrt{(a_{TS}^2 + b_{TS}^2 + c_{TS}^2 + d_{TS}^2)}} \tag{16}$$

$$\begin{aligned} &ALS^2m =< N, V, cR > \\ &R =< P, G, S, \theta_R, E > \\ &V =< x, y, z, \theta_V > \\ &TS = \langle [c_1 : i_1 \cdots c_n : i_n], co_{TS} \rangle \end{aligned} \tag{17}$$

V. Experimentation

As mentioned in Subsection III-B, several professional techniques, methods, tools and devices are currently useful in determining the road elements. Such operations are not problematic tasks for road professionals. To show the easiness of AL-S^{2}m creation, we describe a simplistic experimentation for AL-S^{2}m creation. AL-S^{2}m consists of two parts: (1) delimitation of the nodes or road elements (simple roads, roundabouts and intersection) as polygons (described in Subsection V-A), (2) and the placement of TSs on AL-S^{2}m (given in Subsection V-B).

A. *Delimitating Road Elements*

For delimitating a road element (node), we operate with a pedestrian to bind such a road. A pedestrian equipped with AL-S^{2}m system (Android device endowed with AL-S^{2}m software) operates as follows: (1) starts from an initial point SP on the road edge, (2) moves then to surround the road, (3) to return finally to the starting point SP. AL-S^{2}m system tracks the pedestrians movement and saves the coordinates of its different positions at time intervals. Firstly, AL-S^{2}m determines the coordinates of the point SP using GPS (Global Positioning System). Then, using the accelerometers and the gyro-meters of the android device, AL-S^{2}m system uses a dead reckoning process to compute, and record, the coordinates of such positions. These positions constitute the vertices of the polygon surrounding the road.

The main challenge in tracking by using the dead reckoning lies in the inaccuracies of the values provided by MEMS sensors especially when the transporter of the device is a pedestrian. In our case, we used an approach based on the pedestrian steps. The traditional approach to measure distances using pedestrian steps uses a constant length of such steps. It counts the number of the steps and multiplies it by a given constant measure (one meter, for instance). Such approach increases the errors in the computed distances especially with false steps. To detect steps, such approaches use the accelerometers variations. However, some variations (when sensors return non-zero values whereas there is no movement) are not causes of steps (see Fig. 9).

In contrast, AL-S^{2}m detects false steps. It considers only the variations of the accelerations between two specified values (noted top and bottom, in Fig. 9). The top and bottom values depends on the used device that must be determined by experimenting such device. AL-S^{2}m discards the variations that do not spend from bottom to top as false steps.

In addition, AL-S^{2}m uses steps with variable lengths. It does not count the number of steps to multiply it by a constant length but it rather computes the length of each step (See Fig. 10) and sums such lengths. To compute the step length, we use the values given by the accelerometers and the time interval dt.

Steps with variable lengths and the detection of false steps reduced significantly the errors in the pedestrian path. Minimizing the errors in addition to the small distances covered by the pedestrian associated to the fact that our application does not require high precision, led to acceptable results (see Fig. 10).

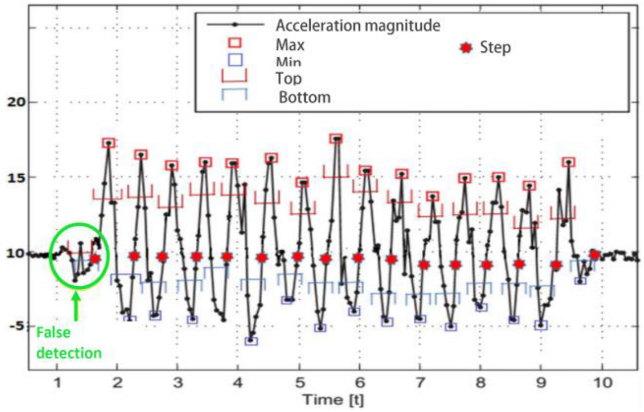


Fig. 9: Steps detection by an Android device. A higher acceleration (greater than or equal to Top) followed by a lower one (less than or equal to bottom) indicates a one-step whereas other variations between such values indicate false steps.

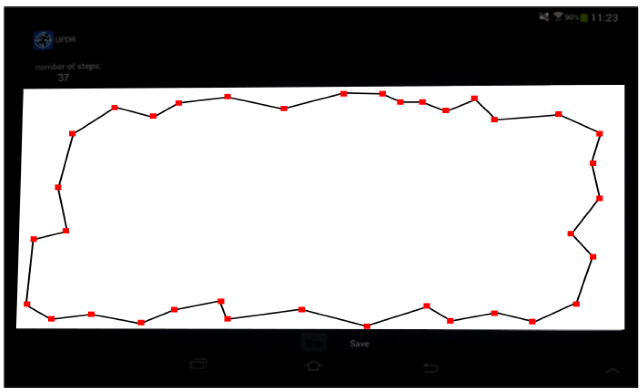


Fig. 10: A screenshot (Tablet with Android operating system) containing a projection of the skew polygon into two dimensions: each two consecutive red points represent a side of the polygon. Note that, at the closing point of the polygon, we have a relatively long segment, which is due to cumulating errors of the sensors.

B. Placement of Traffic Signs

Placing a given TS in AL-S^{2}m uses two interfaces (see Subsection III-B). For the first, a traffic agent moves to a position on the right side of a given road and requests AL-S^{2}m (by typing on the Android device equipped with AL-S^{2}m system) to insert the given TS. By using GPS, AL-S^{2}m determines the position of the traffic agent (indeed, that of the device) which will constitute the TS coordinates as in the case of the hard TS (HTS). Such position allows determining the road R to which belongs TS. Finally, the system inserts the given TS in AL-S^{2}m.

To insert a TS, AL-S^{2}m firstly derives the coefficients of the plan TS (pTS, see (13) and Fig. 6) using the TS coordinates and the rectangular cuboid (G) of R (see (3)). Then, AL-S^{2}m displays a list of TSs and requests the traffic agent to choose the one he (she) would insert. Then, it inserts, in AL-S^{2}m, the new TS as part of R. If the traffic agent chooses a variable TS, AL-S^{2}m demands the parameters of such TS: the conditions and the associated simple TSs. Otherwise, it adds the true value as the TS condition.

The second is made directly on AL-S^{2}m. An agent requests AL-S^{2}m to insert a given TS at a given position of a road R. The traffic agent can specify a distance from a given point (start, end) in R. AL-S^{2}m determines then the coordinates where it must insert TS. It inserts then such TS into AL-S^{2}m as described in the previous paragraph (deriving the coefficients of pTS, choosing a TS, specifying the conditions).

VI. Conclusion and Further Work

In this paper, we described and discussed AL-S^{2}m, a specific roadmap. We presented a virtualization of the road elements. Also, we proposed an original map-matching method and improved TSs detection by an original representation of TSs as plans. Finally, we showed how one can implement AL-S^{2}m, and we indicated how it can create AL-S^{2}m. Our experiments show that AL-S^{2}m implementation is easy.

Creating and exploiting as well as updating AL-S^{2}m are simple, thanks to the simplicity of the used representation (a topology of polygons). The exploitation of AL-S^{2}m needs a simple device with accelerometers, gyro-meters and GPS sensors only. Unlike other roadmaps, AL-S^{2}m takes into account variable TSs and greatly improves the map-matching thanks to the road representation as a polygon. The simplification of AL-S^{2}m updates and the possibility of timely broadcasting such updates allow also to take into account temporary TSs.

In this paper, we have not detailed the AL-S^{2}m dissemination. We intend to explore such problem shortly.

Acknowledgment

The author would like to thank Mohamed Amine LAAMARI, Yasser Moussa BERGHOUT, Mohamed Salah ABDOU and Walid Guerza that implement some parts of the experimentation as part of their master dissertations. I would like also to thank Yazid LAHLOUHI for helping me correcting some mistakes.

References

[1] Bahlmann, C.; Zhu, Y.; Ramesh, V.; Pellkofer, M.; & Koehler, T. A System for Traffic Sign Detection, Tracking and Recognition Using Color, Shape, and Motion Information Proceedings of the IEEE on Intelligent Vehicles Symposium, 6-8 June 2005 pages 255-260, 2005.

[2] Borenstein, J. & Feng, L. Measurement and correction of systematic odometry errors in mobile robots IEEE Transactions on Robotics and Automation 12(6) pages 869-880, 1996.

[3] Broggi, A.; Cerri, P.; Medici, P.; Porta, P. P. & Ghisio, G. Real time road signs recognition Proceedings of IEEE Intelligent Vehicles Symposium, Istanbul, Turkey, June 13-15, 2007 pages 981-986, 2007.

[4] Chang, C. T., Gorissen, B., & Melchior, S. Fast oriented bounding box optimization on the rotation group SO (3,) ACM Transactions on Graphics 30(5) pages 122:1-122:16, 2011.

[5] de la Escalera, A.; Moreno, L. E.; Salichs, M. A. & Armingol, J. M. Road traffic sign detection and classification IEEE Transactions on Industrial Electronics 44(6) pages 848-859, 1997.

[6] Fang, C.-Y.; Chen, S.-W. & Fuh, C.-S. Road-Sign Detection and Tracking IEEE Transactions on Vehicular Technology 52(5) pages 1329-1341, 2003.

[7] Jamshidi, H.; Lukaszewicz, T.; Kashi, A.; Berghuvud, A.; Zepernick, H. & Khatibi, S. Fusion of digital map traffic signs and camera-detected signs Proceedings of 5th International Conference on Signal Processing and Communication Systems (ICSPCS), Outrigger Waikiki on the Beach, Honolulu, HI, USA, December 12-14, 2011 pages 1-7, 2011.

[8] Kim, J.; Kim, S.; Lim, K.; Choi, Y. & Byun, H. Real-time traffic sign detection with vehicle camera images Proceedings of the 7th International Conference on Ubiquitous Information Management and Communication (ICUIMC13), Kota Kinabalu, Malaysia, January 1719, 2013 pages 20:1-20:4, 2013.

[9] Lauer, M. & Stein, D. A Train Localization Algorithm for Train Protection Systems of the Future IEEE Transactions on Intelligent Transportation Systems 16(2) pages 970-979, 2015.

[10] Ficco, M.; Palmieri, F. & Castiglione, A. Supporting First Responders Localization during Crisis Management The 9th International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing, Blumenau, Brazil, July 8-9, 2015.

[11] Mathias, M.; Timofte, R.; Benenson, R. & Van Gool, L. Traffic sign recognition How far are we from the solution? Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN13), Dallas, USA, August 4-9, 2013 pages 1-8, 2013.

[12] Miura, J.; Kanda, T.; & Shirai, Y. An active vision system for real-time traffic sign recognition Proceedings of the IEEE conference on intelligent Transportation Systems, Dearborn, USA, October 1-3, 2000 pages 5257, 2000.

[13] OROURKE, J. Finding minimal enclosing boxes International Journal of Computer & Information Sciences, 14(3) pages 183199, 1985.

[14] Paclk, P.; Novoviov, J.; Pudil, P. & Somol, P. Road sign classification using Laplace kernel classifier Pattern Recognition Letters 21(13-14) Pages 1165-1173, 2000.

[15] Quddus, M. A.: Ochieng, W. Y. & Noland, R. B. Current map-matching algorithms for transport applications: State-of-the art and future research directions Transportation Research Part C: Emerging Technologies 15(5) pages 312-328, 2007.

[16] Quddus, M. A.; Ochieng, W. Y.; Zhao, L. & Noland, R. B. A general map matching algorithm for transport telematics applications GPS solutions 7(3) pages 157-167, 2003.

[17] Smaili, C.; El Najjar M. E. & Charpillet F. A Road Matching Method for Precise Vehicle Localization Using Hybrid Bayesian Network Journal of Intelligent Transportation Systems 12(4) pages 176-188, 2008.

[18] Tao, Z.; Dian-Ge, Y.; Jiang-Tao L. & Xiao-Min, L. A Trajectory-Based Map-Matching System for the Driving Road Identification in Vehicle Navigation Systems Journal of Intelligent Transportation Systems: Technology, Planning, and Operations, Article in Press, DOI: 10.1080/15472450.2015.1015847, 2015.

[19] Williams, R. The Geometrical Foundation of Natural Structure: A Source Book of Design Dover Publications, Inc., pages 34-51, 1979.

An Overview of Surface Tracking and Representation in Fluid Simulation

Listy Stephen
Master of Technology, Computer Science and Engineering
VJEC, Kannur University, India

Anoop Jose
Asst. Professor, Computer Science and Engineering
VJEC, Kannur University, India

***Abstract*—Realism in fluid animation can be achieve with physics based techniques and is the best among other approaches. Now, this area constitutes hot researches. There are number of mechanisms evolved with the advent of both hardware and software technologies. Most of the fluid simulation methods described with or without a clear surface representation. This paper focused on a quantitative survey of various fluid surface tracking and representation techniques. Suitable tracking schemes with the hybrid fluid simulation approach may give mind blowing visual effects for various applications.**

Keywords*—*Fluid simulation; Physics based animation; Realism; Surface representation; Tracking

I. Introduction

Physics based fluid simulation is not merely a new idea in the field 3D animation. Creation of fluid animation on the basis of physics laws normally called fluid simulation. Most of the techniques in the respective area, are the numerical approximation of fundamental Naiver-Stroke equation, which has some relation with the Computational Fluid Dynamics (CFD). But the applied idea is quite different from CFD. Here, it is appropriate to avoid the light weighted forces acting on the fluid body, in order to reduce the overhead of the system. Surface tracking and its representation is an attractive and enthusiastic charge field in fluid simulation. Many researching fields including simulation and rendering of free surfaces still need to solve this problem efficiently.

The advected surface by a general velocity field frequently changes its topology. Due to this reason, the moving surfaces are typically defined implicitly with the zero set of a scalar field rather than explicit representation of a mesh for instance. Deformable and rigid object surfaces are often represented by using an explicit triangle mesh. They are comparatively simple and efficient data structures and able to sent directly to Graphic processor units for rendering instances. Furthermore, the explicit representation makes direct Lagrangian simulation of soft bodies, are simply modifies the vertex positions. With high deformations, local re-sampling might be required. In contrast from the solid surface tracking methods, the free surface of a liquid constantly changes its topology by splitting and merging with itself. Processing with the explicit representations of the surface is a tricky one under this situation. Modification of the triangle mesh would need whenever a self intersection occurs.

Given, a surface representation and a velocity field at time t and building a representation of the surface at time $t + \Delta t$ can be phrased as a surface tracking problem. Varied fluid phenomena such as fire, smoke or clouds etc do not posses a clear interface. But liquids have some hope here. It seems, for the intricacy of fluid motion, including merging, diverging, foam and bubble generation, the triangle meshes are not well suited for the effective surface representation.

II. Surface Tracking Methods

Over the past few years various surface tracking methods were raised and moved on hand in hand with different fluid simulation problems. Tracking will provide compelling advantage for the simulation methods, like detailed and visually pleasing result. Herewith mentioned an overview of such existing techniques.

A. Classical Metaballs

Visualization of deformable implicit surfaces is a challenging topic, as it is aimed at representing a whole nonrigid objects, ranging from soft bodies to water and gaseous phenomena. Metaballs are widely used type of implicit surface, invented by Blinn in the early period, [16] used for achieving fluid-like appearances.

The concept of metaballs is closely related to the idea of smoothed particle hydrodynamics (SPH) [17], a convention used for simulating fluids as clouds of particles. Both methods apply smooth scalar functions, that map points in space to a mass density. These scalar functions, considered as smoothing kernels, basically represent point masses that are smoothed out over a tiny volume of space, like a Gaussian blur in 2D image processing. Moreover, SPH-simulated fluids are visualized quite naturally as metaballs. The proposed techniques for visualizing metaballs rely heavily on the SPH method and referred as fluid atoms. They are animated on the CPU either by free-form animation techniques or by physics-based simulation. Moreover, the dynamics of the fluid atoms are interactively determined, hence preprocessing of the fluid animation sequence not possible in this case. Metaballs produces very blobby appearance.

B. Front Tracking Method

Apart from the idea of Classical Metaballs, the Front-Tracking method by Tryggvason, Bunner and Esmaeeli

[13] explicitly discretize the free surface using particles and stored a connectivity list between them [14]. It is highly complex to hold the connectivity list when the free surface experiences extensive deformations or topological changes.

Generally a front structure consists of points connected by elements. The linked list structure is used to store both the points and the elements (front objects), that contains pointers to the previous object and the next object in the list. The order of the list is completely arbitrary and has no connection to the actual order on the interfaces. Usage of a linked list makes the addition and removal of objects particularly simple. Only one information is stored in each point, that is the point coordinates. On the other side, the elements contain most of the front information. Each element knows about - the elements that are connected to the same endpoints; the points that it is connected to; the jump in density across the element; the surface tension; and any other quantities, that are required for a particular simulation. The elements provides direction and for a given front, all elements must have the same direction.

Deformation and stretches occurred, when the front moves. That causes the resolution of some parts of the front become insufficient, while other parts crowded with front elements. Additional elements must either be added to maintain accuracy when the separation of points becomes extra large or the points must be redistributed to maintain adequate resolution. Generally it is favorable to remove small elements. In addition to reducing the total number of elements used to represent the front, element removal usually preventing the formation of " twists and jerks " much smaller than the grid size. While the restructuring of the front makes codes, that uses more complex explicit tracking than front capturing codes. Usage of a suitable data structure made many of the crucial operations relatively straight forward. In early two-dimensional computations, the position of all interface points were adjusted either at every time step or at every few time step to preserving nearly uniform distance. However this is not necessary and able to remove points where needed. Restructuring is relatively simple in two-dimensions. Here large element were split up by inserting a point and delete an element through removing a point. Sometimes it is needed to put a new point at the mid-point between the old end-points of an element using linear interpolation. This is usually account for the curvature of the front by using a higher order interpolation, particularly significant where the surface tension is large. Non-smooth parts of the front leads to a large pressure variations. A simple Legendre interpolation usually works well in this case. For three-dimensional flows, the restructuring is more complicated. Not only are there several different ways to add and delete points, but other manners of the front, such as the connectivity and the shape of the elements. Adding and deleting elements can be done in a number of ways in 3D. Large elements and longest edges need to split into two and replace both this elements and the one sharing the long edge by four new elements. Likewise, elements are deleted, two at a time by merging the shortest edge into a point. Sometimes it is necessary to reconnect the points by swapping edges to make the elements better shaped. In many simulations this step is not following, since it is possible to show that a combination of element insertion and deletion will have approximately the same result.

Often cited shortcomings of Front Tracking are its algorithmic complexity in tracking surfaces of 3D flows, hardness in robustly handling interface merging and breakup particularly in 3D. The reason for these difficulties are the necessity to logically connect interface elements and record changes in connectivity during interface operations ie element addition, deletion and reconnection. In two dimensions these difficulties are actually minor and the implementation of a robust connectivity algorithm is fairly laid on the line. Nonetheless moving to three dimensions, the algorithmic complexity of connectivity increases fiercely and particularly in interface reconnection during topology changes.

C. Point Set Method

Surface tracking methods compelling to take care, when the free surface undergoes large deformations or topological changes. Torres and Brack-Bill [12] have suggested a method to avert this adversity, known as the point-set method. This mechanism varied from the standard front-tracking process, where the indicator function is calculated from the surface normals. Here the information about a surface of unconnected interfacial points extracted by, first constructing an indicator function $I(x)$ and the interfacial points embedded within a computational grid. With a given grid and the location of the interfacial points, an indicator function were constructed, which labels points $I(x)$ lying inside and outside of the interface. This is to distinguish phase or material from another. Zero and one would be the default values to an indicator function for an ideally thin interface. Admitting this approach expel the front tracking method from its dependence on point connectivity. The point regeneration algorithm is one of the complex and computationally expensive process so that, it may create extra burden for the system.

D. Marching Cube Algorithm

The primitive and most popular idea used in solving the contouring problem is marching cubes. This method have a tendency to create ill-shaped triangles which could fixed to some extend with a manner called dual contouring [8], also preserving sharp boundaries and provides adaptive contouring. Normal estimates at edge crossings are important in dual contouring and is sensitive to inaccuracies in it. Boissonnat and Oudot [7] presented a contouring technique which uses Delaunay triangulation to create possibly good triangulations. Additionally, this approach appears to be extraordinarily expensive for something which must run at every time step. Another alternative provided by Hilton [10], called marching triangles which takes a surface-based rather than volume-based approach of contouring. Marching triangles requires incomparably less computation time and fewer triangles. So it can produce good-quality triangles than marching cubes. This would be considered as, one of the bold advantage. The method is not ensured to create closed and manifold meshes in the presence of sharp or thin features, which proves a negative outlook.

Marching cubes algorithm used in Kunimatsu et al's work [9], which maintaining the inherent capability of dealing with topological changes. There are two primary steps in this approach for surface construction problem. Locating the surface corresponding to a user specified value and create triangles constitutes in the first step and then ensure a quality image of the surface by calculate the normals to the surface at each vertex of each triangle. Marching cubes uses a divide-and-conquer algorithmic approach to locate the surface in a logical cube generated from eight pixels. The algorithm determines how the surface intersects this cube, then moves or marchs to the next cube. Finding the surface intersection in a cube by assign one to a cube's vertex, if the data value at that vertex greater than or equals the value of the surface construction. The vertices being inside or on the surface. Cube vertices with values below the surface receive a zero and are outside the surface. The surface intersects those cube edges where one vertex is outside the surface and the other is inside the surface. These concepts helps to determine the topology of the surface within a cube and finding the location of the intersection later. Since there are eight vertices in each cube and two slates inside and outside, only $2^8 = 256$ ways a surface can intersect the cube. These 256 cases enumerated to create a table to look up surface-edge intersections, provide the labeling of a cube vertices.

E. Screen Space Mesh Approach

Muller et al. [15] introduced a powerful approach for the generation of surfaces defined by the boundary of a three-dimensional point cloud. A depth map with internal and external silhouettes was first generated in screen space. Then it is used to construct a 2D screen space triangle mesh with a technique derived from Marching Squares. The algorithm only generates surface where visible, such that view-dependent level of detail came for free and interesting visual effects, possible by filtering in screen space.

The advantages of the screen space mesh are like this. Resolving the parts of surface which are close to the camera with more triangles than distant parts, yielding camera-dependent level of detail. Since it operates in two dimensions, a method derived from Marching Squares can be employed. Which is substantially faster than the 3D Marching Cubes algorithm. In contrast to other screen space approaches such as ray tracing or point splatting, fast standard triangle shading hardware can be used for state-of-the-art forward shading of the surface and occlusion culling, since the mesh can easily be transformed back into world space.

F. VOF and SOLA

Topological changes of the fluid surface were efficiently handled in VOF method [11], which usually walk-in with the marching cubes algorithm. Fundamental technique only uses one scalar value, that is the volume of fluid for one cell. The metric estimate the exact location of the liquids and calculate the total volume of fluid inside the simulation domain. A light weight version of the basic solution algorithm (SOLA) used in the MAC method. SOLA failed to treat free surfaces. An extended version, SOLA-SURF is also available, that uses surface height function method. The significant qualities like simplicity and flexibility of the SOLA codes make them best foundations for the development of more sophisticated implementations. An alternate mesh version of the SOLA code named as SOLA-VM, was chosen as a fundamental for representing the VOF technique. The new experimental version called SOLA-VOF. Many developments have been made and the basic technique has grown through applications to a broad class of problems. Several other attempts combined the SOLA-SURF code with a different interface tracking techniques based on VOF concept.

G. Level Contour Reconstruction Method

The consolidation of the VOF method and the marching cubes algorithm used in Kunimatsu et als attempt [6], is matched with the level contour reconstruction [5], which carrying the inherent potential of handling with topological changes.

The proposed concepts of researchers and their implementation of the interface reconstructions appeared to be quite simple and plain-dealing. Due to the interface stretches and deforms greatly in the simulations, it is prime to add and remove the interface elements during the solving period. Topology change is an inherent feature of both boiling flows and splashing liquids. The Level contour reconstruction of interfaces must be allowed to recombine when either parts of the similar interface or parts of two separate interfaces came together. Interface reconstruction procedure replaces additional burden of all of these particle addition, deletion, and topology changes.

H. Level Set Method

Liquid surfaces can be efficiently represented using level set method and dynamic implicit surface, proposed by Osher and Fedkiw [1]. A signed distance field $\Phi(x)$ explicitly point out the liquid surface and that could defined as the shortest distance from position x to the liquid surface. So that surface of the liquid featured by the zero contour of signed distance field. Positive $\Phi(x)$ symbolizes x is outside the fluid region while negative $\Phi(x)$ identify the inside. Velocity field u(x), which derived from the Navier-Stokes equations used for simulation purpose. Solving the level set equation using the upwind schemes generates the implicit liquid surface and the victim equations is :

$$\frac{\partial u}{\partial t} + (u.\nabla\Phi) = 0$$

Several useful properties are constitutes on signed distance function such as surface normal and curvature, that are resolved by :

$$N = \nabla\Phi$$
$$k = \nabla^2\Phi$$

Dynamic level set is used to track the fluid surfaces. Evolution of liquid surface volume tracked by attaching the particles directly to the surface at initial position and then just move them around in the velocity field. Requirement of adding extra particles is when the surface becomes over sparsely resolved, and removing the additional ones as the surface folds, or splashes back over itself. A flipside method which is possibly similar, exception in not using the particles. This method was developed by Osher and Sethian and is called the level set method. Super accurate fifth order WENO scheme is applied in solving the level set equations, but it still deteriorates from high smoothing artifacts and volume loss. This problem soothed by coupling particles with level sets. Particles assigning at both sides of the interface would eventually improved the level set based method. Applied particles able to revise the level set values of each frame, that would manage the detailed surface features as well as the total volume of fluid. This approach was named as particle level set (PLS) due to the presence of the auxiliary particles. Another mechanism is Marker level set method (MLS) which help to track the dynamic liquid surface. Here only marker particles were seeds at the interface, that obtaining more power full and highly accurate results than PLS. Moreover, the surface markers helps the Marker level set method to supervise non-diffusive surface texture advection.

I. Semi-Lagrangian Contouring

Stam introduced Semi-Lagrangian methods [3] that have been extensively used in computer graphics community to solve the nonlinear advection term of Navier-Stokes equations. Surface tracking method is build on the top this approach. The Semi-Lagrangian Contouring (SLC) [2] proposed by Bargteil, is different from the PLS. The SLC modifies its signed distance field by Semi-Lagrangian fashion and hence the triangle meshes extracted by Marching Cube algorithm in every time step. Additionally, it benefits from implicit surface and explicit polygon meshes with volume conservation, adaptive resolution, also an easy surface property convection.

Here explicit surface representation begins with each time-step. Fluid simulator is able to build up both distance function and a velocity field. The zero set of the field creates new surface. Backward tracing through the flow field helps to obtain the value of the field function. A special structure called distance tree (balanced octree subdivision of the spatial domain), which is widely used in the implementations. Underlined benefits of this structures includes, fast spatial index and fast approximate signed-distance function, which are cumulatively simplified the whole mesh related procedures by assisting the contouring algorithm.

The greatness of the SLC is the ability to track surface properties, such as texture coordinates, color, simulation variables etc accurately at trivial additional cost. It provides a mapping between surfaces at adjacent time-steps. If vertex v in the current mesh maps to point p in the old mesh and some surface property was stored at p, this property can be copied to v. Tracking of surface properties done on this way.

J. Surface Tracking With Error Compensation

When the free surface is physically valid (means that when the normal is anti-parallel with the pressure gradient), the energy is minimized. The surface normal obtained from the surface tracker and the pressure gradient from the fluid simulation, so that any deviation from the minimum energy state causes an unphysical conflict between the fluid simulation and surface tracker. This error can be scale down by following the direction of steepest descent of the energy function [26]. The energy gradient is the derivative of energy function with respect to its free variables. Here the surface tracker is over detailed and under constrained, hence only need of surface tracker is to adjust the control variables. Approximation of divergence free motion done through constraining the surface tracker adjustments to be local rotations. Thus, degrees of freedom made as the orientations of the surface tracker's normal n.

III. Other Methods And Comparative Analysis

Several other methods adapted the high-resolution surface conform to the low-resolution physics, instead of adapting the fluid simulation to the surface tracker. According to Wojtan and Turk [18], and Kim et al. [19], high-frequency visual artifacts can be removed using smoothing algorithms. Another attempt was by Yu and Turk [20]. They tried to use anisotropic smoothing kernels to bias the loss of surface detail. Same effects also provided by Williams [21] and Thurey et al. [22] for volume-loss artifacts with bi-Laplacian smoothing. Such smoothing approaches provide better result in small doses. But they quashed up many surface details by applying too much enthusiasm. Results physically incorrect surface motions. Small-scale viscous flows may support smoothing with over-damped surface tension and it may not be appropriate for inviscid liquid phenomena. To handle the lack of detailed surface motion when combining a low-resolution simulation and a detailed surface, Thurey et al. [25] Bojsen-Hansen [23] and Yu et al. [24] were proposed a high-resolution dynamic surface waves. These methods masked the high resolution surface artifacts with rippling motions, but limited with unnecessary restrictions like height-field, shallow water, and constant wave speed assumptions. Other invention by Wojtan et al. [18] removes topological inconsistencies in the surface tracker by re-sampling the surface, but suffers from the problem of surface noise.

To visualize the implicit surfaces, normally surface particles are not used. Most of the methodologies are following marching cubes algorithm or ray tracing. Discretization of 3D volume into a grid of cells possible with the former method and later one shoots rays from the viewer at the surface to calculate the color and depth of the surface at every pixel. The third, Point-based method is not a common mechanism for visualizing implicit surfaces. Because of the high computational cost and need of large numbers of particles for visualization. Ray tracing is expensive as well but are often effortless to implement on CPU based hardware. Therefore it is more obvious option for offline rendering. Particle systems are good for exploiting temporal coherence. Once the positions of surface particles on a fluid surface are established at a specific time, they have to be moved only a small range to represent the fluid surface a fraction of time later. Marching cubes and ray tracing cannot take care of this property.

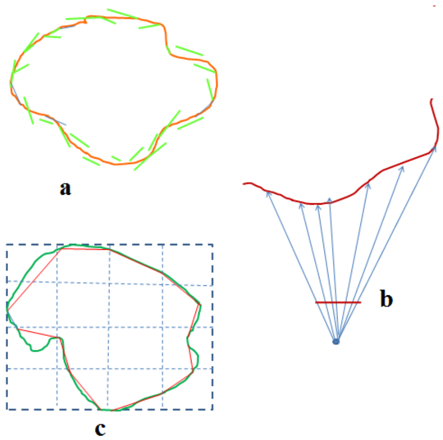


Fig. 1: *Implicit surface representations : a. Marching cube, b. Ray tracing, c. Point based method*

Surface visualization seems like offloaded to graphics hardware with the advancement in the growth of GPU processing power. Furthermore, the applied parallelism of current graphic processor units allows, much faster processing over time. That dominates GPU-run methods over CPU-run methods. Parallelism can also be applied in CPU in some extent with special tools, such as Intel TBB. Doesn't mean all surface visualization techniques are easily modified to work in a parallel environment. For instance, an iterative optimization that walks over the surface by visiting adjacent grid cells is not suitable for parallelization. Its complexity is therefore worser than the point-based method, which has to frequently modify all surface particle positions and the number of particles linearly related to the fluid surface area.

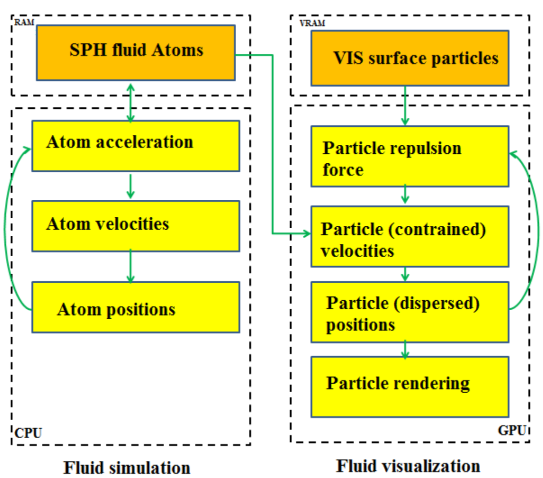


Fig. 2: *Surface visualization on GPU*

Figure 2 illustrate the point based surface visualization by nvidia developers [27]. This method offload the procedures from CPU to GPU and enhances the efficiency.

IV. Conclusion And Future Work

Novel methods for tracking the surface of liquids were surveyed in this paper. It is difficult to propose a best method, that is suitable for all applications ranging from entertainment field to scientific field. The computational cost, appearance and speed of processing are different with each mechanisms. Due to this fact, it is hard to suggest an excellent classification based on standard metrics. Efficient fluid simulation with clear surface representation might be capable of lighten the realism in animation.

Fluid simulation and surface tracking are two different type of problems. Simulation of fluid body can be done without tracking, but in order to add minute details, it will help a lot. In the other sense it can be used as a deformation tool. This survey gives a solid idea of how to carefully choose a particular surface tracking method to complete the hybrid fluid simulation approaches. A fruitful classification is a possible extensions of this survey, that can be considering as future work.

Acknowledgment

The authors would like to thank the reviewers for their useful suggestions to improve the earlier drafts of this manuscript and also thank to Jasmine Hydrose, National Institute of technology Manama, Bahrain.

References

[1] Stanley Osher, Ronald Fedkiw *Level set methods and dynamic implicit surfaces* Springer-Verlag New York, Inc., 2003.

[2] Adam W. Bargteil, Tolga G. Goktekin, James F. O Brien, and John A. Strain *A semi-Lagrangian contouring method for fluid simulation*, 25(1), 2006.

[3] Jos Stam *Stable fluids* ACM Press, Addison-Wesley Publishing Co. New York, NY, USA, 121-128. DOI=10.1145/311535.311548 1999

[4] Foster N, and Fedkiw R. *Practical animation of liquids*, In SIGGRAPH 01: Proceedings of the 28th annual conference on Computer graphics and interactive techniques, ACM, New York, NY, USA, 23C30, 2001.

[5] Shin S, and Juric D, *Modeling three-dimensional multiphase flow using a level contour reconstruction method for front tracking without connectivity*, J. Comput. Phys., 180, 427C470 (2002).

[6] Kunimatsu A, Watanabe Y, Fujii H, et al, *Fast simulation and rendering techniques for fluid objects*, Computer Graphics Forum (Eurographics 2001 Proc.), 20(3), 357C367 (2001).

[7] BOISSONNAT, J. D. AND OUDOT. *Provably good surface sampling and approximation*, In SGP-03: Proceedings of the Eurographics/ACM SIGGRAPH Symposium on Geometry Processing. Eurographics Association, Aire-la-Ville, Switzerland,2003 918.

[8] Tao Ju, Frank Losasso ,Scott Schaefer ,Joe Warren *Dual contouring of hermite data* ACM Transactions on Graphics (TOG) - Proceedings of ACM SIGGRAPH 2002 TOG Homepage Volume 21 Issue 3, July 2002 Pages 339-346

[9] Kunimatsu A, Watanabe Y, Fujii H, et al, *Fast simulation and rendering techniques for fluid objects* , Computer Graphics Forum (Eurographics 2001 Proc.), 20(3), 357C367 (2001).

[10] Hilton and Illingworth and Delaunay, *Marching Triangles: Delaunay Implicit Surface Triangulation*, CiteSeer publications 1997.

[11] Hirt C W, and Nichols B D, *Volume of Fluid (VOF) method for the dynamics of free boundaries* , J. Comput. Phys., 39, 201C255 (1981).

[12] Torres D J, and Brackbill J U, *The point-set method: front-tracking without connectivity*, J. Comput. Phys., 165, 620-644 (2000).

[13] Tryggvason G, Bunner B, Esmaeeli A, et al. *A front tracking method for the Computations of multiphase flow*, J. Comput. Phys., 169, 708-759 (2001).

[14] Unverdi S, and Tryggvason G, *A front tracking method for viscous, incompressible, multifluid flows*, J. Comput. Phys., 100 25 (1992).

[15] Muller M, Schirm S, and Duthaler S. 2007. *Screen space meshes.* In SCA 07: Proceedings of the 2007 ACM SIGGRAPH/Eurographics symposium on Computer animation, Eurographics Association, Airela Ville, Switzer-land, Switzerland, 9C15.

[16] Blinn, James F. 1982. *A Generalization of Algebraic Surface Drawing.*, In ACM Transactions on Graphics 1(3), pp. 235256.

[17] Matthias Mller and David Charypar and Markus Gross, *Particle-Based Fluid Simulation for Interactive Applications*, 2003

[18] WOJTAN, C., AND TURK, G. 2008. *Fast viscoelastic behavior with thin features.*, ACM Transactions on Graphics (SIGGRAPH) 27, 3, 47:147:8

[19] KIM, D., SONG, O.-Y., AND KO, H.-S. 2009. *Stretching and wiggling liquids.*, ACM Transactions on Graphics (SIGGRAPH Asia) 28, 5, 120:1120:7.

[20] YU, J., AND TURK, G. 2010. *Reconstructing surfaces of particle based fluids using anisotropic kernels.*, In Proceedings of the ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA), 217225.

[21] WILLIAMS, B. 2008. *Fluid surface reconstruction from particles.*, Masters thesis, The University Of British Columbia.

[22] THUREY , N., WOJTAN, C., GROSS, M., AND TURK, G. 2010. *A multiscale approach to mesh-based surface tension flows.*, ACM Transactions on Graphics (SIGGRAPH) 29, 4, 48:148:10

[23] BOJSEN-HANSEN, M. 2011. *A Hybrid Mesh-Grid Approach for Efficient Large Body Water Simulation.*, Masters thesis, Aarhus University

[24] YU, J., WOJTAN, C., TURK, G., AND YAP, C. 2012. *Explicit mesh surfaces for particle based fluids.*, Computer Graphics Forum (Eurographics) 31, 2, 815824.

[25] WOJTAN, C., THUREY , N., GROSS, M., AND TURK, G. 2010. *Physics-inspired topology changes for thin fluid features.*, ACM Transactions on Graphics (SIGGRAPH) 29, 4, 50:150:8.

[26] Morten Bojsen-Hansen and Chris Wojtan. 2013. *Liquid surface tracking with error compensation.*, ACM Trans. Graph. 32, 4, Article 68 (July 2013), 13 pages. DOI=10.1145/2461912.2461991 http://doi.acm.org/10.1145/2461912.2461991

[27] Kooten,Gino van den Bergen Alex Telea *Chapter 7 :Point-Based Visualization of Metaballs on a GPU*, Eindhoven link: http://http.developer.nvidia.com/GPUGems3/gpugems3/ch07.html

Runtime Analysis of GPU-Based Stereo Matching

Christian Zentner
School of Software Engineering
Tongji University
Shanghai, China

Yan Liu
School of Software Engineering
Tongji University
Shanghai, China

***Abstract*—This paper elaborates on the possibility to leverage the highly parallel nature of GPUs to implement more efficient stereo matching algorithms. Different algorithms have been implemented and compared on the CPU and the GPU in order to show the speedup gained by moving the computation to the graphics card. The results were evaluated for accuracy using the test available on the Middlebury website for stereo vision. An assessment of the runtime performance was done by a script which examined the runtime behaviour of the individual steps of the stereo matching algorithm.**

***Keywords*—stereo matching; GPU computing; runtime analysis; computer vision; image processing**

I. Introduction

Stereo matching is a technique used to generate a depth map from at least two simple images of the same scene. The resulting depth map is an image which represents the distances from the camera to the objects shown in the original image for every pixel. The information provided by the resulting depth maps can then be used to analyse the scene and support other methods like object tracking algorithms. In our case, the maps will be used in combination with a mobile robot, to support its visual navigation within a room.

The stereo matching process works similar to how human eyes would perceive the scene. Due to the difference in the location of the two eyes, we can see the scene from two different view points. Objects closer to the eye will appear shifted by a larger degree than objects within a longer distance. Based on this shift, the brain can calculate the depth of the objects we see and produce a three dimensional image. For the purpose of the stereo matching algorithm, we only calculate the depth information and skip the three dimensional image.

An integral part of the matching procedure is the correct setup and calibration of the cameras. Analogous to the example with the human eyes, the camera setup needs to consist of at least two cameras. Setups with more than two cameras are possible as well and can increase the precision of the stereo matching algorithm by providing more than one pair of sample images. The cameras have to be positioned in a certain way, so they both record the same scene but from a slightly different position. The distance between the cameras is crucial — if they are too close, the objects are not shifted far enough to precisely calculate the depth, if they are too far from each other, the shift is so large that some objects are not contained in both images. The orientation of the cameras needs to be adjusted as well, so that they provide the same view of the scene.

A. GPU Computing

Graphics hardware has become an integral part of modern computers. It is mainly used to accelerate multidimensional graphical computations, which are usually needed to display complex computer games or for graphical animations of the operating system. A common characteristic of this kind of computations is their huge amount of smaller, often independent calculations, which have to be performed on a regular basis and within a short time interval, in order to create a smooth visual presentation.

The reason why a graphics card can handle such a large amount of computations better than a single CPU is its highly parallel nature, which is perfectly suited for this problem. Instead of a single computing unit, a graphics card consists of several thousands of computing units which can perform many computations in parallel. Since graphical computations are often performed independently on a per-pixel or per-vertex basis, the graphics card can perform a lot of these at the same time, whereas the CPU would have to calculate them sequentially, one at a time.

Graphical computations, however, are not the only kind of computations that could benefit from having such a large array of computing units as offered by the graphics card. In fact, graphics cards have more recently become popular for performing general purpose computations, which is often referred to as GPU computing. An important area of application of GPU computing is image processing, where independent calculations are performed in parallel on every pixel.

For the navigation of our mobile robot, it is crucial to get a most recent snapshot of the current environment. Otherwise, decisions, based on the outdated spatial information, will not take the current position of the objects around the robot into account and might result in fatal collisions. Since stereo matching is an image processing technique, an implementation on the GPU should improve the runtime of the process and allow for a more accurate navigation.

A drawback worth mentioning is the relatively high initialisation cost of GPU programs. The data needed for the following calculations has to be transferred to the memory of the graphics card before the actual computation can begin. Depending on the size of the data in relation to the total amount of calculations, this comparatively slow data transfer can ruin the runtime benefit of the parallel computing units. However, the actual impact of the data transfer is hard to estimate, and the best way to investigate the runtime behaviour of GPU-based applications is to run and compare them with their CPU-based counterparts.

II. MATCHING PROCESS

A. Preprocessing

Once the cameras are set up, they are ready to take a pair of images for further processing. Before these can be used by the matching algorithm, they need to be calibrated and rectified. The calibration is needed to remove distortions that are specific to a camera due to minimal irregularities in the build process. These distortions can be removed by applying a transformation on the image taken by that camera. The parameters for the transformation need to be determined by analysing the cameras distortion. This can be done using the Camera Calibration Toolbox for Matlab [2].

The next step is the rectification of the images. The matching algorithm will include a step where a certain element of one image will be searched for in the other image. To simplify this search, the images are transformed in a way that all objects that are on a plane, which is spanned by a point in the room and the coordinates of the two cameras, occur on the same line in both images taken by the cameras. This step is crucial for the following algorithms as it simplifies the search for a given object dramatically. Instead of searching across the whole image, the given object will now be within the same line as in the original image.

B. Algorithms

The matching algorithms work by analysing the so-called disparity, the shift of objects depicted in the images. Images closer to the cameras will show a larger shift when the two images are compared, whereas objects with a longer distance to the cameras will only show little deviation. To designate a certain disparity to a given pixel, the matching algorithm has to find the same pixel in the second image. However, the information contained in a single pixel is very limited, which makes it difficult to accurately determine the disparity. More information can be extracted from neighbouring pixels or even by taking the image as a whole into account. Stereo matching algorithms are hence divided into local and global methods.

The more information the algorithm uses, the more accurate are the results. Thus, global methods tend to produce much better depth maps than their local counterparts. This, however, comes at the cost of runtime, because global methods have to compute a much larger amount of information. For the purpose of providing up-to-date depth maps, this paper will focus on local methods, due to their significantly faster execution.

Local stereo matching algorithms consist of three basic steps and can include further enhancements to increase the quality of the generated depth maps. For the assessment of the runtime behaviour of different algorithms, a simple and a more sophisticated version of each of the three basic steps have been implemented on the GPU and also on the CPU. Additional enhancements have been neglected. A complete stereo matching algorithm is obtained by combining algorithms for each step. This modular approach allows for different setups which can be used to evaluate the impact of certain parts of the algorithm.

1) Likeliness: The first step is to calculate the likeliness of pixels at a given disparity in the second image compared to each pixel in the first image. To find a pixel with a certain disparity, we simply need to shift the x coordinate by the value of the disparity. This works because of the rectification step which was part of the preprocessing. It will assure that a point in the scene will be mapped to the same line in both images, regardless of the disparity. The actual likeliness is defined by a measure which differs depending on the algorithm used. This has to be performed for every disparity up to a certain maximum.

The simple version, that has been implemented for this step, makes use of a sum of squared differences as the measurement for likeliness. As a little enhancement, the maximum difference is limited by a constant, which keeps the influence of outliers within sane limits [6]. Given a pair of images I and J, the likeliness (or also called cost) for a given pixel $\mathbf{p}$, a disparity of $\mathbf{d} = (d \quad 0)^T$ and a limit of c_{max} can be calculated as follows:

$$c_{ssd}(\mathbf{p}, \mathbf{d}) = \min \left(\sum_{i=R,G,B} (I_i(\mathbf{p}) - J_i(\mathbf{p} - \mathbf{d}))^2, c_{max} \right)$$

For structured surfaces, the results produced by this simple method are quite good, but it falls short for plain areas with little colour differences.

A more advanced version of this step actually combines two different algorithms to exploit their benefits and cancel out each others drawbacks. One value is determined by using the absolute differences of the colour channels of the pixels, which works similar to the simple version:

$$c_{ad}(\mathbf{p}, \mathbf{d}) = \frac{1}{3} \sum_{i=R,G,B} |I_i(\mathbf{p}) - J_i(\mathbf{p} - \mathbf{d})|$$

Then, another value is calculated using the Census transform, which compares the intensity of a pixel with its neighbours. The results of the comparison for each image are stored in a bit string, where each bit encodes whether the intensity of the neighbour was greater or less the intensity of the pixel of interest. The final value c_{cen} is then calculated as the Hamming distance between the two bit strings, which is the amount of bits that are different. Eventually, these two values are normalised using a function ρ, which takes the value for the likeliness c and a parameter λ to adjust the influence of outliers. The normalised values are then accumulated to yield the final result:

$$\rho(c, \lambda) = 1 - \exp\left(-\frac{c}{\lambda}\right)$$

$$c_{adc}(\mathbf{p}, \mathbf{d}) = \rho(c_{ad}(\mathbf{p}, \mathbf{d}), \lambda_{ad}) + \rho(c_{cen}(\mathbf{p}, \mathbf{d}), \lambda_{cen})$$

By combining two different measures, this version performs well under most conditions.

2) Grouping: For the second basic step, more information about the neighbourhood of the given pixel is being incorporated. Here, the algorithm defines a certain area around the pixel and uses the average likeliness of this area as the value for the pixel. The strategy for deciding which pixels should be included into this area is very important and has a great influence on the detail of the depth map.

The naive approach does not pay any attention to the objects in the image. It simply defines a window of size $b_x \times b_y$

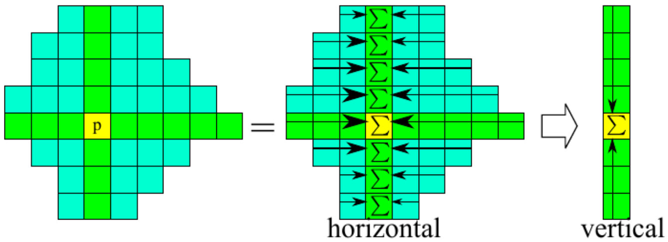


Fig. 1: Calculation of the average likeliness using cross-based aggregation (Mei *et al.*, 2011 [5]).

over which the average likeliness a of the included pixels with likeliness $c(\mathbf{p}, \mathbf{d})$ is calculated:

$$a_{box}(\mathbf{p}, \mathbf{d}) = \frac{1}{2b_x + 2b_y + 2} \sum_{x=p_x-b_x}^{p_x+b_x} \sum_{y=p_y-b_y}^{p_y+b_y} c\left(\begin{pmatrix} x \\ y \end{pmatrix}, \mathbf{d}\right)$$

Since this algorithm does not recognise objects within the image, it might average over object boarders and include elements with a large visual difference. Because of this, details will be lost and the result is not very accurate.

An improved version of this step is called cross-based aggregation, as described by Zhang *et al.* [15]. This approach defines the area, used for the aggregation, by extending into all directions from the pixel, for which the value is being aggregated. This creates arms of pixels to be included in the area, as depicted in figure 1. The length of these arms is limited by a maximum length and a threshold for the intensity difference. A further enhancement can be introduced by defining a second length and threshold. If an arm passes this second length, the second and more strict threshold should be used instead. The values chosen for the maximum length of the arms during the test are 17 and 34 pixels, combined with a threshold of 20 and 6 respectively. This algorithm, particularly with the mentioned enhancement, produces much cleaner results than the naive version, because it takes object boarders into account and tries to preserve details in the image.

3) Evaluation: Eventually, the previously computed values need to be evaluated, in order to estimate the correct disparity for the given pixel. The possibly correct disparity is the one, where the two pixels of both images have the least difference, i.e. the smallest aggregated likeliness value. This step can be used to apply further optimisations, the simplest approach, however, does not apply any optimisation and only selects the disparity which yields the smallest value for the previous step. This version is called winner-takes-all.

An approach, which performs a further optimisation, is called scan-line optimisation. This algorithm generates four new values for each pixel and each disparity and uses their average as the actual value at the given disparity. These four values are calculated based on the values of the pixels along all four directions originating from the given pixel. When all the averages for all disparities for one pixel are determined, the final result is again determined using the winner-takes-all algorithm.

III. Implementation

All of the mentioned algorithms have been implemented on a regular CPU in C++, as well as on the GPU using the OpenCL framework. The CPU versions follow a straight forward implementation, processing the pixels sequentially on a single core. To make full use of the parallel programming paradigm of the graphics hardware, a more complicated approach than for the CPU version is required

Luckily, except for the scan-line optimisation, all employed algorithms can mostly be executed independently and in parallel on a per-pixel basis. For the scan-line optimisation, each scan-line has to be calculated sequentially, which does not allow for a per-pixel division. However, for this algorithm, the disparities for a pixel can be calculated independently.

Another important aspect of GPU programming is the memory access. The graphics card features memory regions with different access speeds. The faster local memory regions are, however, only accessible by computing units of a certain group. That makes the actual division of computing units an important factor for the total runtime of the program, because accessing the main memory region is significantly slower.

The cross-based aggregation, for example, was grouped in a way, that one group is responsible for calculating one arm. The data needed for this arm can be loaded into the local memory, which reduces the access to the main memory to zero. By grouping the computing units this way, two effects could be witnessed. As the first improvement, due to the more organised way of grouping, the amount of lines of source code could be reduce to a third of the earlier, naive implementation. Furthermore, the runtime was reduced by ten times. This shows how important the memory access pattern for computations that need to read a lot of data is.

IV. Results

A. Accuracy

Before the assessment of the runtime performance, the quality of the depth maps, produced by the algorithms, will be tested and compared to other common stereo matching algorithms. This should guarantee, that the results for the runtime tests reflect the behaviour of algorithms that can be used by actual applications, which are not solely designed and optimised for their speed of execution.

The accuracy is tested using the Middlebury website [8], which provides an online test for stereo matching algorithms. A set of sample stereo images, that were used for previous publications [6] and [7], are provided, and the resulting depth maps will automatically be compared with laser scans of the actual environment. Figure 2 shows some of the images used for the test and their results. For each input image pair, one of the input images is shown, followed by the laser-scanned depth map of the scene and the resulting images for the combination of simple algorithms and for the advanced versions.

The evaluation shows the average error and a ranking position in a table containing other implementations. The error is calculated according to the paper on the evaluation of stereo correspondance algorithms [6] and is basically the root-mean-square error of the calculated depth compared to the laser-scanned depth for each pixel. Even the combination of the

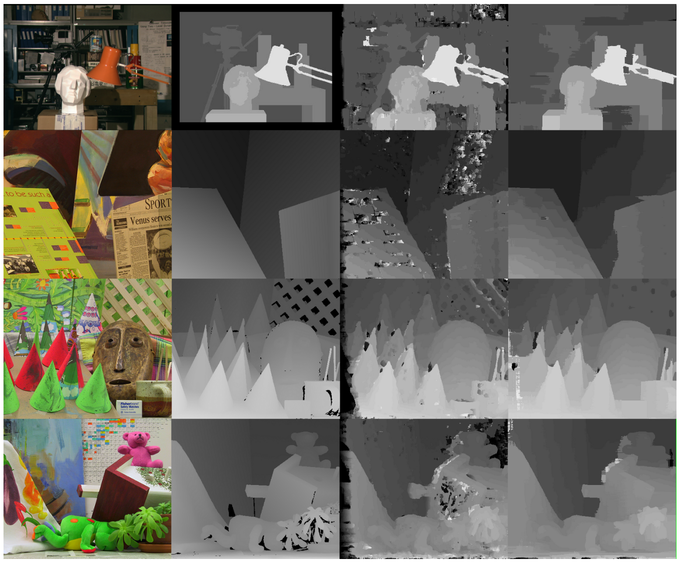


Fig. 2: Test images from the Middlebury website and the resulting depth maps. From left to right: One of the input images (from top to bottom: Tsukuba, Venus, Cones and Teddy), laser-scanned depth map, result for the combination of simple algorithms (sum of squared differences, box-based aggregation and winner takes all), result for the combination of advanced algorithms (absolute differences & census transform, cross-based aggregation and scan-line optimisation).

Image	SSD [ms]	BOX [ms]	WTA [ms]	Σ [ms]
Tsukuba	5.97	111.35	2.73	120.05
Venus	15.97	304.36	6.03	326.37
Cones	31.53	617.59	10.86	659.98
Teddy	31.44	617.21	10.85	659.50
Image	**ADC [ms]**	**CBA [ms]**	**SLO [ms]**	**Σ [ms]**
Tsukuba	686.11	255.25	225.60	1166.96
Venus	1905.53	507.04	875.40	3287.97
Cones	3735.04	567.88	2871.59	7174.50
Teddy	3736.18	764.99	2814.76	7315.93

TABLE I: Runtime on an Intel Core i5-2500K CPU for the combinations SSD (sum of squared differences), BOX (box-based aggregation) and WTA (winner takes all) as well as ADC (absolute differences & census transform), CBA (cross-based aggregation), and SLO (scan-line optimisation).

most simple algorithms is able to compete against some of the methods with low ranking. However, the quality is far from what the high ranked methods can produce, at a relatively high average error of 18.2%. The combination of the more advanced algorithms shows much better results, reducing the average error to 9.3%. Thus, the algorithm gains a place in the middle field of the table. Though these results are much better, and the resulting depth maps mostly provide clear and useful information, there is still a huge gap compared to the best ranked methods.

B. Runtime

Finally, the runtime performance of the CPU implementation is compared to its GPU counterpart. For this purpose, a given combination of algorithms were run for 20 times and their performance, shown in tables I and II, was analysed and averaged by a script. The chosen configurations feature the simple sub-algorithms, SSD (sum of squared differences), BOX (box-based aggregation) and WTA (winner takes all), as well as the more advanced versions, ADC (absolute differences & census transform), CBA (cross-based aggregation) and SLO (scan-line optimisation).

Image	SSD [ms]	BOX [ms]	WTA [ms]	Σ [ms]
Tsukuba	3.53	11.32	1.64	16.50
Venus	9.48	36.02	3.32	48.82
Cones	18.87	75.72	5.69	100.28
Teddy	18.87	75.67	5.69	100.23
Image	**ADC [ms]**	**CBA [ms]**	**SLO [ms]**	**Σ [ms]**
Tsukuba	9.82	17.15	13.75	40.72
Venus	24.79	62.91	29.46	117.16
Cones	74.42	137.92	54.56	266.91
Teddy	74.41	139.52	54.53	268.46

TABLE II: Runtime on a NVIDIA GeForce GTX 570 GPU for the combinations SSD (sum of squared differences), BOX (box-based aggregation) and WTA (winner takes all) as well as ADC (absolute differences & census transform), CBA (cross-based aggregation), and SLO (scan-line optimisation).

The script read the timings, produced by the algorithms for each step of the stereo matching process, and computed for each of them, and for the total run of the combination, the average result and the standard deviation. This allows to compare the runtime of every part of the algorithm and give a more detailed analysis of the benefit of implementing these on the GPU. All tests were performed on a system with an Intel Core i5-2500K @ 3.30 GHz processor and an NVIDIA GeForce GTX 570 graphics card.

The results show that the algorithms responded differently on the implementation choice. While the simpler algorithms, like SSD and WTA, could only gain a speedup by a factor of around 2, other algorithms, like the combination of the absolute differences and the census transform, could profit from the parallelism a lot, showing a speedup by up to a factor of 50 to 75 compared to the CPU version. This also has an impact on the accumulated runtime of the whole stereo matching methods.

Figure 3 gives an overview of the total runtime of the stereo matching for different images from the Middlebury website. The chosen algorithms are the same combinations as in the tables, using the simpler version SBW (SSD, BOX & WTA) and the more advanced version ACS (ADC, CBA & SLO). Note that the axis for the runtime is logarithmic. As for the comparison of the individual parts, the total runtime of the advanced version shows a larger speedup than the less sophisticated implementation. This might be related to the higher initialisation cost of GPU-based programs.

V. Conclusion

This paper provides a rough introduction to stereo matching on the graphics hardware and compares the speedup gained in comparison to a similar CPU implementation. The evaluation can be used as a basic guidance for other developers to decide, whether GPU computing is a suitable option for heavy computations within their projects, and as a starting point for further investigations. Especially, time-critical and more complex operations could benefit a lot from an implementation

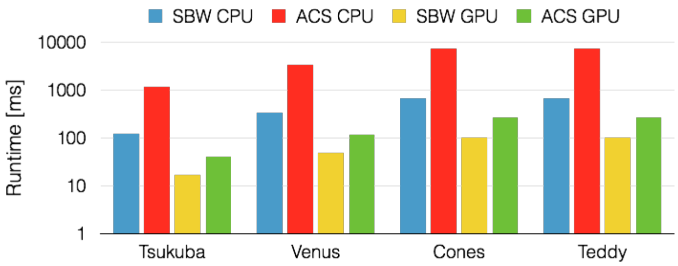


Fig. 3: Comparison of the total runtime of the combinations SBW (sum of squared differences, box-based aggregation and winnter takes all) and ACS (absolute differences & census transform, cross-based aggregation and scan-line optimisation).

on the GPU, given the underlying algorithm can be parallelised in some way.

The test results for the accuracy of the implemented algorithms show that there is still a large space for improvements. However, the produced depth maps of the advanced algorithms are by all means sufficient for most applications, and should suffice for the purpose of this paper. The runtime comparison shows that some of the algorithms can benefit significantly from the highly parallel nature of the graphics hardware. Even algorithms like the scan-line optimisation, which are hard to implement efficiently in a parallel manner, given their partly serial computations, seem to gain a huge performance boost on the GPU.

These observations, in relation with the relatively low cost and high availability of graphics hardware and the increasing interest and popularity of GPU computing, which also pushes the development of frameworks and development tools, certainly justify the use of graphics cards for stereo matching algorithms. Indeed, many recently developed stereo matching methods make use of GPU computing and show that it is possible to obtain high quality depth maps within a short runtime using this technology [5].

For our mobile robot, the presented GPU implementation provides for a much more accurate navigation based on very recent depth maps. Another benefit of the short runtime of this algorithm is that other parts of the whole process, that is responsible for the robot's navigation, could be allowed to consume more time in order to improve their accuracy. This allows for a more flexible fine-tuning of the whole navigation process, to find the perfect balance between using most up-to-date data and using more accurate algorithms.

The only drawback is the higher level of complexity needed to implement efficient algorithms on the GPU. For more sophisticated methods, a GPU implemetation, that exploits the offered level of parallelism to a sane extend, needs to pay attention to many pitfalls that do not pose any obstacles for a CPU version. Especially, finding a performant way of parallelising an algorithm and optimising memory accesses are an important and sometimes difficult task. However, implementing an efficient CPU-based algorithm also requires for complex mechanisms. Eventually, the huge runtime speedup on the graphics hardware makes GPU computing an alternative that is worth to be explored.

References

[1] M. Bleyer and M. Gelautz, "A layered stereo matching algorithm using image segmentation and global visibility constraints," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 59, no. 3, pp. 128–150, 2005.

[2] J.-Y. Bouguet, *Camera calibration toolbox for matlab*. [Online]. Available: http://www.vision.caltech.edu/bouguetj/calib%5C_doc/index.html (visited on Nov. 25, 2012).

[3] S. Chambon and A. Crouzil, "Similarity measures for image matching despite occlusions in stereo vision," *Pattern Recognition*, vol. 44, no. 9, pp. 2063–2075, 2011.

[4] N. Lazaros, G. C. Sirakoulis, and A. Gasteratos, "Review of stereo vision algorithms: From software to hardware," *International Journal of Optomechatronics*, vol. 2, no. 4, pp. 435–462, 2008.

[5] X. Mei, X. Sun, M. Zhou, S. Jiao, H. Wang, and X. Zhang, "On building an accurate stereo matching system on graphics hardware," *GPUCV: ICCV Workshop on GPU in Computer Vision Applications*, 2011.

[6] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[7] ——, "High-accuracy stereo depth maps using structured light," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, vol. 1, pp. 195–202, 2003.

[8] ——, *Middlebury stereo vision page*. [Online]. Available: http://vision.middlebury.edu/stereo/ (visited on Nov. 25, 2012).

[9] T. Schöpping, "Entwicklung eines stereo-visionverfahrens für einen mobilen roboter," Bachelor thesis, Technische Fakultät der Universität Bielefeld, Bielefeld, 2011.

[10] O. Schreer, *Stereoanalyse und Bildsynthese*. Berlin: Springer, 2005, ISBN: 3-540-23439-X.

[11] L. D. Stefano, M. Marchionni, and S. Mattoccia, "A fast area-based stereo matching algorithm," *Image and Vision Computing*, vol. 22, no. 12, pp. 983–1005, 2004.

[12] O. Veksler, "Reducing search space for stereo correspondence with graph cuts," *Proceedings of British Machine Vision Conference*, vol. 2, pp. 709–718, 2006.

[13] K.-J. Yoon and I. S. Kweon, "Adaptive support-weight approach for correspondence search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 650–656, 2006.

[14] C. Zentner, "Gpu-basiertes stereo-matching auf einem mobilen roboter," Bachelor thesis, Technische Fakultät der Universität Bielefeld, Bielefeld, 2012.

[15] K. Zhang, J. Lu, and G. Lafruit, "Cross-based local stereo matching using orthogonal integral images," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, no. 7, pp. 1073–1079, 2009.

A Multiple-Objects Recognition Method Based on Region Similarity Measures: Application to Roof Extraction from Orthophotoplans

Abdellatif El Idrissi*, Youssef El Merabet†, Yassine Ruichek‡, Raja Touahni †, Abderrahmane Sbihi*, Cyril Meurie§ and Ahmed Moussa*

*Communication and Information Technologies Laboratory,Abdelmalek Essaadi University, 90000 Tangier, Morocco
†LASTID Laboratory, Department of Physics, Faculty of Sciences, Ibn Tofail University, B.P 133, 14000 Kenitra, Morocco.
‡IRTES-SeT Laboratory, Technology University of Belfort-Montbeliard, 13 Ernest Thierry-Mieg Street, 90010, Belfort,France
§Univ Lille Nord de France, F-59000 Lille, IFSTTAR, LEOST, F59650, Villeneuve d'Ascq,France.

***Abstract*—In this paper, an efficient method for automatic and accurate detection of multiple objects from images using a region similarity measure is presented. This method involves the construction of two knowledge databases: The first one contains several distinctive textures of objects to be extracted. The second one is composed with textures representing background. Both databases are provided by some examples (training set) of images from which one wants to recognize objects. The proposed procedure starts by an initialization step during which the studied image is segmented into homogeneous regions. In order to separate the objects of interest from the image background, an evaluation of the similarity between the regions of the segmented image and those of the constructed knowledge databases is then performed. The proposed approach presents several advantages in terms of applicability, suitability and simplicity. Experimental results obtained from the method applied to extract building roofs from orthophotoplans prove its robustness and performance over popular methods like K Nearest Neighbours (KNN) and Support Vector Machine (SVM).**

Keywords*—*Object recognition; Region Similarity Measure; Texture; Feature extraction; Orthophotoplans

I. INTRODUCTION

Nowadays, automatic object recognition has become a topic of growing interest for computer vision community. For instance, automatic extraction of man-made objects such as buildings and roads in urban areas has gained significant attention for photogrammetric researchers community over the last decade. This problem is usually considered when we talk about high-level image processing in order to produce numerical or symbolic information [1], [2]. In this context, several approaches have been proposed in the literature. First, one can cite interactive methods that need user interaction in order to extract desired targets or objects of interest from images. Generally, this category of methods has been introduced to alleviate the problems inherent to fully automatic segmentation which seems to never be perfect. These methods endeavour to divide an image into two segments: "object" and "background". The interactivity consists in imposing certain hard constraints for segmentation by pointing out certain pixels (seeds) that absolutely have to be part of the object and certain pixels that have to be part of the background.

Boykov and Jolly proposed an interactive graph cuts (IGC) for interactive image segmentation [3]. The segmentation is performed by the min-cut/max-flow algorithm. User scribbles extract color information that will be used thereafter as hard constraints. Rother et al. in [4] presented an iterative algorithm called GrabCut by simplifying user interaction. Their method combines image segmentation using graph cut and Gaussian Mixture Models (GMMs) based statistical models. A very useful segmentation benchmark, with a platform implementing important algorithms, has recently been proposed by McGuinness and Connor [5]. The authors compared many algorithms such as IGC [3], seeded region growing (SRG) [6], simple interactive object extraction (SIOX) [7] and binary partition tree (BPT) [8], in order to provide a good coverage of the various techniques currently available for foreground extraction, as stated in [5].

The SIOX [7] algorithm is also based on color information and has recently been integrated into the popular imaging program GIMP as "Foreground Selection Tool". The BPT [8] algorithm is based on hierarchical region segmentation, exploiting user interaction to split and merge regions in the tree. Bai and Sapiro [9] proposed a method based on fast kernel density estimation [10] for color statistics, improving geodesic distance-based approach described in [11].

Ning et al. [12] have recently proposed a novel maximal similarity based region merging (MSRM) mechanism for interactive image segmentation. The key idea of MSRM is to perform region merging between adjacent regions by exploiting an effective representation of color statistics basing on (quantized) color histograms computed from the regions. First, the input image is segmented using the mean shift segmentation algorithm. User must then indicate the location and region of the object to be extracted and background by using strokes as markers. Finally, a maximal-similarity based region merging mechanism is used in order to separate the object of interest from the background image while relying on the help of markers introduced by the user. A similar algorithm also based on maximal similarity based region merging has been proposed in [13]. The difference is that this algorithm considers regions as seeds and takes the regions as growth units for region growing (i.e. merging of adjacent regions).

These methods give generally good results, but depending on the degree of user interaction. Thus, they become not suitable for images with high resolution containing too many objects of interest such as aerial and satellite images.

To address these issues, another category of methods, namely semi-automatic or automatic methods was developed. These methods are not only devoted to be applied on aerial or satellite images, but also on any kind of images ranging from simple single intensity images and color images, to laser and stereo images. A considerable number of methods from this category first tend to inspire from techniques introduced in pattern recognition and machine learning domains. In [14], Tso and Mather reported some classification methods used in remote sensing and which are pixel-based approaches like K Nearest Neighbours (KNN), maximum likelihood method and Support Vector Machines (SVM). Several variants or methods were developed for improving SVM method. Mountrakis and al. in [15] wrote a review of methods based on SVM in remote sensing field. They highlighted that SVM based methods are particularly considered in the remote sensing field due to their ability to generalize well even with limited training samples. It took place that SVM still outperformed best odd neural networks [15]. In the papers [16], [17], the authors show that neural networks can also be used for object recognition. Kinnunen and al. presented in [18] a method based on self-organization to deal with unsupervised object discovery. It is based on similar techniques that use bag of features approach and clustering to automatically classify image data. In their method, they replace clustering step by a self-organizing map.

Some authors tried to combine different methods. For instance, the authors of [19] combine KNN and SVM. Another method is what has been proposed by [19] using KNN, SVM and Geometric Moment Invariants (GMI). Introduced by Hu M.K. in [20] and used in several methods as in [21], [22], [23], GMI has been chosen to extract image features like rotation, scale and translation (RST)-invariant. Mathematical morphology has been also used to detect objects of interest. Soille and Pesaresi in [24], [25] developed a method to extract roads. It consists in two stages: a pre-processing one, in order to remove noise from the image, and a processing phase in which a structuring element is defined according to the shape of the object to be extracted. Roughly similar to ours, a recent method is presented by Ahmadi et al. in [26]. The authors adapted the active contour or Snakes model, originally introduced by Kass et al. [27], to automatically extract urban building boundaries. For that, the knowledge about the buildings is incorporated by the user into the system by introducing some pixel values of points inside building boundaries as training data. The system can then make a difference between buildings and background in the image.

Another class of methods consists in joining to classification algorithms prior information like height data or Light Detection and Ranging (LIDAR) data to detect objects of interest. Examples of this class of methods are the works of Halla and Brenner in [28], and Zhao and Trinder [29] who utilized height data and morphological operators for buildings extraction. Following this idea, Samadzadegan et al. proposed a novel approach for objects recognition, based on neuro-fuzzy modeling. They extract structural, textural and spectral information and integrate them in a fuzzy reasoning process to which learning capability of neural networks is introduced [30]. Zimmermann et al. produced Digital Surface Model (DSM) data from stereo images. In this model, multiple cues, colour segmentation, edge detection, texture segmentation and blob detection are combined. They then used the model to detect building roofs using slope and aspect operators [31]. Miliaresis and Kokkas developed in [32] a method for extracting a class of buildings using Digital Elevation Models (DEMs) derived from Lidar data. The method is based on geomorphometric segmentation principles with k-means to obtain a set of clusters formed by background and foreground objects represented on the basis of elevation and slope. Lafarge et al. presented in [33] an automatic building extraction method from DEM based on an object approach. They start by applying marked point process tools to realize a rough approximation of building footprints, which are then regularized by improving connection between neighboring elements and detecting roof height discontinuities.

Methods based on the notion of interest points that allow representing characteristics of targeted objects should also be mentioned here [34], [35], [36]. In [34], Lowe proposed to extract distinctive invariant features from images and used them as key points for matching different views of a sought object using a fast nearest-neighbour algorithm. Similarly, in [36], Berg et al. proposed an algorithm dealing with the problem of deformable shape matching by defining a cost function that measures similarity of corresponding geometric blur point descriptor and geometric distortion between corresponding feature points. Recently, in [37], Liu et al. reported a series of other methods that use the same idea (key points based) on discriminative parts. This means that an object may be represented by local parts which allow to distinguish it from others. These methods can be divided into two classes, the class with methods that integrate selection of discriminative parts with model construction, and the class with methods that separate the two processes [38], [39].

Regarding the related literature, a large number of the above methods present several shortcomings. The methods of the first category require numerous initializations and manual interaction which is very time-consuming when there are many object instances. Methods from the second category are most of the time context-dependent and are sensitive to noise. In the third category, and as stated in [26], the mentioned investigations have stressed to introduce height data in the context of aerial or satellite images to automatically extract buildings. This leads to high computational efforts and makes the approach requiring significant technological resources for data production and processing.

In this work, we propose a new method which is simple but copes with those drawbacks and robustly extract objects of interest. The relevance of the proposed technique could be expressed through the following advantages. First, the method allows automatic extraction of objects of interest and performs without any user interaction. Second, by this method, it becomes possible to accurately detect multiple objects in the same time from a given image. Third, one can achieve robust results under various real-world conditions when considering for example, complex images for which both foreground and background regions have similar colors. Fourth, The method does not require height data or any prior information to recognize the difference between buildings and other background objects. Fifth, the method can be applied in several fields like medical image processing (e.g cancer cell recognition) and remote sensing image processing (e.g vegetation and buildings detection). In this paper, we are especially interested in extracting building roofs from orthophotoplans.

The remainder of the paper is organized as follows. In section II, we explain the proposed methodology and describe its main steps in details. Experiments and both qualitative and quantitative evaluations are presented in section III. A comparative analysis with other methods is also reported in this section. Section IV concludes the paper and addresses future works with the aim of enhancing the performance of the proposed method.

II. Proposed method

A. General description

The proposed method incorporates two major stages: off-line and on-line stages. In the off-line stage, two knowledge databases must be created in order to robustly avoid user interaction. The first one contains representative and distinctive textures of objects to be extracted. The other database is composed of textures picked up from objects that represent background in the image. As an illustrative example, for the application of building roof extraction from aerial images, the first knowledge database $\mathbf{B_{obj}}$ will be constructed with m distinctive textures of building roofs while the second one $\mathbf{B_{back}}$ will be constructed with n distinctive textures of other objects such as vegetation, road, forest, etc. These two databases are provided considering some examples of images. Having these two knowledge databases $\mathbf{B_{obj}}$ and $\mathbf{B_{back}}$ as reference, it is possible to automatically extract building roofs from any aerial image (orthophotoplan, in this study case). Figure 1 illustrates an example of knowledge databases used in this work. It should be noted here that more specific details either on used data or on how to construct those knowledge databases will be provided at the experimental result section (see section III). In the on-line stage, the object extraction

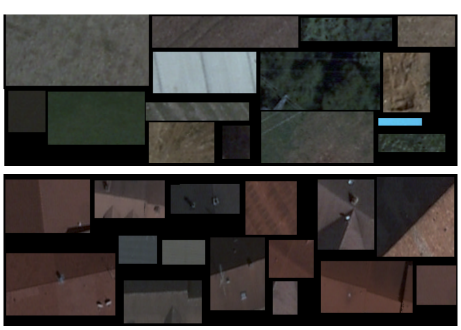


Fig. 1: Example of knowledge databases used in this work. From top to bottom: knowledge database $\mathbf{B_{back}}$ of background (vegetation, road, forest, etc) and knowledge database $\mathbf{B_{obj}}$ of building roofs (red and non-red rooftop buildings).

process is performed. To do this, We begin by over-segment the original image into many small and homogeneous regions. This is called a low-level processing step. In this paper, we have used SRM algorithm [40] (cf. section II-B) as a tool of segmentation. Having a segmented image, the following task is a high-level processing step that consists in extracting features characterizing regions of both segmented image and constructed knowledge databases. In this work, RGB color histogram features (cf. section II-C)are used. The question that arises then is how can we measure the similarity between those regions. Several well-known goodness-of-fit statistical metrics using RGB color histogram features exist in the literature. In this work, the Bhattacharyya descriptor is adopted to accomplish this operation (cf. section II-D).

Once similarity measure is evaluated for all regions, each one of them can be classified as a part of an object of interest or rather as a part of the background of the image (see section II-E). Finally, object contours are delineated keeping only regions labelled as object of interest (building roof in this case).

Figure 2 resumes the general flowchart of the proposed building-detection method.

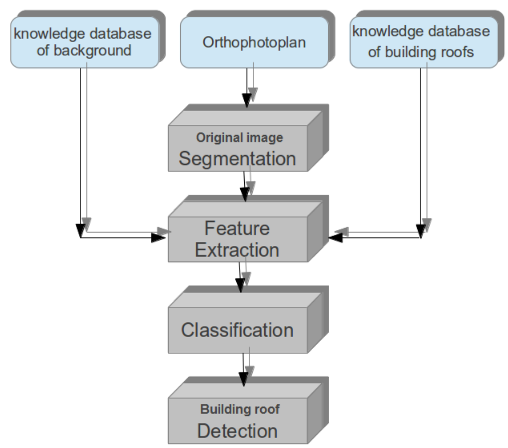


Fig. 2: General flowchart of the proposed building-detection method.

B. Initial segmentation using Statistical Region Merging

The low-level processing step consists in over-segmenting the input image into many small and homogeneous regions with the same properties. The goal of this initial segmentation is to avoid the under-segmentation problem and thus correctly extract all significant regions where boundaries coincide as closely as possible with the significant edges present in the image. Of course, there are many low level segmentation methods in the literature which can achieve that. One can cite Mean shift, Jseg unsupervised segmentation algorithm [41], watershed, Turbopixels [42], Statistical Region Merging (SRM) [40], etc. In this paper, authors have chosen SRM algorithm to obtain the initial segmentation of the input image. Particular advantages of using this algorithm for dealing with large images are that SRM dispenses dynamical maintenance of region adjacency graph (RAG), it allows defining a hierarchy of partitions and it runs in linear-time by using bucket sorting algorithm while transversing the RAG. In addition, the SRM segmentation method not only considers spectral, shape and scale information, but also has the ability to cope with significant noise corruption and handle occlusions (Fig. 3).

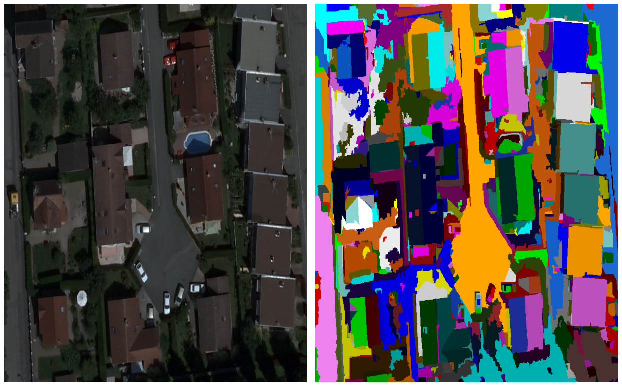


Fig. 3: Example of segmentation result using Statistical Region Merging (SRM) method. From left to right: Original image and its SRM segmentation result.

C. Region representation

In this stage of the method, we dispose of a segmented image obtained via the SRM algorithm. It is still a challenging problem to accurately extract the object contours from this image because only the segmented regions are calculated and no information estimation on their content, which is necessary for the extraction process, is yet available. The main goal consists in classifying each segmented region as target object or background. For this purpose, we need first to join the strategy adopted by many authors and which characterizes the regions using suitable descriptors.

It appears from the literature that there are several aspects that could be considered for representing a region such as edge [43], texture [44], shape, size or color. For the present purpose, the most appropriate information is color. In fact, region texture, which can be understood as repeatedly occurring local patterns in images and its arrangement rules, are unfortunately difficult to describe; Also, the same difficulties can be faced regarding shape and edge. Moreover, region size, although it can be measured simply by computing the number of pixels, it doesn't allow a unique distinction of objects of interest since they can have different sizes from an image to an other or simply they can have the same size as other objects belonging to the background of the image. Hence, color information which can be tackled using simply by computing its mean value or its histogram is an effective parameter to describe statistical information of object color distribution. Note that region histograms are local histograms and they reflect local features in images. Therefore, we exploit color histogram to represent all regions of the segmented image and those of the constructed knowledge databases.

In this purpose, each color channel is at first uniformly quantized into *l*=16 levels; afterwards, the color histogram of each region is calculated within the feature space of $l \times l \times l$ = 4096 bins. Obviously, quantization reduces the information regarding the content of regions and it is used as trade off when one wants to reduce processing time. The RGB color space is used in order to perform these computations.

Now that we have defined the feature adopted for characterizing the regions, the key issue is to determine similarity between regions of the segmented image and those of the constructed knowledge databases. For that, a similarity measure rule $\varrho(R,Q)$ between two regions R and Q should be defined basing on their color histograms.

D. Similarity measure rules

The most similarity measures commonly used are based on vector space model, i.e. taking image region features as points in the vector space, through the calculation of close degree of two points to measure the similarities between the image region features. Common similarity measures include Minkowski measure, histogram intersection method [45], second type distance [46], Bhattacharyya coefficient [47], and log-likelihood ratio statistic [48], etc. For regions R and Q, using the notation $\varrho(R,Q)$ for representing the similarity between regions R and Q, the larger ϱ is, the larger similarity between region R and Q we will get. Denote by $Hist_R^i$ the normalized histogram of a region R, the superscript i represents its i^{th} element. $z = l \times l \times l = 4096$ represents the feature space.

Examples of similarity measures are given as follows:

- Minkowski measure:

$$\varrho(R,Q) = \left(\sum_{i=1}^{z} |Hist_R^i - Hist_Q^i|^p \right)^{\frac{1}{p}}, \quad (1)$$

where $p = 1, 2 or \infty$;

- Euclidean distance:

$$\varrho(R,Q) = \sqrt{\sum_{i=1}^{z} (Hist_R^i - Hist_Q^i)^2}, \quad (2)$$

which is a Minkowski measure with p=2.

- Quadratic distance metric:

$$\varrho(R,Q) = \sqrt{\sum_{i=1}^{z} [(Hist_R^i - Hist_Q^i)^T A (Hist_R^i - Hist_Q^i)]}, \quad (3)$$

A is the bin-similarity matrix;

- Histogram intersection method:

$$\varrho(R,Q) = \frac{\sum_{i=1}^{z} min\{Hist_R^i, Hist_Q^i\}}{\sum_{i=1}^{z} Hist_R^i} \quad (4)$$

- Bhattacharyya coefficient:

$$\varrho(R,Q) = \sum_{i=1}^{z} \sqrt{Hist_R^i . Hist_Q^i} \quad (5)$$

In this work, authors adopted Bhattacharyya coefficient, which represents the cosine of angle between the unit vectors

$$(\sqrt{Hist_R^1}, ..., \sqrt{Hist_R^z})^T$$

and

$$(\sqrt{Hist_Q^1}, ..., \sqrt{Hist_Q^z})^T$$

This choice is due to its ability to simulate very well the similarity value of vector shape. The higher the Bhattacharyya coefficient between regions R and Q is, the higher the similarity between them is. That is to say their histograms are very similar and the angle between the two histogram vectors is very small. Certainly, two similar histograms do not necessarily involve that the two corresponding regions are perceptually similar. Nevertheless, coupling with the proposed classification process introduced in the next section II-E, Bhattacharyya similarity works well in the proposed approach.

It should be mentioned that a histogram is a global descriptor of a local region and it is robust to noise and small variations. Given that the Bhattacharyya coefficient is the inner product of two histogram vectors, this coefficient is thus robust to noise and small variations too. It has been used in [12], [13] for user interaction based image segmentation. Unlike theses methods, the proposed one aims for multiple extraction of objects of interest using two constructed knowledge databases without any need for user to provide markers input usually necessary for region merging process.

E. Classification process

At this stage of the method, we aim to determine which of the two classes (objects of interest or background) will be affected to the regions composing the initial SRM segmentation result, which we denote $\mathbf{M_{SRM}}$. For this end, candidate regions of $\mathbf{M_{SRM}}$ that have maximal similarity with the regions of the knowledge database $\mathbf{B_{obj}}$ of objects and those having maximal similarity with the regions of the knowledge database $\mathbf{B_{back}}$ of background are identified. Once all regions of $\mathbf{M_{SRM}}$ are classified, this leads directly to extracting the desired objects (e.g. building roofs). The proposed object extraction method can be summarized as in algorithm 1. As one can state, the similarity rule is very simple but it is efficient for the classification process. Note that the mean values of similarity moy^R_{obj} and moy^R_{back} are inversely proportional to the value of k, i.e. the higher the value of k is, the lower the mean value of similarity is. If this is the case, a dispersion of the mean values of similarity is obtained, which involves obtaining false classification result of the regions of $\mathbf{M_{SRM}}$. Besides, the k value has an important impact on the quality of results. For the purpose of keeping a significant similarity mean value, avoid the dispersion phenomenon and hence obtain good classification results, the two values moy^R_{obj} and moy^R_{back} are only calculated on the k first values of the sorted similarity vectors V^R_{obj} and V^R_{back} respectively. Although the similarity mean values moy^R_{obj} and moy^R_{back} are sensitive to outliers of k value, we empirically found that there is a range of values where the classification results remain stable. This optimum range is determined experimentally using a trial and error approach. Once this parameter is determined, it keeps the same value for test images. In this work, k is adjusted at 7.

Algorithm 1 Multiple objects extraction algorithm

Require: $I \leftarrow$ input image.
 $\mathbf{B_{obj}} \leftarrow$ Knowledge database of objects of interest (building roofs).
 $\mathbf{B_{back}} \leftarrow$ Knowledge database of background (vegetation, road, forest, etc)

```
(over)Segment I into regions through SRM algorithm in order to obtain the set M_SRM of segmented regions.
Calculate the RGB color histogram features for all regions of M_SRM and for those composing the two constructed knowledge databases B_obj and B_back.
for each candidate region R ∈ M_SRM do
    Calculate the similarity vector V^R_obj = {ϱ(R,Q_i); (Q_i)_{i=1..m} ∈ B_obj} between R and B_obj. ϱ(R,Q_i) is the similarity between the region R and the region Q_i ∈ B_obj.
    Calculate the similarity vector V^R_back = {ϱ(R,Q_j); (Q_j)_{j=1..n} ∈ B_back} between R and B_back. ϱ(R,Q_j) is the similarity between the region R and the region Q_j ∈ B_back.
    Get the order of V^R_obj and V^R_back by decreasing sorting;
    Calculate moy^R_obj = (Σ_{i=1}^{k} ϱ(R,Q_i)) / k, k ≤ m, the mean of the k first elements of V^R_obj.
    Calculate moy^R_back = (Σ_{j=1}^{k} ϱ(R,Q_j)) / k, k ≤ n, the mean of the k first elements of V^R_back.
    if (moy^R_obj ≥ moy^R_back) then
        The region R maximizes the similarity with B_obj, it is then classified as a part of building roof.
    else
        The region R maximizes the similarity with B_back, it is then classified as a part of background.
    end if
end for
The classification process is finished.
return The final segmentation map.
```

III. Experimental results

In this section, we are interesting in assessing the ability of the proposed building-extraction strategy to deal with multiple detection of building roofs from orthophotoplans. As pointed out in the introduction, the proposed building-extraction algorithm runs automatically without any user interaction. To avoid each time calculating region features from the two constructed knowledge databases and thus reduce the computation time, an alternative consists in calculating them once and for all and save them in a binary file. Thus, the process of similarity measure is performed using this binary file and no more the two knowledge databases.

A. Material description

1) Study area and knowledge databases: Data used in this research to evaluate the accuracy of the proposed algorithm are aerial images, particularly orthophotoplans. Several images were acquired for the region of Belfort city situated on the north-eastern of France in 2003, from a hot air balloon. Their spatial resolution is 16 cm/px. These images cover a wider area, where appear complex and multiple objects of different classes, various shadows, occlusions, multiple colors and textures and some terrain height variability. Namely, targeted objects, that are roofs of buildings, are often red and rarely non-red. In addition, they may differ according to their exposure to the sun and so they could change in terms of contrast and luminance. Therefore, these differences should be taken in consideration during the step of the construction of the knowledge databases.

This construction is explicitly performed by selecting a number of distinctive textures representing both the roofs and

TABLE I: Distinctive textures used to build object and background databases.

	Vegetation	Roads	Floors	Swimming Pools	Red roofs	Non-red roofs	Total
Object database	-	-	-	-	7	6	13
Background database	5	5	4	1	-	-	15

the background of the image. For this purpose, and as one can see on the second row of table I, a total of thirteen roof textures have been picked up from original images. Among them, seven are red and six others are gray or somewhat black. Thus, differences in contrast and brightness of objects of interest should be token into account. As for background database, we had taken a total of fifteen textures that belong to the background of the images. Five textures, related to vegetation and roads, are token for each category; four textures are selected from floors whereas only a single texture was kept to represent pools found on ortophotoplan images (cf. last row of tableI).

2) Test images: A set of six images is considered to evaluate the performance of the proposed roof extraction method. These images have been extracted from a huge original image like those described at the previous paragraph (cf. III-A1). To achieve that, the captured scenes should be varied in order to have a set of images that exhibit various conditions and increasing levels of difficulty. Having this in mind, the following criteria were selected: the number of roofs within the scene, their size and color, and finally the degree of discrepancies between roofs and the background. The first row of figure 4 shows four test images. The two other test images, which are enlarged, are shown in figure 5.

B. Accuracy assessment of the method

We begin by a qualitative evaluation of the proposed method using representative test images. Figure 4 illustrates the results of roof detection on the set of processed images. In the upper row of this figure, we show the original images; in the midst, the segmented images are given and in the lower row the corresponding building roof extraction where the final detected building boundaries drawn with red color are superimposed upon the original images. Basing on visual evaluation of the results, one can state that the developed approach demonstrates excellent accuracy in terms of building boundary extraction; this means that the majority of the building roofs present in the images are detected with good boundary delineation. Indeed, this method gives reliable results across complex environment composed of buildings presenting red and non-red rooftop, road areas, vegetation, etc. The images of figures 4.a, 4.b, 4.c and 4.d include several building rooftops and road areas with same color and texture, the proposed approach is able to successfully distinguish between them.

However, as one can see from the experimental results of figure 4, due to radiometric similarity between building roofs and image background, some false or imperfect detections can be generated. In fact, although we obtained notably accurate multiple detection of building roofs, the proposed method missed some part of buildings when the contrast between their rooftop and the background is low. Also, some vegetation areas are extracted as part of buildings because of their radiometric characteristics which are similar. In figure 5, some of building parts that have not been extracted are pointed out by yellow ellipses while some false detections are pointed out by green ellipses.

As for quantitative evaluation, we use measures widely employed in evaluating effectiveness. They constitute a useful and accepted tool in the object recognition field [49]. Within the orthophotoplans used in this work, 100 buildings were first manually delineated. Then, they are used as a reference building set to assess the accuracy of the automated building extraction. The extraction results and reference ones are compared pixel-by-pixel. Each pixel in the images is categorized as one of four possible outcomes:

1) True positive (TP): Both manual and automated methods label the pixel belonging to building.
2) True negative (TN): Both manual and automated methods label the pixel belonging to background.
3) False positive (FP): The automated method incorrectly labels the pixel as belonging to building.
4) False negative (FN): The automated method does not correctly label the pixel truly belonging to building.

To examine detection performance, the number of pixels that fall into each of the four categories TP, TN, FP, FN are determined, and the following measures are computed:

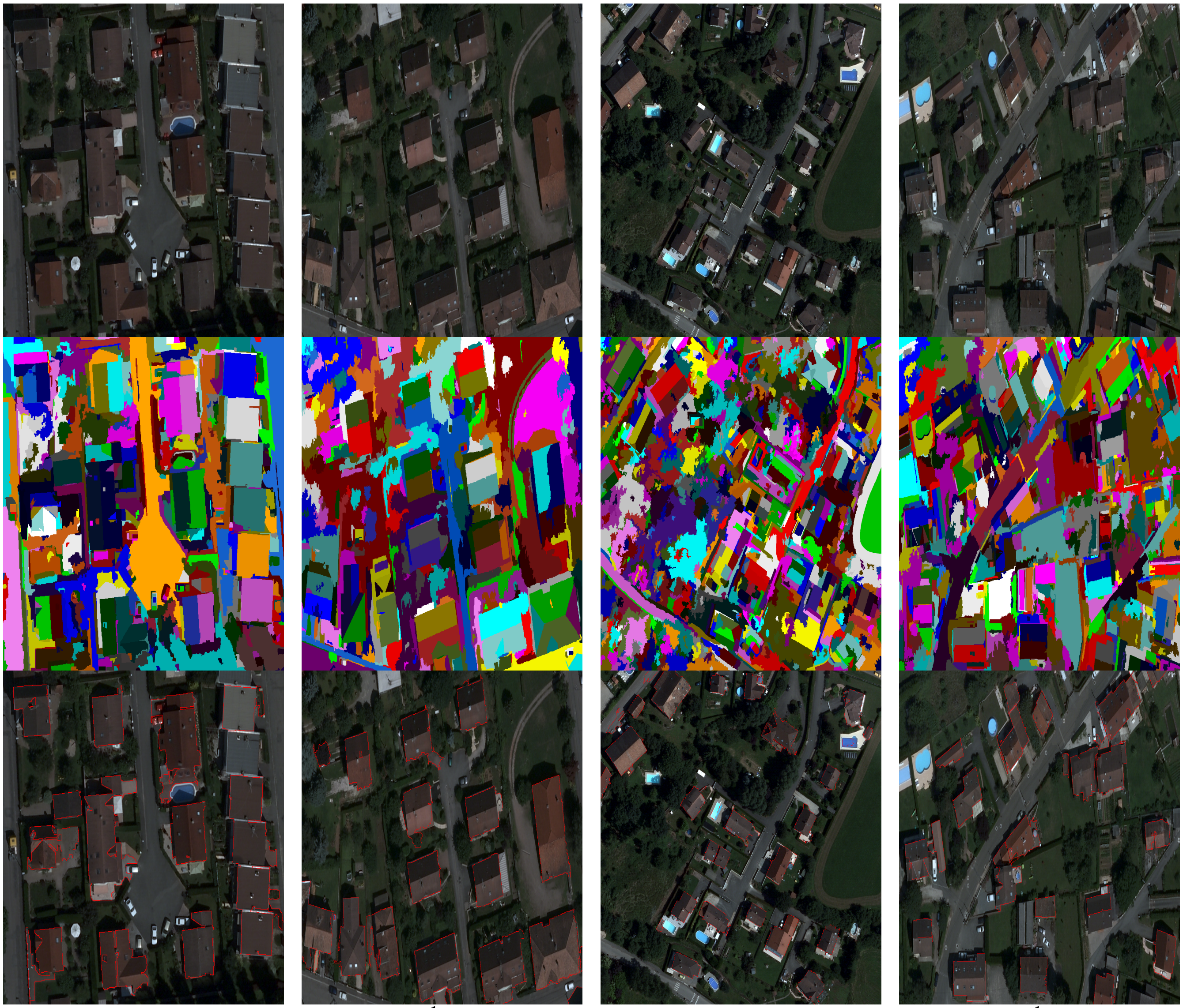
$$\begin{aligned}\text{Branching Factor (B.F)} &= \frac{FP}{TP}\\ \text{Miss Factor (M.F)} &= \frac{FN}{TP}\end{aligned}$$

Detection Percentage (D.P) = 100.TP$\frac{}{TP+TN}$

Quality Percentage (Q.P) = 100.TP$\frac{}{TP+FP+FN}$

The interpretation of the above measures is as follows. The detection percentage denotes the percentage of building pixels correctly labelled by the automated process. The branching factor is a measure of the commission error where the method incorrectly labels background pixels as building. The more accurate the detection is, the closer the value is to zero. The miss factor measures the omission error where the method incorrectly labels building pixels as background. These quality metrics are closely related to the boundary delineation performance of the building extraction method. The quality percentage in turn, measures the absolute quality of the extraction and is the most stringent measure. To obtain 100% quality, the extraction algorithm must correctly label every building pixel ($FN = 0$) without mislabelling any background pixel ($FP = 0$).

The results of the quality assessment of the method for the images, illustrated in figures 4 and 5, are given in table II. The last row of the table gives the average values obtained with all the orthophotoplans used in this work. The



a Image 1 | b Image 2 | c Image 3 | d Image 4

Fig. 4: Automatic extraction of multiple building roofs from the set of processed images. From top to bottom: original images, SRM segmentation results, and corresponding building roof extraction.

values obtained on the set of the processed images confirm the claims mentioned above regarding the performance of the proposed approach. Effectively, the results show that the building-extraction approach is quite successful for extracting the buildings from orthophotoplans with the D.P and Q.P average values of 93.91% and 85.30%, respectively. In addition to this, the branching factor and the miss factor average values were found to be 0.111 and 0.067 , respectively.

Also, we have transcript these comparison results in terms of Receiver Operating Characteristics (ROC) graphs [50]. In machine learning, ROC graphs are used as a useful technique for visualizing and selecting classifiers based on their performance. ROC graphs are two-dimensional graphs in which **True Positive Rate (TPR)** (also called *recall* or *sensitivity*) is plotted on the Y axis and **False Positive Rate (FPR)** (also called *false alarm rate*) is plotted on the X axis. These measures are computed using the four outcomes mentioned above as

TABLE II: The quality assessment results of the building extraction.

Images / measures	B.F	M.F	D.P	Q.P
Fig.4.a	0.1745	0.1116	90	77.75
Fig.4.b	0.14	0.0289	97.19	85.58
Fig.4.c	0.1804	0.0303	97.06	82.59
Fig.4.d	0.0871	0.1126	89.87	83.35
Fig.5.a	0.0184	0.0789	93.37	92.66
Fig.5.b	0.0710	0.0414	96.01	89.88
Avr./100 building roofs	**0.1118**	**0.0673**	**93.91**	**85.30**

a. Image 5

b. Image 6

Fig. 5: Examples of some building parts that have not been extracted (yello ellipses) and some false detections (green ellipses).

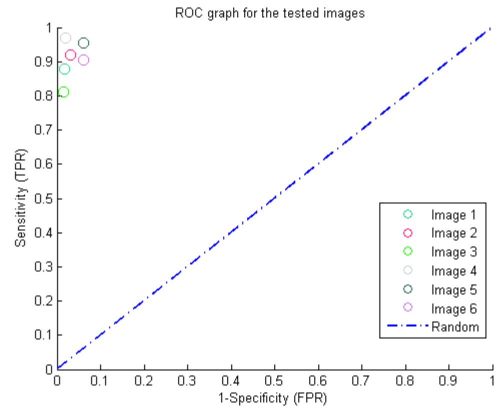


Fig. 6: The ROC graph of the test images using the proposed method.

follows:

$$TPR = \frac{TP}{TP + FN} \tag{6}$$

$$FPR = \frac{FP}{FP + TN} \tag{7}$$

As general rule in the context of discrete classifiers, the most important point in ROC space is the upper left corner, point of coordinate (0,1) which represents perfect classification. Informally, a point in ROC space is better than another if the first one is located into the north-west side of the second one(TPR is higher, FPR is lower, or both). Points above the diagonal dividing the ROC space represent good classification results (better than random), while points below the diagonal represent poor results (worse than random). In sum, the closer the ROC plot is located at the top-left border of ROC space, the more accurate are the results.

Figure 6 shows the accuracy of the proposed method applied on the test images. Basing on this ROC graph, all measures are on the top-left side, indicating hence perfect detection performance on this test set.

C. Comparative evaluation and discussion

To provide further evaluation of the performance of the developed method and considering that the proposed approach is mainly based on a simple supervised classification technique, we propose in this section to compare it with two popular supervised classification methods of the literature, within the same framework of building detection problem. The classification algorithms we have selected for this comparison are Support Vector Machine (SVM) and K Nearest Neighbours (KNN). These supervised classification methods are well known and often used for image classification purposes [14]. The two methods have several control parameters. For KNN, k nearest neighbors and distance are the key words in this algorithm. The principal parameters for SVM algorithm are the type of SVM, the type of kernel function and the degree in the kernel function. For furthermore details concerning these settings, we refer to the paper written by Chang and al. [51]. In this comparison with KNN and SVM classifiers, training samples are given by the color histograms of regions of the two knowledge databases whose labels are known and the test samples are given by the color histograms of regions of the test segmented images.

To obtain a meaningful comparison, each algorithm must be tested considering many possible combinations of input parameters. In effect, for each classification method, we consider its performance and correctness, as measured by its success rate calculated by the measures previously detailed, as well as its stability with respect to changes in parameter settings and with respect to all tested images.

The parameter setting that we have used here was set empirically through manual checking of the recognition results and is reported on table III.

Figure 7 illustrates an example of building extraction results obtained under different possible combinations of input parameters, as mentioned in Table III, for both KNN and SVM algorithms. The visual analysis shows that KNN_0 (using euclidean distance and k=1) and SVM_2 (using ν-SVC as type of SVM, sigmoid as type of kernel function and a degree in kernel function equal to 4) give good detection results. Note that, despite several combinations of parameters, neither of the two tested algorithms outperforms the proposed method. In fact, the approach is quite successful for extracting the buildings from the images (the extraction result match

TABLE III: Experimental parameter values related to the used methods (KNN and SVM) for comparison.

Methods	Parameters	Used Values			
		KNN_0	KNN_1	KNN_2	KNN_3
KNN	K:	1	3	1	3
	Distance:	Euclidian	Euclidian	Cityblock	Cityblock
		Used Values			
		SVM_0	SVM_1	SVM_2	SVM_3
SVM	Type of SVM:	C-SVC	ν-SVC	ν-SVC	ν-SVR
	Type of kernel function:	Radial basis function	Linear	Sigmoid	Polynomial
	Degree in kernel function:	3	4	4	1

TABLE IV: Quality assessment obtained for all the methods (KNN and SVM under different parameters, and the proposed method).

(%)	KNN				SVM				Our method
	KNN_0	KNN_1	KNN_2	KNN_3	SVM_0	SVM_1	SVM_2	SVM_3	
Q.P	42.34	47.48	41.63	29.96	58.79	68.53	70.11	66.33	85.59
D.P	73.69	67.71	69.58	30.86	64.12	78.51	81.27	78.54	97.19
B.F	1.00	0.63	0.97	0.09	0.14	0.19	0.20	0.23	0.14
M.F	0.36	0.48	0.44	2.24	0.56	0.27	0.23	0.27	0.02

most closely the ground truth) whereas the KNN and SVM algorithms lead to many false positives on road and vegetation areas and false negatives within buildings accompanied by a loss of several parts of roofs. The performance evaluation of the tested classification methods has been summarized in table IV and figure 8 resumes it with a graphical representation. The measures used to assess the quality of detection for this comparative analysis are the same as those used within the experiments addressed above to evaluate the proposed method separately (cf. section III-B).

Table IV shows the quantitative indicators based results obtained for each method. For KNN, we note that the optimal result in terms of quality detection (QD) is obtained using the second configuration KNN_1 with a value of 47.5%, while for the SVM, quality detection indicator reaches a value of 70% by the configuration SVM_2. The proposed method shows higher performance reaching 85.6% as a value for quality detection indicator. In relation to the percentage of detection (DP), the optimal values are 73.7% and 81.27% respectively for KNN and SVM against 97.19% for our method. To get an idea on missed parts in detection results, we rely to the omission factor (MF) which reaches 0.36 and 0.23 for KNN (KNN_0) and SVM (SVM_2) respectively. This measure attains 0.02 for the proposed method showing thus its superiority to the other methods in terms of the percentage of pixels that are not classified as building.

These results confirm the efficiency of the decision rule that the proposed method uses. For instance, unlike KNN method, which classifies an item by a majority vote of its neighbours, (i.e. the test item is assigned to the class most common among its k nearest neighbours taken from the training samples), considering instead the maximal average over the k similarity measures of regions from both object and background knowledge databases allows this new method to outperform KNN method.

By considering figures 9 and 10, one can see how ROC graphs show the out-performance of the developed method comparatively to the variants of KNN and SVM classifiers.

As for the computation time, It should be noted that the proposed method requires in average 5 seconds for extracting objects of interest (building roofs in the current application) from images of about 1500 by 1000 pixels working on a machine of 2.75 MHZ(CPU) and a memory of 3 GO (RAM). Besides, it depends on the number of regions of the segmented image.

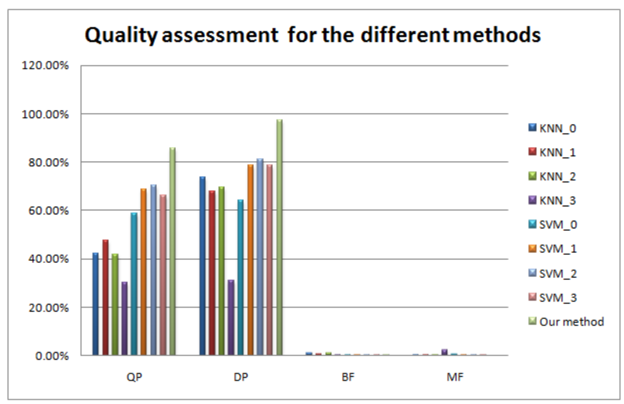


Fig. 8: Graphical representation of table IV.

IV. Conclusion and future works

In this paper, we have presented an efficient method for automatic and accurate multiple objects extraction from images. Unlike interactive methods, the proposed one requires no user interaction. The method involves two knowledge databases where the first one is constructed with several significant textures of objects to be extracted and the second one is composed with textures representing background. After an over-segmentation of the original image, the segmented regions are classified as objects or background using a region similarity measure and the constructed knowledge databases. The proposed method is evaluated for building roof extraction from orthophotoplans, which is a very challenging problem

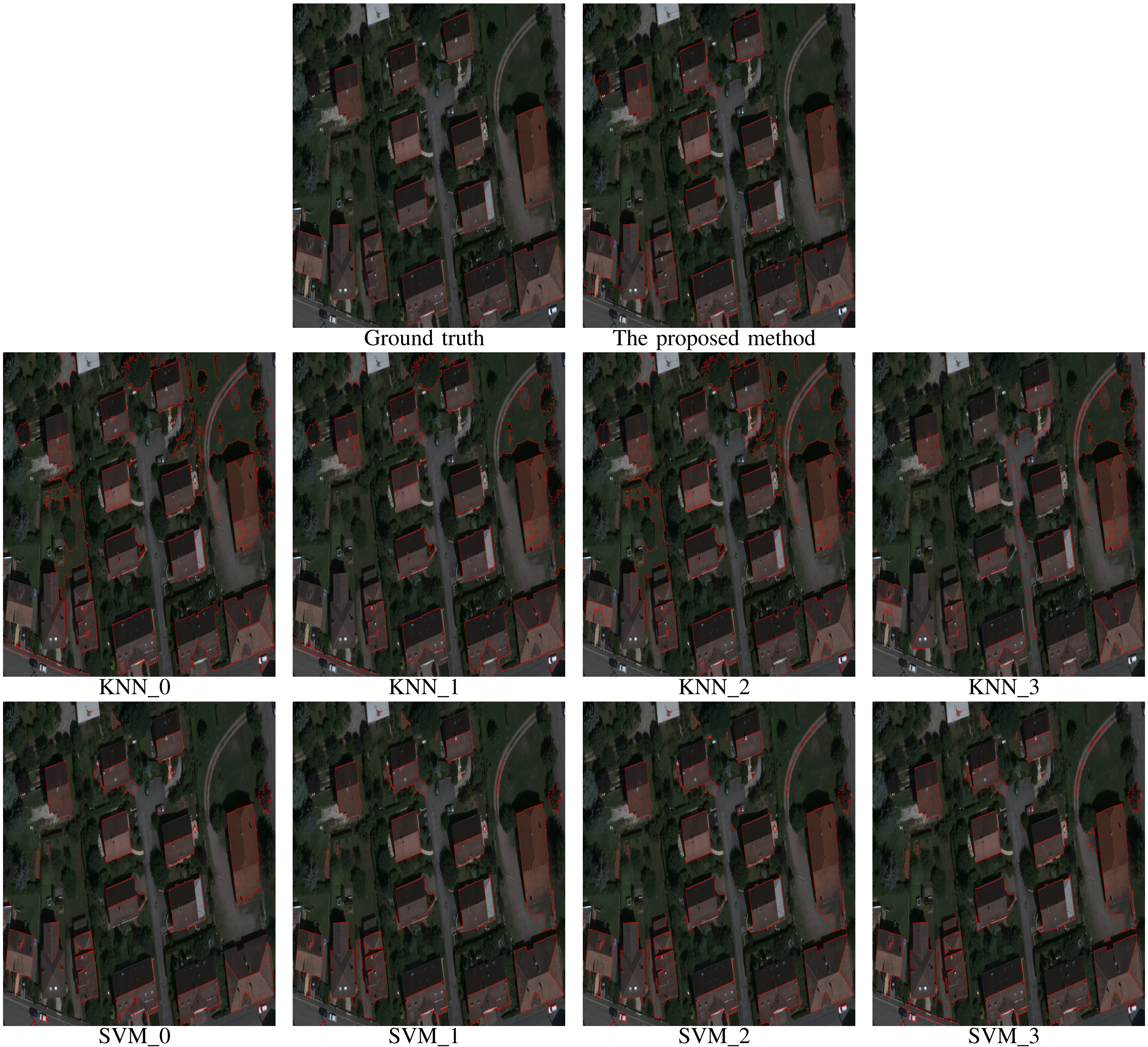


Fig. 7: Comparison between KNN, SVM and the proposed method. First row: the ground truth image and the roof extraction results using the proposed method; second row: extraction results by KNN method under different parameters (see table III); last row: extraction results by SVM method under different parameters (see table III).

because of the complexity of scenes with a large number of different objects (buildings, roads, vegetation, etc.). The evaluation consisted also of a comparison analysis between the proposed method and popular ones (KNN and SVM).

In order to improve the proposed method, there are several open questions that we still need to explore. First, the color histogram features are calculated using the RGB color space. The orthophotoplan images in our possession contain a certain heterogeneity in terms of lights, illumination changes, shadows, etc, what constitutes a breeding ground for false detections. To overcome these drawbacks and hence reduce the effect of illumination and limit the artefacts of the acquired image, studying and evaluating different color spaces and/or colorimetric invariants seems to be an interesting way forward [52], [53]. In addition, the proposed object-extraction method enables flexible feature descriptor integration. Thus, we propose to study the effect of other region characteristics on the quality of the results. One can cite the Local Binary Patterns (LBP) texture operator which is a powerful structural model of texture analysis [54]. Also, we think that it could be possible to estimate analytically the value of the parameter k involved in the similarity computation. To evaluate the genericity of the proposed method, we envisage to apply it for other image types (medical,...).

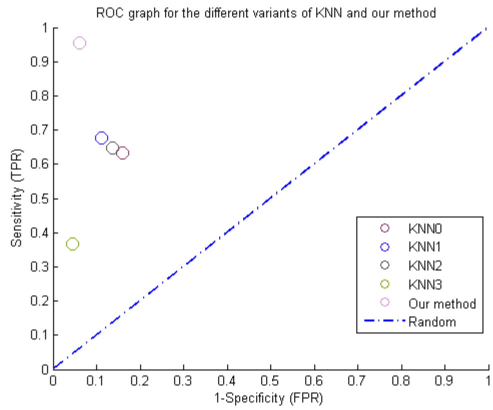


Fig. 9: The ROC graph comparing variants of KNN and the proposed method.

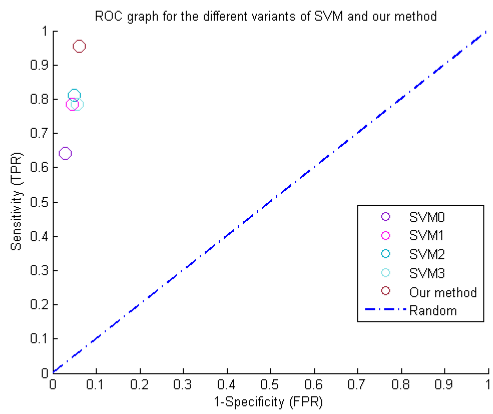


Fig. 10: The ROC graph comparing variants of SVM and the proposed method.

Acknowledgment

The authors would like to thank the "Communauté d'Agglomération Belfortaine" for making available the "orthophotoplan" images on which the tests have been performed.

References

[1] Pueyo, O.; Patow, G. Structuring urban data. *The Visual Computer*, **2014**, *30, Issue 2*, 159-172

[2] Sadgal,M.; El Fazziki,A.; Ait Ouahman, A. *Aerial image processing and object recognition*, *The Visual Computer* **2005** , *21, Issue 1-2*, 118-123

[3] Boykov,Y.; Jolly, M. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *IEEE Intl. Conf. on Comput. Vision* **2001**. *1*, 105-112

[4] Rother, C.; Kolmogorov, V.; Blake, A. GrabCut -Interactive Foreground Extraction using Iterated Graph Cuts. *ACM Transactions on Graphics* **2004**. *23(3)*, 309-314

[5] McGuinness, K.; O'Connor, N. E. A comparative evaluation of interactive segmentation algorithms. *Pattern Rec.* **2010**, *43*, 434-444

[6] Adams, R.; Bischof, L. Seeded region growing.*IEEE Trans. Pattern Anal. Mach. Intell.* **1994**, *16*, 641-647

[7] Friedland, G.; Jantz, K.; Rojas, R. SIOX: simple interactive object extraction in still images. *IEEE Intl. Symposium on Multimedia* **2005**. 253-259

[8] Adamek, T.: Using contour information and segmentation for object registration, modeling and retrieval. *PhD thesis*, Dublin City Univ **2006**

[9] Bai, X.; Sapiro,G. Geodesic matting: a framework for fast interactive image and video segmentation and matting.*Intl. J. Comput.* **2009**,*82*, 113-132

[10] Yang,C.; Duraiswami, R.; Gumerov, N. A.; Davis, L.: Improved fast gauss transform and efficient kernel density estimation. *IEEE Intl. Conf. on Comput. Vision* **2003**. *1*, 664-671

[11] Protiere, A.; Sapiro, G.: Interactive image segmentation via adaptive weighted distances. *IEEE Transactions on Image Processing* **2007**.*16*, 1046-1057

[12] Ninga, J.; Zhanga, L.; Wub, D. Z. C.: Interactive image segmentation by maximal similarity based region merging.*Pattern Recognition* **2010**.*43* , n 2. 445-456

[13] Chen, M.; Hou, Y.; Zeng, X.; Lu,Y.: A multi-label interactive image segmentation method based on region merging. *IEEE International Conference on Mechatronic Science, Electric Engineering and Computer* **2011**, Jilin, China. 2336-2340

[14] Tso, B.; Mather, P.M. : Classification Methods for Remotely Sensed Data, 2nd ed. ; *CRC Press: Boca Raton*, FL, USA **2009**

[15] Giorgos, M.; Jungho, I.; Caesar, O. : Support vector machines in remote sensing: A review.*ISPRS Journal of Photogrammetry and Remote Sensing* **2011**, *66*, 247-259

[16] Knerr, S.; Personnaz, L.; Dreyfus, G. : Single-layer learning revisited: a stepwise procedure for building and training a neural network. *In: Neurocomputing: Algorithms, Architectures and Applications. In: NATO ASI Series* , Springer, **1990**

[17] Bischof, H.; Schneider, W.; Pinz, A.J. : Multispectral classification of Landsat-images using neural networks. *IEEE Trans. Geosci. Remote Sens.* **1992**, *30*, 482-490

[18] Teemu, K.; J. Kristian, K.; Lasse, L.; Heikki, K. : Unsupervised object discovery via self-organisation. *Pattern Recognition Letters* **2012**,*33*, 2102-2112

[19] Muralidharan,R.; Chandrasekar,C. : Object Recognition using SVM-KNN based on Geometric Moment Invariant. *International Journal of Computer Trends and Technology* **2011**, 215-220

[20] Hu, M.K. : Visual pattern recognition by moments invariants. *IRE Trans. Information Theory* **1962**, *8*, 179-87

[21] Dudani, S.; Breeding, K.; McGhee, R.: Aircraft identification by moments invariants. *IEEE Transactions on Computers* **1977**, *26(1)*, 39-45

[22] Flusser, J.; Suk, T.: A moment based approach to registration of image with affine geometric distortion. *IEEE Transactions on Geoscience and Remote Sensing* **1994**, *32(2)*,382-387

[23] Flusser, J. On the independence of rotation moment invariants. *Pattern Recognition* **2000**, *33(9)*,1405-1410

[24] Soille, P.; Pesaresi, M. : Advances in mathematical morphology applied to geoscience and remote sensing. *IEEE Transactions on Geosciences and Remote Sensing* **2006**, *40*, 2042-2055

[25] Giada, S.; De Groeve, T.; Ehrlich, D.; Soille, P. Information extraction from very high resolution satellite imagery over Lukole refugee camp, Tanzania. *Int. J. Remote Sensing* **2002** , *24* , N 22, 4251-4266

[26] Salman, A.; , Valadan, M.J.Z.; Hamid, E.;Hamid, A. M.; Ali, M. : Automatic urban building boundary extraction from high resolution aerial images using an innovative model of active contours. *International Journal of Applied Earth Observation and Geoinformation* **2010**, *12*, 150-157

[27] Kass, M.; Witkin, A.; Terzopoulos, D. Snakes: active contour models. *International Journal of Computer Vision* **1998**, *1*, 321-331

[28] Halla, Brenner, N.C. Extraction of building and trees in urban environment. *ISPRS Journal of Photogrammetry & Remote Sensing* **1999**, *54*, 130-137

[29] Zhao, B.; Trinder, J.C. Integrated-approach-based automatic building extraction. *International Archives of Photogrammetry and Remote Sensing XXXIII* **2000**, *Part B3*

[30] Samadzadegan, F.; Azizi, A.; Hahn, M.; Lucas, C. Automatic 3D object recognition and reconstruction based on neuro-fuzzy modeling. *ISPRS Journal of Photogrammetry & Remote Sensing* **2005**, *59*, 255-277

[31] Zimmermann, P. A new framework for automatic building detection

analyzing multiple cue data. *International Archives of Photogrammetry and Remote Sensing XXXIII* **2000**, *Part B3*

[32] Miliaresis, G.; Kokkas, N. : Segmentation and object-based classification for the extraction of the building class from LIDARDEMs. *Computers & Geosciences* **2007**, *33*, 1076-1087

[33] Lafarge, F.; Descombes, X.; Zerubia, J.; Pierrot-Deseilligny, M. Automatic building extraction from DEMs using an object approach and application to the 3D-city modeling. *ISPRS Journal of Photogrammetry & Remote Sensing* **2008**,*63*, 365-381

[34] Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **2004**, *60(2)*, 91-110

[35] Berretti, S.; Ben Amor, B.; Daoudi,M.; Del-Bimbo, A. 3D facial expression recognition using SIFT descriptors of automatically detected keypoints. *The Visual Computer* **2011**, *27, Issue 11*, 1021-1036.

[36] Berg,A.;Berg, T.; Malik, J. Shape matching and object recognition using low distortion correspondence. *in: Proc. International Conference on Computer Vision and, Pattern Recognition* **2005**,26-33

[37] Ying-Ho, L.; Anthony, J.T.L.; Fu, C. : Object recognition using discriminative parts.*Computer Vision and Image Understanding* **2012**, *116*, 854-867

[38] Lazebnik, S.; Schmid, C.; Ponce, J. Beyond bags of features: spatial pyramid matching for recognizing natural scene categories, in: *Proc. International Conference on Computer Vision and, Pattern Recognition*, 2169-2178 **2006**

[39] Amores,J.; Sebe, N.; Radeva, P. Fast spatial pattern discovery integrating boosting with constellations of contextual descriptors. in: *Proc. International Conference on Computer Vision and Pattern Recognition* **2005**, *2*,769-774

[40] Nock,R.; Nielsen, F.: Statistical region merging. *IEEE Trans. Pattern Anal. Machine Intelligence* **2004**. *26*, n 11, 1452-1458

[41] Deng, Y.; Manjunath, B.S.: Unsupervised Segmentation of Color-Texture Regions in Images and video. *IEEE Trans. Pattern Anal. Mach. Intell.* **2000**, *23*, 800-810

[42] Levinshtein, A.; Stere, A.; Kutulakos, K. N.; Fleet, D. J.; Dickinson, S. J.; Siddiqi, K.: Turbo Pixels: Fast Super pixels Using Geometric Flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2009**. *31*, 2290-2297

[43] Birchfield,S.: Elliptical head tracking using intensity gradients and color histograms.*IEEE Conference on Computer Vision and Pattern Recognition* **1998**. 232-237

[44] Ojala, T.; Pietikainen, M.; Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Transactions on Pattern Analysis and Machine Intelligence* **2002**.*24*,971-987

[45] Swain, M.J.; Ballard, D.H.: Color indexing.*International Journal of Computer Vision* **2002**. *7*, 11-32

[46] Stricker, M.; Orengo, M.: Similarity of color images. *SPIE Storage and Retrieval for Image and Video Databases* **1995**. *3*, 381-392

[47] Kailath, T.: The divergence and Bhattacharyya distance measures in signal selection. *IEEE Transactions on Communications Technology* **1967** . *15*,52-60

[48] Fukunaga, K.: Introduction to Statistical Pattern Recognition. *Second ed., Academic Press* **1990**

[49] Shufelt, J. A. Performance evaluation and analysis of monocular building extraction from aerial imagery. *IEEE Trans. Pattern Anal. Machine Intell.* **1999**,*21*,311-326

[50] Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters*, **2006**, *27* , 861874.

[51] Chih-Chung, C.; Chih-Jen, L. : LIBSVM : a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology* **2011**. 2:27:1–27:27 . Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

[52] El-Merabet, Y.; Meurie, C.; Ruichek, Y.; Sbihi, A.; Touahni, R.: Orthophotoplan segmentation based on regions merging for roof detection. *S&T/SPIE Electronic Imaging 2012 - Image Processing: Machine Vision Applications V*, SPIE 8300, Burlingame USA

[53] El-Merabet, Y.; Meurie, C.; Ruichek, Y.; Sbihi, A.; Touahni, R.: Orthophotoplans segmentation and colorimetric invariants for roof detection. *LNCS, 16th International Conference on Image Analysis and Processing (ICIAP)*, Ravenna Italy, 394-403 **2011**

[54] Abdou-Berbar, M. Three robust features extraction approaches for facial gender classification. *The Visual Computer*, **2014**, *30, Issue 1*, 19-31