IJACSA
WHERE WISDOM SHARES
International Journal of Advanced Computer Science and Applications

Special Issue

Artificial Intelligence

ISSN 2156-5570(Online)
ISSN 2158-107X(Print)

SAI
www.ijacsa.thesai.org

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS

A Publication of
The Science and Information Organization

International Journal of Advanced Computer Science and Applications

Special Issue on Artificial Intelligence

Scope of this Issue:

The range of topics covered by Special Issue on Artificial Intelligence includes the following areas:

- ✓ ***Artificial Neural Networks***
- ✓ ***Fuzzy Logic Theory***
- ✓ ***Neuro-fuzzy systems***
- ✓ ***Evolutionary Algorithms***
- ✓ ***Semantic Networks***
- ✓ ***Artificial Life***
- ✓ ***Expert Systems***
- ✓ ***Computer Vision***
- ✓ ***Machine Learning***
- ✓ ***Signal/Image processing***
- ✓ ***Data retrieval***
- ✓ ***Data mining***
- ✓ ***Fusion techniques***
- ✓ ***Optimization***
- ✓ ***Texture Analysis***
- ✓ ***Hybrid Applications***

IJACSA Special Issue Guest Editor

Dr. Taiwo Ayodele
CEO,
Infonetmedia

Dr. Taiwo Ayodele received his PhD degrees from the Department of Electronics and Computer Engineering, University of Portsmouth, United Kingdom.

He is currently the CEO of Infonetmedia, UK as well as Research Collaborator for Infonomics Society, UK.

He has many years of Teaching and Research Experience and his research background is mostly in the areas of Artificial Intelligence, Knowledge Management, Machine Learning, Neural Networks, E-health, E-Learning & E-Learning Security, Email Management and has a deep interest in planning and executing major IT projects.

Dr Taiwo has published several peer-reviewed technical papers in Journals/Conferences/Book Chapters. Dr. Taiwo also serves as a Chair/Organiser of Workshop on Application of Artificial Intelligence for Email Management, a Reviewer for International Conference on World Congress on Internet Security (WorldCIS), Digital Information Management (ICDIM), Journals of Engineering and Technology Research (JETR) and several other well known journals and conferences.

He also has been awarded Acknowledgement of Contribution to i-Society, 2010 by Infononomic Society. He also has strong industry experiences, having worked in numerous high-tech companies, including Telefonica I + D, Madrid, Spain.

His research interests are in data mining, artificial intelligence, mobile and wireless computing and web-based applications. He has published more than 100 papers in various journals and conferences at national and international level. He is a member of IEEE, IEEE Computer Society, IAEng, and IACSIT

IJACSA Editorial

From the Desk of Managing Editor…

It is a pleasure to present our readers with the Special Issue on Artificial Intelligence of International Journal of Advanced Computer Science and Applications (IJACSA). What is particularly attractive and significant in the present issue is the range of applications that is covered. It is a timely and refreshing addition to the knowledge base in the practical application of Artificial Intelligence.

Artificial intelligence in the future will churn out machines and computers, which are much more sophisticated than the ones that we have today. For example, the speech recognition systems that we see today will become more sophisticated and it is expected that they will reach the human performance levels in the future. It is also believed that they will be able to communicate with human beings, using both text and voice, in unstructured English in the coming few years.

The development of meaningful artificial intelligence will require that machines acquire some variant of human consciousness. Systems that are able to demonstrate conclusively that they possess self-awareness, language skills, surface, shallow and deep knowledge about the world around them and their role within it will be needed going forward. However the field of artificial consciousness remains in its infancy. The early years of the 21st century should see dramatic strides forward in this area however.

By breaking up AI research into more specific problems, such as computer vision, speech recognition and automatic planning, which had more clearly definable goals, scientists managed to create a critical mass of work aimed at solving these individual problems.

The journal's Special Issue on Artificial Intelligence reports results achieved; proposals for new ways of looking at AI problems which includes demonstrations of effectiveness. There is a mixture of foundational and applied papers describing mature work involving computational accounts of aspects of intelligence.

On behalf of the Journal we wish to extend our sincere thanks to our Guest Editors for his precious time and hard work.

We hope to continue exploring the always diverse and often astonishing fields in Advanced Computer Science and Applications.

Thank You for Sharing Wisdom!

Managing Editor
IJACSA
Special Issue on Artificial Intelligence
05 September 2011
editorijacsa@thesai.org
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)

CONTENTS

Parts of Speech Tagging for Afaan Oromo

Getachew Mamo Wegari
Information Technology Department
Jimma Institute of Technology
Jimma, Ethiopia

Million Meshesha (PhD)
Information Science Department
Addis Ababa University
Jimma, Ethiopia

***Abstract*—The main aim of this study is to develop part-of-speech tagger for Afaan Oromo language. After reviewing literatures on Afaan Oromo grammars and identifying tagset and word categories, the study adopted Hidden Markov Model (HMM) approach and has implemented unigram and bigram models of Viterbi algorithm. Unigram model is used to understand word ambiguity in the language, while bigram model is used to undertake contextual analysis of words.**

For training and testing purpose 159 sentences (with a total of 1621 words) that are manually annotated sample corpus are used. The corpus is collected from different public Afaan Oromo newspapers and bulletins to make the sample corpus balanced. A database of lexical probabilities and transitional probabilities are developed from the annotated corpus. These two probabilities are from which the tagger learn and tag sequence of words in sentences.

The performance of the prototype, Afaan Oromo tagger is tested using tenfold cross validation mechanism. The result shows that in both unigram and bigram models 87.58% and 91.97% accuracy is obtained, respectively.

Keywords-Natural Language processing; parts of speech tagging; Hidden Markov Model; N-Gram; Afaan Oromo.

I. INTRODUCTION

At the heart of any natural language processing (NLP) task, there is the issue of natural language understanding. However, the process of building computer programs that understand natural language is not straightforward. As explained in [1], natural languages give rise to lexical ambiguity that words may have different meanings, i.e. one word is in general connected with different readings in the lexicon. Homograph, the phenomenon that certain words showing different morpho-syntatic behavior are identically written. For instance, the word 'Bank' has different meanings; Bank (= financial institute), Bank (= seating accommodation), etc.

In other words, words match more than one lexical category depending on the context that they appear in sentences. For example, if we consider the word miilaa 'leg' in the following two sentences,

Lataan kubbaa miilaa xabata. 'Lata plays football'.

Lataan miilaa eeraa qaba. 'Lata has long leg'.

In the first sentence, miilaa 'leg' takes the position of adjective to describe the noun kubbaa 'ball'. But in the second sentence, miilaa is a noun described by eeraa 'long'.

Besides ambiguity of words, inflection and derivation of the language are other reasons that make natural language understanding very complex. For instance, tapha 'play' contains the following inflection in Afaan Oromo language.

tapha-t ' she plays'

tapha-ta 'he plays'

tapha-tu 'they play'

tapha-ta-niiru ' they played'

tapha-chuu-fi 'they will play'

In the above particular context suffixes are added to show gender {–t, --ta}, number { –tu/--u} and future {--fi}.

To handle such complexities and use computers to understand and manipulate natural language text and speech, there are various research attempts under investigation. Some of these include machine translation, information extraction and retrieval using natural language, text to speech synthesis, automatic written text recognition, grammar checking, and part-of-speech tagging. Most of these approaches have been developed for popular languages like English [3]. However, there are few studies for Afaan Oromo language. So, the study presents the investigation of designing and developing an automatic part-of-speech tagger for Afaan Oromo language.

II. PART-OF-SPEECH TAGGING

Part-of-speech (POS) tagging is the act of assigning each word in sentences a tag that describes how that word is used in the sentences. That means POS tagging assigns whether a given word is used as a noun, adjective, verb, etc. As Pla and Molina [4] notes, one of the most well-known disambiguation problems is POS tagging. A POS tagger attempts to assign the corresponding POS tag to each word in sentences, taking into account the context in which this word appears.

For example, the following is tagged sentence in Afaan Oromo Language.

Leenseen\NN kaleessa\AD deemte\VV 'Lense went yesterday'.

In the above example, words in the sentence, Leensaan kaleessa deemte, are tagged with appropriate lexical categories of noun, adverb and verb respectively. The codes NN, AD, VV are tags for noun, adverb and verb respectively. The process of tagging takes a sentence as input, assigns a POS tag to the word

or to each word in a sentence or in a corpus, and produces the tagged text as output.

There are two efficient approaches that have been established to develop part-speech-tagger [14].

A. Rule based Approach

Rule based taggers use hand coded rules to determine the lexical categories of a word [2, 13]. Words are tagged based on the contextual information around a word that is going to be tagged. Part-of-speech distributions and statistics for each word can be derived from annotated corpora - dictionaries. Dictionaries provide a list of word with their lexical meanings. In dictionaries there are many citations of examples that describe a word in different context. These contextual citations provide information that is used as a clue to develop a rule and determine lexical categories of the word.

In English language, for instance, a rule changes the tag from modal to noun if the previous word is an article. And the rule is applied to a sentence, the/art can/noun rusted/verb. Brill's rules tagger conforms to a limited number of transformation types, called templates. For example, the rule changes the tag from modal to noun if the previous word is an article, corresponds to template. The following table shows sample template that is used in Brill's rule tagger [2].

TABLE I. SAMPLE TEMPLETE BRILL'S RULE

Rules	Explanation
alter(A, B, prevtag(C))	Change A to B if preceding tag is C
alter(A, B, nexttag(C))	Change A to B if the following tag is C

Where, A, B and C represent lexical categories or part-of-speech.

B. Stochastic Approach

Most current part-of-speech taggers are probabilistic (stochastic). It is preferred to tag for a word by calculating the most likely tag in the context of the word and its immediate neighbors [15, 16]. The intuition behind all stochastic taggers is a simple generalization of the 'pick the most-likely tag for this word' approach based on the Bayesian framework. A stochastic approach includes most frequent tag, n – gram and Hidden Markov Model [13].

HMM is the statistical model which is mostly used in POS tagging. The general idea is that, if we have a sequence of words, each with one or more potential tags, then we can choose the most likely sequence of tags by calculating the probability of all possible sequences of tags, and then choosing the sequence with the highest probability [17]. We can directly observe the sequence of words, but we can only estimate the sequence of tags, which is 'hidden' from the observer of the text. A HMM enables us to estimate the most likely sequence of tags, making use of observed frequencies of words and tags (in a training corpus) [14].

The probability of a tag sequence is generally a function of:

- the probability that one tag follows another (n-gram); for example, after a determiner tag an adjective tag or a noun tag is quite likely, but a verb tag is less likely. So in a sentence beginning with the run…, the word 'run' is more likely to be a noun than a verb base form.
- The probability of a word being assigned a particular tag from the list of all possible tags (most frequent tag); for example, the word 'over' could be a common noun in certain restricted contexts, but generally a preposition tag would be overwhelmingly the more likely one.

So, for a given sentence or word sequence, HMM taggers choose the tag sequence that maximizes the following formula [14]:

P(word/tag) * P(tag/previous n tags)

Most frequent tag (likelihood)

N-gram (a prior)

III. AFAAN OROMO

Afaan Oromo is one of the major languages that is widely spoken and used in Ethiopia [6]. Currently it is an official language of Oromia state. It is used by Oromo people, who are the largest ethnic group in Ethiopia, which amounts to 34.5% of the total population according to the 2008 census [19].

With regard to the writing system, since 1991 Qubee (Latin-based alphabet) has been adopted and become the official script of Afaan Oromo [12]. Currently, Afaan Oromo is widely used as both written and spoken language in Ethiopia. Besides being an official working language of Oromia State, Afaan Oromo is the instructional medium for primary and junior secondary schools throughout the region and its administrative zones. It is also given as the department in five universities in Ethiopia. Thus, the language has well established and standardized writing and spoken system [7].

IV. RELATED RESEARCHES

To use computers for understanding and manipulation of Afaan Oromo language, there are very few researches attempted. These attempts include text-to-speech system for Afaan Oromo [8], an automatic sentence parser for Oromo Language [9] and developing morphological analyzer for Afaan Oromo text [10].

There are also other related researches that were conducted on other local language. Specially on Amharic language, two researches were conducted on POS tagging by [5] and [11], but to the best of our knowledge there is no POS tagging research conducted for Afaan Oromo language.

V. APPLICAION OF THE STUDY

The output of POS tagger has many applications in many natural language processing activities [4]. Morpho-syntactic disambiguation is used as preprocessor in NLP systems. Thus,

the use of a POS tagger simplifies the task of syntactic or semantic parsers because they do not have to manage ambiguous morphological sentences. Thus parsing cannot proceed in the absence of lexical analysis, and so it is necessary to first identify and determine part-of-speech of words.

It can also be incorporated in NLP systems that have to deal with unrestricted text, such as information extraction, information retrieval, and machine translation. In this modern world, huge amount of information are available on the Internet in different languages of the world. To access such information we need machine translator to translate into local languages. To develop a machine translation system, the lexical categories of the source and target languages should be analyzed first since a translator translates, for example, nouns of the source language to the nouns of the target language. So, POS tagger is one of the key inputs in machine translation processes.

A word's part-of-speech can further tell us about how the word is pronounced. For instance, the word 'content' in English can be a noun or an adjective. It is pronounced as 'CONtent' and 'conTENT' respectively. Thus, knowing part-of-speech can produce more natural pronunciations in a speech synthesis system and more accuracy in a speech recognition system [8].

All these applications can benefit from POS tagger to improve their performance in both accuracy and computational efficiency.

VI. METHODOLOGY

A. Algorithm Design and Implementation

HMM approach is adopted for the study since it does not need detail linguistic knowledge of the language as rule based approach [14]. Viterbi algorithm is used for implementing the tagger.

The Viterbi algorithm is a dynamic programming algorithm that optimizes the tagging of a sequence, making the tagging much more efficient in both time and memory consumption. In a naïve implementation it would calculate the probability of every possible path through the sequence of possible word-tag pairs, and then select the one with the highest probability. Since the number of possible paths through a sequence with a lot of ambiguities can be quite large, this will consume a lot more memory and time than necessary [18].

Since the path with highest probability will be a path that only includes optimal sub paths, there is no need to keep sub paths that are not optimal. Thus, the Viterbi algorithm only keeps the optimal sub path of each node at each position in the sequence, discarding the others.

B. Test and Evaluation

The prototype tagger is tested based on the sample test data prepared for this purpose. The performance evaluation is analyzed based on correctly tagged once by the prototype tagger.

The performance analysis is using tenfold cross validation. Ten fold cross validation divides a given corpus in to ten folds. And nine folds are used for training and the tenth fold is used for testing. It provides an unbiased estimate of value of prediction error and preferred for small sample corpus [20].

VII. AFAAN OROMO TAGSET AND CORPUS

A. Afaan Oromo Tagsets

Since there is no tagset prepared for natural language processing purpose for Afaan Oromo language, seventeen tags have been identified for the study as indicated in Table II.

TABLE II. TAGSETS

Tags	Description
NN	A tag for all types of nouns that are not joined with other categories in sentences.
NP	A tag for all nouns that are not separated from postpositions.
NC	A tag for all nouns that are not separated from conjunctions.
PP	A tag for all pronouns that are not joined with other categories.
PS	A tag for all pronouns that are not separated from postpositions.
PC	A tag for all pronouns that are not separated from conjunctions.
VV	A tag for all main verbs in sentences.
AX	A tag for all auxiliary verbs.
JJ	A tag for all adjectives that are separated from other categories.
JC	A tag for adjectives that are not separated from conjunction.
JN	A tag for numeral adjectives.
AD	A tag for all types of adverbs in the language.
PR	A tag for all preposition/postposition that are separated from other categories.
ON	A tag for ordinary numerals.
CC	A tag for all conjunctions that are separated from other categories.
II	A tag for all introjections in the language.
PN	A tag for all punctuations in the language.

B. Corpus

The collected corpus for the study was manually tagged by experts of linguists in the field. The tagging process is based on the identified tagset and corpus that is manually tagged, considering contextual position of words in a sentence. This tagged corpus is used for training the tagger and evaluates its performance. The total tagged corpus consists of 159 sentences (the total of 1621 tokens).

VIII. THE LEXICON

Lexicon was prepared from which the two probabilities are developed for the analysis of the data set.

TABLE III. SAMPLE OF LEXCON

words	NN…	PP…	VV…	JJ…	AD…	Total
nama	2	0	0	1	0	3
Yeroo	0	0	0	0	9	9
.	.	.	.	.	.	.
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Total	334	100	351	226	81	1621

A. Lexicon probability

The lexical probabilities have been estimated by computing the relative frequencies of every word per category from the training annotated corpus. All statistical information, that enables to develop probabilities, are derived automatically from a hand annotated corpus (the lexicon).

For instance, the lexical probability of the word *Oromoon* tagged with *NN* is calculated as:

C(Oromoon, NN) = 7

C(NN) = 334

So, P(Oromoon/NN) = C(Oromoon, NN)/C(NN)

= 7/334

= 0.0206

Where, C and P are count of and Probability, respectively.

TABLE IV. SAMPLE LEXICAL PROBABILITY

Words with given lexical probability	Probability
P(Oromoon/NN)	0.0206
P(jedhaman/VV)	0.0052
P(kabajaa/AD)	0.02174
P(ayyaanichaafi/NC)	0.11111
P(amma/AD)	0.04348
P(yeroo/AD)	0.10869

B. Transition Probability

In transitional probabilities, the information of one part-of-speech category preceded by other categories is developed from training lexicon corpus. For this study, bigram is used. Bigram considers the information of the category (t-1) preceded the target category (t).

That means, P(t/t-1), where t is – part-of-speech category.

For example, C($S) = 157

C(NN,$S) = 79

P(NN/$S) = C(NN, $S)/C($S)

= 79/157

= 0.5032

TABLE V. SAMPLE TRANSITION PROBABILITY

Bigram Category	Probability
P(NN/$S)	0.5032
P(VV/$S)	0.0063
P(NN/VV)	0.1538
P(NN/PN)	0.0063
P(JJ/NN)	0.2695
P(JJ/$S)	0.1465
P(PP/NN)	0.1018

IX. AFAAN OROMO PARTS OF SPEECH TAGGER

The tagger learns from the two probabilities to label appropriate tag to each word in sentences. The tagger for the study is developed from Viterbi algorithm of hidden Markov model.

A. Performance Analysis of the tagger

TABLE VI. AVERAGE TAGGER RESULTS

Unigram	Bigram
87.58%	91.97%

In the performance analysis, the tagger is repeatedly trained and tested following tenfold cross validation.

The algorithms of the tagger are tested with a corpus of 146 Afaan Oromo words in average in each test set and that is trained on the training set of 1315 words, and the result of each test are compared with a copy of the test set that is hand annotated. As a result, the results of the experiments for both bigram and unigram algorithms show an accuracy of 91.97% and 87.5% correctly tagged words in average respectively.

With this corpus, the distributions of accuracy performance in both models are not as far from each other. The maximum variation in the distribution of bigram and unigram models is 8.97 and 11.04 respectively. If the corpus is standardized, this variation will reduce since standardized corpus consist relatively complete representative of words for the language and fair distribution of words in training set and test are observed.

In bigram model, the statistical accuracy is performed more than unigram model. Bigram model uses probability of contextual information besides the highest probability of categories given a word in a sentence to tag the word. The difference accuracy rate from bigram to unigram is 4.39% with this dataset.

This indicates, contextual information (the position in which the word appear in sentence) affects the determination of word categories for Afaan Oromo language.

ACKNOWLEDGMENT

The support of Jimma University, Ethiopia, is greatly acknowledged. The authors would also like to acknowledge the help and support from Tigist Mazgabu, Dr. Mengesha Mamo and Mr. Fituma Tefera.

REFERENCES

[1] Hermann Helbig. Knowledge representation and the semantics of natural language. Springer-Verlg Berlin Heidelberg, Germany, 2006.

[2] James Allen. Natural language Understanding. The Benjamin/Cummings Publishing company, Redwood City, Canada, 1995

[3] Gobinda G. Chowdhury. Natural Language Processing: Department of Computer and Information Sciences, University of Strathclyde, Glasgow G1 1XH, UK, http://www.cis.strath.ac.uk/cis/research/publications/papers/strathcis_publication_320.pdf

[4] Ferran Pla and Antonio Molina. Natural Language Engineering: Improving part of- speech tagging using lexicalized HMMs_ 2004. Cambridge University Press, United Kingdom, 2004

[5] Mesfin Getachew. Automatic part-of-speech tagging for Amharic language an experiment using stochastic Hidden Markov Approach. MSc. Thesis. School of Graduate Studies, Addis Ababa University, 2001.

[6] Abara Nefa. Long Vowels in Afaan Oromo: A Generative Approach. M.A. Thesis. School of Graduate Studies, Addis Ababa University, 1988. Unpublished.

[7] Kula K. T., Vasudeva Varma and Prasad Pingali. Evaluation of Oromo-English Cross-Language Information Retrieval. In IJCAI 2007 Workshop on CLIA, Hyderabad (India), 2007.

[8] Morka Mekonnen. Text to speech system for Afaan Oromo. MSc Thesis. School of Graduate Studies, Addis Ababa University, 2001.Unpublished

[9] Diriba Magarsa. An automatic sentence parser for Oromo language. MSc Thesis. School of Graduate Studies, Addis Ababa University, 2000. Unpublished

[10] Assefa W/Mariam. Developing morphological analysis for Afaan Oromo text. MSc Thesis. School of Graduate Studies, Addis Ababa University, 2000. Unpublished

[11] Yenewondim Biadgo. Application of multilayer perception neural network for tagging part-of-speech for Amharic language. MSc Thesis. School of Graduate Studies, Addis Ababa University, 2005. Unpublished

[12] Gumii Qormaata Afaan Oromo. Caasluga Afaan Oromo. Komoshinii Aadaafi Tuurizimii Oromiyaa, 1996. Unpublished

[13] Pierre M. Nugues. An Introduction to Language Processing with Perl and Prolog. Springer-Verlag Berlin Heidelberg, Germany, 2006

[14] Daniel Jurafsky and James H. Martin. Speech and Language Processing: An Introduction to Natural Speech Recognition. Prentice-Hall, Inc., 2000.

[15] Sandipan Dand, Sudeshna Sarkar, Anupam Basu. Automatic Part-of-Speech Tagging for Bengali: An Approach for Morphologically Rich Languages in a Poor Resource Scenario. Department of Computer Science and Engineering Indian Institute of Technology Kharagpur, 2007

[16] Frank Van Eynde. Part-of-speech Tagging and Lemmatisation for the Spoken Dutch Corpus, Center for Computational Linguistics Maria-Theresiastraat 21 3000 Leuven, Belgium, 2000

[17] Roger Garside and Nicholas Smith. A Hybrid Grammatical Tagger: CLAWS4, http://ucrel.lancs.ac.uk/papers/HybridTaggerGS97.pdf

[18] Simon STÅHL. Part-of-Speech Tagger for Swedish, Computer Science, Lund University, 2000

[19] Census report: Ethiopia's population now 76 million. December 4th, 2008. http://ethiopolitics.com/news

[20] Yoshua Bengio and Yves Grandvalet. No Unbiased Estimator of the Variance of K-Fold Cross-Validation. Dept. IRO, Universit´e de Montr´eal C.P. 6128, Montreal, Qc, H3C 3J7, Canada, 2004 http://www.faqs.org/faqs/ai-faq/neural-nets/part3/section-12.html

Speaker Identification using Row Mean of Haar and Kekre's Transform on Spectrograms of Different Frame Sizes

Dr. H B Kekre
Senior Professor, Computer Dept.,
MPSTME, NMIMS University,
Mumbai, India.

Vaishali Kulkarni
Associate Professor, Electronics and Telecommunication,
MPSTME, NMIMS University,
Mumbai, India.

***Abstract*—In this paper, we propose Speaker Identification using two transforms, namely Haar Transform and Kekre's Transform. The speech signal spoken by a particular speaker is converted into a spectrogram by using 25% and 50% overlap between consecutive sample vectors. The two transforms are applied on the spectrogram. The row mean of the transformed matrix forms the feature vector, which is used in the training as well as matching phases. The results of both the transform techniques have been compared. Haar transform gives fairly good results with a maximum accuracy of 69% for both 25% as well as 50% overlap. Kekre's Transform shows much better performance, with a maximum accuracy of 85.7% for 25% overlap and 88.5% accuracy for 50% overlap.**

Keywords-Speaker Identification; Spectrogram; Haar Transform; Kekre's Transform; Row Mean; Euclidean distance

I. INTRODUCTION

Humans recognize the voice of someone familiar, and are able to match the speaker's name to his/her voice. This process is called Speaker Identification. Speaker Identification falls under the broad category of Speaker Recognition [1] – [3], which covers Identification as well as Verification. Speaker identification determines which registered speaker provides a given utterance from amongst a set of known speakers (also known as closed set identification). Speaker verification accepts or rejects the identity claim of a speaker (also known as open set identification). Speaker Identification task can be further classified into text-dependent or text-independent task [4] – [6]. In the former case, the utterance presented to the system is known beforehand. In the latter case, no assumption about the text being spoken is made, but the system must model the general underlying properties of the speaker's vocal spectrum. In general, text-dependent systems are more reliable and accurate, since both the content and voice can be compared [3], [4]. With a large number of applications like voice dialing, phone banking, teleshopping, database access services, information services, voice mail, security systems and remote access to computers etc., the automated systems need to perform as well or even better, than humans [7] – [10]. Work on Speaker Identification started as early as 1960. Since then many techniques such as filter banks [11], formant analysis [12], auto-corelation [13], instantaneous spectra covariance matrix [14], spectrum and fundamental frequency histograms [15], linear prediction coefficients [16] and long term averaged spectra [17] for feature extraction have been implemented. Some of recent works on speaker identification depend on classical features including *cepstrum* with many variants [4], sub-band processing technique [18 - 21], Gaussian mixture models (GMM) [22], linear prediction coding [23, 24], wavelet transform [25 - 27] and neural networks [26 - 28].A lot of work in this regard has been done. But still there is lack of understanding of the characteristics of the speech signal that can uniquely identify a speaker.

We have proposed speaker identification using power distribution in the frequency domain [29], [30]. We have also proposed speaker recognition using vector quantization in time domain by using LBG (Linde Buzo Gray), KFCG (Kekre's Fast Codebook Generation) and KMCG (Kekre's Median Codebook Generation) algorithms [31 – 33] and in transform domain using DFT (Discrete Fourier Transform), DCT (Discrete Cosine Transform) and DST (Discrete Sine Transform) [34].

The concept of row mean of the transform techniques has been used for content based image retrieval (CBIR) [35 – 38]. This technique also has been applied on speaker identification by first converting the speech signal into a spectrogram [39]. We have proposed Speaker Identification using row mean of DFT, DCT, DST and Walsh Transforms on the speech signal [40] – [41].

In this paper we have proposed a different approach by using the spectrograms. Row mean of Haar and Kekre's Transforms are taken on the spectrogram of the speech signal taken on different frame sizes. The generalized block diagram of the speaker identification system is shown in Fig. 1. As shown in Fig. 1, the reference signals in the database are first converted into their spectrograms and then the transforms are applied. The feature vectors are extracted and stored. The test signal to be identified is similarly processed and the feature vector is matched with the feature vectors stored in the database. The feature vector of the speaker in the database which gives the minimum Euclidean distance with the test signal is declared as the speaker identified. Section II describes the process of converting the speech signal into a spectrogram. The Haar and Kekre's transforms have been explained in section III. In Section IV, the feature vector

extraction is explained. Results are discussed in section V and conclusion ion section VI.

II. SPECTROGRAM GENERATION

The first step in the speaker identification system is to convert the speech signal into a spectrogram. A spectrogram is a time-varying spectral representation [42] (forming an image) that shows how the spectral density of a signal varies with time. Spectrograms have been used for speaker Identification since a very long time [43 - 45]. Spectrograms are usually created in one of two ways: approximated as a filterbank that results from a series of bandpass filters (this was the only way before the advent of modern digital signal processing), or calculated from the time signal using the short-time Fourier transform (STFT). Creating a spectrogram using the STFT is usually a digital process. Digitally sampled data, in the time domain, is broken up into chunks, which usually overlap, and Fourier transformed to calculate the magnitude of the frequency spectrum for each chunk. Each chunk then corresponds to a vertical line in the image; a measurement of magnitude versus frequency for a specific moment in time. The spectrums or time plots are then "laid side by side" to form the image or a three-dimensional surface. This is done using the following steps:

1. The speech signal is first divided into frames, (of sizes 32, 64, 96, 128, 160, 192, 224, 256, 292 or 320) with an overlap of 25% or 50%.
2. These frames are arranged column wise to form a matrix. E.g. if the speech signal is a one dimensional signal of 44096×1. We divide this into frames of 256 samples each with an overlap of 25% between consecutive frames i.e. overlap of 64. These frames are then arranged column wise to form a matrix of dimension 256×229.
3. Discrete Fourier Transform (DFT) is applied to this matrix column wise.
4. The spectrogram is then plotted as the squared magnitude of this transform matrix.

Fig. 2 (a) shows a sample speech signal from the database. Fig.2 (b) shows the spectrogram plotted for the speech signal of fig. 2(a) with a frame size of 256 with a overlap of 25% between adjacent samples. Fig. 2 (c) shows the spectrogram plotted for the same speech signal with an overlap of 50% between adjacent samples. Fig. 3 shows the spectrograms generated for three different speakers for two iterations. Fig. 3a & Fig 3b are the spectrograms of the same speaker 1 for the same sentence for two iterations. Similarly Fig. 3c & Fig. 3d are for speaker 2 and Fig. 3e & Fig. 3f are for speaker 3. As can be seen from the spectrograms there is much similarity between two iterations of the same speaker whereas the spectrograms of different speakers vary. Looking at these features we decided to implement the row mean technique.

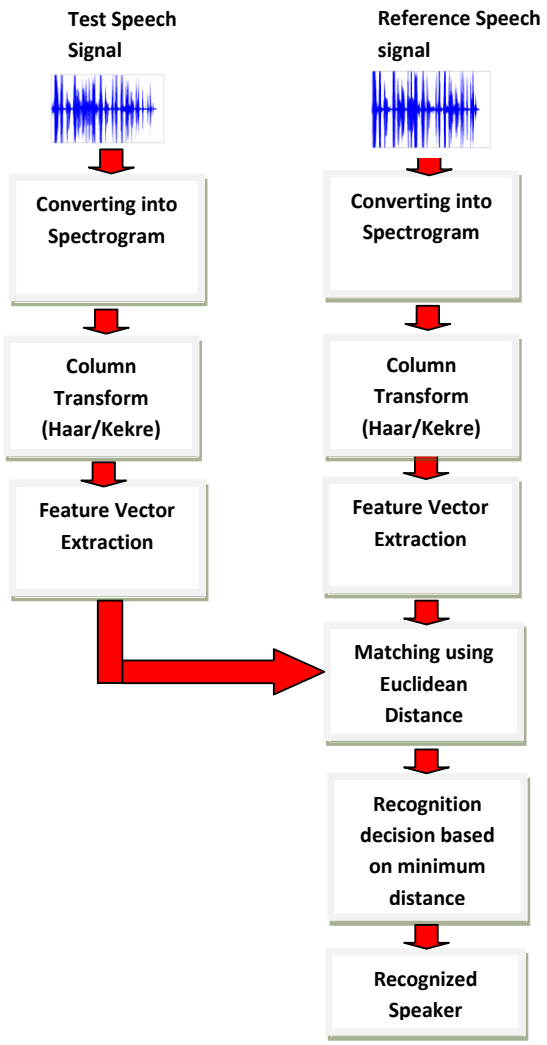
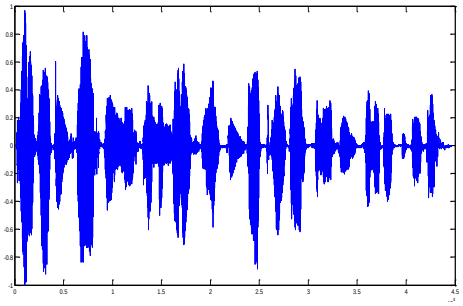


Figure 1. Speaker Identification System

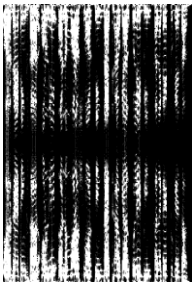
Fig. 2 (a) shows a sample speech signal from the database. Fig.2 (b) shows the spectrogram plotted for the speech signal of fig. 2(a) with a frame size of 256 with a overlap of 25% between adjacent samples. Fig. 2 (c) shows the spectrogram plotted for the same speech signal with an overlap of 50% between adjacent samples.

Fig. 3 shows the spectrograms generated for three different speakers for two iterations. Fig. 3a & Fig 3b are the spectrograms of the same speaker 1 for the same sentence for two iterations.

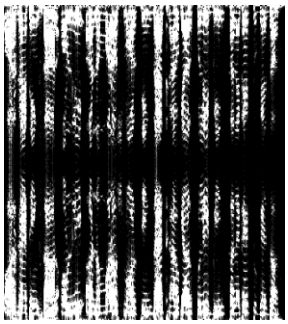
Similarly Fig. 3c & Fig. 3d are for speaker 2 and Fig. 3e & Fig. 3f are for speaker 3. As can be seen from the spectrograms there is much similarity between two iterations of the same speaker whereas the spectrograms of different speakers vary. Looking at these features we decided to implement the row mean technique.



a. Sample Speech signal



b. Spectrogram of frame size 256 with a 25% overlap



c. Spectrogram of frame size 256 with a 50% overlap

Figure 2. Speech Signal and its Spectrogram

III. TRANSFORM TECHNIQUES

For the present work on Speaker Identification, we have used two transform techniques, namely Haar transform and Kekre's Transform. The two transforms have been explained in this section.

A. Haar Transform

This sequence was proposed in 1909 by Alfréd Haar [46]. Haar used these functions to give an example of a countable orthonormal system for the space of square-integrable functions on the real line [47, 48]. The Haar transform is derived from the Haar matrix. The Haar transform is separable and can be expressed in matrix form as:

$$[F] = [H]\,[f]\,[H]^T \quad (1)$$

Where f is an NxN image, H is an NxN Haar transform matrix and F is the resulting NxN transformed image. The transformation H contains the Haar basis function $h_k(t)$ which are defined over the continuous closed interval t Є [0,1].

The Haar basis functions are

- When k=0, the Haar function is defined as a constant

$$h_0(t)=1/\sqrt{N} \quad (2)$$

- When k>0 , the Haar function is defined by

$$h_k(t) = \frac{1}{\sqrt{N}} \begin{cases} 2^{p/2} & (q-1)/2^p \le t < (q-0.5)/2 \\ -2^{p/2} & (q-0.5)/2^p \le t < q/2^p \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

Where $0 \le p < \log 2N$ and $1 \le q \le 2p$

The N Haar functions can be sampled at t = (2n+1) Δ, where Δ = T/ (2N) and n = 0, 1, 2, 3…, N-1 to form an N x N matrix for discrete Haar transform. For example, when N=4, we have

$$\mathbf{H}_4 = \frac{1}{2}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & 0 & 0 \\ 0 & 0 & \sqrt{2} & -\sqrt{2} \end{bmatrix} \quad (4)$$

For N=8

$$\mathbf{H}_8 = \frac{1}{\sqrt{8}}\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & -2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & -2 \end{bmatrix} \begin{matrix} \varphi_0(t) \\ \psi_0(t) \\ \psi_{1,0}(t) \\ \psi_{1,1}(t) \\ \psi_{2,0}(t) \\ \psi_{2,1}(t) \\ \psi_{2,2}(t) \\ \psi_{2,3}(t) \end{matrix} \quad (5)$$

B. Kekre's Transform

Kekre Transform matrix [49, 51] can be of any size NxN, which need not have to be in powers of 2 (as is the case with most of other transforms including Haar Transform). All upper diagonal and diagonal values of Kekre's transform matrix are one, while the lower diagonal part except the values just below diagonal are zero. Generalized N×N Kekre Transform Matrix can be given as in (6).

The formula for generating the term K_{xy} of Kekre's transform matrix is given by (7).

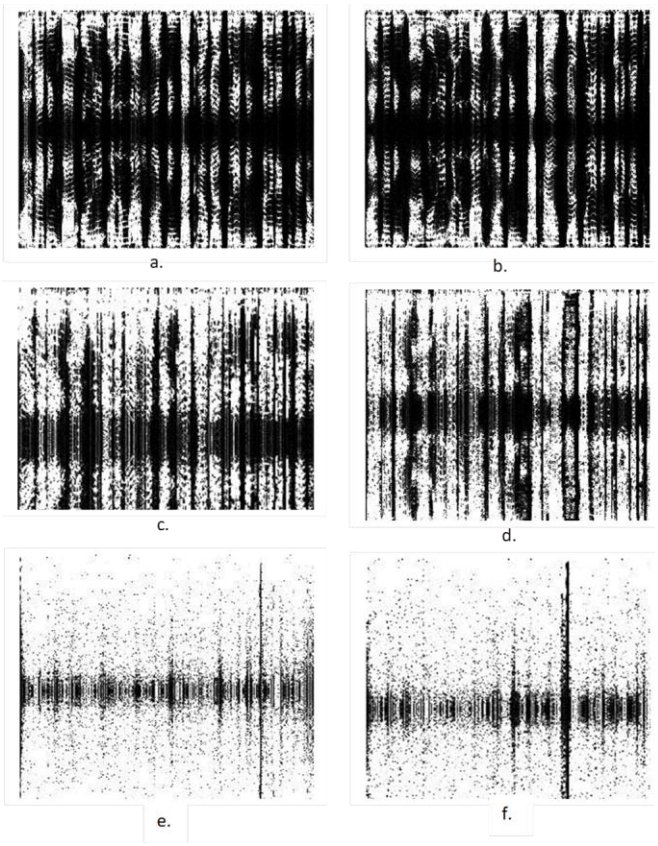


Figure 3.Spectrograms for three different speakers. a & b for speaker 1,c & d for speaker 2 and e & f for speaker 3 for the text , "All great things are only a number of small things that have carefully been collected together".

Fig. 4 shows eight waveforms generated using (3) for N=8. Writing this in matrix form we get 8x8 Haar matrix.

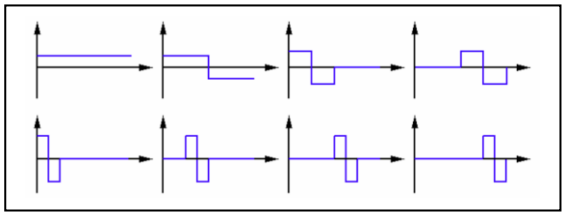


Figure 4. Eight waveforms generated using (3) for N=8

$$K_{NxN} = \begin{bmatrix} 1 & 1 & 1 & .. & 1 & 1 \\ -N+1 & 1 & 1 & .. & 1 & 1 \\ 0 & -N+2 & 1 & .. & 1 & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & .. & 1 & 1 \\ 0 & 0 & 0 & .. & -N+(N-1) & 1 \end{bmatrix} \quad (6)$$

$$Kxy = \begin{cases} 1 & ; x \le y \\ -N + (x-1) & ; x = y + 1 \\ 0 & ; x > y + 1 \end{cases} \quad (7)$$

Kekre's Transform has been used for image retrieval [50].

Also Kekre's Wavelet transform ($N^2 \times N^2$) has been developed using (N×N) Kekre's Transform Matrix [51].

IV. FEATURE VECTOR EXTRACTION

The procedure for feature vector extraction is given below:

1. Column Transform (Haar or Kekre's Transform) is applied on the spectrogram of the speech signal.
2. The mean of the absolute values of the rows of the transform matrix is then calculated.
3. These row means form a column vector (M×1) where M is the number of rows in the transform matrix).
4. This column vector forms the feature vector for the speech sample.
5. The feature vectors for all the speech samples are calculated for different values of n and stored in the database. Fig. 5 shows the Feature Vector generation technique.

Figure 5. Feature Vector Generation from Spectrogram

V. RESULTS

A. Database Description

The speech samples used in this work are recorded using Sound Forge 4.5. The sampling frequency is 8000 Hz (8 bit, mono PCM samples).

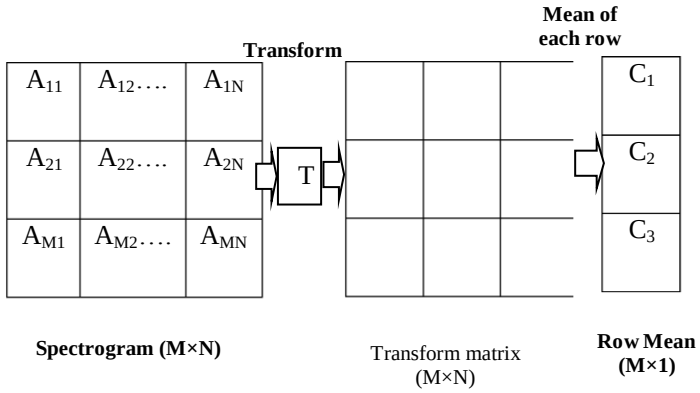


Table I shows the database description. Five iterations of four different sentences of varying lengths are recorded from each of the speakers. Twenty samples per speaker are taken. For text dependent identification, four iterations of a particular sentence are kept in the database and the remaining one iteration is used for testing.

TABLE I. DATABASE DESCRIPTION

Parameter	Sample characteristics
Language	English
No. of Speakers	42
Speech type	Read speech
Recording conditions	Normal. (A silent room)
Sampling frequency	8000 Hz
Resolution	8 bps

B. Experimental Results

The feature vectors of all the reference speech samples are stored in the database in the training phase. In the matching phase, the test sample that is to be identified is taken and similarly processed as in the training phase to form the feature vector. The stored feature vector which gives the minimum Euclidean distance with the input sample feature vector is declared as the speaker identified.

C. Accuracy of Identification

The accuracy of the identification system is calculated as given by (8).

$$Accuracy\ (\%) = \frac{number\ of\ matches}{number\ of\ samples\ tested} \times 100 \quad \mathbf{x\ 100} \tag{8}$$

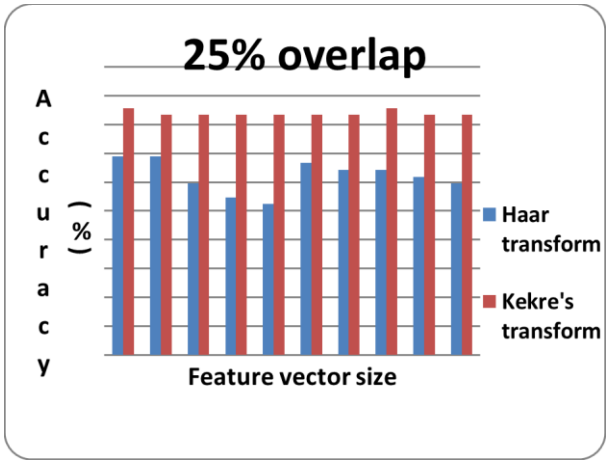


Figure 6. Performance comparison of Haar and Kekre's Transform for a overlap of 25%

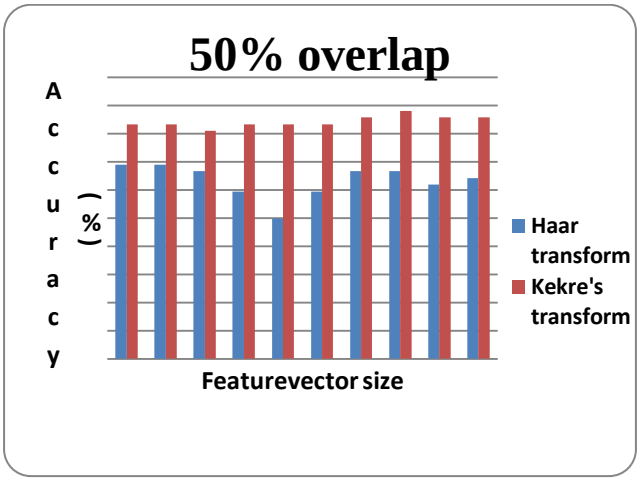


Figure 7. Performance comparison of Haar and Kekre's Transform for a overlap of 50%

Fig. 6 shows the results obtained by using the two transforms for an overlap of 25% between the adjacent frames while creating the spectrograms of the speech signals. As can be seen from the graphs, the Haar transform gives an average performance of around 60%. The maximum accuracy is obtained for a feature vector size of 32 and 64 (69%)and minimum for a feature vector size of 160 (51%). As the feature vector size is increased further, the accuracy drops and there is no improvement.

For Kekre's transform, the average accuracy is around 83.33%, with a maximum accuracy of around 85% for a feature vector size of 256 and minimum accuracy for feature vector size of 64 to 160 (82%). Fig.7 shows the results obtained by using the two transforms for an overlap of 50% between the adjacent frames while creating the spectrograms of the speech signals. Here also Haar transform gives an average accuracy of around 60%. The behavior of haar transform for both the cases is much similar.

For Kekre's transform, the average accuracy is slightly more i.e. 85.7%. The maximum accuracy is 88.5% for a feature vector size of 256. Overall Kekre's transform gives much better results as compared to Haar transform.

VI. Conclusion

In this paper we have compared the performance of Haar and Kekre's transforms for speaker identification for two different cases (25% and 50% overlap). Haar transform gives an average accuracy of around 60% for both the cases. Accuracy does not increase as the feature vector size is increased from 64 onwards.

Kekre's transform gives an accuracy of more than 80% for both the cases. The maximum accuracy obtained for Kekre's transform 88.5% for a feature vector size of 256. The present study is ongoing and we are analyzing the performance on other transforms.

References

[1] Lawrence Rabiner, Biing-Hwang Juang and B.Yegnanarayana, "Fundamental of Speech Recognition", Prentice-Hall, Englewood Cliffs, 2009.

[2] S Furui, "50 years of progress in speech and speaker recognition research", ECTI Transactions on Computer and Information Technology, Vol. 1, No.2, November 2005.

[3] D. A. Reynolds, "An overview of automatic speaker recognition technology", Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP'02), 2002, pp. IV-4072–IV-4075.

[4] Joseph P. Campbell, Jr., Senior Member, IEEE, "Speaker Recognition: A Tutorial", Proceedings of the IEEE, vol. 85, no. 9, pp. 1437-1462, September 1997.

[5] F.Bimbot, J.-F. Bonastre, C. Fredouille, G. Gravier, I. Magrin-Chagnolleau, S. Meignier, T. Merlin, J. Ortega-García, D.Petrovska-Delacrétaz, and D. A. Reynolds, "A tutorial on text-independent speaker verification," EURASIP J. Appl. Signal Process., vol. 2004, no. 1, pp. 430–451, 2004.

[6] S. Furui. Recent advances in speaker recognition. AVBPA97, pp 237--251, 1997.

[7] D. A. Reynolds, "Experimental evaluation of features for robust speaker identification," IEEE Trans. Speech Audio Process., vol. 2, no. 4, pp. 639–643, Oct. 1994.

[8] Tomi Kinnunen, Evgeny Karpov, and Pasi Fr¨anti, "Realtime Speaker Identification", ICSLP2004.

[9] Marco Grimaldi and Fred Cummins, "Speaker Identification using Instantaneous Frequencies", IEEE Transactions on Audio, Speech, and Language Processing, vol., 16, no. 6, August 2008.

[10] Zhong-Xuan, Yuan & Bo-Ling, Xu & Chong-Zhi, Yu. (1999). "Binary Quantization of Feature vectors for robust text-independent Speaker Identification" in IEEE Transactions on Speech and Audio Processing Vol. 7, No. 1, January 1999. IEEE, New York, NY, U.S.A.
[11] S. Pruzansky, "Pattern-matching procedure for automatic talker recognition", J.A.S.A., 35, pp. 354-358, 1963.
[12] G.R.Doddington, "A method of speaker verification", J.A.S.A., 49,139 (A), 1971.
[13] P.D. Bricker, et. al., "Statistical techniques for talker identification", B.S.T.J., 50, pp. 1427-1454, 1971.
[14] K.P.Li, et. al., "Experimental studies in speaker verification using a adaptive system", J.A.S.A., 40, pp. 966-978, 1966.
[15] B. Beek, et. al., "Automatic speaker recognition system", Rome Air Development Center Report, 1971.
[16] M.R.Sambur, Speaker recognition and verification using linear prediction analysis, Ph. D. Dissert., M.I.T., 1972.
[17] S. Furui, et. al., "Talker recognition by long time averaged speech spectrum", Electronics and Communications in Japan, 55-A. pp. 54-61, 1972.
[18] Besacier, L., J.F. Bonnastre and C. Fredouille, 2000.Localization and Selection of Speaker-Specific Information with Statistical Modeling. Speech Communications. 31: 89-106.
[19] Besacier, L. and J.F. Bonnastre, 2000. Subband Architecture for Automatic Speaker Recognition. Signal Processing. 80: 1245-1259.
[20] Damper, R.I. and J.E. Higgins, 2003. Improving Speaker Identification in Noise by Subband Processing and Decision Fusion. Pattern Recognition Letters. 24: 2167-2173.
[21] Sivakumaran, P., A.M. Ariyaeeinia and M.J. Loomes, 2003. Subband Based Text-dependent Speaker Verification. Speech Communications. 41: 485-509.
[22] Reynolds, D.A, T.F. Quatieri and R.B. Dunn., 2000. Speaker Verification Using Adapted Gaussian Mixture Models. Digital Signal Processing. pp: 19-4.
[23] Bassam A. Mustafa, B. Y. Thanoon and S.D. Al- Shamaa., 2005. A Database System for Speaker Identificatoin. Proceedings of The 2nd International Conference on Information Technology. Al-Zaytoonah University of Jordan. May 2005.
[24] Mohd Saleem, A. M., K. Mustafa and I. Ahmad., 2005. Spoken Word of German Digits Uttered by Native and non Native Speakers. Proceedings of The 2nd International Conference on Information Technology. Al-Zaytoonah University of Jordan. May 2005.
[25] Prina Ricotti, L., 2005. Multitapring and Wavelet Variant of MFCC in Speech Recognition. IEE Proceedings on Vis. Image Signal Process., pp: 29- 35.
[26] Dokur, Z. and T. Olmz., 2003. Classification of Respiratory Sounds By using An Artificial Neural Networks. International Journal of Pattern Recognition and artificial Intelligence. 4: 567-580.
[27] Abduladheem A., M.A. Alwan, and A.A. Jassim, 2005. Hybrid Wavelet-Network Neural/FFT Nural Phoneme Recognition. Proceedings of The 2nd International Conference on Information Technology. Al-Zaytoonah University of Jordan, May 2005.
[28] Farrell, K.R., R.J. Mammone and K.T. Assalah., 1994. Speaker Recognition Using Neural. Networks and Conventional Classifiers. IEEE Trans. on Speech and Audio Proc. pp: 194-205.
[29] Dr. H B Kekre, Vaishali Kulkarni,"Speaker Identification using Power Distribution in Frequency Spectrum", Technopath, Journal of Science, Engineering & Technology Management, Vol. 02, No.1, January 2010.
[30] Dr. H B Kekre, Vaishali Kulkarni, "Speaker Identification by using Power Distribution in Frequency Spectrum", ThinkQuest - 2010 International Conference on Contours of Computing Technology", BGIT, Mumbai,13th -14th March 2010.
[31] H B Kekre, Vaishali Kulkarni, "Speaker Identification by using Vector Quantization", International Journal of Engineering Science and Technology, May 2010.
[32] H B Kekre, Vaishali Kulkarni, "Performance Comparison of Speaker Recognition using Vector Quantization by LBG and KFCG", International Journal of Computer Applications, vol. 3, July 2010.
[33] H B Kekre, Vaishali Kulkarni, "Performance Comparison of Automatic Speaker Recognition using Vector Quantization by LBG KFCG and KMCG", International Journal of Computer Science and Security, Vol: 4 Issue: 5, 2010.
[34] H B Kekre, Vaishali Kulkarni, "Comparative Analysis of Automatic Speaker Recognition using Kekre's Fast Codebook Generation Algorithm in Time Domain and Transform Domain", International Journal of Computer Applications, Volume 7 No.1. September 2010.
[35] Dr. H.B.Kekre, Sudeep D. Thepade, Akshay Maloo "Performance Comparision of Image Retrieval using Row Mean of Transformed Column Image", International Journal on Computer Science and Engineering Vol. 02, No. 05, 2010, 1908-1912
[36] Dr.H.B.Kekre,Sudeep Thepade "Edge Texture Based CBIR using Row Mean of Transformed Column Gradient Image", International Journal of Computer Applications (0975 – 8887) Volume 7– No.10, October 2010
[37] Dr. H.B.Kekre, Sudeep D. Thepade, Akshay Maloo "Eigenvectors of Covariance Matrix using Row Mean and Column Mean Sequences for Face Recognition", International Journal of Biometrics and Bioinformatics (IJBB), Volume (4): Issue (2)
[38] Dr. H.B.Kekre, Sudeep Thepade, Archana Athawale, "Grayscale Image Retrieval using DCT on Row mean, Column mean and Combination", Journal of Sci., Engg. & Tech. Mgt. Vol 2 (1), January 2010
[39] Dr. H. B. Kekre, Dr. T. K. Sarode, Shachi J. Natu, Prachi J. Natu "Performance Comparison of Speaker Identification Using DCT, Walsh, Haar on Full and Row Mean of Spectrogram", International Journal of Computer Applications (0975 – 8887) Volume 5– No.6, August 2010
[40] Dr. H B Kekre, Vaishali Kulkarni "Comparative Analysis of Speaker Identification using row mean of DFT, DCT, DST and Walsh Transforms", International Journal of Computer Science and Information Security, Vol. 9, No.1, January 2011.
[41] Dr. H B Kekre, Vaishali Kulkarni, Sunil Venkatraman, Anshu Priya, Sujatha Narashiman, "Speaker Identification using Row Mean of DCT and Walsh Hadamard Transform", International Journal on Computer Science and Engineering, Vol. 3, No.1, March 2011.
[42] S. Haykin, editor, Advances in Spectrum Analysis and Array Processing, vol.1, Prentice-Hall, 1991.
[43] Tridibesh Dutta, "Text dependent speaker identification based on spectrograms", Proceedings of Image and vision computing, pp. 238-243, New Zealand 2007
[44] Azzam Sleit, Sami Serhan, and Loai Nemir, "A histogram based speaker identification technique", International Conference on ICADIWT, pp. 384-388, May 2008.
[45] Tridibesh Dutta and Gopal K. Basak, "Text dependent speaker identification using similar patterns in spectrograms", PRIP'2007 Proceedings, Volume 1, pp. 87-92, Minsk, 2007.
[46] Haar, Alfred, "ZurTheorie der orthogonalenFunktionensysteme". (German), MathematischeAnnalen, volume 69, No. 3, 1910, pp. 331-371.
[47] Minh N. Do, Martin Vetterli, "Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance", IEEE Transactions On Image Processing, Volume 11, Number 2, pp.146-158, February 2002.
[48] Charles K. Chui, "An Introduction to Wavelets", Academic Press, 1992, San Diego, ISBN 0585470901.
[49] Dr. Kekre H. B. and Thepade Sudeep, "Image Retrieval using Non-Involutional Orthogonal Kekre's Transfrom", International Journal of MultiDisciplinary Research And Advnces in Engineering,IJMRAE, Vol.1, No.I,Novenber 2009,pp189-203.
[50] Dr. H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathamesh Verlekar and Suraj Shirke, "Kekre Transform over Row Mean, Column Mean and Both Using Image Tiling for Image Retrieval", International Journal of Computer and Electrical Engineering, Vol.2, No.6, December, 2010 1793-8163.
[51] Dr. H.B.Kekre, Archana Athawale,Deepali Sadavarti, "Algorithm to Generate Kekre's Wavelet Transform from Kekre's Transform", International Journal of Engineering Science and Technology Vol. 2(5), 2010, 756-767.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's

NMIMS University, Vile Parle(w), Mumbai, INDIA. He has guided 17 Ph.D.s, 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networks. He has more than 350 papers in National / International Conferences / Journals to his credit. Recently fifteen students working under his guidance have received best paper awards. Recently five research scholars have received Ph. D. degree from NMIMS University Currently he is guiding seven Ph.D. students. He is member of ISTE and IETE.

Vaishali Kulkarni has received B.E in Electronics Engg. from Mumbai University in 1997, M.E (Electronics and Telecom) from Mumbai University in 2006. Presently she is pursuing Ph. D from NMIMS University. She has a teaching experience of around 10 years. She is Associate Professor in telecom Department in MPSTME, NMIMS University. Her areas of interest include networking, Signal processing, Speech processing: Speech and Speaker Recognition. She has 15 papers in National / International Conferences / Journals to her credit.

Forecasting the Tehran Stock Market by Artificial Neural Network

(CasestudyMobarakeh-steelCo.)

Reza Aghababaeyan, TamannaSiddiqui, NajeebAhmadKhan
Department of Computer Science
Jamia Hamdard University
New Delhi- India

***Abstract*— One of the most important problems in modern finance is finding efficient ways to summarize and visualize the stock market data to give individuals or institutions useful information about the market behavior for investment decisions. The enormous amount of valuable data generated by the stock market has attracted researchers to explore this problem domain using different methodologies. Potential significant benefits of solving these problems motivated extensive research for years. In this paper, computational data mining methodology was used to predict seven major stock market indexes. Two learning algorithms including Linear Regression and Neural Network Standard feed-**forward **back prop (FFB) were tested and compared. The models were trained from four years of historical data from March 2007 to February 2011 in order to predict the major stock prices indexes in the Iran (Tehran Stock Exchange). The performance of these prediction models was evaluated using two widely used statistical metrics. We can show that using Neural Network Standard feed-forward back prop (FFB) algorithm resulted in better prediction accuracy. In addition, traditional knowledge shows that a longer training period with more training data could help to build a more accurate prediction model. However, as the stock market in Iran has been highly fluctuating in the past two years, this paper shows that data collected from a closer and shorter period could help to reduce the prediction error for such highly speculated fast changing environment.**

Keywords- Data mining*; *Stock Exchange; Artificial Neural Network; Matlab.

I. INTRODUCTION

Data mining called knowledge discovery in databases, in computer science, the process of discovering interesting and useful patterns and relationships in large volumes of data. The field combines tools from statistics and artificial intelligence (such as neural networks and machine learning) with database management to analyze large digital collections, known as data sets. An artificial neural network (ANN), usually called neural network (NN), is a mathematical model or computational model that is inspired by the structure and/or functional aspect of biological neural networks. Neural Networks (NN) as Artificial Intelligence method has become very important in making stock market predictions, as it has proved to be more advantages than the other methods. Since then lot of research was carried out using different topologies of Neural Networks. According to Wong, Bodnovich and Selvi [1] the most frequent areas of neural networks applications are production operations (53.5%) and finance (25.4%). In finance, it is more specific to stock market predictions. When compared to the other methods NN outperformed and the accuracy rate ranges from 68% to 90% [9]. The popularity of these methods is mainly due to the benefits outnumber the limitations it has. Different NN methods were used for optimal feature selection to generating buy and sell signals, the more popular being the former.

In this paper, we anticipated Mobarakeh-Steel Co. try. The high level was in Tehran Stock Exchange. We used data from 15 March 2007 until 14 February 2011 for training the neural Network and from15 February 2011 until 30th we have performed experiments using MATLAB and we got 97 % results, which are very encouraging.

In stock market when brokers want to sell or buy stock, they mostly depend on technical trading rules. Robert Edward and John Magee have [10], defined technical trading rules as "the science of recording the actual history of trading (price changes, volume of transaction, etc.) in a certain stock or in 'The averages' and then deducing from that pictured history the probable future trend". Different artificial intelligence methods were used to optimize the prediction by successful selection of trading rules.

II. PROPOSED APPROACH

A researcher originates primary data for the specific purpose of the problem at hand, but secondary data are data that have already been collected for other purposes. Secondary data can be classified as internal and external. Our research strategy is the analyses are of secondary data. For conducting our research, types of data are needed:

1) Companies' stock prices, which are an internal secondary data, gathered from the financial databases.

2) 2. Existing Approaches to Cash Forecasting.

Techniques used for cash forecasting can be broadly classified.

a) Time -series Method

b) Factor analysis method

c) Expert system approach

We have used Factor analysis method for cash forecasting because by using this method the results are better than of two above methods.

Among the 20 text files provided by Tehran Stock Exchange Service Company (TSESC), Mobarakeh-Steel Co. Intra day price and their corresponding date and time during years 2007 till 2011 are chosen to be given to the split and merge algorithm. Before implementing, the segmentation algorithm in R Programming Language, Mobarakeh-Steel Co. text file should be read by the program. The program reads 1904 intra day prices (data points) for this company during years 2007 and 2011, which is equal to 2266.

TABLE 1. Data Sets

Training Set	**From 15nd march 2007 to14th February 2011 for training the neural**
Validation Set	**From 15 February 2011 to 30th February 2011for validating the neural**

Architecture of the Model:

Design of right architecture involves several important steps:

a) Selecting the number of layers

b) Basic decision about the amount of neurons to be used in each layer

c) Choosing the appropriate neurons' transfer functions.

If there are not enough neurons in each layer, the outputs will not be able to fit all the data points (under- fitting). On the other hand, if there are too many neurons in each layer, oscillations may occur between data points (over-fitting). Therefore, a topology study was conducted in order to find the most appropriate architecture neural network to fit Cash forecasting parameters. There are several combinations of neurons and layers.

III. EXPERIMENTAL

The Propose approach has been implemented through Matlab Software.

A. *Number* of Layers

The model is a three layer feed-forward neural network and was trained using fast back propagation algorithm because it was found to be the most efficient and reliable means to be used for this study. Table 2 shows a comparison of the two algorithms.

B. Input Layer Size

Number of input neurons depends upon:

1) Number of cash withdrawal factors included in the model

2) Way these factors are encoded

TABLE 2. Selection of Algorithms

Function	Technique	Time(s)
TRAINBP	**Back propagation**	**59**
TRAINBPX	**Fast Back propagation**	**40**

Mainly calendar effects are included as parameters affecting cash withdrawal in this model, Total number of input neurons needed in this model hence is seven, each representing the values of an individual variable at a particular instant of time.

C. *Output Layer Size*

In this model, only one output unit is needed for indicating the value of forecasted cash.

D. Optimal Hidden Layer Size

There is no easy way to determine the optimal number of hidden units without training using number of hidden units and estimating the error of each. The best approach to find the optimal number of hidden units is trial and error. In practice, we can use either the forward selection or backward selection to determine the hidden layer size.

Forward selection: Starts with choosing an appropriate criterion for evaluating the performance of the network. Then we select a small number of hidden neurons; record its performance i.e. forecast accuracy.

Next, we slightly increase the hidden neurons, train and test until the error is acceptably small or no significant improvement is noted.

Backward selection: Starts with a large number of hidden neurons and the decreases the number gradually.

For this study, the forward selection approach was used to select the size of hidden layer and best result was with 10 Neurons in hidden layer, as evident from the following Table 3.

TABLE 3. Optimal Neurons in Hidden Layer

No. of Neurons in Hidden Layer	Mean Forecast Error for 60datapoints	Sum Squared error at 5000epochs
7	**9.675**	**0.004343**
8	**9.85726**	**0.004153**
9	**8.51395**	**0.002881**
10	**9.52748**	**0.003768**
11	**9.39397**	**0.003565**
12	**8.56169**	**0.003645**

Optimal Transfer Function As shown in table 4, for best transfer function, tansig in hidden (at 10 neurons) and logsig in output layer were found to be optimal.

TABLE 4. Optimal Transfer Function

Transfer function	Mean Forecast Error for 60datapoints	Sum Squared error at 5000epochs
Logsig - purelin	11.0786	0.007697
Tansig - purelin	9.90157	0.004647
Logsig - logsig	7.38031	0.002591
Tansig - tansig	9.53892	0.004560
Logsig - tansig	12.2373	0.011609
Tansig - logsig	6.74556	0.002246

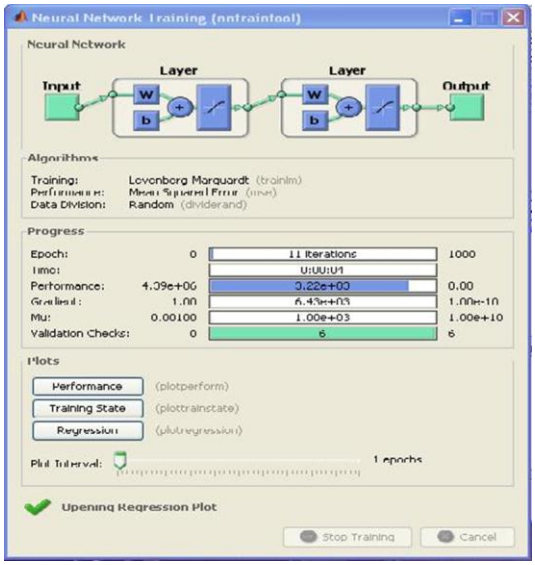


Figure1, Neural Network Training' Shown in the figure 1 Mobarakeh-steel co. data training, that data training was 15nd march 2007 to 14th February 2011 for Training.

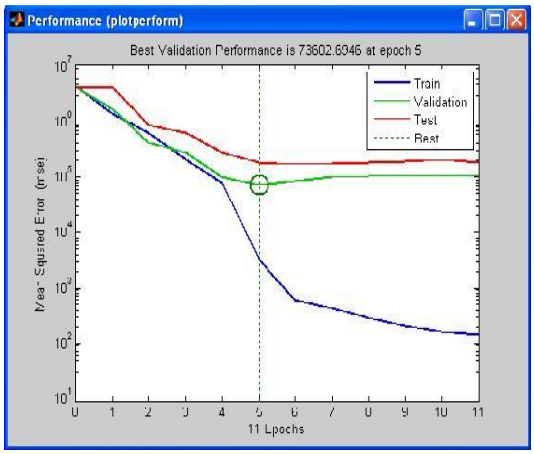


Figure2, Performance

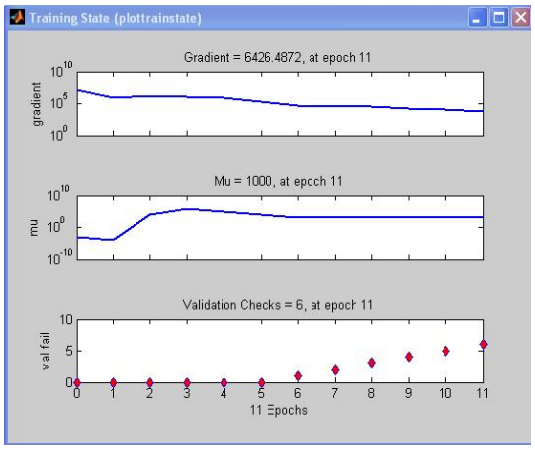


Figure3, Training State 'Shown in the figure 3 Training State (Plot train state), that Gradient= 6426.4872 at epoch 11. Mu = 1000, at epoch 11. Validation checks= six, at epoch 11.

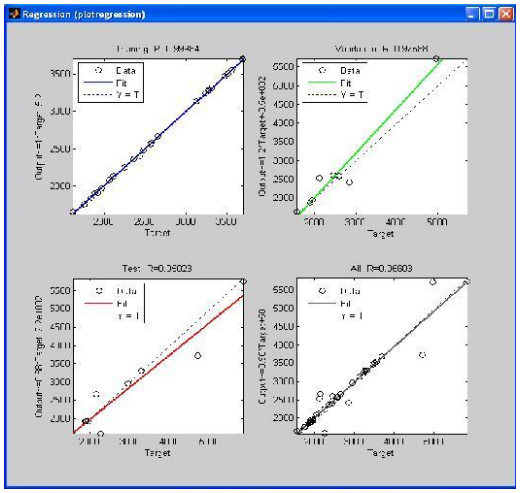


Figure4, regression (plot regression) 'Shown in the figure 4 Regression= 0.96603

TABLE 5. Real Price of mobarakeh-steel co

1904	1890	1845	2061	2435	2949	3128	3557
3279	3273	3470	3275	3310	3514	4696	5819
4957	3689	3689	1889	1965	1926	1754	1766
1882	2122	1560	1920	1920	2118	1948	1626
1895	1967	2253	2848	2462	2367	2153	2591
2652	2580	2475	2564	2564	3241	2266	

Shown in the figure above Real Price of Mobarakeh-steel co. from 15nd march 2007 to 14 February 2011.

In this model, only one output unit is needed for indicating the value of forecasted cash. Shown in the table 6 that percentage predicted for Real Data. Shown in the table 7 that Prices Predicted from Mobarakeh-steel Co.

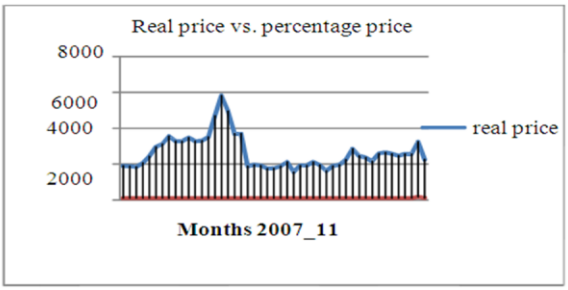


Figure 5, Real price vs. percentage price.

TABLE 6. Percentage of Prediction

1826	1823	1815	2063	2354	2894	3087	3493
3241	3235	3419	3243	3275	3446	4663	5349
4947	3634	3634	1912	1939	1925	1789	1791
1815	2114	1697	1905	1905	2092	1928	1723
1885	1945	2240	2863	2453	2323	2082	2573
2649	2560	2431	2540	2540	5497	2713	

TABLE 7. Price Prediction of mobarakeh-steel Co.

96	97	98	100	97	98	99	98
99	99	99	94	99	98	99	92
100	98	99	101	99	100	102	102
97	100	109	99	99	99	99	106
100	99	100	101	100	98	97	99
100	99	98	99	99	169	117	

IV. Results and Discussions

Results: In this work, we anticipated Mobarakeh-Steel Company's real price data. The high level was in Tehran Stock Exchange. We used data from 15 March 2007 till 14th February 2011 for training the neural Network and from 15 February 2011 until 30 February 2011 for validating the network. After the network was created by MATLAB, the results were 97 percent, which were very encouraging for this research work.

V. Conclusion

An Overview of Study:

Stock markets have been studied over and over again to extract useful patterns and predict their movements. Mining textual documents and time series concurrently, such as predicting the movements of stock prices based on the contents of the intraday price, is an emerging topic in data mining and text mining community. Stock price trend forecasting based solely on the technical and fundamental data analysis enjoys great popularity. However, numeric time series data only contain the event and not the cause why it happened.

Textual data such as news articles have richer information, hence exploiting textual information especially in addition to numeric time series data increases the quality of the input and improved predictions are expected from this kind of input rather than only numerical data. Information about company's report or breaking news stories can dramatically affect the share price of a security.

In order to make the prediction model, the research process should be implemented consists of different steps including data collection, data preprocessing, alignment, feature and document selection, document representation, classification and model evaluation. With the prediction model (feed-forward back prop) we can conclude that our prediction model out performs the random labeling. The prediction model will notify the up or down of the stock price movement when an upcoming pieces of news is released, and 83 percent of time can predict correctly.

This can be very beneficial for individual and corporate investors, financial analysts, and users of financial news. With such a model, they can foresee the future behavior and movement of stock prices; take correct actions immediately and act properly in their trading to gain more profit and prevent loss.

References

[1] B.K. Wong, T.A. Bonovich, Y. Selvi, "Neural Network Applications in Business: A Review and Analysis to the literature (1988-95), Decision Support Systems, Vol. 19, Pp301-320, 1997.

[2] C. Moler, the creator of MATLAB (December 2004). "The Origins of MATLAB". Retrieved April 15, 2007.

[3] C. Moler, "MATLAB Programming Language". Retrieved 2010-12-17.

[4] C. Moler in a Math works newsletter Cleve Moler, the creator of MATLAB (2000). "MATLAB Incorporates LAPACK". Retrieved December 20, 2008.

[5] D.R. Cooper, P.S. Schindler, 2003. Business Research Methods. 8thed .New York: Mc Graw-Hill.

[6] D. Spiel man, (2004-02-10). "Connecting C and Matlab". Yale University, Computer Science Department. Retrieved 2008-05-20.

[7] J.Kazama, T.Makino, J. Tsujii, "Support Vector Networks. Machin Learning" (1995), 20(3), p.273-297

[8] G. Amos (2004). MATLAB: An Introduction with Applications 2nd Edition. John Wiley & Sons. ISBN 978-0-471-69420-5.

[9] M. Zelaic, "Neural Network Applications in Stock Market Prediction –

A Methodology Analysis" Proc. of 9th Intl'Conf. Information and Intelligent Systems, 1998.

[10] R. Edward, J. Magee, 'Technical Analysis of Stock trends'. Seventh Edition 1997.

[12] R. Goering, "Matlab edges closer to electronic design automation world," EE Times, 10/04/2004

[13] Wikipedia,"MATLAB technical documentation". Mathworks.com. Retrieved 2010-06-07.

[14] J. Stafford, "The Wrong Choice: Locked in by license restrictions,"SearchOpenSource.com, 21 May 2003.

AUTHORS PROFILE

Reza Aghababaeyan, PhD Student of Computer Science, New Delhi-India 2011.

Tamanna Siddiqui, Ph.d of Computer Science New Delhi-India (my supervisor).

Najeeb Ahmad Khan, Master of Computer Science New Delhi-India.

A Comparison Study between Data Mining Tools over some Classification Methods

Abdullah H. Wahbeh, Qasem A. Al-Radaideh, Mohammed N. Al-Kabi, and Emad M. Al-Shawakfa
Department of Computer Information Systems
Faculty of Information Technology, Yarmouk University
Irbid 21163, Jordan

***Abstract*- Nowadays, huge amount of data and information are available for everyone, Data can now be stored in many different kinds of databases and information repositories, besides being available on the Internet or in printed form. With such amount of data, there is a need for powerful techniques for better interpretation of these data that exceeds the human's ability for comprehension and making decision in a better way. In order to reveal the best tools for dealing with the classification task that helps in decision making, this paper has conducted a comparative study between a number of some of the free available data mining and knowledge discovery tools and software packages. Results have showed that the performance of the tools for the classification task is affected by the kind of dataset used and by the way the classification algorithms were implemented within the toolkits. For the applicability issue, the WEKA toolkit has achieved the highest applicability followed by Orange, Tanagra, and KNIME respectively. Finally; WEKA toolkit has achieved the highest improvement in classification performance; when moving from the percentage split test mode to the Cross Validation test mode, followed by Orange, KNIME and finally Tanagra respectively.**

Keywords-component; data mining tools; data classification; Wekak; Orange; Tanagra; KNIME.

I. INTRODUCTION

Today's databases and data repositories contain so much data and information that it becomes almost impossible to manually analyze them for valuable decision-making. Therefore, humans need assistance in their analysis capacity; humans need data mining and its applications [1]. Such requirement has generated an urgent need for automated tools that can assist us in transforming those vast amounts of data into useful information and knowledge.

Data mining is the process of discovering interesting knowledge from large amounts of data stored in databases, data warehouses, or other information repositories. Data mining involves an integration of techniques from multiple disciplines such as database and data warehousing technology, statistics, machine learning, high-performance computing, pattern recognition, neural networks, data visualization, information retrieval, image and signal processing, and spatial or temporal data analysis [2]. Data mining has many application fields such as marketing, business, science and engineering, economics, games and bioinformatics.

Currently, many data mining and knowledge discovery tools and software are available for every one and different usage such as the Waikato Environment for Knowledge Analysis (WEKA) [3] [4], RapidMiner [5][6], Clementine [6], Rosetta, Intelligent Miner [1] etc. These tools and software provide a set of methods and algorithms that help in better utilization of data and information available to users; including methods and algorithms for data analysis, cluster analysis, Genetic algorithms, Nearest neighbor, data visualization, regression analysis, Decision trees, Predictive analytics, Text mining, etc.

This research has conducted a comparison study between a number of available data mining software and tools depending on their ability for classifying data correctly and accurately. The accuracy measure; which represents the percentage of correctly classified instances, is used for judging the performance of the selected tools and software.

The rest of the paper is organized as follows: Section 2 summaries related works on data mining, mining tools and data classification. Section 3 gives a general description on the methodology followed and provides a general description of the tools and software under test. Section 4 reports our experimental results of the proposed methodology and compares the results of the different software and tools used. Finally, we close this paper with a summary and an outlook for some future work.

II. RELATED WORKS

King and Elder [7] have conducted an evaluation of fourteen data mining tools ranging in price from $75 to $25,000. The evaluation process was performed by three kinds of user groups: (1) four undergraduates; who are inexperienced users in data mining, (2) a relatively experienced graduate student, and (3) a professional data mining consultant. Tests were performed using four data sets. To test tools flexibility and capability, their output types have varied: two binary classifications (one with missing data), a multi-class set, and a noiseless estimation set. A random two-thirds of the cases in each have served as training data; the remaining one-third was test data. Authors have developed a list of 20 criteria, plus a standardized procedure, for evaluating data mining tools. The tools ran under Microsoft Windows 95, NT, or Macintosh 7.5 operating systems, and have employed Decision Trees, Rule

Induction, Neural Networks, or Polynomial Networks to solve two binary classification problems, a multi-class classification problem, and a noiseless estimation problem. Results have provided a technical report that details the evaluation procedure and the scoring of all component criteria. Authors also showed that the choice of a tool depends on a weighted score of several categories such as software budget and user experience. Finally, authors have showed that the tools' price is related to quality.

Carrier and Povel [8] have described a general schema for the characterization of data mining software tools. Authors have described a template for the characterization of DM software along a number of complementary dimensions, together with a dynamic database of 41 of the most popular data mining tools. The business-oriented proposal for the characterization of data mining tools is defined depending on the business goal, model type, process-dependent features, user interface features, system requirements and vendor information. Using these characteristics, authors had characterized 41 popular DM tools. Finally; authors have concluded that with the help of a standard schema and a corresponding database, users are able to select a data mining software package, with respect to its ability, to meet high-level business objectives.

Collier et al. [9] have presented a framework for evaluating data mining tools and described a methodology for applying this framework. This methodology is based on firsthand experiences in data mining using commercial data sets from a variety of industries. Experience has suggested four categories of criteria for evaluating data mining tools: performance, functionality, usability, and support of ancillary activities. Authors have demonstrated that the assessment methodology takes advantage of decision matrix concepts to objectify an inherently subjective process. Furthermore, using a standard spreadsheet application, the proposed framework by [9] is easily automatable, and thus easy to be rendered and feasible to employ. Authors have showed that there is no single best tool for all data mining applications. Furthermore, there are a several data mining software tools that share the market leadership.

Abbott et al. [10] have compared five of the most highly acclaimed data mining tools on a fraud detection application. Authors have employed a two stage selection phase preceded by an in-depth evaluation. For the first stage, more than 40 data mining tools/vendors were rated depending on six qualities. The top 10 tools continued to the second stage of the selection phase and these tools were further rated on several additional characteristics. After selecting the 10 software packages, authors have used expert evaluators and re-rated each tool's characteristics, and the top five tools were selected for extensive hands-on evaluation. The selected tools and software were Clementine, Darwin, Enterprise Miner, Intelligent Miner, and PRW. The tools and software properties evaluated included the areas of client-server compliance, automation capabilities, breadth of algorithms implemented, ease of use, and overall accuracy on fraud-detection test data. Results have showed that the evaluated five products by authors would all display excellent properties; however, each may be best suited for a different environment. Authors have concluded that Intelligent Miner has the advantage of being the current market leader. Clementine excels in support provided and in ease of use. Enterprise Miner would especially enhance a statistical environment. Darwin is best when network bandwidth is at a premium. Finally, PRW is a strong choice when it's not obvious what algorithm will be most appropriate, or when analysts are more familiar with spreadsheets than UNIX.

Hen and Lee [1] have compared and analyzed the performance of five known data mining tools namely, IBM intelligent miner, SPSS Clementine, SAS enterprise miner, Oracle data miner, and Microsoft business intelligence development studio. 38 metrics were used to compare the performance of the selected tools. Test data was mined by various data mining methods ranging from different types of algorithms that are supported by the five tools, these includes classification algorithms, regression algorithms, segmentation algorithms, association algorithms, and sequential analysis algorithms. Results have provided a review of these tools and have proposed a data mining middleware adopting the strengths of these tools.

III. The Comparative Study

The methodology of the study constitute of collecting a set of free data mining and knowledge discovery tools to be tested, specifying the data sets to be used, and selecting a set of classification algorithm to test the tools' performance. Fig. 1 demonstrates the overall methodology followed for fulfilling the goal of this research.

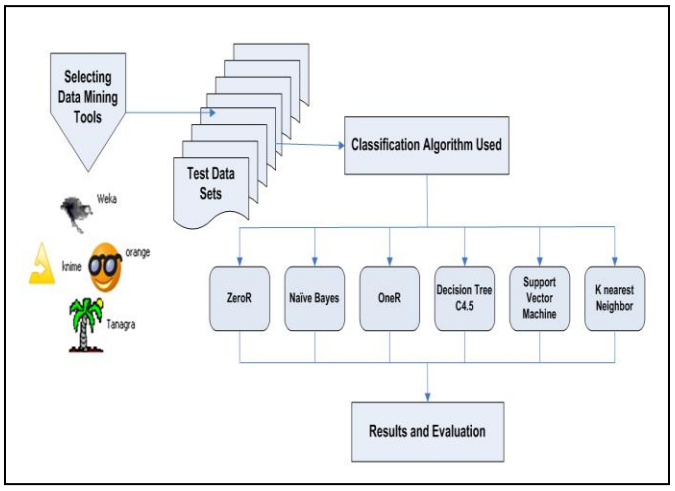


Figure 1. Methodology of Study.

A. Tools Description

The first step in the methodology consists of selecting a number of available open source data mining tools to be tested. Many open data mining tools are available for free on the Web. After surfing the Internet, a number of tools were chosen; including the Waikato Environment for Knowledge Analysis (WEKA), Tanagra, the Konstanz Information Miner (KNIME), and Orange Canvas.

- WEKA toolkit [12] is a widely used toolkit for machine learning and data mining that was originally developed at the University of Waikato in New Zealand. It contains a

large collection of state-of-the-art machine learning and data mining algorithms written in Java. WEKA contains tools for regression, classification, clustering, association rules, visualization, and data pre-processing. WEKA has become very popular with the academic and industrial researchers, and is also widely used for teaching purposes.

- Tanagra is free data mining software for academic and research purposes. It offers several data mining methods like exploratory data analysis, statistical learning and machine learning. The first purpose of the Tanagra project is to give researchers and students easy-to-use data mining software. The second purpose of TANAGRA is to propose to researchers an architecture allowing them to easily add their own data mining methods, to compare their performances. The third and last purpose is that novice developers should take advantage of the free access to source code, to look how this sort of software was built, the problems to avoid, the main steps of the project, and which tools and code libraries to use for. In this way, Tanagra can be considered as a pedagogical tool for learning programming techniques as well [13].
- KNIME (Konstanz Information Miner) is a user-friendly and comprehensive open-source data integration, processing, analysis, and exploration platform. From day one, KNIME has been developed using rigorous software engineering practices and is currently being used actively by over 6,000 professionals all over the world, in both industry and academia. KNIME is a modular data exploration platform that enables the user to visually create data flows (often referred to as pipelines), selectively execute some or all analysis steps, and later investigate the results through interactive views on data and models [14].
- Orange is a library of C++ core objects and routines that includes a large variety of standard and not-so-standard machine learning and data mining algorithms, plus routines for data input and manipulation. This includes a variety of tasks such as pretty-print of decision trees, attribute subset, bagging and boosting, and alike. Orange also includes a set of graphical widgets that use methods from core library and Orange modules. Through visual programming, widgets can be assembled together into an application by a visual programming tool called Orange Canvas. All these together make the Orange tool, a comprehensive, component-based framework for machine learning and data mining, intended for both experienced users and researchers in machine learning who want to develop and test their own algorithms while reusing as much of the code as possible, and for those just entering who can enjoy in powerful while easy-to-use visual programming environment [15].

B. Data Set Description

Once the tools have been chosen, a number of data sets are selected for running the test. For bias issues, several data sets have been downloaded from the UCI repository [16]. Table 1 shows the selected and downloaded data sets for testing purposes as shown in the table, each dataset is described by the data type being used, the types of attributes; whether they are categorical, real, or integer, the number of instances stored within the data set, the number of attributes that describe each dataset, and the year the dataset was created. Also, the table demonstrates that all the selected data sets are used for the classification task which is the main concentration of this paper.

These data sets were chosen because they have different characteristics and have addressed different areas, such as the number of instances which range from 100 to 20,000. Also, the number of attributes; which range from 5 to 70, and the attribute types; where some data sets contain one type while others contain two types. Such characteristics reflect different dataset shapes where some data sets contain a small number of instances but large number of attributes and vice versa.

TABLE1: UCI DATA SET DESCRIPTION

Data Set Name	Data Type	Default Task	Attribute Type	# Instances	# Attributes
Audiology (Standardized)	Multivariate	Classification	Categorical	226	69
Breast Cancer Wisconsin (Original)	Multivariate	Classification	Integer	699	10
Car Evaluation	Multivariate	Classification	Categorical	1728	6
Flags	Multivariate	Classification	Categorical, Integer	194	30
Letter Recognition	Multivariate	Classification	Integer	20000	16
Nursery	Multivariate	Classification	Categorical	12960	8
Soybean (Large)	Multivariate	Classification	Categorical	638	36
Spambase	Multivariate	Classification	Integer, Real	4601	57
Zoo	Multivariate	Classification	Categorical, Integer	101	17

C. Data Classification

Data classification is a two-step process: in the first step; a classifier is built describing a predetermined set of data classes or concepts. This is the learning step (or training phase), where a classification algorithm builds the classifier by analyzing or "learning from" a training set made up of database tuples and their associated class labels. In the second step, the model is used for classification; the predictive accuracy of the classifier is estimated using the training set to measure the accuracy of the classifier. The accuracy of a classifier on a given test set is the percentage of test set tuples that are correctly classified by the classifier. The associated class label of each test tuple is compared with the learned classifier's class prediction for that tuple. If the accuracy of the classifier is considered acceptable, the classifier can be used to classify future data tuples for which the class label is not known [2].

1. Classification Algorithm Description

After selecting the data sets, a number of classification algorithm are chosen for conducting the test. Many classification algorithms mentioned in literature are available for users such as Naïve Bayes (NB) algorithm [17] [18], K Nearest Neighbor (KNN) algorithm [18] [19] [20], Support Vector Machine (SVM) algorithm [21], and C4.5 algorithm [22. For testing purposes, we selected well known classifiers that are almost available in every open source tool, namely; Naïve Bayes (NB) classifier, One Rule (OneR) classifier, Zero Rule (ZeroR) classifier, Decision Tree Classifier; which is represented by the C4.5 Classifier, Support Vector Machine (SVM) classifier, and the K Nearest Neighbor (KNN) classifier.

2. Evaluation of Classificatiuon Algorithms

For evaluation purpose, two test modes were used; the k-fold Cross Validation (k-fold CV) mode and the Percentage Split (also called Holdout method) mode. The k-fold CV refers to a widely used experimental testing procedure where the

database is randomly divided into k disjoint blocks of objects, then the data mining algorithm is trained using k-1 blocks and the remaining block is used to test the performance of the algorithm; this process is repeated k times. At the end, the recorded measures are averaged. It is common to choose k = 10 or any other size depending mainly on the size of the original dataset.

In percentage split (Holdout) method, the database is randomly split into two disjoint datasets. The first set; which the data mining system tries to extract knowledge from, is called the training set. The extracted knowledge may be tested against the second set which is called the test set. In machine learning, to have the training and test sets, it is common to randomly split a dataset under the mining task into two parts. It is common to have 66% of the objects of the original database as a training set and the rest of objects as a test set [23].

The accuracy measure refers to the percentage of the correctly classified instances from the test data. The goal of testing, using the two modes, is to check whether there is an improvement in the accuracy measure when moving from the first test mode to the second test mode for all tools. Once the tests are carried out using the selected data sets, then using the available classifiers and test modes, results are collected and an overall comparison is conducted in order to determine the best tool for the classification purposes.

IV. Experiments and Evaluations

To evaluate the selected tools using the given datasets, several experiments were conducted. This section presents the results obtained after running the four data mining tools using the selected data sets described in Table 1.

A. Experiments Setup and Preliminaries

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation. Please keep your affiliations as succinct as possible (for example, do not differentiate among departments of the same organization). This template was designed for two affiliations.

As for experiments' setup, all tests were accomplished as follows: the holdout method has used 66% of each data set as training data and the remaining 34% as test data while the Cross Validation method used k = 10. The accuracy measure is used as a performance measure to compare the selected tools.

After running the four tools, we have obtained some results regarding the ability to run the selected algorithms on the selected tools. All algorithms ran successfully on WEKA; the six selected classifiers used the nine selected data sets.

As for the Orange tool, all classification techniques run successfully, except the OneR classifier; which is not implemented in Orange. For KNIME and Tanagra, Table 2 showed that some of the algorithms are unable to run some of the selected data sets. We noticed that this is due to one of the following three reasons; the first one is that the classifier is unable to run against the dataset because it is a multi-class data set and the classifier is only able to deal with binary classes; which are referenced in the tables with entry (MC). The second reason is that the classifier is unable to run the selected dataset because it contains discrete values and the algorithm is unable to deal with such kind of values; referenced in tables with (D) entry. The third reason is that the tool itself does not have an implementation for some classifiers; this reason is referenced in tables with not applicable (NA) entry.

We can notice that the One Rule algorithm (OneR) has no implementation in KNIME, Tanagra and Orange. Also, the ZeroR has no implementation in KNIME and Tanagra tools, and hence, it is referenced in tables as NA. On the other hand, tables shows that the K Nearest Neighbor (KNN) algorithm does not run against part of the data sets such as Audiology, Car, Nursery, and the SoyBean data sets because they contain some discrete values where the KNN algorithm cannot deal with. Finally, the Support Vector Machine does not run against any data sets; except the Breast-W and SpamBase data sets. This is because the other data sets are either containing a multi class data set and/or containing discrete values.

TABLE 2: ABILITY TO RUN SELECTED ALGORITHMS ON KNIME AND TANAGRA

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	OK	OK	OK	OK	OK	OK	OK	OK	OK
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	OK	OK	OK	OK	OK	OK	OK	OK	OK
SVM	MC/D	OK	MC/D	MC	MC	MC/D	MC/D	OK	MC
KNN	D	OK	D	OK	OK	D	D	OK	OK
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

* OK: Algorithm Run Successfully. NA: Algorithm has no Implementation. D: Discrete Value. MC: Multi Class

B. Evaluating the Performance of the Algorithms

For performance issues, Table 3 shows the results after running algorithms using WEKA toolkit. For the NB classifier, the accuracy measure has ranged between 44%-97%, while the OneR classifier accuracy has ranged between 5%-92%. For the C4.5 and SVM classifiers, results were almost the same for all data sets; where it ranged between 49%-96%. The KNN classifier has achieved accuracy measure values between 59%-98%. Finally, the ZeroR classifier has achieved the lowest accuracy measure for all of the data sets with accuracy measures ranging between 4%-70%.

TABLE 3: THE ACCURACY MEASURES GIVEN BY WEKA TOOL USING PERCENTAGE SPLIT.

	Audiology	Breast-W	Car	Flags	LETTERS	Nursery	SoyBean	SpamBase	Zoo
NB	71.43%	94.96%	87.59%	43.94%	64.47%	90.63%	90.56%	78.02%	97.14%
OneR	42.86%	92.02%	69.56%	4.55%	16.82%	70.41%	39.06%	77.83%	37.14%
C4.5	83.12%	95.38%	90.99%	48.48%	85.47%	96.48%	90.56%	92.20%	94.29%
SVM	84.42%	95.38%	93.37%	59.09%	81.13%	92.83%	93.99%	90.54%	94.29%
KNN	58.44%	95.38%	90.65%	51.52%	93.57%	97.53%	89.70%	89.27%	77.14%
ZeroR	27.27%	63.87%	69.56%	34.85%	3.90%	32.90%	13.30%	60.58%	37.14%

For the Orange toolkit, results are shown in Table 4. The NB classifier has achieved accuracy measures ranging between 52%-96%. The OneR classifier has no results as it has no implementation. For the C4.5 classifier, results have ranged between 51%-96%, while the SVM and KNN classifiers have achieved measures between 55%-97% and 56%-96% respectively. Finally, the ZeroR classifier has achieved the lowest measures for almost all data sets with values ranging between 4%-70%.

TABLE 4: THE ACCURACY MEASURES GIVEN BY ORANGE TOOL USING PERCENTAGE SPLIT.

	Audiology	Breast-W	Car	Flags	LETTERS	Nursery	SoyBean	SpamBase	Zoo
NB	70.13%	96.22%	86.90%	51.52%	61.44%	90.04%	92.24%	89.51%	88.24%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	72.73%	95.80%	91.50%	51.52%	85.50%	95.87%	79.74%	89.96%	91.18%
SVM	54.55%	95.38%	94.39%	56.06%	74.72%	97.07%	89.22%	92.26%	88.24%
KNN	76.62%	94.54%	88.78%	56.06%	95.84%	92.76%	92.67%	85.29%	85.29%
ZeroR	24.68%	65.55%	70.07%	34.85%	4.06%	33.33%	13.36%	60.61%	41.18%

Table 5 shows results achieved using the KNIME toolkit; for the NB classifiers results have ranged between 42%-95%. On the other hand, the ZeroR and OneR classifiers have no results because they have no implementation. The C4.5 classifier has achieved accuracy ranging between 43%-97%. The SVM and KNN classifiers did not run using some of the data sets because of the presence of one of the three reasons mentioned before; however, these classifiers have achieved measures between 67%-98% and 26%-97% respectively.

Finally, for the Tanagra tool, results are shown in Table 6 where the NB classifier has achieved an accuracy ranging between 60% and 96%. ZeroR and OneR classifiers have no results; because they have no implementation. C4.5 classifier has achieved results between 39% and 96%. On the other hand, SVM and KNN did not run using all the data sets as happened with KNIME; however, they both have achieved results ranging between 91%-97 and 29%-99% respectively.

TABLE 5: THE ACCURACY MEASURES GIVEN BY KNIME TOOL USING PERCENTAGE SPLIT

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	53.20%	95.00%	86.10%	42.40%	62.90%	90.50%	85.40%	89.80%	82.90%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	68.80%	95.00%	93.50%	43.10%	85.30%	96.70%	66.10%	91.10%	94.30%
SVM	MC/D	97.90%	MC/D	MC	MC	MC/D	MC/D	67.00%	MC
KNN	D	96.60%	D	25.80%	95.00%	D	D	80.90%	45.70%
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

TABLE 6: THE ACCURACY MEASURES GIVEN BY TANAGRA TOOL USING PERCENTAGE SPLIT

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	66.23%	95.80%	87.24%	63.64%	59.59%	90.74%	89.70%	87.54%	88.57%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	81.82%	92.44%	89.97%	39.39%	86.34%	96.32%	90.56%	90.73%	88.57%
SVM	MC/D	96.64%	MC/D	MC	MC	MC/D	MC/D	90.73%	MC
KNN	D	98.74%	D	28.79%	94.75%	D	D	79.17%	82.86%
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 7 shows the results obtained after running the algorithms against test data sets using the second test mode with 10-folds-CV. As shown in the table, the NB classifier has achieved accuracy measures ranging between 56%-96% while the OneR classifier has achieved measures ranging between 17%-93%. Both C4.5 and SVM classifiers have achieved accuracy ranging between 59%-97% and 61%-97% respectively. Accuracy measures ranging between 57%-98% were achieved using the KNN classifier. ZeroR rule classifier has achieved the lowest accuracy measures ranging between 4%-70%.

TABLE 7: THE ACCURACY MEASURES GIVEN BY WEKA USING 10-FOLDS-CV.

	Audiology	Breast-W	Car	Flags	LETTERS	Nursery	SoyBean	SpamBase	Zoo
NB	73.45%	95.99%	85.53%	55.15%	64.12%	90.32%	92.97%	79.29%	95.05%
OneR	46.46%	92.70%	70.02%	4.64%	17.24%	70.97%	39.97%	78.40%	57.43%
C4.5	77.87%	94.56%	92.36%	59.28%	87.98%	97.05%	91.51%	92.98%	92.08%
SVM	81.85%	97.00%	93.75%	60.82%	82.34%	93.08%	93.85%	90.42%	96.04%
KNN	62.83%	96.71%	93.52%	57.22%	95.52%	98.38%	90.19%	90.42%	95.05%
ZeroR	25.22%	65.52%	70.02%	35.57%	4.07%	33.33%	13.47%	60.60%	40.59%

Table 8 shows the accuracy measures using the Orange toolkit with 10-folds CV. As the table demonstrates, the NB classifier has achieved an accuracy measure ranging between 58%-97%. On the other hand, OneR has no accuracy measures because it has no implementation. For the C4.5 and SVM classifiers, the accuracy measures have ranged between 54%-96% and 64%-98% respectively. The KNN classifier has achieved accuracy measures ranging between 58%-96%. However, ZeroR has achieved the lowest measures ranging between 4%-70%.

TABLE 8: THE ACCURACY MEASURES GIVEN BY ORANGE USING 10-FOLDS-CV.

	Audiology	Breast-W	Car	Flags	LETTERS	Nursery	SoyBean	SpamBase	Zoo
NB	73.10%	97.14%	85.70%	58.24%	60.01%	90.29%	93.86%	89.31%	91.18%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	76.13%	95.85%	93.58%	54.03%	87.96%	96.57%	89.61%	90.64%	94.18%
SVM	64.23%	96.57%	95.54%	66.47%	76.58%	97.78%	93.27%	85.79%	92.09%
KNN	79.21%	95.71%	88.42%	57.61%	96.48%	92.63%	81.55%	88.71%	96.09%
ZeroR	25.24%	65.52%	70.02%	35.58%	4.06%	33.33%	13.18%	50.02%	40.46%

For the KNIME toolkit, Table 9 shows the results obtained using 10-folds CV as the test mode. The results of Table 9 shows that the NB classifier has achieved accuracy measures ranging between 52%-95%, the OneR and ZeroR classifiers have no accuracy measures because they have no implementation. The C4.5 classifier has achieved accuracy measures ranging between 55%-97%. Finally, both SVM and KNN classifiers have problems running some of the data sets, however, they have achieved accuracy measures ranging between 67%-96% and 33%-98% respectively.

TABLE 9: THE ACCURACY MEASURES GIVEN BY KNIME USING 10-FOLDS-CV.

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	59.30%	94.80%	85.80%	51.50%	61.60%	90.30%	91.20%	89.90%	88.10%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	70.70%	94.30%	93.50%	54.50%	87.50%	97.30%	72.00%	91.30%	93.10%
SVM	MC/D	96.30%	MC/D	MC	MC	MC/D	MC/D	67.30%	MC
KNN	D	97.50%	D	33.00%	95.40%	D	D	80.90%	71.30%
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

Table 10 shows the accuracy measures achieved using Tanagra with 10-folds CV test mode. As this table demonstrates, the NB classifier has achieved accuracy measures that have ranged between 63% and 96%. For the OneR and ZeroR classifiers, Tanagra has no implementation for such classifiers. On the other hand, the SVM and KNN classifiers have not achieved accuracy measures for all data sets because of some reasons; however, they have achieved accuracy measures that ranged between 90%-97% and 25%-97% respectively. Finally, the C4.5 classifier has achieved accuracy measures ranging between 57%-96%.

TABLE 10: THE ACCURACY MEASURES GIVEN BY TANAGAR USING 10-FOLDS-CV.

	Audiology	Breast-W	Car	Flags	Letters	Nursery	SoyBean	SpamBase	Zoo
NB	70.00%	95.80%	84.30%	62.63%	59.98%	89.91%	89.85%	88.28%	93.00%
OneR	NA	NA	NA	NA	NA	NA	NA	NA	NA
C4.5	71.36%	93.33%	86.45%	56.84%	85.84%	95.83%	90.24%	91.54%	88.00%
SVM	MC/D	96.96%	MC/D	MC	MC	MC/D	MC/D	89.98%	MC
KNN	D	96.81%	D	25.26%	95.76%	D	D	79.00%	92.00%
ZeroR	NA	NA	NA	NA	NA	NA	NA	NA	NA

C. Performance Improvement

In this section, it is worth to measure the effect of using different evaluation methods for the tools under study. Fig. 2 shows the performance improvements in accuracy when moving from the percentage split test mode to the 10-folds CV mode. This figure demonstrates that WEKA toolkit has achieved the highest improvements in accuracy with a 32 accuracy measures increase, when moving from the percentage split test to the CV test. Orange toolkit on the other hand, has achieved the second highest improvement with a 29 accuracy measures increase, when moving from the percentage split test to CV test. Finally, both KNIME and Tanagra toolkits have achieved the lowest improvements with 12 and 8 accuracy measures increase respectively.

In addition, Fig. 2 shows that the KNIME toolkit has achieved the best rate in terms of the number of accuracy measures decreased; only 4 accuracy measures are decreased when moving from the percentage split test to the CV test in KNIME. For the Orange and WEKA toolkits, the number of accuracy measures decreased where 6 and 7 respectively Finally, the Tanagra toolkit has achieved the least rate with the number of 9 accuracy measures decrease when moving from the percentage split test to the CV test.

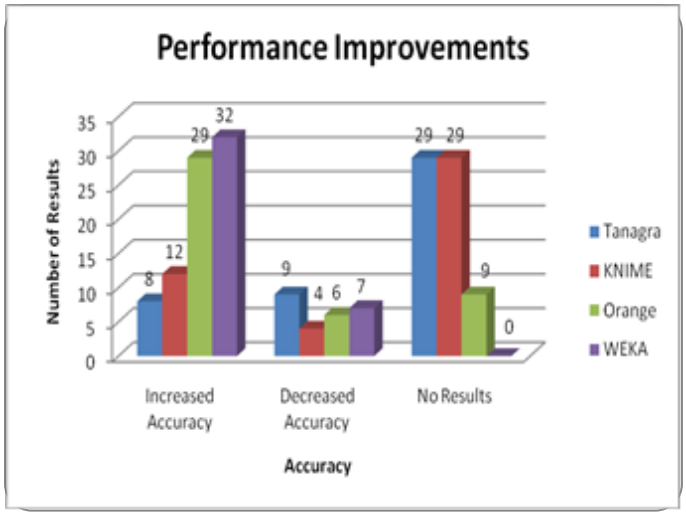


Figure 2. *Performance Improvements*.

Finally, when comparing the four tools in terms of the number of tests that produced no accuracy; the Tanagra and KNIME toolkits have achieved the highest number of tests with no accuracy measures; with a 29 measures each. For the Orange toolkit, only 9 tests have no accuracy measures. On the other hand, WEKA toolkit has 0 tests with no accuracy measures.These results showed that no tool is better than the other to be used for a classification task, this is may be due to the kind of data sets used, or maybe there are some differences in the way the algorithms were implemented within the tools themselves (for example the SVM classifier implemented in WEKA and Orange can handle the problem of multiclass data sets; which is not the case in Tanagra and KNIME that were designed to handle only two class problems).

In terms of applicability (the ability to run a specific algorithm on a selected tool), the WEKA toolkit has achieved the highest applicability, since it is able to run the six selected classifiers using all data sets. Orange Canvas toolkit has scored the second place in terms of applicability, since it run five classifiers out of the six selected classifiers with no ability to run the OneR Classifier. Finally; the KNIME and Tanagra toolkits have both achieved the lowest applicability with the ability to run two classifiers namely; NB and C4.5 on all data sets completely, and partially using another two classifiers namely; SVM and KNN classifiers, while it has no ability to run the last two classifiers namely; OneR and ZeroR classifiers.

In terms of performance improvements, we can judge that WEKA and Orange toolkits have achieved the highest improvements with a 32 and 29 values increased respectively and only 7 and 6 values decreased respectively. On the other hand, the KNIME and Tanagra toolkits have achieved the lowest improvements with 12 and 8 values increased in accuracy respectively and 4 and 9 values decreased respectively.

V. Conclusion and Future Work

This research has conducted a comparison between four data mining toolkits for classification purposes, nine different data sets were used to judge the four toolkits tested using six classification algorithms namely; Naïve Bayes (NB), Decision Tree (C4.5), Support Vector Machine (SVM), K Nearest Neighbor (KNN), One Rule (OneR), and Zero Rule (ZeroR). This study has concluded that no tool is better than the other if used for a classification task, since the classification task itself is affected by the type of dataset and the way the classifier was implemented within the toolkit. However; in terms of classifiers' applicability, we concluded that the WEKA toolkit was the best tool in terms of the ability to run the selected classifier followed by Orange, Tanagra, and finally KNIME respectively.

Finally; WEKA toolkit has achieved the highest performance improvements when moving from the Percentage Split test mode to the Cross Validation test mode followed by Orange, KNIME, and then Tanagra Respectively. As a future research, we are planning to test the selected data mining tools for other machine learning tasks; such as clustering, using test data sets designed for such tasks and the known algorithms for clustering and association.

References

[1] Goebel, M., Gruenwald, L., A survey of data mining and knowledge discovery software tools, ACM SIGKDD Explorations Newsletter, v.1 n.1, p.20-33, June 1999 [doi>10.1145/846170.846172].

[2] Han, J., Kamber, M., Jian P., Data Mining Concepts and Techniques. San Francisco, CA: Morgan Kaufmann Publishers, 2011.

[3] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann P., Witten, I., H., The WEKA data mining software: an update, ACM SIGKDD Explorations Newsletter, v.11 n.1, June 2009 [doi>10.1145/1656274.1656278].

[4] Hornik, K., Buchta, C., Zeileis, A., Open-Source Machine Learning: R Meets Weka, Journal of Computational Statistics - Proceedings of DSC 2007, Volume 24 Issue 2, May 2009 [doi>10.1007/s00180-008-0119-7]. Hunyadi, D., Rapid Miner E-Commerce, Proceedings of the 12th WSEAS International Conference on Automatic Control, Modelling & Simulation, 2010.

[5] Al-Harbi, S., Almuhareb, A., Al-Thubaity, A., Khorsheed, S., and Al-Rajeh, A., "Automatic Arabic Text Classification", 9th International journal of statistical analysis of textual data, pp. 77-83, 2008.

[6] King, M., A., and Elder, J., F., Evaluation of Fourteen Desktop Data Mining Tools, in Proceedings of the 1998 IEEE International Conference on Systems, Man and Cybernetics, 1998.

[7] Giraud-Carrier, C., and Povel, O., Characterising Data Mining software, Intelligent Data Analysis, v.7 n.3, p.181-192, August 2003

[8] Carey, B., Marjaniemi, C., Sautter, D., Marjaniemi, C., A Methodology for Evaluating and Selecting Data Mining Software, Proceedings of the Thirty-second Annual Hawaii International Conference on System Sciences-Volume 6, January 05-08, 1999.

[9] Abbot, D. W., Matkovsky, I. P., Elder IV, J. F., An Evaluation of High-end Data Mining Tools for Fraud Detection, IEEE International Conference on Systems, Man, and Cybernetics, San Diego, CA, October, pp. 12--14, 1998.

[10] Hen, L., E., and Lee, S., P., Performance analysis of data mining tools cumulating with a proposed data mining middleware, Journal of Computer Science: 2008.

[11] WEKA, the University of Waikato, Available at: http://www.cs.waikato.ac.nz/ml/weka/, (Accessed 20 April 2011).

[12] Tanagra – a Free Data Mining Software for Teaching and Research, Available at: http://eric.univ-lyon2.fr/~ricco/tanagra/en/tanagra.html, (Accessed 20 April 2011).

[13] KNIME (Konstanz Information Miner), Available at: http://www.knime.org/, (Accessed 20 April 2011).

[14] Orange – Data Mining Fruitful and Fun, Available at: http:// http://orange.biolab.si/, (Accessed 20 April 2011).

[15] UCI Machine Learning Repository, Available at: http:// http://archive.ics.uci.edu/ml/, (Accessed 22 April 2011).

[16] Flach, P., A., Lachiche, N., Naive Bayesian Classification of Structured Data, Machine Learning, v.57 n.3, p.233-269, December 2004.

[17] Heb, A., Dopichaj, P., Maab, C., Multi-value Classification of Very Short Texts, KI '08 Proceedings of the 31st annual German conference on Advances in Artificial Intelligence, pp. 70-77, 2008,

[18] Zhou, S., Ling, T., W., Guan, J., Hu, J., Zhou, A., Fast Text Classification: A Training-Corpus Pruning Based Approach. Proceedings of the Eighth International Conference on Database Systems for Advanced Applications, pp.127, 2003.

[19] Pathak, A., N., Sehgal M., Christopher, D., A Study on Selective Data Mining Algorithms, IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 2, March 2011.

[20] Li, Y., Bontcheva, K., dapting Support Vector Machines for F-term-based Classification of Patents, Journal ACM Transactions on Asian Language Information Processing, Volume 7 Issue 2, June 2008.

[21] Al-Radaideh, Q., Al-Shawakfa, E., Al-Najjar, M., I., Mining Student Data Using Decision Trees, The 2006 International Arab Conference on Information Technology (ACIT2006), December 19-21, 2006.

[22] Al-Radaideh, Q., The Impact of Classification Evaluation Methods on Rough Sets Based Classifiers, Proceedings of the 2008 International

Arab Conference on Information Technology (ACIT2008). University of Sfax, Tunisia. December 2008.

AUTHORS PROFILE

Abdullah H. Wahbeh is a lecturer in the department of Computer Information Systems at Yarmouk University in Jordan. He obtained his Master and bachelor degrees in Computer Information Systems (CIS) from Yarmouk University, Irbid-Jordan. His research interests include: data mining, web mining and information retrieval.

Qasem A. Al-Radaideh is an Assistant Professor of Computer Information Systems at Yarmouk University. He got his Ph.D. in Data Mining field from the University Putra Malaysia in 2005. His research interest includes: Data Mining and Knowledge Discovery in Database, Natural Language Processing, Arabic Language Computation, Information Retrieval, and Websites evaluation. He has several publications in the areas of Data Mining and Arabic Language Computation. He is curently the national advisor of Microsoft students partners program (MSPs) and MS-Dot Net Clubs in Jordan.

Mohammed N. Al-KABI is an Assistant Professor in the Department of Computer Information Systems at Yarmouk University. He obtained his Ph.D. degree in Mathematics from the University of Lodz (Poland). Prior to joining Yarmouk University, he spent six years at the Nahrain University and Mustanserya University (Iraq). His research interests include Information Retrieval and Search Engines, Data Mining, and Natural Language Processing.

Emad M. Al-Shawakfa is an Assistant Professor at the Computer Information Systems Department at Yarmouk University since September 2000. He holds a PhD degree in Computer Science from Illinois Institute of Technology (IIT) – Chicago, USA in the year 2000. His research interests are in Computer Networks, Data Mining, and Natural Language Processing. He has several publications in these fields and currently working on others.

SOM Based Visualization Technique For Detection Of Cancerous Masses In Mammogram

S.Pitchumani Angayarkanni M.C.A,M.Phil,(Ph.d)
Assistant Professor
Department of Computer Science
Lady Doak College
Madurai

Dr.V.Saravanan
H.O.D,Department of M.C.A
Karunya Deemed University
Coimbatore, Tamil Nadu

***Abstract*— Breast cancer is the most common form of cancer in women. An intelligent computer-aided diagnosis system can be very helpful for radiologist in detecting and diagnosing micro calcifications patterns earlier and faster than typical screening programs. In this paper, we present a system based on gabor filter based enhancement technique and feature extraction techniques using texture based segmentation and SOM(Self Organization Map) which is a form of Artificial Neural Network(ANN) used to analyze the texture features extracted. SOM determines which texture feature has the ability to classify benign, malignant and normal cases. Watershed segmentation technique is used to classify cancerous region from the non cancerous region. We have investigated and analyzed a number of feature extraction techniques and found that a combination of ten features, such as Cor-relation, Cluster Prominence, Energy, Entropy, Homogeneity, Difference variance, Difference Entropy, Information Measure, and Normalized are calculated. These features gives the distribution of tonality information and was found to be the best combination to distinguish a benign micro calcification pattern from one that is malignant and normal. The system was developed on a Windows platform. It is an easy to use intelligent system that gives the user options to diagnose, detect, enlarge, zoom, and measure distances of areas in digital mammograms. Further Using Linear Filtering Technique and the Texture Features as Mask are convolved with the segmented image .The tumor is detected using the above method and using watershed segmentation, a fair segmentation is obtained The artificial neural network with unsupervised learning together with texture based approach leads to the accuracy and positive predictive value of each algorithm were used as the evaluation indicators. 121 records acquired from the breast cancer patients at the MIAS database. The results revealed that the accuracies of texture based unsupervised learning has 0.9534 (sensitivity 0.98716 and specificity 0.9582 which was detected thorough the ROC. The results showed that the gabor based unsupervised learning described in the present study was able to produce accurate results in the classification of breast cancer data and the classification rule identified was more acceptable and comprehensible.**

Keywords- Image Enhancement; Gabor Filter; Texture Features; SOM; ROC.

I. INTRODUCTION

Breast cancer is one of the major causes for the increased morality cause many among women especially in developed countries. It is second most common cancer in women. The World Health Organization's International estimated that more than 1,50,000 women worldwide die of breast cancer in year. In India , breast cancer accounts for 23% of all the female cancer death followed by cervical cancer which accounts to 17.5% in India. Early detection of cancer leads to significant improvements in conservation treatment[1]. However, recent studies have shown that the sensitivity of these systems is significantly decreased as the density of the breast increased while the specificity of the systems remained relatively constant. In this work **w**e have developed automatic neuron genetic algorithmic approach to automatically detect the suspicious regions on digital mammograms based on asymmetries between left and right breast image.

Diagnosing cancer tissues using digital mammograms is a time consuming task even highly skilled radiologists because mammograms contain low signal to noise ratio and a complicated structural background. Therefore in digital mammogram, there is still a need to enhance imaging, where enhancement in medical imaging is the use of computers to make image clearer. This may aid interpretation by humans or computers. Mammography is one of the most promising cancer control strategies since the cause of cancer is still unknown[2].

Radiologist turn to digital mammography as an alternative diagnostic method due to the problems created by conventional screening programs. A digital mammogram is created when conventional mammogram is digitized; through the use of a specific mammogram is digitizer or a camera, so it can be processed by the computer. Image enhancement methods are applied to improve the visual appearance of the mammograms. Initially the mammogram image is read from the dataset and partial filter (Combination of Low and high Pass filter) is applied to remove the noise from the image.

Gabor filter with texture based segmentation and SOM based clustering techniques were used in efficient detection of cancerous masses in mammogram MRI images . The selected features are fed to SOM Network hybrid to detect the texture feature estimation variable which varies highly for normal, benign and malignant tumor[2] . The watershed algorithm is used to segment the cancerous region based on the feature which plays a vital role in classification and the Receiver Operating Characteristic (ROC) analysis is performed to evaluate the performance of the feature selection methods In this paper various steps in detection of microcalcification such as i) Preprocessing and enhncement using Histogram

Equalization and Gabor filter ii) Texture based Segmentation iii) SOM based Visualization iv) Watershed segmentation V) TIC – TOC method to find how efficiency in automatic detection system.

II. Image Acquisition

Mammograms used in this research are retrieved from the MIAS(Mammographic Image Analysis Society) database. The databases contains 161 mammogram records including normal and microcalcification cases.

The Mammography Image Analysis Society (MIAS), which is an organization of UK research groups interested in the understanding of mammograms, has produced a digital mammography database (ftp://peipa.essex.ac.uk)[2]. The data used in these experiments was taken from the MIAS. The database contains left and right breast images for 151 records, is used. Its quantity consists of 302 images, which belong to three types such as Normal, benign and malignant. There are 200 normal, 44 benign and 58 malignant (abnormal) images.

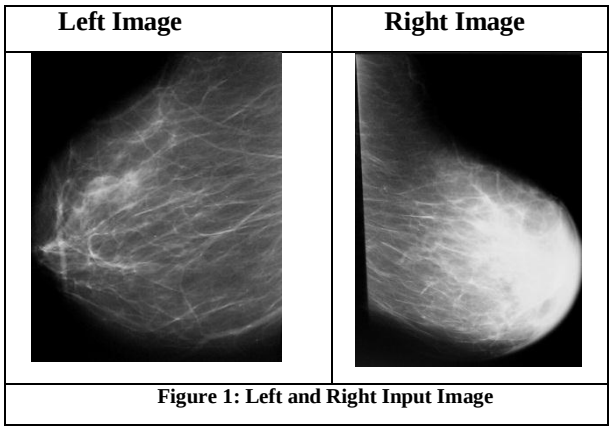


Figure 1: Left and Right Input Image

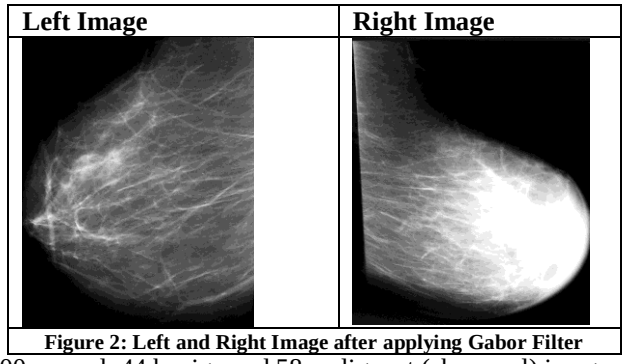


Figure 2: Left and Right Image after applying Gabor Filter

III. Preprocessing & Enhancement Techniques

One of the most important problems in image processing is denoising. Usually the procedure used for denoising, is dependent on the features of the image, aim of processing and also post-processing algorithms. Preprocessing is the first phase of image analysis. The purpose of preprocessing is to improve the quality of the image being processed. It makes the subsequent phases of image processing easier. Median Filter is used for smoothening to retain the edge strengths.

IV. Gabor Filter

Gabor filter is a linear filter used for edge detection. Frequency and orientation representations of Gabor filter are similar to those of human visual system, and it has been found to be particularly appropriate for texture representation and discrimination[3]. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.

Its impulse response is defined by a harmonic function multiplied by a Gaussian function. Because of the multiplication-convolution property (Convolution theorem), the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of the harmonic function and the Fourier transform of the Gaussian function.

The filter has a real and an imaginary component representing orthogonal directions. The two components may be formed into a complex number or used individually.

Complex

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma)=\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)\exp\left(i\left(2\pi\frac{x'}{\lambda}+\psi\right)\right)$$

Real

Imaginary

$$g(x,y;\lambda,\theta,\psi,\sigma,\gamma)=\exp\left(-\frac{x'^2+\gamma^2y'^2}{2\sigma^2}\right)\sin\left(2\pi\frac{x'}{\lambda}+\psi\right)$$

where

$$x'=x\cos\theta+y\sin\theta$$
$$y'=-x\sin\theta+y\cos\theta$$

Gabor wavelet filters smooth the image by blocking detail information. Mass detection aims to extract the edge of the tumor from surrounding normal tissues and background. PSNR, RMS, MSE, NSD, ENL value calculated for each of 121 pairs of mammogram images clearly shows that gabor wavelet filter when applied to mammogram image leads to best Image Quality[4].

The orientation and scale can be changed in this program to extract texture information. Here 3 scales and 4 orientation was used.

Table 1: Estimated Values

PSNR	RMS	NSD	ENL	MES
87.65	2.97	4.55	89.89	8.83

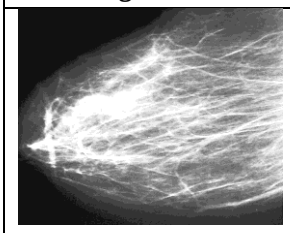
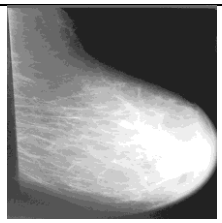
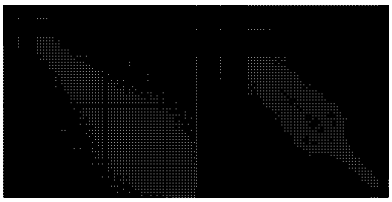
Left Image	Right Image
	

Figure 3: Left and Right Image after applying Histogram Equalization

V. TEXTURE BASED APPROACH

Dhawan et al has specified image structural features for classification as benign and malignant. The textural features are extracted using GLCM(gray level co-occurrence matrix) . They specify specific textural characteristics such as homogenetity, contrast, entropy,energy and regularity of the structure within the image [27,36,37,39,43,86]. The Image is divided into 3*3 pixel values and the following parameters were evaluated .Correlation, Cluster Prominence, Energy, Entrpy, Homogenity, Diff variance,Diff[4]. Entropy and Infinity measure.It was found that the region which is malignant has energy and homogeneity value=0.5 to 1, infinity measure between 0.5 – 1.An image of size 268*100 is divided into 3*3 kernel window and for a total of 289 subimages the texture values were evaluated and clustered based on the above estimated values.



A) Left Image B)Right Image

Figure 4: Texture based Segmentation

Texture Analysis analyzes the textures extracted and finds which aids in detect ability of tumor and it also helps us to find the relationship between the textures extracted. Texture Analysis can be done by using SOM (Self Organization Map)

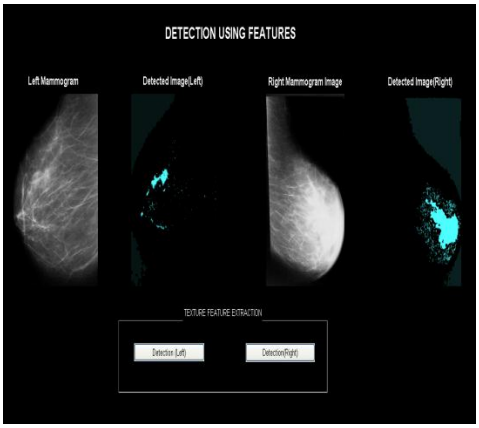


Figure 5 :Feature detection

Correlation	$\sum\sum(i,j)-\mu_x\mu_y/\sigma_x\,\sigma_y$ where μ_x, μ_y and σ_x σ_y are the means and std deviations of ρx and ρy, the partial probability density fucntions
Entropy	$-\sum_i\sum_j \rho(i,j)\log(\rho(i,j))$
Difference Entropy	$-\sum_{i=0}^{N_{0-i}} \rho_{x-y}(i)\log\{\rho_{x-y}(i)\}$
Difference Variance	Variance of ρ_{x-y}
Information Measure	HXY-HXY1/max{HX,HY} Where HX and HY are entropies of ρx and ρy HXY1= $-\sum_i\sum_j \rho(i,j)\log\{\rho_x(i)\,\rho_y(j)\}$ HXY2$_=$ $-\sum_i\sum_j \rho_x(i)p_y(j)\log\{\rho_x(i)\,\rho_y(j)\}$
Energy	$\sum_i(\sum_j(1+(i-j)^2))$
Homogeneity	$\sum_i\left(\sum_j p(i,j)/(1+(i-j)^2)\right)$
Normalized	$\sum_i\sum_j \frac{p(i,j)}{1+\lvert(i-j\rvert}$

VI. SELF ORGANIZING MAP

A self-organized map (SOM)[8]-type of artificial neural network that is trained using unsupervised learning to produce a two-dimensional, representation of the training samples, called a map (Kohonen Map). SOMs operate in two modes: training and mapping. Training builds the map using input examples. It is a competitive process, also called vector quantization. Mapping automatically classifies a new input vector[6].

A self-organizing map consists of components called nodes or neurons. Associated with each node is a weight vector of the same dimension as the input data vectors and a position in the map space. The usual arrangement of nodes is a regular spacing in a hexagonal or rectangular grid.

The self-organizing map describes a mapping from a higher dimensional input space to a lower dimensional map space. The procedure for placing a vector from data space onto the map is to find the node with the closest weight vector to the vector taken from data space and to assign the map coordinates of this node to our vector.

Illustration of 'Training'

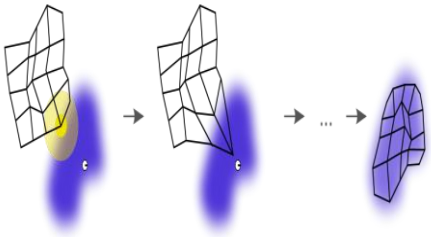


Figure 6 Training Process

The blue blob is the distribution of the training data, and the small white disc is the current training sample drawn from that distribution. At first (left) the SOM nodes are arbitrarily positioned in the data space[5]. The node nearest to the training node (highlighted in yellow) is selected, and is moved towards the training datum, as (to a lesser extent) are its neighbors on the grid. After much iteration the grid tends to approximate the data distribution (right).

Mapping

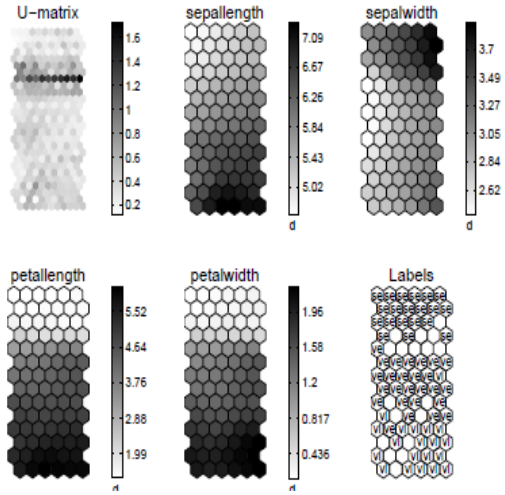


Figure 7 Mapping Visualization

The trained samples are mapped depending upon the values of the samples. Clustering is differentiated using different colors. color map can be HSV, Grayscale, RGB etc.

This approach uses SOM toolbox to perform training and mapping. This toolbox contains GUI based application which is easy to use implementation. Toolbox can be used to preprocess data, initialize and train SOMs using a range of different kinds of topologies, visualize SOMs in various ways, and analyze the properties of the SOM and data.

SOM toolbox in this approach has helped to visualize the relationship between the features and also how the feature varies for different types of cases like (Benign, Malignant and Normal).

From this project it is found that 'Information Measure related to Cor-relation' varies for the above specified cases during Mapping and it is also found that Energy and Entropy are oppositely cor-related.

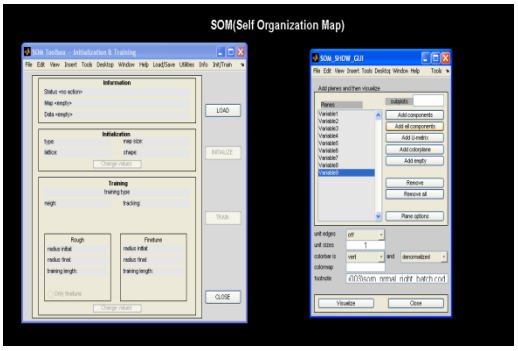


Figure 8:SOM Toolbox

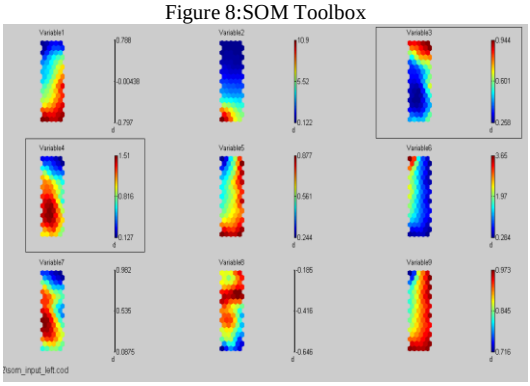


Figure 9:Visualization of texture parameter analysis using SOM

Analysis:

- In case of **Normal**, there is no difference in the clustering and we can find that 'Energy' and 'Entropy' are oppositely related
- In case of **'Benign'**, there is a difference in the clustering of 'Information Measure related to cor relation' and we can also find that 'Energy' and 'Entropy' are oppositely related.
- In case of **'Malignant'**, there is a difference in the clustering of 'Information Measure related to cor relation' and we can also find that 'Energy' and 'Entropy' are oppositely related

VII. Watershed Segmentation

This algorithm segments regions of an image. When tumor detected image is passed to the watershed transform, tumor region is shown in a different color. Watershed Transform is usually well suited for bubble or metallographic images but when combined with Linear Filtering Methods, it has been proved that it can segment regions of any shape.

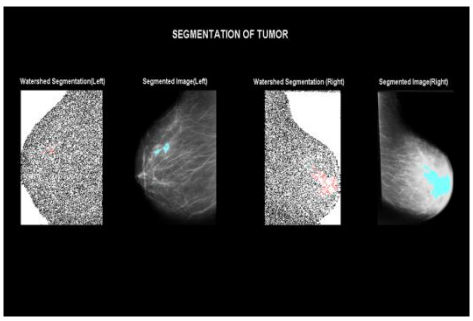


Figure 10: Segmentation of Malignant tumor using watershed transform

VIII. Results and Discussions

From this project, one can analyze that

1. Using SOM, depending on the type of clustering, one can find that 'Energy' and 'Entropy' are oppositely related.' Information Measure related to cor-relation' differs for benign and malignant cases during Mapping.
2. Watershed Transform was usually well suited for bubble or metallographic images but when combined with Linear Filtering Methods, it has been proved in this project that it can segment regions of any shape[7,8].
3. The Time Taken(CPU Time) by the above technique was calculated using Tic Toc method and it was found that it take 3'50''ms.The efficiency of the technique was compared with other methods and was found that

Table 2 TIC-TOC time estimation

Computation Time for Different Methods		
Methods	**Author and References**	**Computational Time**
Morphological Analysis	Wan Mimi Diyana, Julie Larcher, Rosli Besar	3'20"
Filtering Technique	Proposed Approach	3'50"
Fractal Dimension Analysis	Wan Mimi Diyana, Julie Larcher, Rosli Besar	7'20"
Complete HOS Test	Wan Mimi Diyana, Julie Larcher, Rosli Besar	9'20"

Though the Time Taken is slower than 'Morphological Analysis' Method **(3'20'')** but the efficiency is better than the other methods.

IX. Conclusion

One of the major causes of death among women is due to breast cancer. so early diagnosis through regular screening and timely treatment has been shown to prevent cancer. This paper focuses on the approach to identify the presence of breast cancer mass in mammograms[9,10]. The proposed work utilizes Filtering Techniques for detection and watershed segmentation for segmentation.

Preprocessing with Gabor filter and using the filtering technique with watershed transform is a new approach, using this we have successfully detected the breast cancer masses in mammograms. The results indicate that this system can facilitate the doctor to detect tumor in the early stage of diagnosis process with a fraction of second.

Table 3: Comparison of our method with other methods:

Authors and References	**Methods**	**Detection Rate**
Ferrari and Rengayyan[19]	Directional Filtering with Gabor Wavelets	74.4%
Lau and Bischof[20]	Asymmetry measures	85%
Sallam and Bowyer[21]	Unwrapping Technique	86.6%
Thangavel and Karnan[22]	Metaheuristic approach Redundant Wavelets Wavelets Packets	94.8% 84.79% 87.68%

Proposed Method:

Method	**Detection Rate**
Histogram based Gabor Wavelet approach Texture based approach SOM Based Clustering	96.67% 97.78% 99.78%
Overall Accuracy of Combined approach	**98.07%**

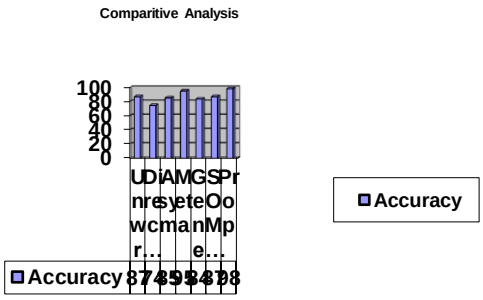


Figure 11: Comparative Study

References

[1] Bosch. A.; Munoz, X.; Oliver.A.; Marti. J., Modeling and Classifying Breast Tissue Density in Mammograms, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on Volume 2, Issue , 2006 Page(s): 1552 – 15582.

[2] Dar-Ren Chena, Ruey-Feng Changb, Chii-Jen Chenb, Ming-Feng Hob, Shou-Jen Kuoa, Shou-Tung Chena, Shin-Jer Hungc, Woo Kyung Moond,Classification of breast ultrasound images using fractal feature, ClinicalImage, Volume 29, Issue4, Pages 234-245.

[3] Suri, J.S., Rangayyan, R.M.: Recent Advances in Breast Imaging, Mammography,and Computer-Aided Diagnosis of Breast Cancer. 1st edn. SPIE (2006)

[4] Hoos, A., Cordon-Cardo, C.: Tissue microarray pro.ling of cancer specimens and cell lines: Opportunities and limitations. Mod. Pathol. 81(10), 1331–1338 (2001)

[5] Lekadir, K., Elson, D.S., Requejo-Isidro, J., Dunsby, C., McGinty, J., Galletly, N.,Stamp, G., French, P.M., Yang, G.Z.: Tissue characterization using dimensionality reduction and .uorescence

imaging. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 586–593. Springer, Heidelberg (2006).

[6]. Dr. WIlliam H. Wolberg, (no date), Breast Cancer Wisconsin Dataset (online) (http://www.radwin.org/michael/projects/learning/about-breast-cancer-wisconsin.html (1July 2005)

[7] J.A.M. van Dijck, L.M. Verbeek, J.H.C.L. Hendriks, The current delectability of breast cancer in a mammographic screening program, Cancer 72 (1933) 1938–1993.

[8] I.W. Hutt, S.M. Astley, C.R.M. Boggis, Prompting as an aid to diagnosis in mammography, in: A.G. Gale, S.M. Astley, D.R. Dance,A.Y. Cairns (Eds.), Digital Mammography, Elsevier, Amsterdam, 1994, pp.389–398.

[9] H. Li, Y. Wang, K.J. Ray Liu, Computerized radiographic mass detection. Part II: decision support by featured database visualization and modular neural networks, IEEE. Trans. Med. Imag. 20 (4) (2001) 302– 313.

[10] Pfarl G. Breast Imagimg Reporting and Data System. 2006; Available from: http://www.birads.at, Access at: 2006Oct24.

[11] Otsu N. A threshold selection method from gray-level histogram. IEEE Transactions on System Man Cybernetics 1979; 9(1): 62-66.

[12] A. R. Abdel-Dayem, M.R. El-Sakka, Fuzzy Entropy Based Detection of Suspicious masses in Digital Mammogram Images. 27th Annual Conference of IEEE, China 2005, September 1-4.

[13] L. Li, W. Qian, L.P. Clarke, Image feature extraction for mass detection using digital mammography: effects of wavelet analysis, Proc. SPIE Med. Imag. 3338 (1998) 168–1176 Med. Phys. 26(3) (1999).

[14] L. Li, W. Qian, L.P. Clarke, R.A. Clark, Improving mass detection by adaptive and multi-scale processing in digitized mammograms, SPIE (Society of Photographic Scientists and Engineer) Conference on Image Processing, San Diego, CA, vol. 3661, February 1999, pp. 490–498.

[15] Silvano DZ, Luigi C, Stefano L. Image thresholding using fuzzy entropies. IEEE Trans. on systems man and cybernetics, Part B 1998; 28(1):15-23.

[16] B. S. Manjunath, W. Y. Ma, Texture features for browsing and retrieval of image data, IEEE Trans. Pattern Anal. & Machine Intell., vol.18, no. 8, pp.837–842, 1996.

[17] Three-Dimensional Technique for Automatic Brain Segmentation of the Ventricles Based on Optimal Histogram Thresholds of MRI, Danmary Sanchez, Malek Adjouadi, Byron Bernal, Nolan Altman, WSEAS TRANSACTIONS on COMPUTERS, Issue 7, Volume 4, July 2005, ISSN 1109-2750.

[18] Comparison of Three Gaussian Mixture Modeling and Spatial Encoding Methods for Segmenting Human brain MRI, Mahmood Zeydabadi-Nejad, Reza A. Zoroofi, Hamid Soltanian- Zadeh, WSEAS TRANSACTIONS on ELECTRONICS, Issue 3, =Volume 1, July 2004, ISSN 1109-9445.

[19] Fuzzy Rule Based Classifiers from Support Vector Learning, Stergios Papadimitriou, Konstantinos Terzidis, Seferina Mavroudi, Lambros Skarlas , Spiridon Likothanasis, Issue 7, Volume 4, July 2005, ISSN 1109-2750.

AUTHORS PROFILE

S.Pitchumani Angayarkanni M.C.A.,M.Phil.,(Ph.D) working in the department of computer science,Lady Doak College, Madurai for the past ten years. My areas of interest are medical image processing, Neural Network & Data Mining.

Improvement of Secret Image Invisibility in Circulation Image with Dyadic Wavelet Based Data Hiding with Run-Length Coded Secret Images of Which Location of Codes are Determined with Random Number

Kohei Arai
Dept. of Information Science,
Saga University
Saga, Japan

Yuji Yamada
Dept. of Information Science
Saga University
Saga, Japan

***Abstract*— An attempt is made for improvement of secret image invisibility in circulation images with dyadic wavelet based data hiding with run-length coded secret images of which location of codes are determined by random number. Through experiments, it is confirmed that secret images are almost invisible in circulation images. Also robustness of the proposed data hiding method against data compression of circulation images is discussed. Data hiding performance in terms of invisibility of secret images which are embedded in circulation images is evaluated with the Root Mean Square difference between the original secret image and extracted one from the circulation images. Meanwhile the conventional Multi-Resolution Analysis (MRA) based data hiding is attempted with a variety of parameters, level of MRA and the frequency component location of which secret image is replaced to it and is compared to the proposed method. It is found that the proposed data hiding method is superior to the conventional method. Also the conventional data hiding method is not robust against circulation image processing.**

Keywords- Dyadic wavelet; Lifting wavelet; Data hiding; Data compression.

I. INTRODUCTION

Wavelet analysis applications are getting more popular in time series analysis, image analysis, information security area, etc.[1],[2]. Data hiding is important for information contents security, in particular, for protection of copy right. Using data hiding methods, some important information such as copyright, signature, etc. can be embedded. Data hiding based on wavelet analysis, in particular, Multi-Resolution Analysis: MRA is widely used. One of the problems on data hiding methods is visibility of the embedded information on the available circulation images [3]-[7]. The other problem is robustness against image processing which is applied to the circulation images including data compressions. It sometime occurs that small amount of information on the embedded image appears on the circulation images slightly due to the embedding mechanism of the data hiding.

In order to improve invisibility of the secret images in the circulation images, run-length coded binarized secret images are used. The locations of the codes after the data compression in one of the frequency component images after the dyadic wavelet transformation [8] are determined with random numbers generated by Merssene Twister of random number generator. After all, reconstructed image (inverse dyadic wavelet transformation) is used for circulation. The original secret images are almost invisible in the circulation images. This paper deals with the current problems on the widely used MRA based data hiding method (Conventional data hiding method). One of the problems is visibility of secret image in the circulation images followed by robustness against circulation image manipulations including image deformation, geometric conversion, data compression, etc. In order to overcome the aforementioned problems, a method for data hiding based on lifting dyadic wavelet transformation with run-length coding of data compression which is applied to secret image together with pixel order exchange is proposed. First, the aforementioned problems of the conventional MRA based data hiding method are discussed followed by the proposed method. Then robustness against JPEG and JPEG 2000 of data compression is discussed in the paper.

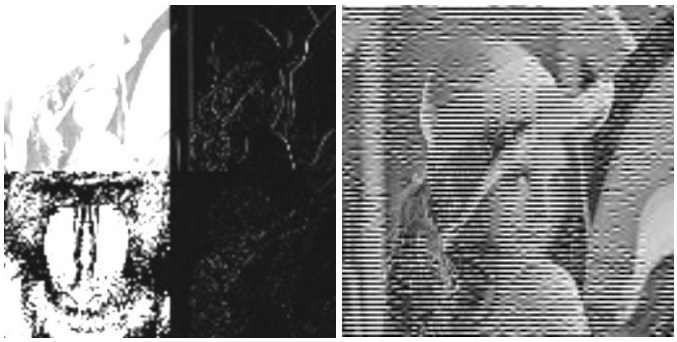
II. PROPOSED METHOD

A. Conventional wavelet based data hiding methods

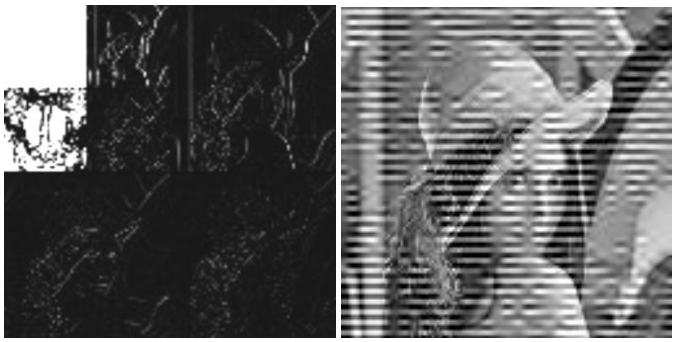
Wavelet utilized MRA allows decompose the image with wavelet coefficients (high and low frequency components) and also the original image can be reconstructed with the decomposed wavelet coefficients perfectly. If the high frequency component is replaced with secret image to be hidden, and if the reconstructed image is circulated to public, then secret image can be extracted from the circulated image if receiver knows which component is replaced with secret image. In this case, secret image has to be invisible in the circulated image. Also even if image processing including geometric conversion (linear transformation) and data compression (non-linear transformation) is applied to the circulated image, secret

image has to be extracted. The aforementioned "invisibility" and "robustness against image processing" are very important for data hiding.

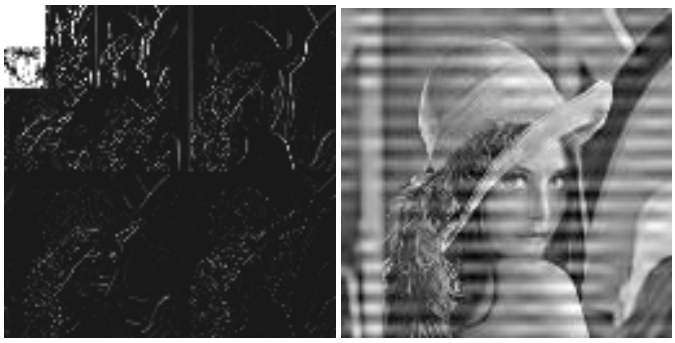
One of the examples of conventional MRA based data hiding is shown in Figure 1. The original image size is 128 by 128 pixels while the secret image size (binary) is 64 by 64 pixels for Figure 1 (a) and 32 by 32 pixels for Figure 1 (c) as well as 16 by 16 for Figure 1 (e), respectively. If the secret image is inserted at the HL (High and Low frequency components in horizontal and vertical directions) component of MRA image, and if reconstructed with the secret image, then secret image is somewhat visible in the reconstructed image which is circulation image to public.



(a) MRA image with secret image (b) Reconstructed image (Level 1)



(c) MRA image with secret image (d) Reconstructed image (Level 2)



(c) MRA image with secret image (d) Reconstructed image (Level 3)

Figure 1: Resultant images of wavelet based MRA which contains the secret image of Mandrill (left bottom), LL (left top), LH (right top), and HH (right bottom) as well as reconstructed image (circulation image) which contains embedded secret image

Invisibility of the secret image depends on base function of wavelet, level of the MRA stages. In general, secret image is getting much invisible in the reconstructed image in accordance with increasing of level of MRA. Furthermore, secret image is much invisible in the reconstructed image when the secret is replaced to HL, LH, or HH frequency components, rather than LL components. In other word, secret image would better to hide at higher frequency components rather than low frequency component because human eyes has low pass filter of frequency response.

Root Mean Square (RMS) difference between the original image and reconstructed image which contains secret image in LH, or HL, or HH with the level of MRA, i based on Daubechies base function with support length of j which is denoted as Db_j is shown in Table 1. In this case, i is 1, 2, and 3 while j is 2, 4, and 8, respectively.

It is obvious that RMS difference is decreased in accordance with increasing of level. It is unclear that relation between RMS difference and support length. The replaced component of HH seems to be better RMS difference in comparison to the other HL, LH components. Some examples of reconstructed images with MRA based data hiding are shown in Figure 2. By comparing these, it is found that the location of frequency component of which the secret image is replaced is the most influencing factor (the best location for replacing secret image is HH) followed by the level of which the secret image is replaced. Meanwhile, support length of Daubechies base function is not so influencing in comparison to the other factors.

TABLE 1 Root Mean Square: RMS difference between the original image and reconstructed image containing secret image (Lv_i: Level of MRA, Db_j: Daubechies base function with support length of j).

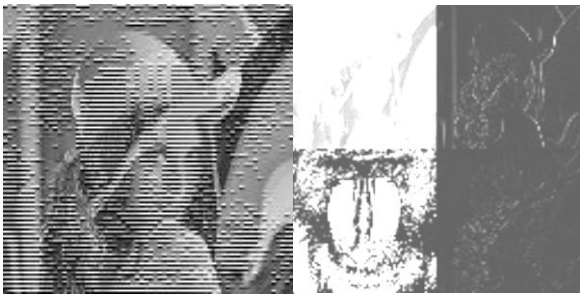
RMS Difference			
Lv_1	Db_2	Db_4	Db_8
LH	78.25	77.17	77.26
HH	78.27	77.26	77.41
HL	78.91	77.64	77.61
L_{V2}	Db_2	Db_4	Db_8
LH	43.57	42.75	43.68
HH	43.8	41.04	43.93
HL	45.2	43.71	44.44
Lv_3	Db_2	Db_4	Db_8
LH	21.27	21.51	20.78
HH	20.94	20.97	20.86
HL	24.84	25.05	23.77

Figure 2 Some examples of reconstructed images of MRA based data hiding (lv_i db_j denotes i level and Daubechies base function with support length of j, respectively)

B. *Robustness against image processing, intensity inversion, up-side-down of geometric conversion, magnification*

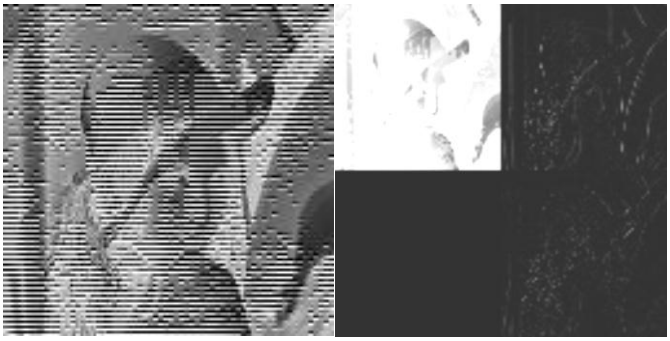
Even if someone applies image manipulation to the circulation images, the secret image has to be extracted.

In order to confirm such robustness of data hiding against image manipulation or processing, intensity inversion, up-side-down of geometric conversion as well as magnification (twice large size) are applied to the circulation image and then extraction of the secret image is attempted. Resultant images with the different level of MRA (1 and 3) are shown in Figure 3. In these cases, Daubechies base function with support length of 8 is utilized for MRA. Although the extracted secret image of Mandrill in the case of magnification is poorly and is better than the other cases. Conventional MRA based data hiding is not robust against image processing applied to circulation images except magnification. Poor quality of the secret image of Mandrill can be seen in the extracted image of Figure 3 (n).



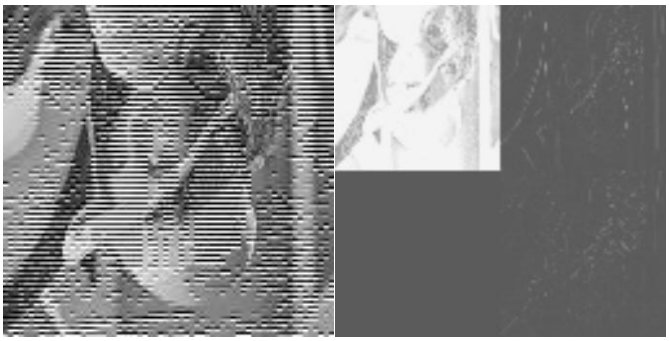
(a)Reconstructed:LH_Lv1_Db8 (b)Extracted:without image process

(c)Reconstructed:LH_Lv1_Db8 (d)Extracted:without image process

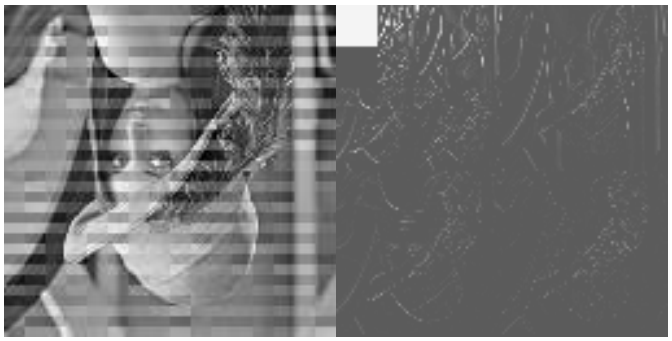


(e)Reconstructed:LH_Lv1_Db8 (f)Extracted image:Intensity inversion

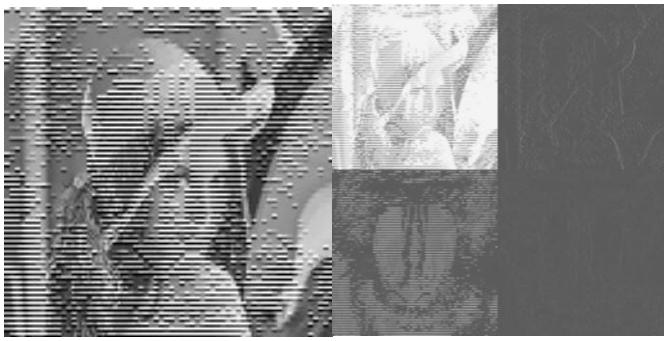
(g)Reconstructed:LH_Lv3_Db8 (h)Extracted image: Intensity inversion



(i)Reconstructed:LH_Lv1_Db8 (j)Extracted image:up-side -down conversion



(k)Reconstructed:LH_Lv3_Db8 (l)Extrcated image:up-side -down conversion



(m)Reconstructed:LH_Lv1_Db8 (n)Extracted image:Magnification

(o)Reconstructed:LH_Lv3_Db8 (p)Extracted image:Magnification

*F*igure 3 Robustness of the proposed data hiding against image processing, intensity inversion, up-side-down of geometric conversion, magnification with the different level of MRA of which the secret image is replaced to it.

C. Proposed data hiding method based on lifting dyadic wavelet transformation with secret image manipulations with run-length coding and pixel order exchanges of permutation using random number

The proposed method for data hiding is based on dyadic wavelet transformation. Dyadic wavelet allows to separate frequency components keeping image size with that of original image. Dyadic wavelet is called as a binary wavelet and has high pass and low pass filter components, {h[k], g[k]} and reconstruction filter {$\underline{h}$[k], $\underline{g}$[k]}. Low and high frequency components, C_n and d_n are expressed as follows,

$$C_n[i]=\Sigma_k h[k]C_{n-1}[i+k2^{n-1}] \quad (1)$$
$$d_n[i]=\Sigma_k g[k]C_{n-1}[i+k2^{n-1}] \quad (2)$$

Then original image is also reconstructed with the low and high frequency components as follows,

$$C_{n-1}[i]=1/2\Sigma_k \underline{h}[k]C_n[i-k2^{n-1}]+\Sigma_k \underline{g}[k]d_n[i-k2^{n-1}] \quad (3)$$

If a new parameter s[m] is employed, then lifting dyadic wavelet is defined as follows,

$$h^{new}[k]=h^{old}[k] \quad (4)$$
$$\underline{h}^{new}[k]=\underline{h}^{old}[k]+\Sigma_m s[-m]\, g^{old}[k-m] \quad (5)$$
$$g^{new}[k]=g^{old}[k]-\Sigma_m s[m]\, h^{old}[k-m] \quad (6)$$
$$\underline{g}^{new}[k]=\underline{g}^{old}[k] \quad (7)$$

Figure 4 shows a schematic process flow of the proposed data hiding based on lifting dyadic wavelet transformation. It is possible to hide the embedded image at the certain location of wavelet transformation images then circulation images containing the embedded image can be reconstructed through inverse wavelet transformation. In this case, although binary secret images are assumed, half tone, colored images are also available for secret images.

First, secret image is binarized. Before binarized secret images are replaced to one of the high frequency component images, run-length coding is applied to secret images in order to improve an invisibility of the secret images in the circulation images. Figure 5 shows schematic process flow of the run-length coding method. The number of pixels in the original binary image is 27 while the number of pixels in the compressed image is just 6 (quantization level is variable).

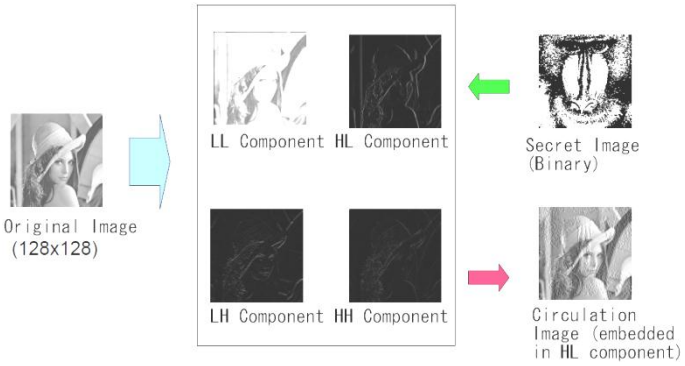


Figure 4 Schematic process flow of the proposed data hiding

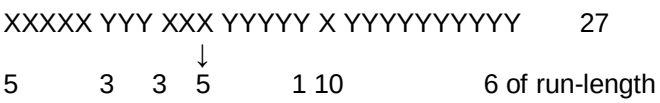


Figure 5 Schematic process flow of run-length coding.

Then run-length coded data are replaced to one of the high frequency components with the pixel order exchanges based on generated random numbers which are generated by Mersenne Twister. Only if the receiver who knows the initial value of random number of Merssene Twister and how to decode run-length coding, then such the receiver can extract the secret images. Thus the copy right holders can assert their copy right through extraction of secret images. Figure 6 shows the process flow of the proposed data hiding (hide the secret image into the original image and create circulation image embedded the secret image then extract the secret image from the circulation image)

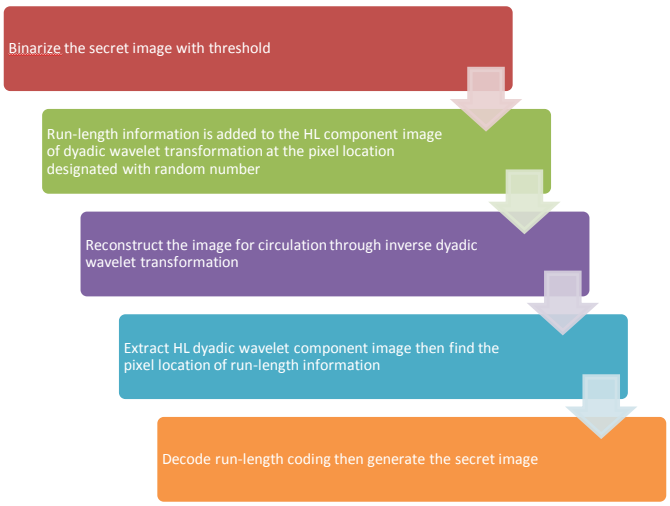


Figure 6 Process flows of the proposed data hiding method

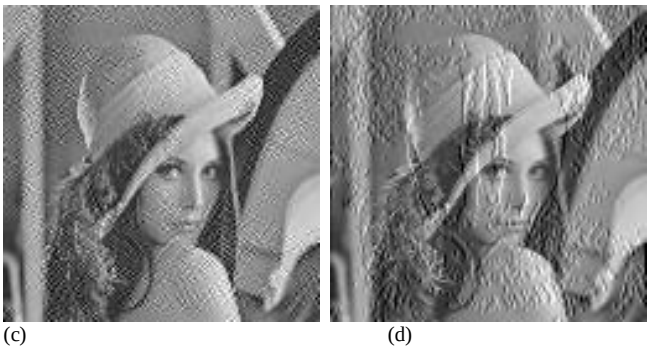
III. EXPERIMENTS

A. *Dyadic wavelet based data hiding methods*

In Figure 7, the secret binary image of Mandrill [9] with a certain threshold is embedded in the HL component of the dyadic wavelet transformed images derived from the original image of Lena [9] with dyadic wavelet transformation. At the left bottom corner of Figure 7, a reconstructed image (image for circulation) is shown. The secret image can be recognizable in the circulation image, unfortunately. In these cases, Daubechies wavelet base function (support length=2) [10] is used. On the other hand, Figure 7 (b),(c),(d) shows reconstructed images of Mandrill of secret image (a) embedded Lena of original images of which the secret image is embedded at the LH, HH, and HL of frequency components, respectively. Image size is not changed for original and dyadic wavelet transformed images.

Table 2 shows RMS difference between original and reconstructed images for dyadic wavelet based data hiding. In this case, RMS difference of Daubechies base function utilized dyadic wavelet is compared to that of Haar base function. The difference is quite obvious that Daubechies base function utilized dyadic wavelet is superior to that of Haar base function.

(a) (b)



(c) (d)

Figure 7 (a) The secret image of Mandrill, (b) Reconstructed image of Mandrill image embedded (LH) Lena image for circulation, (c) Reconstructed image of Mandrill image embedded (HH) Lena image for circulation, (c) Reconstructed image of Mandrill image embedded (HL) Lena image for circulation.

TABLE 2: RMS difference between original and reconstructed images for dyadic wavelet based data hiding.

Lv_1	Haar	Db_2
LH	43.65	24.15
HH	38.90	20.58
HL	42.67	23.42

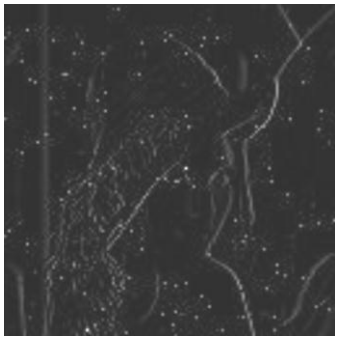
B. *Data hiding method with run-length coding and pixel order exchange based on random number (permutation matrix operation)*

As shown in Figure 7, the binarized secret image of Mandrill is recognizable in the reconstructed image (circulation image). The propose data hiding method is to apply run-length coding to the binarized secret image in concern before replacing the secret image to one of high frequency components, HL, LH and HH. Also bit stream order exchange is applied to the run-length coded compressed data.

As is aforementioned in the reference [6] and [7], invisibility of the secret image is improved remarkably by scanning scheme (permutation of pixel order of the secret image in accordance with random number). Mersenne Twister [11] of random number generator is used for the permutation.

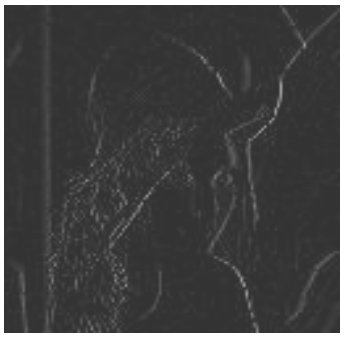
Figure 8 (a) shows LH frequency component image embedded with the binarized secret binary image of Shuttle cock with the threshold of 15 and with permutation by Mersenne Twister of random number generator while Figure 8 (b) shows the reconstructed image for circulation for the case of (a). Meanwhile, Figure 8 (c) shows LH frequency component image embedded with the binarized secret binary image of Shuttle cock with the threshold of 255 and with permutation by Mersenne Twister of random number generator while Figure 8 (d) shows the reconstructed image for circulation for the case of (c). Figure 8 (e) shows the secret image of Shuttle cock. Also Figure 8 (f) is the binarized Shuttle cock. As are shown in Figure 8 (g) and (h), Gray scale image and Binarized Mandrill has too many black pixels so that the Shuttle cock of secret image is used in the experiments.

Table 3 shows the relations between threshold and total data amount of compressed data as well as RMS difference. In accordance with increasing of threshold, the number of black pixels is decreased so that total data amount after the run-length coding is also decreased.



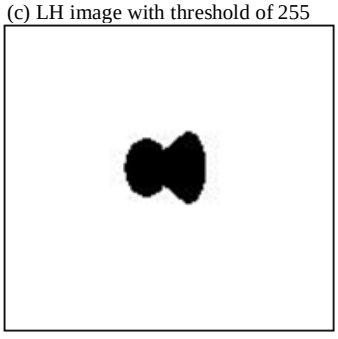
(a) LH image with threshold of 10

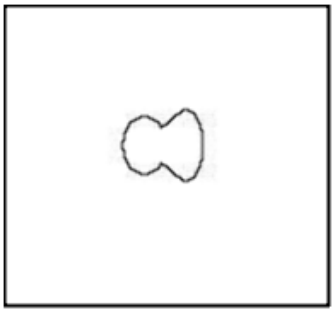
(b) Reconstructed image



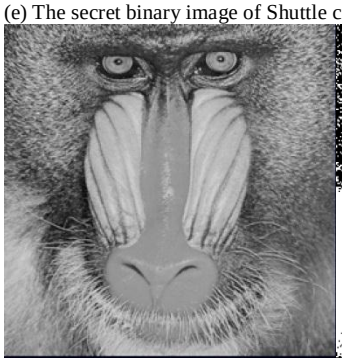
(c) LH image with threshold of 255

(d) Reconstructed image





(e) The secret binary image of Shuttle cock (f) After the binarization



(g) Gray scale image of Mandrill (h) Binarized Mandrill

Figure 8 (a) LH frequency component image embedded with the binarized secret binary image of Shuttle cock with the threshold of 15 and with permutation by Mersenne Twister of random number generator, (b) the reconstructed image for circulation for the case of (a), (c) LH frequency component image embedded with the binarized secret binary image of Shuttle cock with the threshold of 255 and with permutation by Mersenne Twister of random number generator, (d) the reconstructed image for circulation for the case of (c), (e) the secret image of Shuttle cock, (f) the binarized Shuttle cock, (g) Gray scale image of Mandrill and (h) Binarized Mandrill.

TABLE 3: Relations between threshold and total amount of compressed data as well as RMS difference.

Secret image	Threshold	Total data amount	RMS difference
Shuttle Cock	255	133	3.65
	15	1140	1.71

Bit stream order of the compressed data is exchanged based on random number. After that the exchanged bit stream is embedded into LH image. Then image for circulation is reconstructed with HL, HH and LL images together with secret image embedded LH image. The reconstructed image with threshold 15 and 255 which are shown in Figure 8 (b) and (d) are better quality image in comparison to the images which are shown in Figure 6. RMS difference between reconstructed image and the original image is shown in Table 3. RMS difference of 1.71 is negligible comparing to that of 24.15 of dyadic wavelet based data hiding method.

Figure 9 shows the relation between the number of pixels and RMS difference as well as total data amount. The number of pixels implies that the number of black pixels after the binarisation with the designated threshold so that it depends on the threshold.

The number of pixels is increased with decreasing of threshold. As is shown in Figure 9, the minimum of RMS difference is situated at 15 of the number of pixels. The total amount of data in unit of bit means the compressed data amount with run-length coding. The shuttle cock of secret image is binalized and replaced to LH component of dyadic wavelet transformed image results in circulation image. After that, the circulation image is compressed with run-length coding. The compressed data amount depends on the number of pixels of the binalized secret image which depends on threshold. In these cases, the location of frequency component for replacing secret image of HH and LH shows almost same RMS difference and these are better than the RMS difference of HL component. This implies that the location of HH or HL is more appropriate rather than LH component.

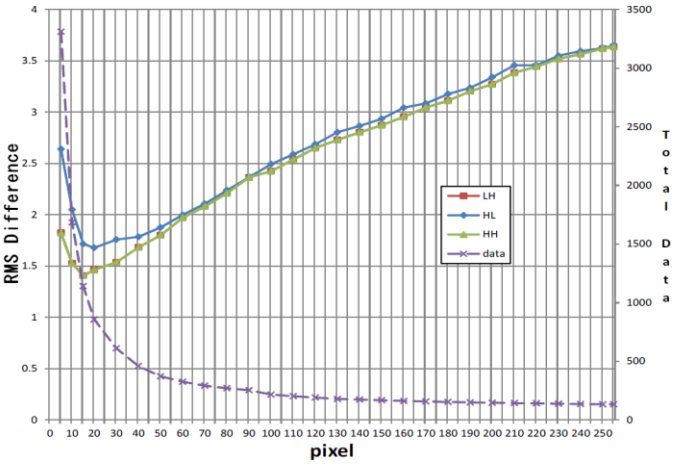


Figure 9 Relation between the number of pixels and RMS difference as well as total data amount.

C. *Robustness of the proposed data hiding method against the data compressions, JPEG and JPEG 2000*

The proposed method is also applied to the other data compression methods of JPEG and JPEG 2000. Figure 10 shows a small portion of decompressed images of JPEG and JPEG 2000 with data compression ratio of 100.

It is obvious that decompressed image of JPEG has block distortions while that of JPEG 2000 has not such distortion. Also much larger mosquito noise is found in the decompressed image of JPEG rather than JPEG 2000 as well. Peak signal-to-noise ratio PSNR of JPEG 2000 as a function of bit rate is greater than that of JPEG for the "Lena" image as is shown in Figure 11.

(a) JPEG (b) JPEG 2000

Figure 10 Enlarged portions of images of decompressed images of JPEG and JPEG 2000 with data compression ratio of 100.

After dyadic wavelet transformation is applied to the original image and HH component of the transformed image is replaced with the permutated Mandrill of secret image with random number, reconstruction image (circulation image) is generated with inverse dyadic wavelet transformation. After that JPEG and JPEG 2000 of data compression is applied to the circulation image. Using the compressed images, secret image can be extracted with decompression and inverse dyadic wavelet transformation together with inversely permutation by the same random number which is generated by Mersenne Twister.

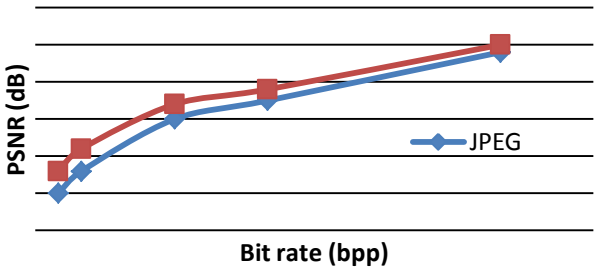


Figure 11 PSNR of JPEG and JPEG 2000 as a function of bot rate.

Figure 12 shows the circulation images with embedded secret image of Mandrill with the data compression by JPEG and JPEG 2000 with the same permutation by the same random number which is generated by Mersenne Twister. In the images, defect due to secret image with permutated by random number is visible because the binalization of parameter of threshold is set at 255. Although both images are quite similar, details are different as is mentioned before.

After the extraction of secret image, RMS difference between the original secret image and the extracted secret image is calculated. The results show that RMS difference for JPEG data compression is 6.38 while that for JPEG 2000 is totally equal to zero because JPEG 2000 is information loss less coding while JPEG is information lossy coding.

(a) JPEG (b) JPEG 2000

Figure 12 Circulation images with embedded secret image of Mandrill with the data compression by JPEG and JPEG 2000 with the same permutation by the same random number which is generated by Mersenne Twister

IV. Conclusions

It is found that MRA based conventional data hiding is not robust against image processing, intensity inversion, up-side-down of geometric conversion, and magnification. Secret image is much visible in the circulation images rather than the proposed lifting dyadic wavelet transformation based data hiding method. It is found the best location of component of MRA resultant images of which the secret image is replaced to it. Also the best level of MRA as well as support length of Daubechies base function for minimization of visibility of the secret image in the circulation images are found.

The proposed data hiding method is robust against data compression, JPEG and JPEG 2000. Also the secret image in circulation images with the proposed method is much invisible rather than the conventional method. Because the proposed method uses random number utilized pixel order exchanges for the secret images together with run-length coding of data compression. In the reconstruction process, scheme and initial value of the random number generator is known by receivers so that it is possible to reconstruct the embedded secret images from the circulation images. In this research, the secret image is binalized. The most appropriate threshold which allows minimization of RMS difference between extracted secret image and the original secret image is also found.

Acknowledgment

The authors would like to thank all the teaching staff and the students in the research group for their valuable comments and suggestions through conducting this research work.

References

[1] Kohei Arai, Fundamental Theory on Wavelet Analysis, Morikita Shuppan Publishing Co. Ltd., 2000.

[2] Kohei Arai, Self Learning on Wavelet Analysis, Kindai-Kagakusha publishing Co. Ltd.. 2006.

[3] Kohei Arai and Kaname Seto, Data hiding method based on Multi-Resolution Analysis: MRA, Visualization Society of Japan, 22, Suppl.No.1, 229-232, 2002.

[4] Kohei Arai and Kaname Seto, Data hiding method with coordinate conversion in feature space, Visualization Society of Japan, 25, Suppl.No.1, 55-58, 2005.

[5] Kohei Arai and Kaname Seto, Improvement of invisibility of secret images embedded in circulate images based on MRA with coordinate conversion and Principal Component Analysis: PCA, Journal of Image and Electronics Society of Japan, 36, 5, 665-673, 2007.

[6] Kohei Arai and Kaname Seto, Improvement of invisibility of secret images embedded in circulate images based on MRA with scanning scheme conversion, Visualization Society of Japan, 29, Suppl.No.1, 167-170, 2009.

[7] Kohei Arai, Improvement of security and invisibility of secret images embedded in circulate images and based on MRA, Report of RIMS-Research Institute for Mathematical Sciences Kyoto University, ISSN188-2818,1684,93-113,2010.

[8] S.Mallat and S.Zhong, ``Characterization of signals from multiscale edges," IEEE Trans. Pattern Anal. Machine Intell., 14, pp.710-732, 1992.

[9] http://vision.kuee.kyoto-u.ac.jp/IUE/IMAGE_DATABASE/STD_IMAGES/ (Accessed on March 11 2011).

[10] Kohei Arai and Leland Jameson, Earth observation satellite data analysis based on wavelet analysis, Morikita-Shuppan Publishing Co., Ltd., 2001.

[11] http://www.math.sci.hiroshima-u.ac.jp/~m-mat/MT/mt.html (Accessed on March 11 2011).

[12] Takagi and Shimoda Edt., Kohei Arai et al., Image Analysis Handbook, Tokyo University Shuppan-Kai, 1991.

[13] M. Weinberger, G. Seroussi, and G. Sapiro, "The LOCO-I lossless image compression algorithm: Principles and standardization into JPEG-LS," IEEE Trans. Image Processing, vol. 9, no. 8, pp. 1309–1324, Aug. 2000, originally as Hewlett-Packard Laboratories Technical Report No. HPL-98-193R1, November 1998, revised October 1999.

[14] http://www.uw-de.com/jpeg2000/(Accessed on March 11 2011).

AUTHORS PROFILE

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona and is Vice Chairman of ICSU/COSPAR Commission A.

Yuji Yamada received BS degree of information science from Saga University in 2010 and is now in the master course of graduate school of Saga university.

Unsupervised Method of Object Retrieval Using Similar Region Merging and Flood Fill

Kanak Saxena
Samrat Ashok Technological Institute
Vidisha

Sanjeev Jain
Madhav Institute of Technology & Science
Gwalior

Uday Pratap Singh
Lakshmi Narain College of Technology
Bhopal

***Abstract*— In this work; we address a novel interactive framework for object retrieval using unsupervised similar region merging and flood fill method which models the spatial and appearance relations among image pixels. Efficient and effective image segmentation is usually very hard for natural and complex images. This paper presents a new technique for similar region merging and objects retrieval. The users only need to roughly indicate the after which steps desired objects boundary is obtained during merging of similar regions. A novel similarity based region merging mechanism is proposed to guide the merging process with the help of mean shift technique. A region *R* is merged with its adjacent regions *Q* if Q has highest similarity with *R* among all Q's adjacent regions. The proposed method automatically merges the regions that are initially segmented through mean shift technique, and then effectively extracts the object contour by merging all similar regions. Extensive experiments are performed on 22 object classes (524 images total) show promising results.**

Keywords- Image segmentationl; similar regions; region merging; mean shift; flood fill.

I. INTRODUCTION

CLASS-SPECIFIC (or category-level) object segmentation is one of the fundamental problems in computer vision. Although a human can delineate the object boundaries with much ease, segmenting images is not as ease for a computer. Its goal to segment an image into regions with each region solely containing object(s) of a class. As object segmentation requires that each segmented region to be a semantic object, it is much more challenging than traditionally object segmentation [1, 2, 3, 4]. There has been a substantial amount of research on image segmentation including clustering based methods, region growing methods [5], histogram based methods [6], and more recent one such as adaptive thresh-hold methods [7], level set methods [8], graph based methods [4, 9] etc.

Despite many years of research, unsupervised image segmentation techniques without human interaction still do not produce satisfactory results [10]. Therefore semi-supervised segmentation methods incorporating user interactions have been proposed [11, 12, 13, 14, 15] and are becoming more and more popular. For instance, in the active contour model (ACM) i.e. snake algorithm [11], a proper selection of initial curve by user lead to a good convergence of the true object contour.

In order to do semantically meaningful image segmentation, it is essential to take priori (e.g. object part configuration [16], or class fragments [17]) information about the image into account.

The low level image segmentation methods, such as mean shift [18, 19], watershed [20] and super pixels [21], usually divide the image into small regions. These low level segmentation methods provide a good basis for the subsequent high level operations, such as region merging. As a popular segmentation technique for color images, mean-shift [19] can have less segmented parts in comparison to watershed and super pixels [15, 21, 22] while preserving well the edge information of the objects. Because of less number of segmentation, the statistical features of each region, which will be exploited by the proposed unsupervised similar region merging method and object detection can be more robustly calculated and then be used in guiding the region merging process.

In this paper, we proposed unsupervised similar region merging method based on initial segmentation of mean shift. The proposed method will calculate the similarity of different regions and merge them based on largest similarity. The object will then extract from the background when merging process ends. Although the idea of region merging is first introduced by [23] this paper uses the region merging for obtaining the contour for object and then extracting desired object from image. The key contribution of the proposed method is a novel similarity based region merging technique, which is adaptive to image content and does not requires a present threshold. With the proposed region merging algorithm, the segmented region will be automatically merged and labeled, when the desired object contour is identified and avoided from background, the object contour can be readily extracted from background. The proposed algorithm is very simple but it can successfully extract the objects from complex scenes.

The rest of the paper is organized as follows; section 2 presents the proposed region merging algorithm. Section 3 performs extensive experiments to verify the proposed method and analysis. Section 4 concludes the paper and section 5 experimental results for different color spaces, different initial segmentation and comparison of proposed method with various existing algorithms.

II. SIMILARITY REGION MERGING

In proposed method, an initial segmentation is required to partition the image into homogeneous region for merging.

For this we use any existing low level image segmentation methods e.g. watershed [20], super-pixel [21], level set [24] and mean-shift [18, 19] can be used for this step. In this paper we use mean-shift method for initial segmentation because it has less over segmentation and well preserve the object boundaries. For the initial segmentation we use the mean shift segmentation software the EDISON system [25] to obtain the initial segmentation map. Fig. 1. shows an example of mean shift initial segmentation. For detailed information about mean shift and EDISON system, please refer to [18, 19, 25, 26]. In this paper we only focus on the region merging.

a b

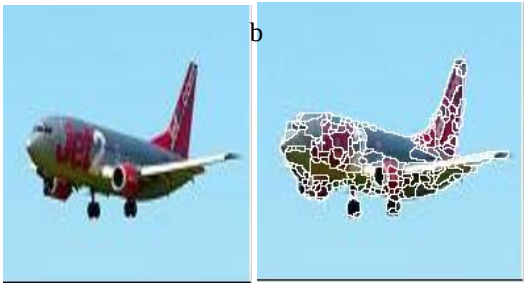


Fig. 1 Original Image Mean shift segmentation

A. Similarity Measure

After mean shift initial segmentation, we have a number of small regions. To guide the following region merging process, we need to represents these regions using some descriptor and define a rule for merging. A region can be described in many aspects, such as texture [27], shape and size [c] and color edge [28] of the regions. Among them color descriptor is very useful for representation of the object color features. In the context of region merging based segmentation, color descriptor is more robust than other feature descriptors because shape and size feature is vary lots while the colors of different regions from the same object will have high similarity. Therefore we use color histogram represent each region in this paper. The RGB color space is used to compute the color histogram of each region in this paper. We uniformly quantize each color channels into 16 levels and then the histogram is calculated in the feature space of 4096 bins. Next problem is how to merge the region based on their color histograms so that the desired object can be extracted. The key issue in region merging is how to determine the similarity between different segmented regions of image so that the similar regions can be merged by some logic control. Therefore we need to define a similarity measure Formula (1) between two regions R and Q to accommodate the comparison between various regions, for this there are some well known statistical metrics such as Euclidean metric, Bhattacharyya coefficient and log-likelihood ratio statistic [29].

Here we use Bhattacharyya coefficient [29, 30, 31, 32] to measure the similarity between two regions say R and Q is:

$$P(R,Q) = \sum_{u=1}^{4096} \sqrt{Hist_R^u Hist_Q^u} \qquad (1)$$

Where $Hist_R$ and $Hist_Q$ are the normalized histogram of R and Q, respectively and superscript u represents the u^{th} element of them.

$$cos\theta = \left(\sqrt{Hist_{R,\ldots\ldots,}^{1} Hist_{R,}^{4096}}\right)^T \left(\sqrt{Hist_{Q,\ldots\ldots,}^{1} Hist_{Q,}^{4096}}\right) \qquad (2)$$

The higher the Bhattacharyya coefficient between R and Q is the higher the similarity between them i.e. smaller the angle θ. The geometric explanation of Bhattacharyya coefficient actually reflects the perceptual similarity between two regions. If two regions have similar contents then their histogram will be very similar, and their Bhattacharyya coefficient will be very high i.e. angle between histogram vectors is very small. Certainly it is possible that two different regions may have different histogram, such case happen very rare. Similarity measure between two regions we use Bhattacharyya similarity which works well in proposed region merging method. The Bhattacharyya descriptor is a very simple yet efficient way to represent similarity between regions. However other color spaces e.g. HSV, YCbCr etc. and other distance measure such as the Chernoff, Euclidean and Manhattan are also be adopted that for the region merging. In section 3 we present examples by using HSV, YCbCr color spaces and Manhattan distance. Results will be similar to those by using the RGB color space and Bhattacharyya descriptor.

B. Similarity Based Merging Rule

It is still a challenging problem to extract accurately the object contour from the background. The conventional region merging methods are merging two adjacent regions whose similarity is above based on threshold [32]. These methods are difficult because of threshold selection. A big threshold will lead to incomplete merging belonging to object, while a small threshold will cause over-merging. Moreover it is difficult to detect when region merging process should stop. Proposed region merging method will start from any random segment part and start automatic region merging process. The entire region will be gradually labeled as either object region or background region. The lazy snapping cutout method proposed in [15], which combine graph cut with watershed based initial segmentation, is actually a region merging method. It is controlled by max-flow method [33]. In this paper we present an adaptive similarity based merging technique of regions either in foreground or in background.

Let Q be the adjacent region of R and denoted by $\overline{S_Q} = \{S_i^Q\}_{i=1,2,\ldots,q}$ its set of Q's adjacent regions. Using Bhattacharyya coefficient calculate similarity among Q's adjacent regions $\overline{S_Q} = \{S_i^Q\}_{i=1,2,\ldots,q}$. Obviously R will be one of the adjacent regions of SQ. If the similarity between R and Q will be maximum then region R will be merged in region Q. We will use merging rule according to the formula defined as:

$$P(R_j, Q) = \max_{i=1,2,\ldots,k} P(R_j, S_i^{Q_i}) \qquad (3)$$

Equation (2) is the merging rule which establish the basis of proposed region merging process. Important advantage of (2) is that it prevents the use threshold for merging control, and the

Bhattacharyya coefficient is the inner product of the two histogram vectors and it is robust to small noise and variations. The automatic region merging process cover the all part of segmented image, and after every step of merging we will whether we want to work on this image or not. Therefore in the automatic region merging process object regions will have high probabilities to be identified as object.

C. The merging process

The whole object retrieval process is working in two stages. In first stage similar region merging process is as follows, our strategy to merge the small segmented image which is start with any randomly selected and merge this with any of its adjacent regions with high similarity. Some two step supervised merging process used in [34, 35] for image pyramid construction. Different from [34, 35] proposed method used image segmentation and it is unsupervised technique of region merging. We will merge segmented image regions with their adjacent regions as: if for each region Q we will set its adjacent regions S_B i=1, 2, …, r. If the similarity between any Rj for any i= j is maximum i.e.

$$P(R_j, Q) = \max_{i=1,2,\ldots,k} P(R_j, S_i^{Q_i}) \quad (4)$$

Then Q and Rj are merged into one region and new region is same leveled by

$$Q = QUR_j \quad (5)$$

The above procedure is implemented iteratively. Note that to each and every iterative step we will see whether the desired object is retrieve or not. Specifically the segmented region is shrinking; we will stop iteration when desired object is found.

After the first stage i.e. when full part of object boundaries or likely to appear which is seems in every step we apply second stage of algorithm for this we select a input point on the object and expand this using four connectivity of pixels by using well known Flood Fill method.

Object Retrieval Algorithm

Input: (1) the image (2) the initial mean shift segmentation of input image
Output: desired object

While there is a merging up to object contour

1. First stage of merging of initial segmented image (by mean shift method) using similar merging rule.
2. After step one number of regions are minimized and again apply similar region merging rule, this is and iterative procedure.
3. After retrieving object contour go to step (4).
4. Apply Region Labeling and after that Flood Fill method on the image obtained in after step 3

Region Labeling (I)

% I: binary Image; I (u, v) =0: background, I (u, v) =1: foreground %

4.1. Let m←2
4.2. for all image coordinates (u, v) do
4.3. if I (u, v) =1 then
4.4. Flood Fill (I, u, v, m)
4.5. m← m+1
4.6. return the labeled image I.

% After region labeling we apply Flood Fill method using Breadth-First Search %

5. FloodFill (I, u, v, label)
   5.1. Create an empty **queue** Q
   5.2. ENQUEUE (Q, (u, v))
   5.3. **While** Q is not empty **do**
   5.4. (x, y)← DEQUEUE (Q)
   5.5. If (x, y) is inside image and I (x, y) =1 then
   5.6. Set I (x, y)= label
   5.7. ENQUEUE (Q, (x+1, y))
   5.8. ENQUEUE (Q, (x, y+1))
   5.9. ENQUEUE (Q, (x-1, y))
   5.10. ENQUEUE (Q, (x, y-1))
   5.11. return

The proposed similar region merging method is an iterative method. After doing stage (1) what is the guarantee that the automatic similarity merging method will converge after a certain extent? To answer this question we will prove a proposition stated below.

Proposition1. The Similarity region merging method in section 2.3 will converge i.e. every region in the image will be merge after a certain extent.

Proof. If a region Q has the maximal similarity with region R then region R will be merged with region Q i.e. Q= QUR, in the first stage of proposed method this procedure is repeatedly and number of segmentation in the image is finite so the desired contour of object is obtained after a certain extent i.e. after kth iteration.

From above analysis we see that the number of regions in image (after mean segmentation) is N (say) it will decrease in the process if iterative region merging. The whole algorithm will stop and all segmented region is in either object or in background.

Therefore proposed algorithm converges and it will be label all the region of image.

III. Experimental Analysis

The proposed similarity region merging method is an unsupervised method, since it will automatically merge the regions and it will labels every regions either object or background.

In section 3.1 we will first shows the unsupervised similarity region merging method qualitatively by several representative examples; in section 3.2 we compare proposed method with well-known hybrid graph model, graph cut and normalized cut; in section 3.3 we test our proposed method for different color spaces and different distance metrics.

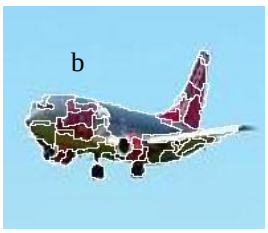
A. Experimental analysis and Results

Fig. 2. shows an example of how unsupervised similarity region merging method extract object contour in complex scene. After initial segmentation by mean shift, automatic segmentation merging starts and after every step we test our merging results and also after which stage of merging we want

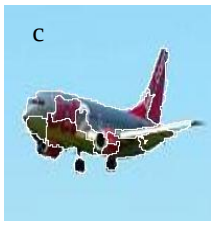
to use flood fill method. Fig. 2(a) is the initial segmented regions cover only small part but representative features of object and background regions. As shown in figure 2 the unsupervised similar region merging steps via iterative implementation.

Fig. 2(a), 2(b), 2(c), 2(d) and 2(e) shows that different steps of well extract object contour from the image and Fig. 2(f) is object mask. Fig. 2(g) shows the extracted object using the two steps object retrieval method.

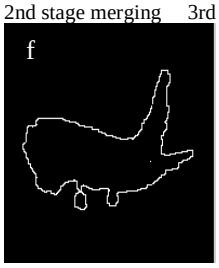
Fig.2 Region merging Process Initial segmentation

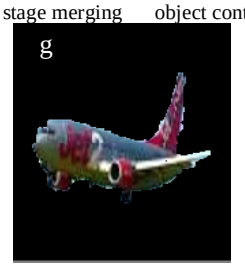


1st stage merging



2nd stage merging 3rd stage merging object contour





object mask object

In the second experiment, we want to separate a bird from background. Fig. 3(a) shows that the initial mean shift segmentation results are serve our segmentation for extraction of object contour from complex background. Fig. 3(b) to 3(e) shows that different step of fully extracted object contour from input image.

Fig. 3(g) shows the extracted object using the two steps object retrieval method.

The execution time object retrieval using unsupervised similar region merging and flood fill depends upon a number of factors, including size of image, the initial mean shift segmentation results etc. We implement unsupervised similar region merging and flood fill algorithm in the MATLAB (R 2008a) 7.6 programming environment and run it on a PC with P4 2.80 GHz CPU and 1.0 GB RAM.

Table 1 shows the running time of proposed method on testing different types of images e.g. bird and airplanes etc.

Table-1

Image	Size of image	Number of regions after initial Segmentation	Running Time (in Sec)
Birds	200 x 200	396	7.0988
Airplanes	200 x 200	338	6.2885
Horses	200 x 200	565	9.03111
Dogs	200 x 200	623	11.4329

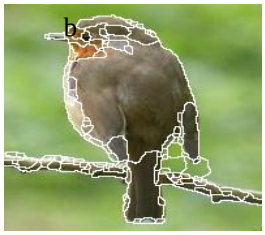
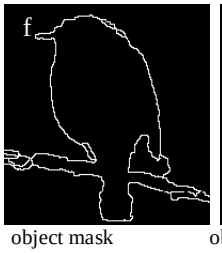


Fig.3 Region merging Process Initial segmentation 2nd stage merging

3rd stage merging 4th stage merging 5th stage merging



object mask object

B. Comparison with HGM and N-cut

In this section we compare the object retrieval method using unsupervised similarity region merging and flood fill method with hybrid graph model (HGM) and well known Normalized cut [4]. Since the original graph cut segmentation is a pixels based method (GC_P) for a fair comparison of proposed method, we extended the original pixel based graph method (GC_P) to a region based graph cut (GCR) i.e. the nodes in the graph are mean shift segmented region instead of original pixels.

Table 2 shows the comparison of the three methods on testing different types of images e.g. bird and airplanes etc. We can see that proposed unsupervised region merging method achieves the best results in comparison to others, while (GCR) performs better result in comparison to (GCP).

It can be seen that (GCR) will miss some object regions and wrongly label background regions as object regions.

Table-2 : Evaluation of results on 12 different class of image

Object Class	No. of images		F_2		
	Total	Special	N-Cut	HGM	Flood

					Fill
Airplane	100	25	0.3051	0.7609	0.7810
Horses	25	15	0.5268	0.8006	0.8123
Birds	25	15	0.6202	0.7443	0.7534
Cat	25	15	0.5904	0.7609	0.7812
Dogs	25	15	0.4404	0.9173	0.9215
Elephants	50	20	0.5540	0.6851	0.7263
Cars	25	15	0.2800	0.7953	0.8146
Flowers	25	15	0.4425	0.6996	0.7321
Women	50	20	0.5898	0.8123	0.8362
Fruits	40	15	0.5830	0.7100	0.7654
Plane	40	15	0.3598	0.7906	0.8431
Average	**380**	**175**	**0.4583**	**0.7743**	**0.8472**

To quantitatively compare the three methods, as shown in table 3, we mutually labeled the desired objects in the test image and took them as ground truth. After this we compute true positive rate (TPR) and false positive rate (FPR) for these segmentation results. The TPR is defined as the ratio of number of correctly classified object pixels to the number of total object pixels, and FPR is defined as the ratio of number of background pixels but classified as object pixels to the number of ground pixels. Obviously, higher the TPR and lower the FPR that method is better.

Table-3

Image	Method	TPR(%)	FPR(%)
Birds	GC_P	94.24	3.32
	GC_R	96.56	3.96
	Flood Fill	98.97	0.69
Airplanes	GC_P	96.23	2.99
	GC_R	97.35	1.74
	Flood Fill	97.59	0.84
Bird	GC_P	92.56	3.51
	GC_R	93.73	3.12
	Flood Fill	94.83	1.36
Dogs	GC_P	84.62	2.64
	GC_R	89.29	2.27
	Flood Fill	92.48	1.13
Horses	GC_P	76.18	2.62
	GC_R	88.63	3.46
	Flood Fill	95.68	1.92
Flower	GC_P	78.57	2.89
	GC_R	89.65	2.08
	Flood Fill	96.62	1.26
Tiger	GC_P	72.43	9.46
	GC_R	79.59	3.59
	Flood Fill	94.68	0.93
Starfish-1	GC_P	79.40	7.44
	GC_R	89.63	3.46
	Flood Fill	96.38	1.29

C. Unsupervised region merging under different color spaces, distance metrics and initial segmentation

Although RGB space and Bhattacharyya distance are used in proposed method, other color spaces and metrics are also used. In this section, we present some example to verify the performance of unsupervised region merging and flood fill method. We first test the effect of color space on the region merging result. In this experiment RGB color space is converted into HSV and YCbCr. The Bhattacharyya coefficient is calculated for the histogram of these color spaces. Fig. 4. shows the unsupervised region merging on the images birds and airplanes and after that we use flood fill method on HSV and YCbCr space for extraction of object.

The Fig. 4(b) shows the initially segmented images in the HSV color space and Fig. 4(c), 4(d) and 4(e) shows the finally segmented object contour and Fig. 4(f) is mask of object and finally Fig. 4(g) shows object retrieve by using unsupervised region merging and after that we use flood fill algorithm for object retrieval. We can see that the results are same as those by using RGB color spaces with Bhattacharyya distance.

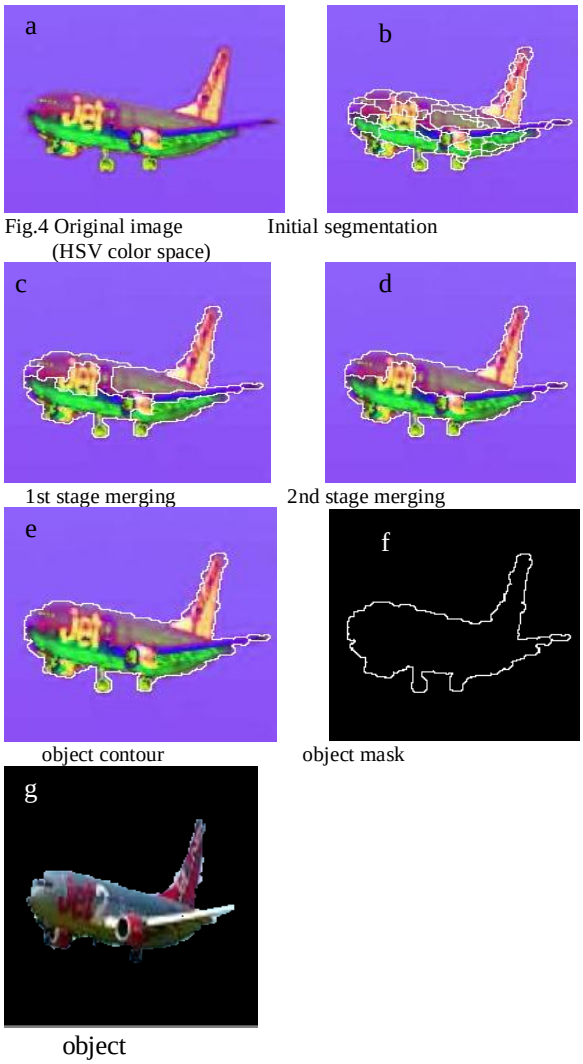


Fig.4 Original image (HSV color space) Initial segmentation

1st stage merging 2nd stage merging

object contour object mask

object

Again we test the effect of distance metric on the segmentation results. In this experiment, RGB, HSV and YCbCr color spaces is used with Euclidean distance, we denote HistR and HistQ are normalized color histogram of two regions R and Q the Euclidean distance between them is defined as:

a

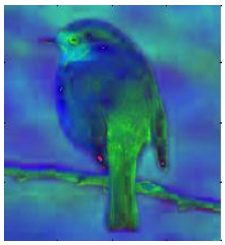
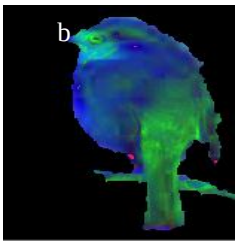


Fig.5 Original image (HSV color space)



object

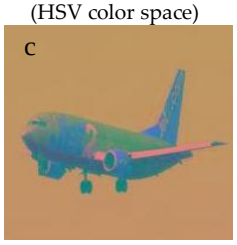
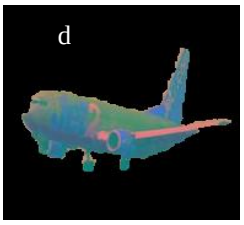


Fig.5 Original image (YCbCr color space)



object

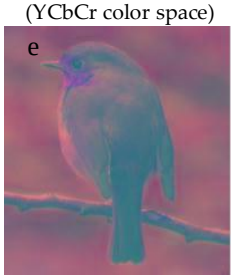


Fig.5 Original image (YCbCr color space)

object

Fig. 5 shows the segmentation results on the images of birds and airplanes. We can see that the results are same as those by Bhattacharyya distance.

At last we test the unsupervised similar region merging method with some other initial segmentation. Besides the mean-shift, watershed [20] and super-pixels [21] are another popular initial segmentation method. Super-pixels partition the images into more over segmentation in comparison to mean shift. In this experiment super-pixel method is used for initial segmentation. Section 5 shows the results of airplanes and birds. It can be seen that watersheds and super-pixel gives the similar results as mean shift.

IV. Conclusion

In this paper proposed a class specific object segmentation method using unsupervised similar region merging and flood fill algorithm. The image is initially segmented using mean-shift segmentation and automatic start of merging with any random segmented region and after each merging we check whether the object contour is obtained or not, if at any particular stage of merging object contour is obtained then use flood fill algorithm and click with mouse which object we want to extract. The proposed scheme is simple yet powerful and it is image content adaptive.

In future we can extract multiple objects from input image by using unsupervised method as well as supervised method by merging similar regions using some metric. Extensive experiments were conducted to validate the proposed method of extracting single object from complex scenes. The proposed method is efficiently exploits the color similarity of the target. The proposed method provides a general region merging framework, it does not depend initially mean shift segmentation method or other color image segmentation methods [20, 24, 25, 36] can also be used for segmentation. Also we can use appending the different object part to obtaining complete object from complex scene, and also we can use some supervised technique also.

V. Experimental Results using Proposed Method *(for different color space and initial segmentation)*

To see how Similar Region Merging Flood Fill produces promising segmentation results in the case that there is a large variation in shape (including position, size, profile and pose) within an object class. We refer to the Fig. 6, 7, 8, 9, 10, 11. The task is to segment an airplane from the background scene. To segment a new image that may contain object of several classes, we use initial mean shift segmentation method to segment the image into K regions in which all containing instance (s) of object class. We assume that K is known a priori for each test image. In this section, we present the example to verify the performance of proposed method under the different color spaces like HSV, YCbCr etc. and for different initial segmentation like watershed [20], super pixels [21].

We first test effect of color space on region merging result. In this experiment RGB color space is converted into HSV,

YCbCr color spaces. Fig. 6. shows the object retrieval from RGB space, where as Fig. 10 and Fig. 11 shows the object retrieval from HSV and YCbCr color spaces respectively. Also Fig. 12. and Fig. 13. shows that object retrieval using different initial segmentation besides mean shift, watershed [20] is another important method of initial segmentation. Different from mean shift it partitions the image into more number of regions. Fig. 12. and Fig. 13.shows the result on bird and horse. Due to large number of regions in the initial segmentation of images using watersheds [20] its running time is more in comparison to mean shift initial segmentation and also results shows that the retrieve object after processing in similar fashion is not good as comparison with mean shift.

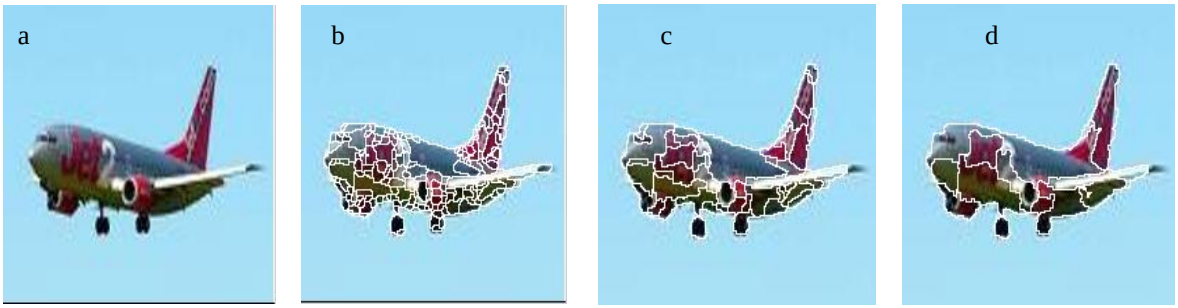
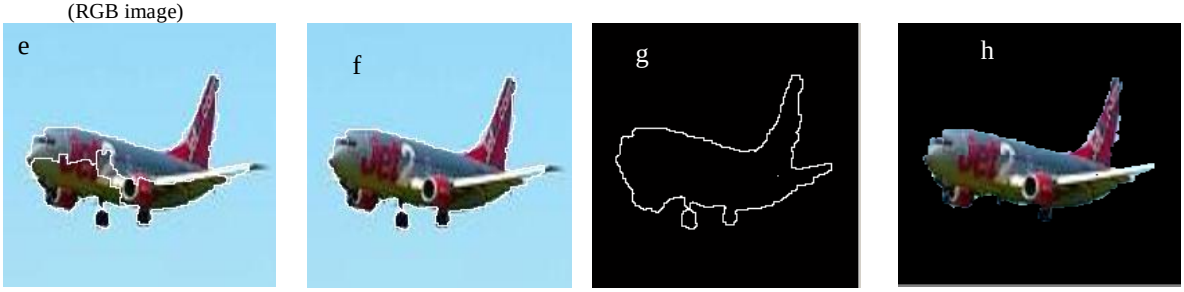


Fig. 6 Original image (RGB image) Mean shift segmentation 1st stage merging 2nd stage merging



3rd stage merging Object contour Object mask Object

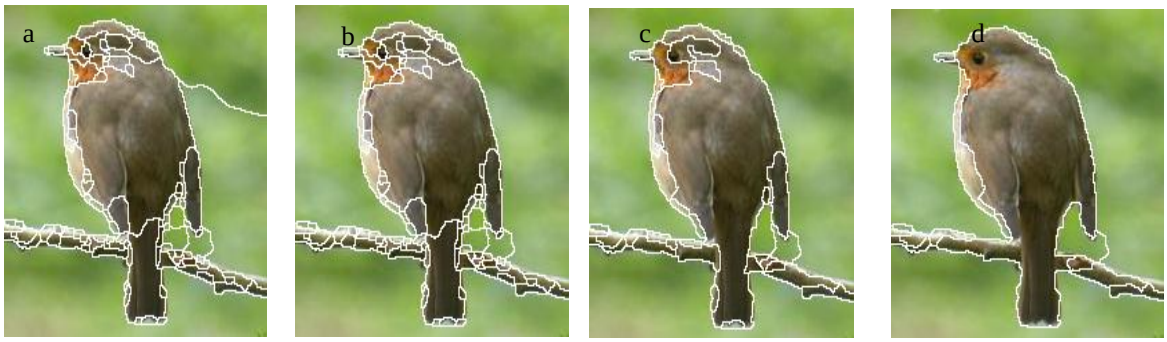
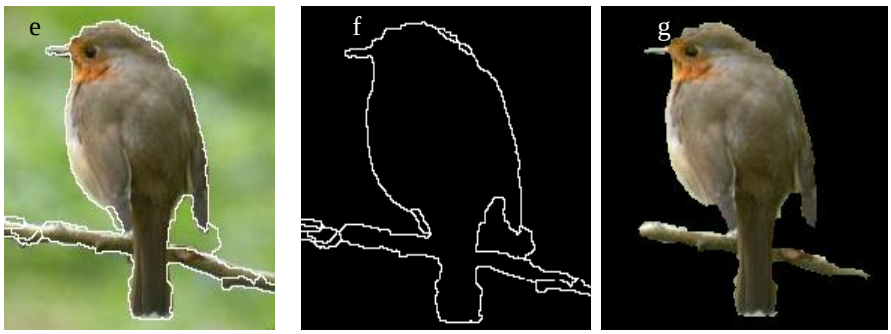


Fig. 7 Mean shift segmentation (RGB Image) 1st stage merging 2nd stage merging 3rd stage merging



3rd stage merging Object mask Object

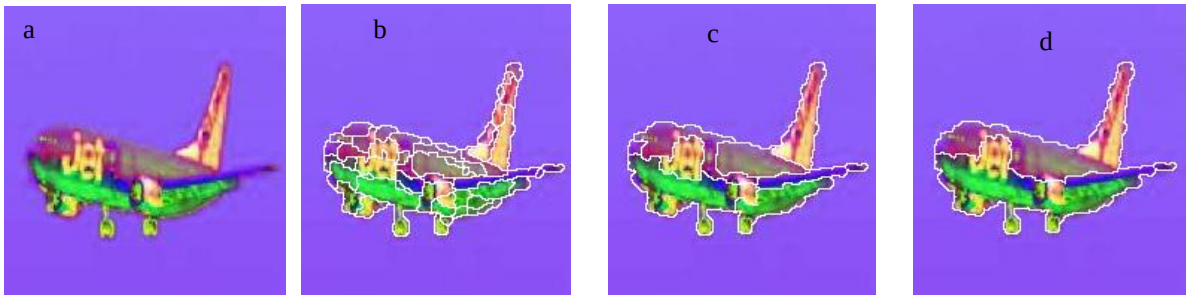
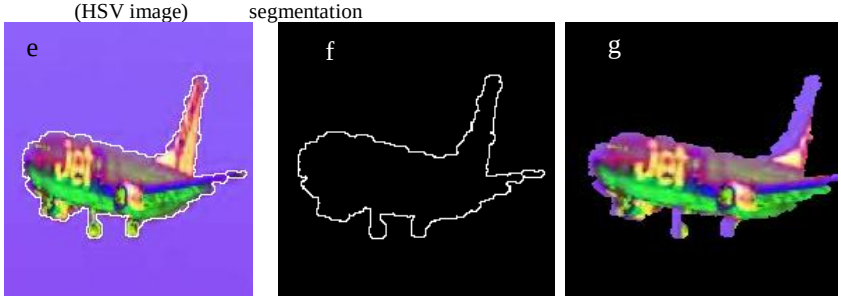


Fig. 8 Original image (HSV image) | Mean shift segmentation | 1st stage merging | 2nd stage merging



Object Contour | Object Mask | Object

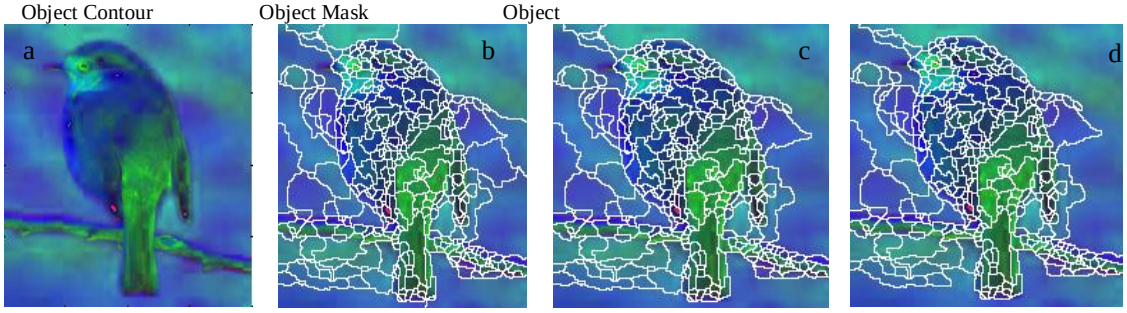
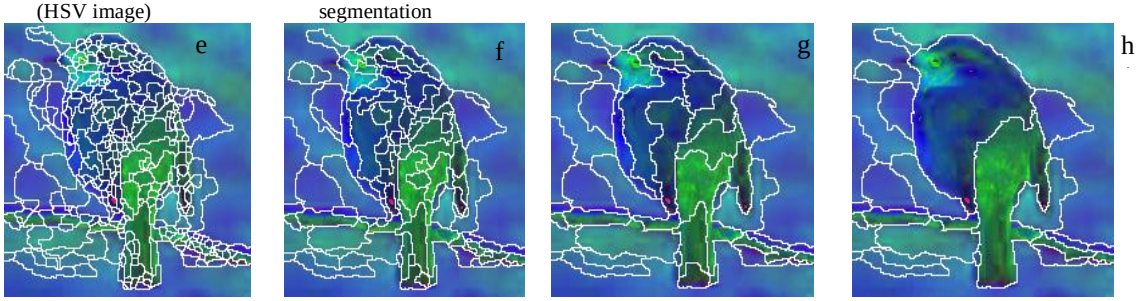
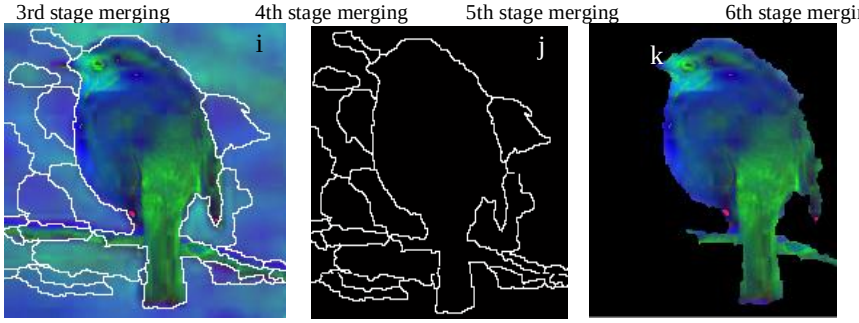


Fig. 9 Original image (HSV image) | Mean shift segmentation | 1st stage merging | 2nd stage merging



3rd stage merging | 4th stage merging | 5th stage merging | 6th stage merging



7th stage merging | Object mask | Object

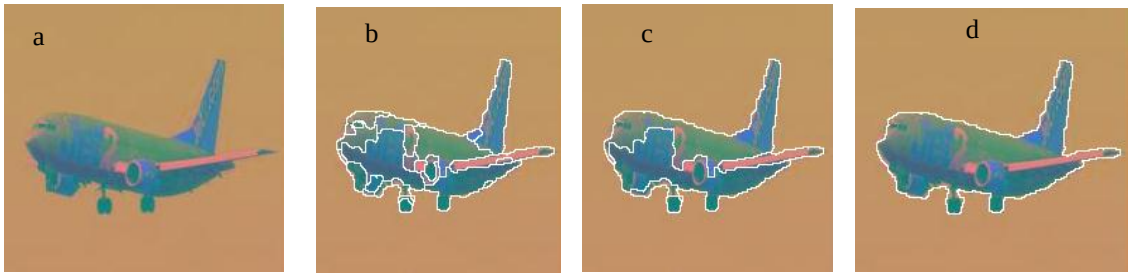
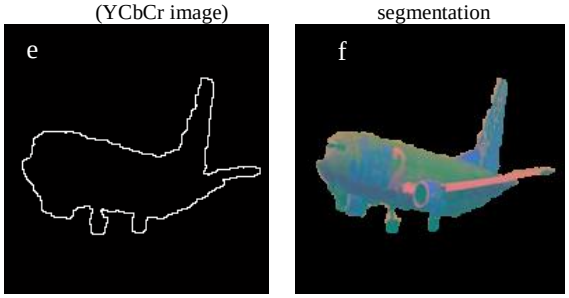


Fig. 10. Original image (YCbCr image) | Mean shift segmentation | 1st stage merging | 2nd stage merging



Object mask | Object

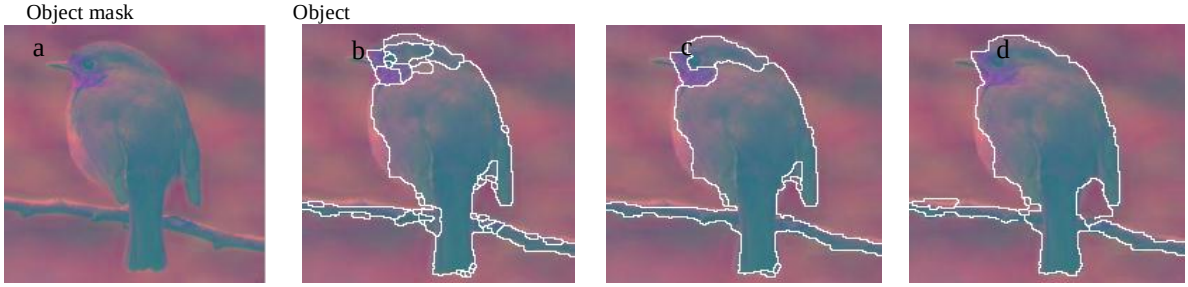
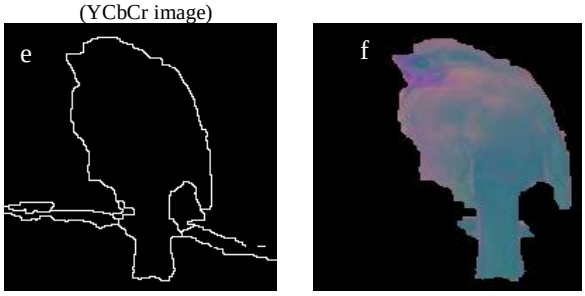


Fig. 11. Original image (YCbCr image) | Mean shift segmentation | 1st stage merging | Object contour



Object mask | Object

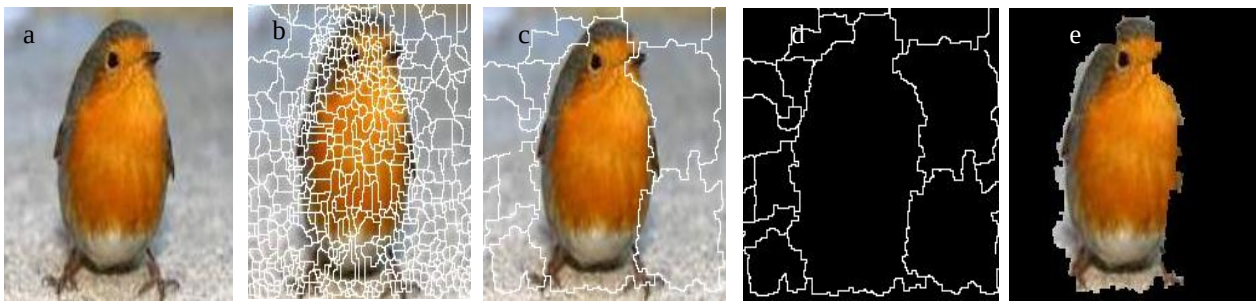


Fig. 12.Original image | Watershed segmentation | Object contour | Object Mask | Object

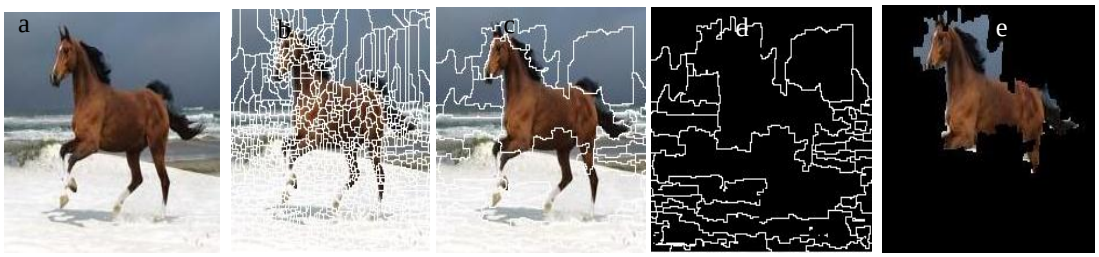


Fig. 13. Original Image Watershed Segmentation Object contour Object Mask Object

REFERENCES

[1]. E. Sharon, A. Brandt, and R. Basri "Segmentation and Boundary Detection using Multiscale Intensity mesurements" Proc. IEEE conf. Computer Vision and Pattern Recognition, pp.469-476, 2001.

[2]. M. Galun, E.sharon, R. Basri and A. Brandt "Texture Segmentation by Multiscale Aggregation of Filter Responses and Shape Elements" Proc. IEEE Int'l conf. Computer Vision and Pattern Recognition, pp. 716-723, 2003.

[3]. E. Sharon, M. Galun, D. Sharon, R. Basri and A. Brandt, "Hierarchy and Adaptivity in Segmenting Visual Scenes," Nature vol. 442, no. 7104, pp. 810-813, june 2006.

[4]. J. Shi and J. Malik, "Normalized Cuts and Image Segmentation," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 22, no. 8, pp. 888-905, August 2000.

[5]. D.A. Forsyth, J.ponce, Computer Vision: A Morden Approch, Prentice-Hall Englewood Cliffs, NJ, 2002.

[6]. N.Bonnet, J.Cutrona, M.Herbin, A no Threshold Histogrm Based Image Segmentation Method, Pattern Recognition vol. 35, no.10, pp. 2319-2322, 2002.

[7]. E.Navon, O.Miller, A.Averbunch, "Color Image Segmentation Based on Adaptive Local Thresholds," Image *and* Vision Computing vol. 23, no.1, pp. 69-85, 2005.

[8]. S. Osher, N. Paragios, "Geometric Level Set Methods in Imaging," Vision and Graphics, Springer, New York, 2003.

[9]. Y. Boykov, G. Funka-Lei, "Graph cuts and efficient n-d image segmentation" International Journal of Computer vision, vol 70. no.2, pp.109-131,2006.

[10]. D. Martin, C.Folwkes, D. Tal, J.Malik, "A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics," Internatinal Conference on Computer Vision, pp. 416-423, 2001.

[11]. M. Kass, A. Witkin, D. Terzopoulos, "Snake: Active Contour Models," International Journal of Computer Vision, vol.1, no.4, pp. 321-331, 1987.

[12]. F. Meyer, S. Beucher, "Morphological Segmentation," Journal of Visual Communication and Image representation, vol.1, no. 1, pp. 21-46, 1990.

[13]. P. Felzenszwalb, D. Huttenlocher, "Efficient Graph Based Image Segmentation,"
International Journal of Computer Vision, vol. 59, no. 2, pp. 167-181, 2004.

[14]. Q. Yang, C.Wang, X. Tang, M. Chang and Z. Ye, "Progressive cut and Image Cutout Algorithm That Models User Intention," IEEE Multimedia vol.14, no.3, pp. 56-66, 2007.

[15]. Y. Li, J. Sun, C. Tang, H. Shum, "Lazy Snapping" SIGGRAPH vol. 23, pp.303-308, 2004.

[16]. S.X. Yu, R. Gross and J.shi, "Concurrent Object Recognition and Segmentation by Graph Partitioning,"
Proc. Neural Information Processing System, pp. 1383-1390, 2002.

[17]. E. Borenstein and S. Ullman, "Class-Specific, Top-Down Segmentation,"
proc. Seveth European Conf. Computer Vision, pp. 109-124, 2002.

[18]. Y. Cheng, "Mean Shift, Mode Seeking and Clustering," IEEE Transaction on Pattern and Machine Intelligence, vol. 17, no.8, pp. 790-799, 1995.

[19]. D. Comaniciu, P. Meer, "Mean Shift: A Robust Approch Towards Feature Space Analysis," IEEE Trnsaction on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pp. 603-619, 2002.

[20]. L. Vincent, P. Soille, "Watersheds in Digital Space: An Efficient Algorithm Based on Immersion Simulations," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 13, no. 6, pp. 583-598, 1991.

[21]. X. Ren, J. Malik, "Learning a Classification Model for Segmentation," *ICCV03*, vol.1, pp. 10-17, Nice, 2003.

[22]. Y. Li, J.Sun, H. Shum, "Vedio Object cut and Paste," SIGGRAPH vol. 24, pp. 595-600, 2005.

[23]. J.Ning, et al., "Interactive Image Segmentation by Maximal Similarity Based Region Merging," Pattern Recognition, vol. , no. , pp., 2009.

[24]. B. Sumengen, "Variational Image Segmentation and Curve Evolution on Natural Images," Ph.D. Thesis, University of California.

[25]. EDISON software. (http://www.caip.rutgers. edu/riul/ research/code.html).

[26]. C. Christoudias, B. Georgescu, P. Meer, "Synergism in Low Level Vision," Proceedings of Internatinal Conference on Pattern Recognition, vol. 4, pp. 150-155, 2002.

[27]. T. Ozala, M. Pietikainen, T. Maenpaa, "Multi-resolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 24, no. 7, pp. 971-987, 2002.

[28]. S. Birchfield, "Elliptical Head Tracking Using Intensity Gradient and Color Histograms," Proceeding of IEEE Conference Computer Vision and Pattern Recognition, pp. 232-237, 1998.

[29]. K. Fukunaga, "Introduction to Statistical Pattern Recognition," Second ed., Academic Press, 1990.

[30]. T. Kailath, "The Divergence and Bhattacharyya Distance Measures in Signal Selection," IEEE Transaction on Communication Technology, vol. 15, no. 1, pp. 52-60, 1967.

[31]. D. Comaniciu, V. Ramesh, P. Meer, "Kernel-Based Object Tracking," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564-577, 2003.

[32]. M. Sonka, V. Hlavac, R. Boyle,"Image Processing, *Analysis and Computer Vision," Thomson*, 2007.

[33]. Y. Boykov, V. Kolmogorov, "An Experimental Comparison of min-cut/max-flow Algorithm for Energy Minimization in Vision," *IEEE Trasaction on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124-1137, 2004.

[34]. P. Meer, Stochastic Image Pyramids, Computer Vision, *Graphics and Image processing (CVGIP)*, vol. 43, no. 3, pp. 269-294, 1989.

[35]. J.M. Jolion, "The Adaptive Pyramid a framework for 2D Image Analysis," Computer Vision, Graphics and Image Processing (CVGIP), *Image Understanding* vol. 55, no.3, pp. 339-348, 1992.

[36]. J.Wang, V. Thiesson, Y.Xu, M. F. Cohen, "Image and Vedio Segmentation by Anisotropic Kernel Mean Shift," *Proceeding of the European Conference on Computer Vision, prague, Czech Republic*, vol. 3022, pp. 238-249, 2004.

Tifinagh Character Recognition Using Geodesic Distances, Decision Trees & Neural Networks

O.Bencharef, M.Fakir, B. Minaoui
Sultan Moulay Slimane University,
Faculty of Science and Technology,
Beni Mellal -- Morocco

B.Bouikhalene
Sultan Moulay Slimane University,
Polydiscilinary faculty,
Beni Mellal – Morocco

***Abstract*—The recognition of Tifinagh characters cannot be perfectly carried out using the conventional methods which are based on the invariance, this is due to the similarity that exists between some characters which differ from each other only by size or rotation, hence the need to come up with new methods to remedy this shortage. In this paper we propose a direct method based on the calculation of what is called Geodesic Descriptors which have shown significant reliability vis-à-vis the change of scale, noise presence and geometric distortions. For classification, we have opted for a method based on the hybridization of decision trees and neural networks.**

Keywords-component ; Tifinagh character recognition; Neural networks ; Decision trees, Riemannian geometry ; Geodesic distances.

I. INTRODUCTION

Recently, computer vision has become one of the most appealing fields of research where shape recognition stands as one of its main pillars.

In the classical scheme of shape recognition, we distinguish basically two major phases: (i) the extraction and (ii) the classification of descriptors. [1][2]

The descriptors extraction can be defined as a particular form of downsizing, which aims to simplify the amount of resources needed to describe a large set of data accurately. Different techniques have been used [3][4][5].

In this paper, we present a new approach for the extraction process which is based on the calculation of geodesic distances within images containing Tifinagh characters. The geodesic distance is one of the basic concepts of Riemannian geometry that comes out in many contexts to compensate the insufficiency of Euclidean geometry. For instance, it is used in mapping to calculate the length of a path on a spherical surface, it is also used for adaptive mesh generation and 3D objects representation [6][7]. The objective is to adapt all these tools in order to use them for Tifinagh character recognition.

To test our approach, we have opted for a classifier based on the hybridization of neural networks (NN) and decision trees.

This paper is organized as follows: section two provides an overview on Tifinagh characters, section three describes some of the basic notions of Riemannian geometry and explains the method we applied, section four emphasizes on the classification process and the last section is dedicated to experimental results.

II. THE TIFINAGH CHARACTERS

Historically, Tifinagh characters were popular with Moroccan theologians under the name "Khath Ramal", that is "sand characters". That was the writing of caravan traders who used it to exchange messages by leaving signs on caravan routes. Tifinagh characters have almost become mystical due to the importance of communication in finding paths during journeys in desert.

Those characters are kept by Saharan community and represent today the ancient writing of "Touaregue". Archeologists have found texts in Tifinagh in different shapes: geometrical, human or even divine. They have also noticed resemblance to other characters from foreign civilizations: Phoenicians, Russian and Aramaic.

According to researchers, the name Tifinagh is compound of two words: Tifi (that is "discovering") and Nagh (that is "one's self"). The Royal institute of Amazigh Culture (ICRAM) has proposed a standardization of Tifinagh characters composed of 33 elements. [8],[9]

Figure 1. Tifinagh characters adopted by the ICRAM

III. EXTRACTION OF GEODESIC DESCRIPTORS

A. Theoretical approach

1) Basic concept of Riemannian geometry

Riemannian geometry was first put forward by Bernhard Riemann in the nineteenth century. It deals with a broad range of geometries which metric properties vary from a point to another. We define Riemannian geometry as the studies of

Riemannian manifolds: smooth manifolds with a Riemannian metric.[10] To better understand this, we present some basic definitions:

- A manifold is a topological space that is locally Euclidean (i.e., around every point, there is a neighboring area that is topologically the same).[11][12]
- An inner product is a generalization of the dot product. In a vector space, it is a way to multiply vectors together, with the result of this multiplication being a scalar. The inner product of two vectors u and v is given by:

$$<u, v>M = {}^{t}u\, M\, v. \quad (1)$$

- The collection of all inner products of a manifold is called the Riemannian metric.

2) Geodesic distance

In a Riemannian metric space (x, M (x)) the length of a path [a,b] is calculated using the parameterization '(t) = a +t ab, where t belongs to [0, 1]. [13]

Then:

$$\ell_{\mathcal{M}}(\mathbf{ab}) = \int_0^1 \|\gamma'(t)\|\, \mathrm{d}t = \int_0^1 \sqrt{{}^t\mathbf{ab}\, \mathcal{M}(\mathbf{a}+t\,\mathbf{ab})\, \mathbf{ab}}\, \mathrm{d}t. \quad (2)$$

The geodesic distance (Dl_M) is the shortest path between two points a and b, or one of the shortest paths if there are many:

$$Dl_M(ab)=\mathrm{Min}(l_M(ab)) \quad (3)$$

B. Proposed method

The proposed extraction process is based on the calculation of geodesic distances between the four geometric extremities of the sought character.

1) Pretreatment

The pretreatment that we have integrated is composed of two standard processes, (i) the noise elimination and (ii) the contour detection.[14] (Figure 2)

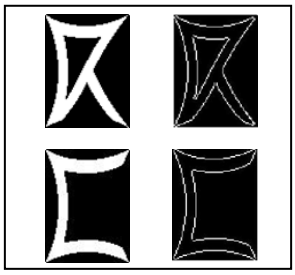


Figure 2. Example of contour detection

2) Extremities detection

In order to detect extremities, we have used an algorithm that browses the character contour and detects the closest points to each of the image angles.

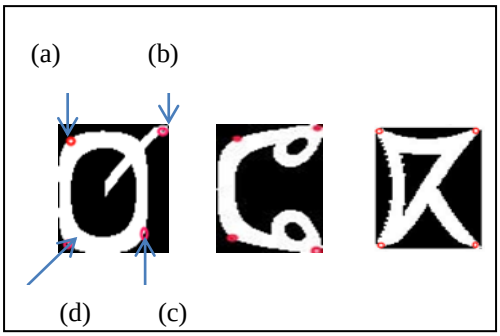


Figure 3. Example of character extremities

3) Geodesic descriptors

We named "Geodesic Descriptors" geodesic distances between the four extremities of the image divided by their Euclidean distances.

Considering:

- Dl_M(xy): the geodesic distance between x and y
- *d*xy: the Euclidean distance between x and y
- a,b,c & d the geometric extremities of each character (Figure 2)

We will name:

- *1st metric descriptor* *D1= DlM(ab) / dab*
- *2nd metric descriptor* *D2= DlM(ac) / dac*
- *3rd metric descriptor* *D3= DlM(ad) / dad*
- *4th metric descriptor* *D4= DlM(bc) / dbc*
- *5th metric descriptor* *D5= DlM(bd) / dbd*
- *6th metric descriptor* *D6= DlM(cd) / dcd*

To compute geodesic distances on a binary image, we have applied an algorithm that uses a scan function where each iteration has sequences that go forward and backward so as to determine the shortest path. The used algorithm considers orthogonal and diagonal pixel distances by using a weight of 1 to orthogonal pixels and a weight of square root of 2 for the diagonal markers.[15][16]

To insure resistance to scale changes of the proposed descriptors, we divided the geodesic distance of each path according to the Euclidian distance.

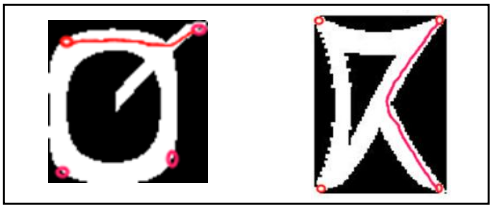


Figure 4. Geodesic distances between a & b for the ''yame'' character and between b & c for the '"yake" character

Table I illustrates the obtained results for the six descriptors used in this article:

Table I. Results for some Tifinagh characters

	D1	D2	D3	D4	D5	D6
	2.12	1.31	1.29	1.21	1.23	1.02
	1.10	1.62	1.13	1.38	1.11	1.40
	2.02	1.41	1.71	1.70	1.43	2.00
	1.07	1.04	1.03	1.41	1.12	1.73

Notice that the proposed descriptors have allowed:

- clear distinction between the tested characters ; and
- Distinction between characters which are geometrically close (obtained by rotation, like "Yars" & "Yass" characters, see Table II).

Table II. Geodesic descriptors for the ''Yars'' & ''Yass'' characters

	1.13	1.48	1.11	1.22	1.40	1.20
	1.23	1.39	1.12	1.20	1.5	1.12

The proposed descriptors have also shown considerable resistance to scale changes. (Table III)

Table III. Metric descriptors calculated for different sizes of the character ''Yass''

	D1	D2	D3	D4	D5	D6
	1.19	1.44	1.07	1.15	141	1.05
	1.22	1.39	1.10	1.18	1.49	1.11
	1.23	1.39	1.12	1.20	1.5	1.12

IV. CLASSIFICATION

At first glance, it seems that the proposed descriptors can distinguish between all Tifinagh characters (Figures 5, 6 & 7). However, confusion still remains when it comes to composed characters (Figure 6) or other characters that have a circular shape (Figure 7).

Figure 5. Characters defined directly by geodesic descriptors

Figure 6. Composed Characters

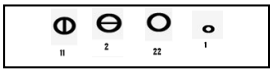


Figure 7. Circular Characters

To deal with these particular cases, we have chosen to operate with a hybrid classifier made of decision trees and neural networks.

On the one hand, decision trees have a discriminatory characteristic which allowed us to separate characters in four classes (Figure 8). On the other hand, neural networks allow character recognition, thanks to their ability to implicitly detect complex nonlinear relationships between dependent and independent variables, and to detect all possible interactions between predictor variables.[17][18].

In practice, we used a multilayer neural network (two layers) with supervised learning, driven by the back propagation of the gradient. This consists in determining the error made by each neuron and then modifying values of weight in order to minimize this error.

For the decision tree, we used the following rules:

- R1: after detecting the number of motifs N in the image. If N >1, then: R22, if not: R21.
- R22: if the size of the first motif if twice (or more) bigger than the size of the second motif, then: N3, if not: N4.
- R21: if the ratio of geodesic distances (D1/D3) is between 0.8 & 1.2 and (D2/D6) is between 0.8 & 1.2, then: N2, if not: N1.

V. EXPERIMENTAL RESULTS

We tested our approach of Tifinagh character recognition on the database "Y. Ouguengay"[12]. This database includes 2175 characters printed in different sizes and writing styles. (Figure 9) Each character will be determined using geodesic descriptors, identification by neural networks and by combined neural networks (using decision trees).

We tested our approach on different characters of the database. Table IV gives an idea about recognition ratios of the database objects.

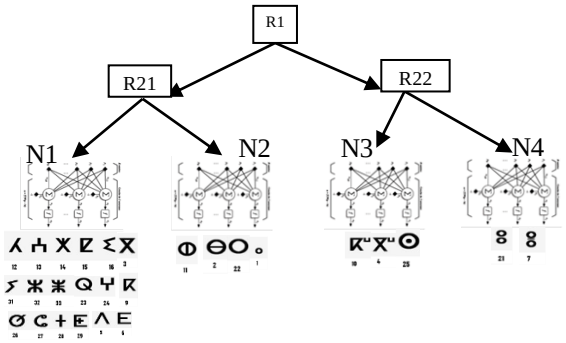


Figure 8. Integrated classification process used to recognize Tifinagh characters

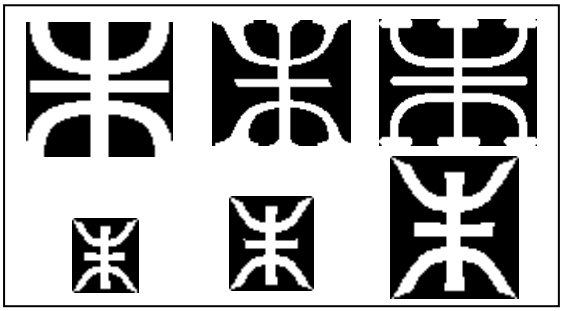


Figure 9: A Tifinagh character from the ''Y.Ouguengay'' database printed in different sizes and writing styles

Table IV. Recognition rate depending on the number of characters to identify

Number of characters to identify	Neural Networks (NN)*	NN & Decision trees*
10	98%	99%
20	93%	95%
25	81%	94%
33	71%	93%

*: Results obtained for centered images

Notice that despite the size of the database which is of 16 samples of each character, the suggested descriptors have proven effective using neural networks. The integration of decision trees has brought the recognition ratios remarkably higher.

In order to test the reliability of our recognition approach, we used it on images presenting different kinds of alterations. As noticed on Table V, recognition ratios are excellent vis-à-vis noise presence and handwritten characters. Ratios are good when it comes to variation of luminosity and changes in scale.

Table V. Recognition rate on images presenting alterations using NN & Decision trees

Alterations	Hm	Lu	Pn	Sc
Recognition Rate	97%	93%	95%	92%

With: Hm: Handwritten characters, Lu : Variation of luminosity, Pn : Presence of noise , Sc : Scale change

VI. CONCLUSION

In this study, we have used the geodesic distances as a new approach for shape descriptors extraction and we have opted for a hybridization of neural networks and decision trees for classification. The robustness of our recognition system was tested and illustrated on a Tifinagh database supplemented by images with different alterations such as the luminance variation, the presence of white Gaussian noise with a variance of 10% and alterations due to handwriting. The recognition system proved efficient as we obtained:

- A recognition rate of 99% for a training set composed of 20 samples of 10 characters; and
- An excellent robustness vis-à-vis the presence of noise and a good robustness in the case of geometric distortion and luminance variation.

The results can be improved by acting on several parameters such as:

- Increasing the training set;
- Parallel use of other shape descriptors; and
- The integration of discriminative characteristics of the different characters.

REFERENCES

[1] Oren Boiman, Eli Shechtman and Michal Irani(2008), In Defense of Nearest-Neighbor Based Image Classification, IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[2] A. Bosch, A. Zisserman, and X. Munoz.(2007), Image classification using random forests and ferns.In ICCV.

[3] F.L. Alt.(1962), Digital pattern recognition by moments, J. ACM, pp. 240–258.

[4] S.A. Dudani,(1977), Aircraft identification by moment invariants, IEEE Trans. Comput.,pp 39–45.

[5] Chee-Way Chonga, P. Raveendranb and R. Mukundan,(2003), Translation invariants of Zernike moments", Pattern Recognition , pp 1765– 1773.

[6] X.Gu ,(2004) Genus Zero Surface conformal apping,IEEE TANSACTIONS ON MEDICAL IMAGING,VOL.23 NO.8.

[7] E.KALSEN & al (2004) Analysis of planar shapes using geodesic paths on shape spaces, IEEE TRANSACTIOS ON PATTERN ANALYSIS AND MACHINEINTELIGENCE , vol .26, No 3 .

[8] A. Rachidi, D. Mammass. (2005), Informatisation de La Langue Amazighe: Méthodes et Mises En OEuvre, SETIT 2005 3rd International Conference: Sciences of Electronic Technologies of Information and Telecommunications March 27-31, 2005 – TUNISIA.

[9] M. Amrouch, Y. Es Saady, A. Rachidi, M. Elyassa, D. Mammass (April 2009), Printed Amazigh Character Recognition by a Hybrid Approach Based on Hidden Markov Models and the Hough Transform, ICMCS'09, Ouarzazate-Maroc.

[10] S. Gallot, D. Hulin, et J. Lafontaine. Riemannian Geometry. Universitext. Springer Verlag, New York, 1990.

[11] A. Fuster, L. Astola and L. Florack, A Riemannian Scalar Measure for Diffusion Tensor Images, Lecture notes in Computer Science, 5702 (2009), pp. 419–426..

[12] X. Pennec, P. Fillard and N. Ayache, A Riemannian Framework for Tensor Computing, Int. J. Computer Vision, 66(1) (2006), pp. 41–66.

[13] Kimmel, R. & Sethian, J. A. (1998). Computing geodesic paths on manifolds, Proceedings of the National Academy of Sciences of the United States of America, Vol. 95, No. 15, pp. 8431-8435

[14] L.D. Cohen. Multiple Contour Finding and Perceptual Grouping Using Minimal Paths. Journal of Mathematical Imaging and Vision, 14(3), 2001 . Presented at VLSM01

[15] J.KERL (2008), Numerical differential geometry in Matlab, Graduate student Colloquium, university of Arizona.

[16] J.A. Sethian et R. Kimmel. Computing Geodesic Paths on Manifolds. Proc. Natl. Acad. Sci., 95(15):8431–8435, 1998.

[17] SIMARD P., STEINKRAUS D., PLATT J. C.(2005), Best Practices for onvolutional Neural Networks Applied to Visual Document Analysis, ICDAR, pp. 958-962

[18] G. R. Dattatreya and L. N. Kanal, " Decision trees in pattern recognition," In Progress in Pattern Recognition 2, Kanal and Rosenfeld (eds.) , Elsevier Science Publisher B.V., 189-239 (1985).

AUTHORS PROFILE

Omar BENCHAREF obtained he's DESS degree in 2007 from the University of Cadi Ayyad Marrakech Morocco. Currently he is a PhD student at the Center of Doctoral Studies in the Faculty of Science and technology of Beni Mellal. His research concerns image processing & Recognition.

Dr Mohamed FAKIR obtained a degree in Master of Electrical Engineering from Nagaoka University of Technology in 1991 and a Ph.D. degree in electrical engineering from the University of Cadi Ayyad, Morocco. He was a team member in Hitachi Ltd., Japan between 1991 and 1994. He is currently a professor at the Faculty of Science and Technology, University Sultan Moulay Slimane, Morocco. His research concerns the recognition and artificial intelligence.

Dr. Belaid BOUIKHALENE obtained a Ph.D. degree in Mathematics in 2001 and a degree of Master in Computer science in 2005 from the University of Ibn Tofel Kenitra, Morocco. He is currently a professor at University Sultan Moulay Slimane, Morocco, His research focuses on mathematics and applications, decision information systems, e-learnig, pattern recognition and artificial intelligence.

Dr. Brahim MINAOUI obtained a Ph.D. degree in physics. He is currently a the Faculty of Science and Technology,professor at University Sultan Moulay Slimane, Morocco, His research focuses on mathematics and applications, decision information systems, recognition, Artificial intelligence & physics.

Motion Blobs as a Feature for Detection on Smoke

Khalid Nazim S. A.,
Research Scholar,
Singhania University,
Rajasthan, India

Dr. M.B. Sanjay Pande,
Professor and Head,
Dept. of Computer Science & Engineering,
VVIET, Mysore, India.

***Abstract*— Disturbance that is caused due to visual perception with the atmosphere is coined as smoke, but the major problem is to quantify the detected smoke that is made up of small particles of carbonaceous matter in the air, resulting mainly from the burning of organic material. The present work focuses on the detection of smoke immaterial it being accidental, arson or created one and raise an alarm through an electrical device that senses the presence of visible or invisible particles or in simple terms a smoke detector issuing a signal to fire alarm system / issue a local audible alarm from detector itself.**

Keywords- Motion blobs; Blob Extraction; Feature Extraction.

I. INTRODUCTION

Smoke**:** Any disturbance that is caused due to visual perception with the atmosphere can be termed as smoke. But on a contrary it can also be defined in many ways such as, the vaporous system made up of small particles of carbonaceous matter in the air, resulting mainly from the burning of organic material, such as wood or coal OR a suspension of fine solid or liquid particles in a gaseous medium OR a cloud of fine particles OR something insubstantial, unreal, or transitory OR a substance used in warfare to produce a smoke screen OR something used to conceal or obscure OR a pale to grayish blue to bluish or dark gray OR smoke is the collection of airborne solid and liquid particulates, gases emitted when a material undergoes combustion or pyrolysis[17,21]. Research in detecting smoke using surveillance cameras has become very active recently. It is now possible to address the problems in traditional smoke detectors based on particle sampling with the aid of video smoke detection namely:

1) Traditional smoke detectors require a close proximity to the smoke.

Video forensic evidence for future fire investigations. The video based detectors can sense:

- Presence of flames within the field of view of the camera.
- Reflected fire light when flames are obstructed.
- Presence of pluming smoke clouds.
- Presence of ambient smoke.
- Unauthorized Intrusion.

Ugur Toreyin et. al., presented a method for smoke detection in video. It is assumed that camera monitoring the scene is stationary. Since the smoke is semi-transparent, edges of image frames start losing their sharpness and this leads to a decrease in the high frequency content of the image. To determine the smoke in the field of view of the camera, the background of the scene is estimated and decrease of high frequency energy of the scene is monitored using the spatial wavelet transformations of the current and the background images[7]. Edges of the scene are especially important because they produce local extrema in the wavelet domain.

2) They usually do not provide information about fire location, size etc.

The most interesting concept of this paper is to differentiate the type of smoke based on the texture or colour such as:

Type 1: White smoke: This occurs due to anti-freeze burning of the piston cylinder. The possible ways of causes are a cracked head, blown head gasket, (warped head), or cracked cylinder block (normally uncommon).

Type 2: Black smoke: Black smoke is oftentimes a result of too much fuel and not enough air in the combustion chamber. In rare cases, it can be caused by weak fuel pressure causing fuel to 'drip' from injectors rather than 'spray'. It can also be caused by weak fire in the combustion chamber.

Type 3: Gray smoke: Gray smoke is caused by brake fluid. It generally means that the brake master cylinder is bad and is getting sucked through the vacuum brake hose. *Type 4:* B*lue smoke:* Blue smoke is generally caused by the burning of oil in the combustion chamber. Normal causes of oil getting into the combustion chamber are weak piston rings, bad valve guides, bad valve seals or plugged up engines where oil is sucked back through PCV system [8, 15, 17, 21].

II. LITERATURE SURVEY

A. Video Based method for Smoke Detection

In video-based smoke detectors, CCTV (Closed-circuit television) cameras can monitor and recognize smoke and flames overlooking large spaces at great distances, while providing video surveillance capabilities as a bonus [23,25]. This shall detect fire in seconds, supply vital situational awareness in the form of live video to remotely located guards, trigger fire alarms and provide vast amounts of pre-recorded

A decrease in the values of local extrema is also an indicator of smoke. In addition, scene becomes grayish when there is smoke and hence this leads to a decrease in chrominance values of pixels. Periodic behavior in smoke boundaries and convexity of smoke regions are also analyzed. All of these clues are combined to reach a final decision. Fire detection algorithms are based on the use of color and motion information in video to detect the flames [12]. However, smoke detection is vital for fire alarm systems when large and open areas are monitored, because the source of the fire and flames cannot always fall into the field of view.

Edges in an image correspond to local extrema in wavelet domain. A Gradual decrease in their sharpness results in the decrease of the values of these extrema. However, these extrema values corresponding to edges do not boil down to zero when there is smoke [11,13]. In fact, they simply loose some of their energy but they still stay in their original locations, occluded partially by the semi-transparent smoke. Independent of the fuel type, smoke naturally decreases the chrominance channels U and V values of pixels. Apart from this, it is well-known that the flicker frequencies of flames are around 10 Hz, this flicker frequency is not greatly affected by either the fuel type or the burner size [5, 12]. As a result, smoke boundaries also oscillate with a lower frequency at the early stages of fire. Another important feature of the smoke that is exploited in this method is that smoke regions have convex shapes [11].

An algorithm for detecting smoke in video was developed which is based on determining the edge regions whose wavelet sub-band energies decrease with time. These regions are then analyzed along with their corresponding background regions with respect to their RGB and chrominance values. The flicker of the smoke and convexity of smoke regions are also set as clues for the final decision. This method can also be used for the detection of smoke in movies and video databases. In addition to this can also be incorporated with a surveillance system monitoring an indoor or an outdoor area of interest for early detection of fire [1,2,4].

R.J. Ferraria et.al, proposed a real-time image processing technique for the detection of steam in video images. The problem of detecting steam is treated as a supervised pattern recognition problem. A statistical Hidden Markov Tree (HMT) model derived from the coefficients of the Dual-Tree Complex Wavelet Transform (DT-CWT) in small (48 × 48) local regions of the image frames is used to characterize the steam texture pattern. The parameters of the HMT model are used as an input feature vector to a Support Vector Machine (SVM) technique, specially tailored for this purpose [6,18]. By detecting and determining the total area covered by steam in a video frame, a computerized image processing system can automatically decide whether if the frame can be used for further analysis. The proposed method was quantitatively evaluated by using a labeled image data set with video frames sampled from a real oil sand video stream. The classifications of results were 90% correct when compared to human labeled image frames. This technique is useful as a pre-processing step in automated image processing systems [10, 16, 23].

Real-time automated image processing systems, used in size analysis, depend on good quality high contrast images in order to correctly segment and measure oil sand fragment size including oversize lumps [6]. According to Ziyou Xiong et.al. When a fire occurs, minimum detection latency is crucial to minimize damage and save lives. Current smoke sensors inherently suffer from the transport delay of the smoke from the fire to the sensor, a video smoke detection system would not have this delay. Further, video is a volume sensor, not a point sensor wherein a point sensor looks at a point in space, which may not be affected by smoke or fire. But a volume sensor potentially monitors a larger area and has much higher probability of successful early detection of smoke or flame.

Video smoke detection is a good option when smoke does not propagate in a "normal" manner, e.g., in tunnels, mines, and other areas with forced ventilation and in areas with air stratification, e.g., hangars, warehouses, etc. Video is also a good option for large, open areas where there may be no heat or smoke propagation to a fixed point e.g., saw mills, petrochemical refineries, forest fires, etc.

B. Background Subtraction

We follow the approach of Stauffer and Grimson [27] i.e., using adaptive Gaussian Mixture Model (GMM) to approximate the background modeling process. This is because in practice multiple surfaces often appear in a particular pixel and the lighting conditions change.

In this process, each time the parameters are updated, the Gaussians are evaluated to hypothesize which are most likely to be part of the background process. Gaussians are grouped using connected component analysis as moving blobs.

C. Flickering extraction

A pixel at the edge of a turbulent flame could appear and disappear several times in one second of a video sequence. This kind of temporal periodicity is commonly known as flickering. Flickering frequency of turbulent flame has shown experimentally to be around 10Hz. Flickering frequency of smoke however, could be as low as (2 ~ 3) Hz for slowly-moving smoke. The temporal periodicity can be calculated using Fast Fourier Transform (FFT), Wavelet Transform or Mean Crossing Rate (MCR). In our system, we have used the Mean Crossing Rate (MCR) method [3].

D. Smoke classification

Blobs with contours are candidates of smoke regions. Features are extracted from them and passed to a smoke classification module for further check. The features that we have used are based on the work by Catrakis et al. in characterizing turbulent phenomena. Smoke [13] and (non-laminar flow) flames [19] are both based on turbulent phenomena. The shape complexity of turbulent phenomena may be characterized by a dimensionless edge/area or surface/volume measure [13,26]. One way, of detecting smoke is to determine the edge length and area, or the surface area and volume of smoke in images or video[15,26].

E. Flame Recognition in Video

Walter Phillips III, Mubarak Shah and Niels da Vitoria Lobo, presented a paper based on an automatic system for fire detection in video sequences. Particle sampling, temperature sampling and air transparency testing are simple methods that are used most frequently today for fire detection. Unfortunately, these methods require a close proximity to the fire. In addition, these methods are not always reliable, as they do not always detect the combustion itself, most of them detect smoke, which could be produced in other ways.

Existing methods of visual fire detection rely almost exclusively upon spectral analysis using rare and costly spectroscopy equipment. This limits fire detection to those individuals who can afford the high prices of the expensive sensors that are necessary to implement these methods. In addition, these approaches are still vulnerable to false alarms

caused by objects that are of the same colour as fire, especially the sun. Healey, 1993 and Foo, 1995 have presented two previous vision-based methods that seem quite promising.

III. Data Collection

An Olympus digital camera with the specification (AF 3x optical zoom 6.5-19.5mm, 7.1 megapixel) is used for collecting the different data sets and we have assumed the camera to be stationary.

The fragrance sticks were used as the source of smoke. While recording the video, initially the still black background is captured for approximately one second and later the smoke is introduced, which was recorded for one more second.

Several such videos were collected and used to find the mean, standard deviation and variance of all the three components or channels of an RGB image (colored image).

The proposed architecture for the Video Based Smoke Detector is as shown below in fig 1 and comprises of the following five stages namely.

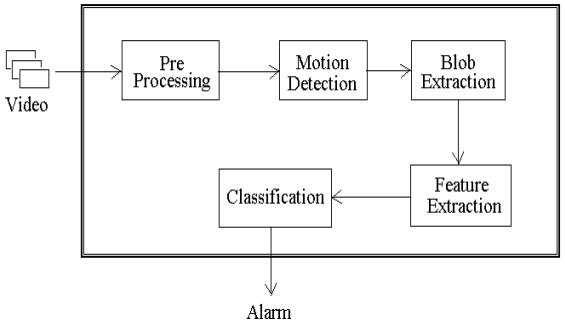


Fig 1: Proposed architecture of Video Based Smoke detector

Stage 1: The preprocessing stage:

In this stage of processing the image is filtered and noise is eliminated. Later the images are segmented for further processing

Stage 2: The Motion detection Stage:

This stage accepts the filtered image as input that involves the detection of moving objects entering the field of view.

Stage 3: The Blob Extraction Stage:

In this stage we make use of a unimodel and multimodel thresholding method for monochrome and color images respectively.

Stage 4: The Feature Extraction Stage:

This stage extracts the features of the input data to a reduced representation set of features, i.e. if the data is suspected to be notoriously redundant with not much of information.

Stage 5: The classification Stage:

This stage involves the classification of the extracted blobs depending on the presence of smoke or not and to raise an alarm subsequently.

IV. Implementation

The proposed architecture for the Video Based Smoke Detector comprises of different stages. The first stage is the *preprocessing* stage where the image will be filtered and noise will be eliminated. The filtered image is then given as input to the *motion detection* stage which involves the detection of moving objects entering the field of view. In *Blob Extraction*, we make use of unimodel thresholding and multimodel thresholding for monochrome and colour images respectively which provides presence of moving objects. The next stage is *Feature Extraction* where the output contains only required information obtained out of the large input data set (which is suspected to be notoriously redundant), this output data will be transformed into a reduced representation to obtain set of features.

The last and the final is the *Classification* stage where the extracted blobs are classified to check the presence of smoke or not.

Stage 1: The Preprocessing Stage

This stage is used to remove the noise present in the video as shown in fig 2 below. First the image is converted from RGB to gray scale. Once the image is converted to. grayscale, the Discrete Fourier Transform is used to transform the image from spatial domain to frequency domain.

For a square image of size (N×N), the two-dimensional DFT (Discrete Fourier Transform) is given by:

$$F(k,l) = \frac{1}{N^2} \sum_{a=0}^{N-1} \sum_{b=0}^{N-1} f(a,b)\, e^{-\iota 2\pi \left(\frac{ka}{N} + \frac{lb}{N}\right)}$$

where *f(a, b)* is the image in the spatial domain and the exponential term is the basis function corresponding to each point *F(k, l)* in the Fourier space.

Stage 2: The Motion detection Stage

As shown in the fig 3 above, after changing the data set, the portion of the frame affected by smoke is white in colour and the background is black.

Stage 3: The Blob Extraction Stage:

Blob extraction is an image segmentation technique that categorizes the pixels in an image as a part belonging to one of many discrete regions.

Fig 2: The Preprocessing Stage

Input frame	Output frame
	

The Motion Detection stage involves detection of moving objects entering the field of view. There are many approaches for motion detection in a continuous video stream. All of them are based on comparing the current video frame with the one from the previous frames or with something that is known as background. One of the most common approaches is to compare the current frame with the previous one. Also another approach is to compare the current frame not with the previous one but with the first frame in the video sequence. So if there were no objects in the initial frame, comparison of the current frame with the first one will give us the whole moving object good results in the cases where there is no guarantee that the first frame will contain only static background.

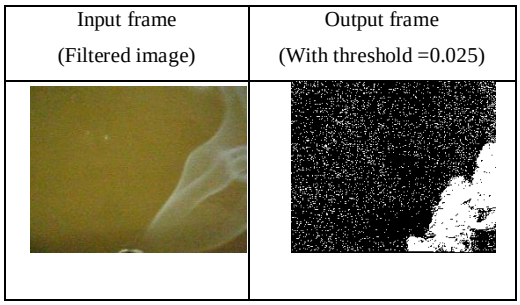


Fig 3: The Motion Detection

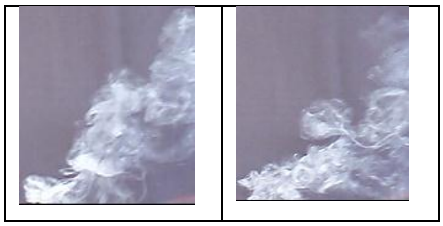


Fig 4: The output of blob extraction

The outcome after the blob extraction and cropping of blobs are as shown above in fig 4.

Blob extraction is generally performed on the resulting binary image from a thresholding step. Blobs may be counted, filtered and tracked. Inconsistent terminology for this procedure exists, including *region labelling, connected-component labelling and blob discovery* or *region extraction.*

Stage4: The Feature extraction Stage:

Feature extraction is a special form of dimensionality reduction. When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant that is much data, but not much information, then the input data will be transformed into a reduced representation set of features called as features vector. Transforming the input data into the set of features is called features extraction. If the features extracted are carefully chosen then it is expected that the features set will extract the relevant information from the input data in order to perform the desired task using this reduced representation instead of the full size input.

Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved. Analysis with a large number of variables generally requires a large amount of memory and computation power or a classification algorithm that over fits the training sample and generalizes poorly to new samples.

Feature extraction is a general term for methods of constructing combinations of the variables.

V. CONCLUSION

A video smoke detection system is termed as a volume sensor than an point delay sensor. A volume sensor potentially monitors a larger area and has much higher probability of successful early detection of smoke or flame. Video smoke detection is a good option when smoke does not propagate in a "normal" manner, for example, in tunnels, mines, and other areas with forced ventilation and in areas with air stratification, for example, hangars, warehouses, etc. Video is also a good option for large, open areas where there may be no heat or smoke propagation to a fixed point e.g., saw mills, petrochemical refineries, forest fires, etc.

In the present work, Video Based Smoke Detection, we process a given video to detect the presence of smoke and store it as sequence of images to a location on the disk. Since working on the video directly is not supported by Mat lab, we first convert the given video into *.avi* format (Audio/Video Interleaved) file format and later these frames were fetched sequentially for the filtering process and written back to the disk.

Work is not done in the field of developing an interface for a device to record the videos and to the hardware that connects to the fire alarm.

The goal of feature extraction is to characterize an object to be recognized by measurements whose values are very similar for objects in the same category and very different for objects in different categories. This leads to the idea of seeking distinguishing features that are invariant to irrelevant transformations of the input. In general, features that describe properties such as shape, color and many kinds of textures are invariant to translation, rotation and scaling.

REFERENCES

[1] Yuan F ,"A fast accumulative motion orientation model based on integral image for video smoke detection," Pattern Recog Lett 29(7):925– 932, 2008.

[2] Cui Y, Dong H, Zhou E ,"An early fire detection method based on smoke texture analysis and discrimination," In: Proceedings of the 2008 congress on image and signal processing, vol 3, CISP'08, pp 95–99, 2008.

[3] Xiong Z, Caballero R, Wang H, Alan MF, Muhidin AL, Peng P-Y, "Video-based smoke detection: possibilities, techniques, and challenges," In: IFPA, fire suppression and detection research and applications—a Technical working conference (SUPDET),orlando,FL,2007.

[4] R.J. Ferraria, H. Zhanga and C.R. Kube, "Real-time detection of steam in video images Pattern recognition," Volume 40,Issue 3, Pages 1148-1159, March 2007.

Extraction of Line Features from Multifidus Muscle of CT Scanned Images with Morphologic Filter Together with Wavelet Multi Resolution Analysis

Kohei Arai
Dept.of Information Science,
Graduate School of Science and Engineering
Saga University
Saga city, Japan

Yuichiro Eguchi
Dept. of Internal Medicine, Medical Faculty
Saga University
Saga city, Japan

Yoichiro Kitajima
Eguchi Hospital
Ogi city, Japan

***Abstract*—A method for line feature extraction from multifidus muscle of Computer Tomography (CT) scanned image with morphologic filter together with wavelet based Multi Resolution Analysis (MRA) is proposed. The contour of the multifidus muscle can be extracted from hip CT image. The area of multifidus muscle is then estimated and is used for an index of belly fat because there is a high correlation between belly fat and multifidus muscle. When the area of the multifidus muscle was calculated from the CT image, the MRA with Daubechies base functions and with the parameter of MRA of level is three would appropriate. After the wavelet transformation is applied to the original hip CT image three times and LLL (3D low frequency components) is filled "0" then inverse wavelet transformation is applied for reconstruction. The proposed method is validated with four patients.**

Keywords-multifidusmuscle; Computer Tomography; wavele; Multi Resulution Analysis; morphological filter.

I. INTRODUCTION

Nonalcoholic fatty liver diseases (NAFLD) are often associated with obesity, insulin resistance, and excessive visceral fat accumulation. The aims of this study were (1) to evaluate the relationship between the severity of fatty liver and visceral fat accumulation in nonalcoholic fatty liver diseases, and (2) to investigate the relationships of fatty liver with biochemical data and insulin resistance [1],[2].

It is effective to prevent the pain that receives the load and the stress easily by daily life, the movement, and labor, pierces, stimulates the multifidus muscle and the iliopsoas muscle of Toge muscle the peripheral nerve of congestion and a tumor, an inflammation, and a stripe film especially sidewise, and generates the sidewise Tsida muscle, interspinales muscles, and to strengthen the inner muscle such as the Tstoge muscles sidewise the tie of each one of the vertebra. Moreover, it is connected with improving the motor function of the entire spine to strengthen the multifidus muscle. Moreover, capacity and the line prime number of the multifidus muscle can give working hard and the standard when it is possible to use as an index of fat in high this and the correlation and celiac, and it recovers from Metaboriccshindororm based on this index.

Although a volume of the multifidus muscle can be estimated with X rays CT image of hip, methods for extraction of contour of the multifidus muscle from the hip CT images is needed. Although there are the conventional edge detection methods (differentiation methods) which can be used for it, it is not easy to extract the contour because image defects affect to the extractions. For instance, detected edges with differentiation used to be disconnected. There are some isolated pixels due to noise on the CT images. Also it is always true that undesired edges are extracted with differentiations. Therefore sophisticated contour extraction methods are necessary [3]-[9].

Because the high frequency component of the hip CT image shows edge components, wavelet based Multi-Resolution Analysis: MRA is proposed to extract the contour of multifidus muscle [10], [11]. Namely, wavelet transformation based on base function is applied to the hip CT image then the original hip CT image is divided into four frequency components, low frequency component for both horizontal and vertical directions (LL), high frequency in horizontal direction and low frequency in vertical direction of component (HL), low frequency in horizontal direction and high frequency in vertical direction of component (LH), and high frequency component for both horizontal and vertical direction (HH). Because MRA is equivalent to the filter bank, the desired edge with different frequency components can be extracted with the different "level" which corresponds to the center frequency of the band-pass filter which can be realized with MRA. Moreover, not only the contour of the multifidus muscle but also Semmot is extracted from the extracted edge. At this time, it was often divided into parts though Semmot was same originally Semmot, and it tended to overvalue the line prime number. To evade this, this paper proposes the Morphological analysis. That is, this validity has been improved by giving Dilation and Erosion to an appropriate frequency edge image and evaluating the line prime number. Through experiments of which the proposed method is applied to patient's CT image, the proposed method is validated.

The following chapter describes the proposed method followed by some experiments with four patients' hip CT

images. Then, finally, conclusions and some discussions are also followed.

II. Proposed Method

A. Multifidus Muscle

The muscle that runs from "Horizontal projection" to one in on between "Thorn projections" of a high-ranking vertebra is called Tstoge muscle sidewise. When the Tstoge muscle is divided further sidewise which is divided into the musculi rotatores, multifidus muscle and half musculus spinalis. Half musculus spinalis extend to the vertebra from 5 to 6. The multifidus muscle extends to the vertebra from 3 to 4, and is located from half musculus spinalis to deep. Moreover, the musculi rotatores extends between one two vertebras, and is located in deep most. The multifidus muscle operates as musculus transversus abdominis shrink, processus spinosus are pulled, and it is rotated to the other side. There is a rotation movement such as turning the waist to turn around back. The former is steady, and the shrinkage of doing of inclination and musculus transversus abdominis is steady the sacrum inclining forward and the latter and increases the tension outside of the chest line of the backbone film, rises pressure in celiac, and after the sacrum, is steady of the lumbar vertebra the multifidus muscle and the iliococcygeus muscle and the coccygeal muscle. That is the position of the sacrum control.

In addition, musculus transversus abdominis increase the tension of-intestines ligament by synchronizing with the outer unit after the film of the chest line of the backbone, and contribute to the close power strengthening of the pelvic band. They are six keep abreast of movements along the rotation movements of six of the circumference of the XYZ axis of coordinate (moment) and the axis of coordinates. It is Tstoge muscle (musculi rotatores, half musculus spinalis, and multifidus muscle) sidewise erector muscle of spine (musculus spinalis, longest muscle, and iliocostal muscle), interspinales muscles, and the Tsida muscles sidewise that are called that being able to move this movement of 12 pieces freely like the mind is a peculiar line of the backbone. It is located in the depths in the muscle with short mileage, and it takes part in thinner movement. Fig. 1 shows the location of multifidus muscle.

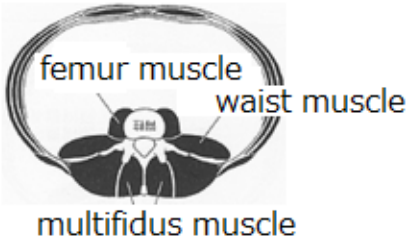
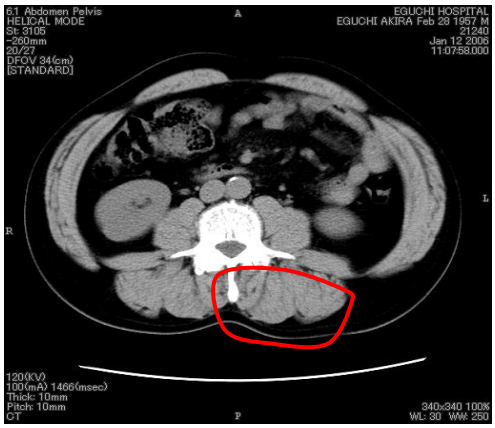


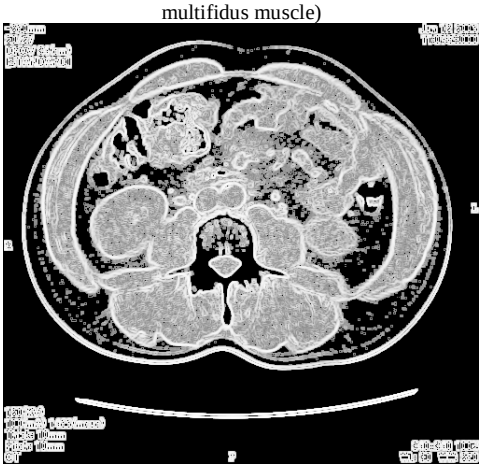
Figure 1 Multifidus muscle

B. Hip CT Image and Edge Extractions

Fig. 2 shows the example of original hip CT image and the extracted edges by hip CT image with Laplasian operator. The edge extracted by the conventional differentiator or second order differentiators such as Sobel and Laplasian operators tends to emphasize not only edges but also noises other than the desired edges as shown in Fig. 2.



(a) Original CT image (Closed area portion with red colored line shows multifidus muscle)

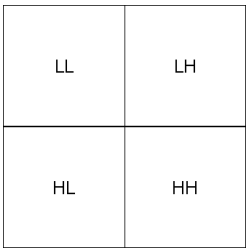


(b) Edge detected image with Laplasian operator

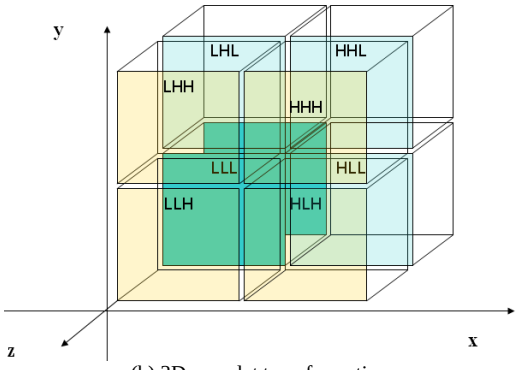
Figure 2 Original image and edge detected image with Laplasian operator.

C. Contour Extraction by WaveletMmulti Resolution Analysis (MRA)

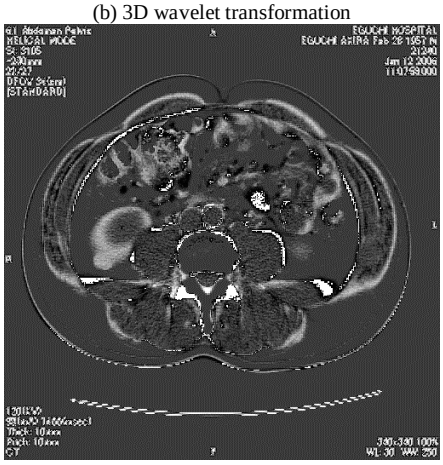
Methods for extraction of contour line of multifidus muscle from the hip CT images by using wavelet based 2D and 3D MRA which are illustrated in Fig. 3 (a) and (b) are proposed. It is essential that the multifidus muscle is three dimensional objects so that the 3D MRA is much appropriate rather than the 2D MRA based method. Using the 2D or the 3D MRA, high frequency components which corresponds to the edges can be extracted. Furthermore, MRA allows reconstruction of original image without any loss.



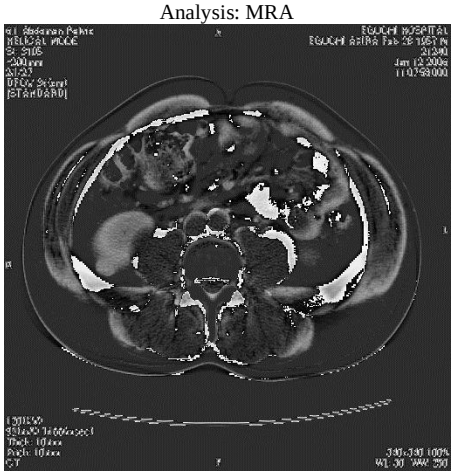
(a) 2D wavelet transformation



(b) 3D wavelet transformation



(c) An example of edge detected image based on 2D Multi-Resolution Analysis: MRA



(d) An example of edge detected image based on 3D Multi-Resolution Analysis: MRA

Figure 3 2D and 3D wavelet transformation and an example of edge detected image based on 2D MRA

If the image is reconstructed with LH, HL, and HH components then edges can be extracted from the 2D of hip CT images. That is the same thing for the 3D MRA based edge detection. When the image is reconstructed without LLL components, then edges are extracted from the 3D of hip CT images. Fig. 3 (c) and (d) shows examples of the edges extracted with the 2D and the 3D MRA based edge detection methods. As the results, the 3D MRA based edge detection method allows many clear edges in comparison to that of the 2D MRA based method. In this case, the level (how many times the 2D or the 3D MRA is applied. This corresponds to which frequency components of edges is desired to extract), and base function of MRA is in concern.

D. Extraction of clear contour with binarization and morphological filtereing

After the reconstruction based on the 2D MRA based and the 3D MRA based methods, binarization with an appropriate threshold is applied to the reconstructed images. There are so many isolated pixels and disconnected contour lines in the binarized images. In order to remove the isolated pixels and connect the disconnected contour lines, morphological filter is applied to the binalized images. Fig. 4 shows the well-known morphological filter components.

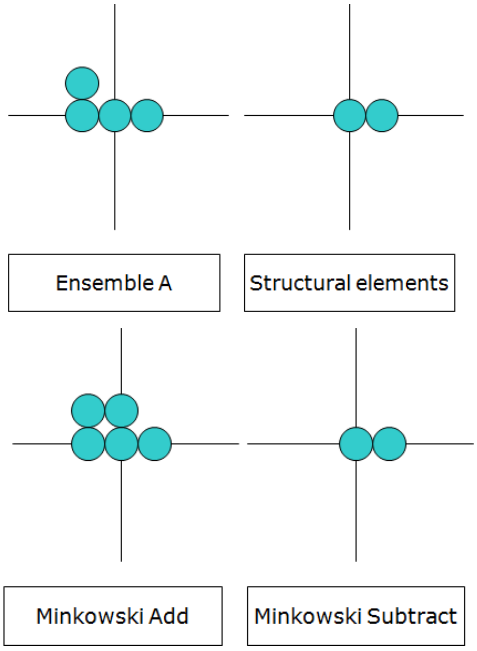
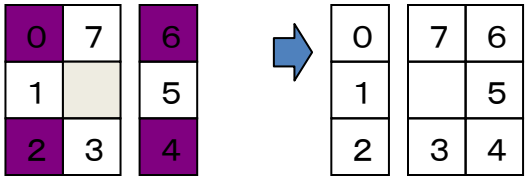
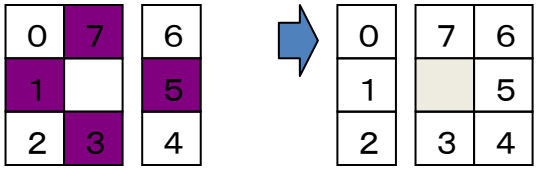


Figure 4 Morphological filter

Fig. 5 also shows typical erosion and dilation of morphological filter. The dilation processing is applied first to the binarized images followed by the erosion processing. Then the disconnected contour lines in the binarized image are connected. Also the erosion processing is applied to the binarized image then isolated pixels are removed.



(a)Dilation



(b) Erosion

Figure 5 Typical erosion and dilation of morphological filter

III. EXPERIEMNTS

A. Experimental conditions

The following four persons are tested,

- Age: 72, 63, 59, 41
- Height(cm)&Weight(kg): 164&71, 170&75, 160&76, 156&62
- BMI: 26.21, 25.95, 29.69, 24.22

B. Cost Functions

The proposed index assumes it is possible to become the index of the function of the multifidus muscle, proposes the size of the multifidus muscle, and has aimed to give the risk such as Metaboriccshindororm and the standard of the recovery degree in the future. Therefore, the correlation with fat in celiac is index necessary and height is necessary.

It is Body Mass Index: BMI, an area of whole fat as a past index, and the area of the hypodermic fat, the area of the waist muscle, the area of the multifid muscle, and the CT worth ratio (ratio of the CT value of the multifidus muscle and the internal organs fat), etc. are proposed as an index. R square value with fat in celiac is high, and the one that it is possible to extract it from hip CT image comparatively easily is selected from among these indices. BMI is division of the weight of the body in the square of the height, and it is possible to examine it extremely easily. If the ratio can worth identify the pixel, CT is comparatively easily computable it though requests from hip CT image besides. Because the area of the waist muscle and the area of the multifid muscle are only coefficients of the pixel, as long as the contour can be extracted, it is comparatively easily computable. Specific of the adipose cell is attended to paniculus adiposus and whole fat with the difficulty as well as fat in celiac. Therefore, the proposal index was evaluated by comparing the result of guessing these past indices by the specialist's judgment with the calculation result of the area of the multifid muscle based on MRA that was the proposal index as the correct answer in this paper. The relation to the index is shown as the judgment result of specialist who requested it from the CT image to 362 patients from 25 to 81 years old from Fig.6, that is, correct answer fat in celiac and so far.

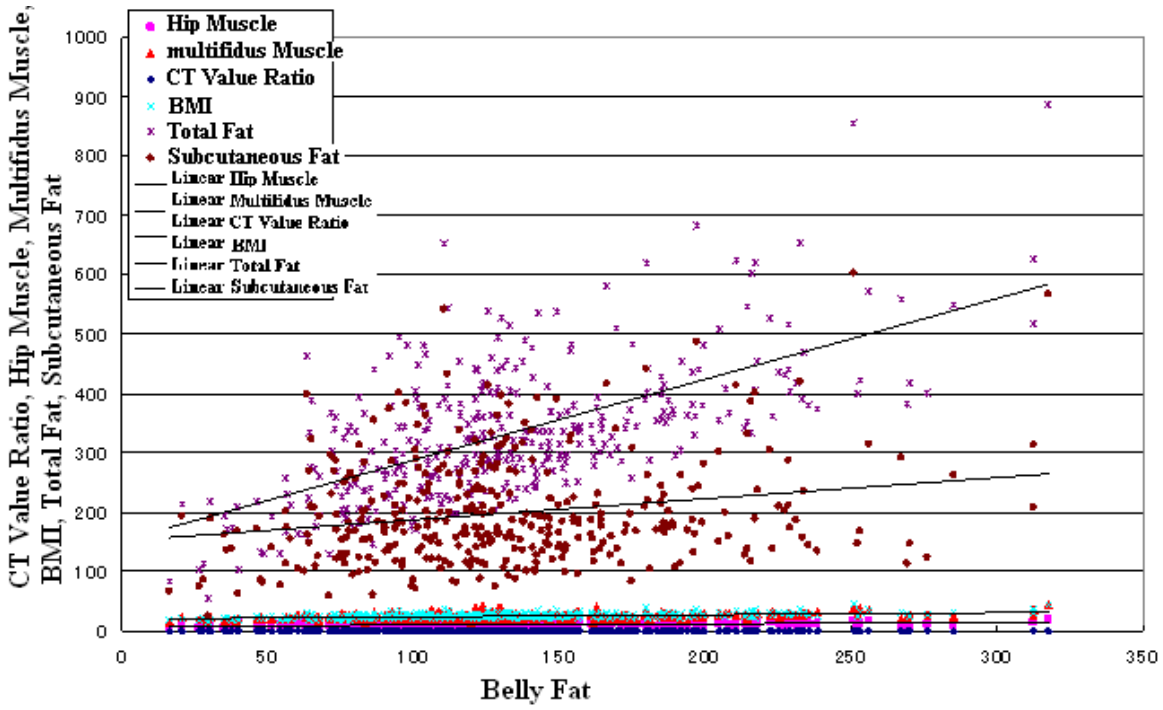


Figure 6 Relations between belly fat and the other measured factors.

R square value is an area of whole fat, an area of the hypodermic fat, and it is 0.408, 0.045, 0.001, 0.178, 0.111, and the correlation with the area of whole fat understands and the highest thing is understood respectively of the CT worth area of the muscle (the ratio and the irresolute attitude of the area of the multifid muscle). However, it was judged that the area of the multifid muscle from which R square value was requested after this without the processing time comparatively lying high was an index of second best because a lot of time and time lay to request this area as shown in the above-mentioned.

C. Base Function and the Level of MRA

The levels were changed up to 1, 2, 3, and 4 by using Haar and Daubechies as a base function, MRA was given to hip CT image, and a contour extraction of the multifidus muscle was tried. Fig. 7 shows one example of the result.

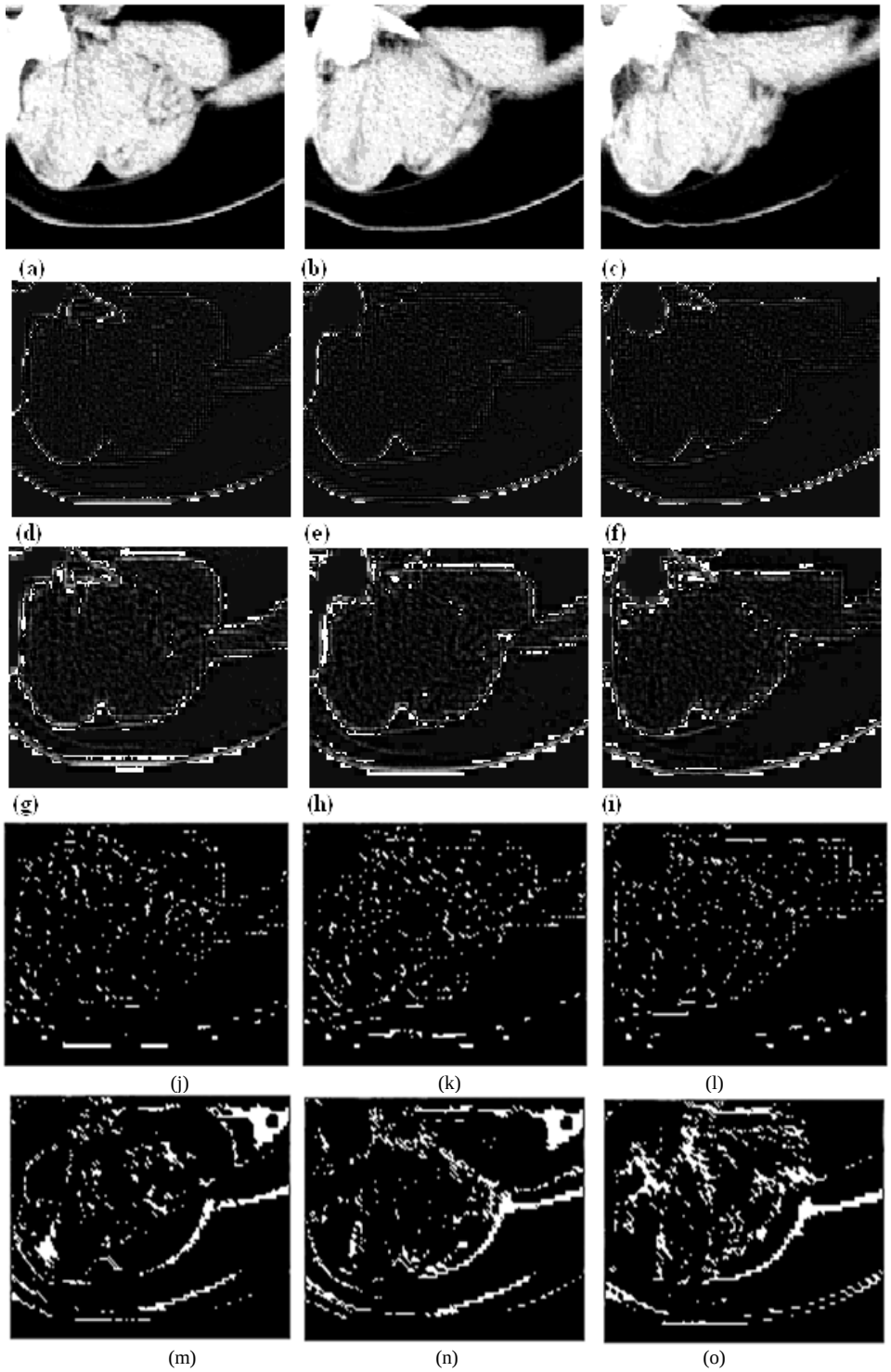


Figure 7 Examples of edge detected images through reconstruction without LL component (2D MRA) and LLL (3D MRA) from 1 to 4 levels of the wavelet transformed images. (a), (b) and (c) are original CT images at the different height (portion). (d), (e) and (f) are edge detected images reconstructed without LL component. (g), (h) and (i) are edge detected images reconstructed without LLL component. Meanwhile, (j), (k), (l) shows the binarized images. On the other hand, (m), (n), (o) shows the resultant images after the morphological filter.

Fig. 7 (a), (b), (c) shows hip CT original picture image in a different part. (d), (e) and (f) are edge detected images reconstructed without LL component. (g), (h) and (i) are edge detected images reconstructed without LLL component.

It is found that the level 2 of MRA parameter is the best in terms of detecting capability of edges from the hip CT images. Also it is found that the Daubechies base function is superior to the Haar base function in terms of edge detecting capability from the hip CT images.

D. Effect of the Morphological Filter

Fig. 7 (j), (k), (l) shows the binarized images while Fig. 7 (m), (n), and (o) shows the resultant images after the morphological filtering is applied to the binarized and reconstructed images. It is easily seen that the contour of the multifidus muscle is extracted almost perfectly.

E. Relations Between the Estimated Multifidus Muscle Areas and the Belly Fat as well as CT Value Ratio and the Belly Fat

Multifidus muscle area can be estimated with the resultant images after the morphological filter. Although belly fat is used to be estimated with BMI in the simplest method and with the CT value ratio, it is usually said that estimation accuracy is not good enough. Meanwhile, the relation between multifidus muscle area which can be estimated with the proposed method is relatively high as shown in Fig. 8 (a) in comparison to the relation between visceral fat and CT value ratio as shown in Fig. 8 (b).

R square values for the relation between visceral fat and multifidus muscle as well as CT value ratio are 0.897, and 0.384, respectively. On the other hand, 3D MRA based edge extraction is much effective (R square value is 0.897) than that of 2D MRA based edge extraction (R square value is around 0.7). Due to the fact that isolated pixels are situated with a couple of pixels and the contour lines are disconnected with a couple of pixel distance so that twice erosions followed by twice dilation are the best conditions for morphological filtering. It is quite obvious that visceral fat is absolutely linked to the well known Metaboriccshindororm which is our major concern. The visceral fat, however, is not easy to estimate. Time consumable image processing is required in general. Turns out, the mutifidus muscle is not so difficult to estimate and is closely related to the Metaboriccshindororm so that we focused the multifidus muscles as an index of the Metaboriccshindororm.

IV. Conclusions

The multifidus muscle area estimation method based on MRA together with morphological filter is proposed. Through the experiments with four patients, it may conclude the followings,

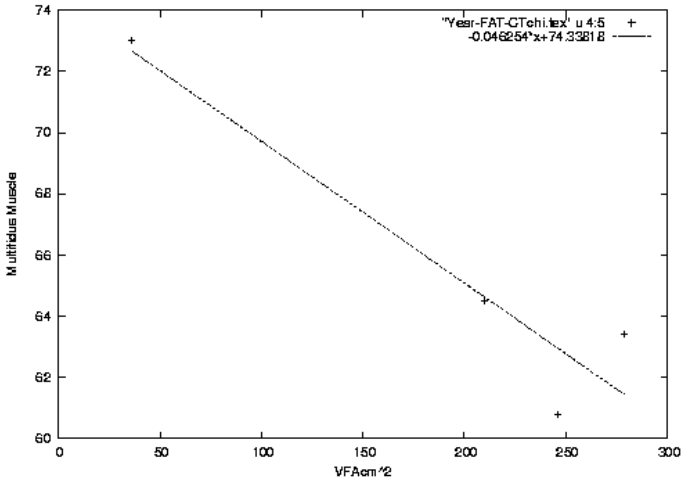
(1) The 3D MRA based edge detection is superior to the 2D MRA based method. Both are superior to the conventional differentiation based edge detection methods.

(2) Daubechies base function is superior to the Haar base function in terms of edge detecting performance. In the case of edge detection from the hip CT image, level 2 of MRA parameter would be the best. This depends on the frequency component of desired edges.

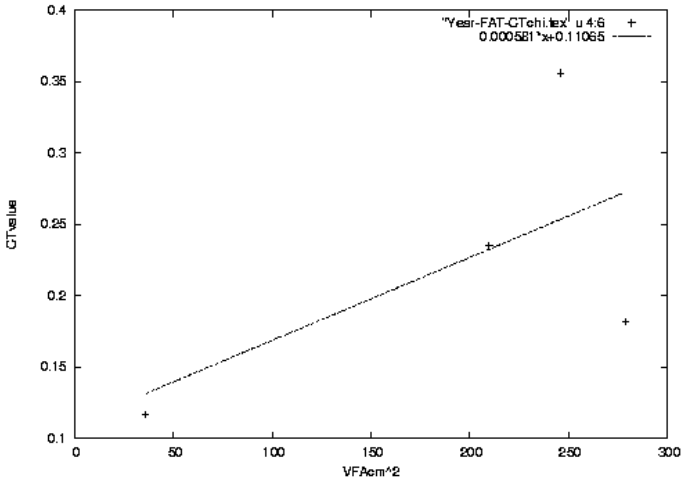
(3) Morphological filter is effective to remove isolated pixels and to connect the detected disconnected contour line of multifidus muscle. In this case twice application of erosion followed by twice application of dilation seems to be the best in terms of connection of disconnected contour line of multifidus muscles.

(4) R square value of the relation between visceral fat and the multifidus muscles area is over 0.8 (significant) so that the multifidus muscles are used for an index of Metaboriccshindororm.

Thus the patients of Metaboriccshindororm are encouraged using the proposed index of multifidus muscles which are relatively easy to estimate from the hip CT images.



(a)The estimated multifidus muscle



(b) CT Value Ratio

Figure 8 Relation among the belly fat and the estimated multifidus muscle as well as CT value ratio.

Acknowledgment

Authors express our gratitude for conducting the experiments by graduate school student Mr. Mitsuo Mochida that cooperates in the experiment.

References

[1] Y. Eguchi, T. Eguchi, T.mizuta, Y.Ide, T. Yasutake, R.Iwakiri, A. Hisatomi, I.Ozaki, K. Yamamoto, Y. Kitajima, Y. Kawaguchi, S. Kuroki, N. Ono, Visceral fat accumulation and insulin resistance are important factorsin nonalcoholic fatty liver disease, J Gastroenterol 2006; 41:462–469, 2006.

[2] Y. Kitajima, Y. Eguchi, E. Ishibashi, S. Nakashito, S. Aoki, S. Toda, T. Mizuta, I. Ozaki, N. Ono, T. Eguchi, K. Arai, R. Iwajiri and K. Fujimoto, Aged-related fat deposition in multifidus muscle could be a marker for nonalcoholic fatty liver disease, Journal of Gastroenterology, Springer 218-224, 2009.

[3] CT performance evaluation committee, Standard concerning performance evaluation of X-ray computer tomography device (the second recommendation) and Japan Association of Medical Practitioners magazine and 88(8)759 - 771, 1989.

[4] Technical committee X rays CT device performance evaluation and examination group, Standard concerning X rays CT device performance evaluation, Journal of Japanese Society of Radiological Technology 47(1)56 - 63, 1991.

[5] K. Suzuki et al., technical problem and day of helical CT system, Journal of Japanese Society of Radiological Technology 52(3)384 - 388, 1996.

[6] K. Tsujioka, Technical problem-performance evaluation of helical CT system, Journal of Japanese Society of Radiological Technology 52(3)389 - 396, 1996.

[7] S. Muramatsu et al., basic performance and day of scanning CT of spiral, Journal of Japanese Society of Radiological Technology 52(1)81 - 85, 1996.

[8] K. Kubota et al., application and day of helical scanning in brain bottom, Journal of Japanese Society of Radiological Technology 52(9)1066, 1996.

[9] S. Kuwahara, Performance evaluation for Hispeed AdvantageRp, Magazine for Radiological Technology of Hokkaido, 56, 51 - 58, 1996.

[10] K. Arai, Basic theory of wave let analysis, Morikitashuppan Publishing Co. Ltd., 2000.

[11] K. Arai, Self-study on wavelet analysis, KindaiKagaku Publishing Co. Ltd., 2006.

AUTHORS PROFILE

Kohei ARAI

Saga University

Saga, Japan

Kohei Arai received a PhD from Nihon University in 1982. He was subsequently appointed to the University of Tokyo, CCRS, and the Japan Aerospace Exploration Agency. He was appointed professor at Saga University in 1990. He is also an adjunct professor at the University of Arizona and is Vice Chairman of ICSU/COSPAR Commission A.

Robust Face Detection Using Circular Multi Block Local Binary Pattern and Integral Haar Features

Dr.P.K.Suri
Dean, Chairman, Professor, CS&A
Kurukshetra University
Kurukshetra
India

Er.Amit Verma
A.P, ECE Department
REBIEBT, Sahauran,
India

***Abstract*— In real world applications, it is very challenging to implement a good detector which gives best performance with great speed and accuracy. There is always a trade-off in terms of speed and accuracy, when we consider performance of a face detector. In the current work we have implemented a robust face detector which uses the new concept called integral Haar histograms with CMBLBP or CSMBLBP (circular multi block local binary operator).Our detector runs for real world applications and its performance is far better than any of the present detector. It works with good speed and enough accuracy with varying face sizes, varying illumination, varying angle, different face expressions, rotation, scaling like challenges which are mostly issues of concern in the domain of face detection. We use Matlab and Image processing tool box for the implementation of the above mentioned technique.**

Keywords- CMBLBP; MBLBP; LBP; Gentle Boosting; Face Detection.

I. SUMMARY OF THE PAPER

This paper presents the novel face detection system using a new technique called circular multi block local binary operator .In the second section, we present basic introduction. Third and fourth sections include LBP and MLBP (Multi block local binary operator). Fifth section gives detail of Boosting. Section sixth describes CMBLBP (circular multi block local binary operator). Section seventh gives Data base, Experiments and Result. Section Eight and nine include conclusion and references

II. INTRODUCTION

Face detection has a wide range of applications in the areas such as automatic face recognition, object detection, face tracking, red-eye removal, face expression recognition, human-machine interaction, surveillance, skin detection[17] etc. In recent years, there has been a vital progress on detection schemes [10-11] based on appearance of faces. To build automatic and robust systems in term of speed and accuracy that can be executed on mobile products or in cameras, very efficient and robust face detection algorithms[1] are required. Most of the systems consider face detection as a class problem with the variable having two dimensions that may be either face or non-face. Some parameters like facial appearance, lighting, expressions, and other factors make this two class problem very complex for differentiating face and non-face. Therefore, there is always a need of such classifiers having good performance characteristics. The most effective method for constructing face and non-face classifiers is learning based approach. For example, neural network-based methods [2], support vector machines [18], etc. In our previous work[20], there has been proposed a very good method using Gabor filter in which 40 Gabor filters are used for a face to be detected and correspondingly feature vector is constructed. The performance is very good in that case but speed is major factor under concern. Although it gave good result in variable illumination, different facial expressions, different skin tone and with different human features, the detector didn't give good result for variable size and rotated face (angle more than 40^o).In addition to the good performance of detector, there always remains a need of fast detector which can be used for live image clips in the real world. The Haar-like features[6][22] (as from figure 1),

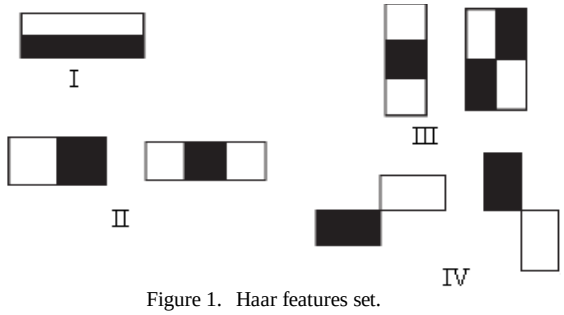


Figure 1. Haar features set.

Encode differences in average intensities between two rectangular regions and these can be calculated rapidly through integral image. Most of the proposed algorithms use pixel values as features. However, they are very sensitive to illumination conditions and noises. Papageorgiou et al. [19] also used Haar-like features. These features are able to extract texture without depending on absolute intensities. Many others [9] proposed variant of Haar features. Viola and Jones proposed [12] an efficient system for evaluating these features which is called an integral image [6] and, they also introduced an efficient scheme for constructing a strong classifier by cascading a small number of distinctive features using Adaboost [3] technique. It consequence comes in terms of increased robustness and higher computational efficiency.

Though Haar-like feature [19] and Gabor techniques as from our previous work[20],provides good performance in extracting textures and cascading architecture, it is still not

suitable for live clips or for real time analysis. Moreover, it is quite hard to detect faces of variable sizes with rotation in angle of face under test.

III. LBP

Local Binary Pattern (LBP)[7] features have performed very well in various applications, including texture classification and segmentation, image retrieval and surface inspection. Here we are using this pattern for face detection[25-26]. The original LBP operator labels the pixels of an image by keeping threshold of the 3x3 neighbourhood of each pixel with the centre pixel value and considering the result as shown in figure 2.

40	42	40
56	39	35
20	40	40

1	1	1
1	c	0
0	1	1

Input(value) Threshold Output (value)

Pattern=$(10110111)_2$ Pattern Value=183

Figure 2. LBP Pattern

The 256-bin histogram of the labels computed over an image can be used as a texture descriptor. Each bin of histogram (LBP code) can be regarded as a micro-texton. Local primitives which are codified by these bins include different types of curved edges, spots, flat areas, etc. The LBP operator has been extended to consider different neighbor sizes. For example, the operator $LBP_{4,1}$ uses 4 neighbors while $LBP_{16,2}$ considers the 16 neighbors on a circle of radius 2 in case of circular local binary operator.

In general, the operator $LBP_{P,R}$ refers to a neighborhood size of P equally spaced pixels on a circle of radius R that form a circularly symmetric neighbor set(as from figure 3). $LBP_{P,R}$ produces 2^P different output values, corresponding to the 2^P different binary patterns that can be formed by the P pixels in the neighbor set. It has been shown that certain bins contain more information than those of others. Therefore, it is possible to use only a subset of the 2^P LBPs to describe the textured images. Ojala et al.[15] defined these fundamental patterns as those with a small number of bitwise transitions from 0 to 1 and vice versa. For example, 00000000 and 11111111 contain 0 transition while 00000110 and 01111110 contain 2 transitions and so on. Accumulating the patterns which have more than 2 transitions into a single bin yields an LBP descriptor.

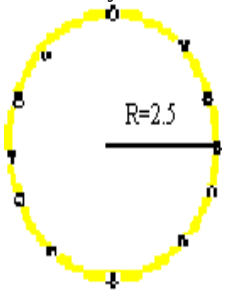
It is observed that uniform patterns account for nearly 90% of all patterns in the (8,1) neighbourhood (as from figure 3) and for about 70% in the (16,2) neighbourhood in texture images. Hence, accumulating the patterns which have more than 2 transitions into a single bin yields an LBP operator, denoted as $LBP_{P,R}^{U2}$ with less than 2^P bins. For example, the number of labels for a neighbourhood of 8 pixels is 256 for the standard LBP but 59 for $LBP^{u2.}$ After labelling an image with the LBP operator, a histogram of the labelled image f_1 (x,y) can be defined as follows:

$$H_i = \sum_{X,Y} I(f_i(x,y) = i), i = 0 \dots \dots n-1 \quad (1)$$

where n is the number of different labels produced by the LBP operator and

$$I(X) = \begin{cases} 1 & A\ is\ true \\ 0 & A\ is\ false \end{cases} \quad (2)$$

This LBP histogram contains information about the distribution of the local micro-patterns, such as edges, spots and flat areas, over the whole image, so can be used to statistically describe image characteristics.



*	*	*
*	*	*
*	*	*

Figure. 3 local binary operator with P,R respectively 12,2.5 and 8,1

The most important out of all considerable properties of LBP features [23-24]is their tolerance against monotonic illumination changes and their computational simplicity at the same time.

IV. MBLBP

Traditional Haar-like rectangle features measure the difference between the average intensities of rectangular regions[8] (Figure 1).

For example, the value of a two-rectangle filter is the difference between the sums of the pixels within two rectangular regions. If we change the position, size, shape and arrangement of rectangular regions, the Haar-like features can capture the intensity gradient at different locations, spatial frequencies and directions. Viola an Jones [6] applied three kinds of such features for detecting frontal faces. By using the integral image, any rectangle filter types, at any scale or location, can be evaluated in constant time period. However, the Haar-like features seem too simple and show some limits [19].

4	4	1		
4	4	1	5	6
6			4	3
3			5	6

4	4	1
4	4	1

Average value 18/6=3

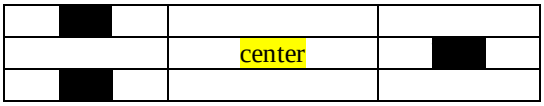


Figure 4. Multi-block LBP feature for image representation.

As shown in the figure 4, the MBLBP features encode rectangular region's intensities by local binary pattern. The resulting binary pattern describe diverse image structures. When compared with the original Local Binary Pattern calculated in a local 3x3 neighbourhood between pixels, we see that MBLBP can capture large scale structure.

V. BOOSTING

Boosting algorithm are used to reduce the number of distinctive rectangle features (to make the size of feature vector smaller) and construct a powerful classifier. It speeds up the computations thereby increases the speed of the detector to find the faces for the given image. But Haar-like rectangle features seem too simple, and the detector often contains thousands of rectangle features for considerable performance. The size of feature vector decides the computation costs both in training as well as test phase. More is the size of feature vector lesser is the speed of the detector. At the resulting stages, weak classifiers based on these features become too weak to improve the classifier's performance. For that, many other methods have been proposed like rotated Haar-like features [19], census transform [5][16], sparse features [4-5], etc. In this paper, we present a new distinctive feature, called Circular-Multi-block Local Binary Pattern (CMBLBP) feature, to represent facial image. The basic idea of CMB-LBP is to encode rectangular regions in the circle by local binary pattern operator [7]. Since it is circular and multi block, our method captures much more area than that by other methods proposed in [12][20]. Original Local Binary Pattern is calculated for 3x3 neighbourhoods, so the area captured by LBP in that case is small. On the other hand, the MB-LBP features can capture large scale structure that may include the dominant features of image structures. The problem with the value of LBP is that its output is just a symbol for representing the binary string. For this non-metric feature value, multi-branch operations are done to calculate weak classifiers. We implement Gentle Adaboost for feature selection and classifier construction. Then a cascade detector is built. Another advantage of CMBLBP is that the number of exhaustive set of CMBLBP features is much smaller. Boosting-based method use Adaboost algorithm to select a significant feature set from the large complete feature set and to enhance its speed, fast Adaboost is further used[13]. We here use gentle boost as a variant of Adaboost. The small feature set of CMBLBP can make this procedure even more simple.

VI. CMBLP

In this paper, we propose a new distinctive rectangle feature from circular region, called CMBLBP. The basic idea behind that is simple difference rule as in Haar[14]. It is changed into encoding circular region from blocks of a rectangle by binary operator. In this method the value of central pixel is subtracted from the value of neighbour pixels. Then, the information is presented without any loss as a joint distribution(as from equation 3) of the value of central pixel and the differences.

$$T = t(g_c, g_o - g_c, \dots \dots g_{p-1} - g_c) \quad (3)$$

Assuming the difference independent of central pixel value, the distribution can be factorized as:

$$T = t(g_c)t(g_o - g_c, \dots \dots g_{p-1} - g_c) \quad (4)$$

Since $T = t(g_c)$ described the overall luminance of an image, which is not related to local image.

So, $T = t(g_o - g_c, \dots \dots g_{p-1} - g_c)$ (5)

Although invariant against gray scale shifts, the differences are effected by scaling only. To achieve invariance with respect to any transformation of the gray scale, only the sign of difference is considered that is,

$$T \approx t(s(g_o - g_c, \dots \dots s(g_{p-1} - g_c)) \quad (6)$$

Here

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (7)$$

After binomial weight assignment (2^P) to each sign that is $s(g_p\text{-}g_c)$,transforming the difference in a neighborhood into a unique LBP code for given block of pattern. The code characterization of image I(x,y)

$$\text{LBP (for a given block of pattern)}_{P,R}(x_c,y_c) = \sum_{P=0}^{P-1} s(g_P - g_C)\, 2^P \quad (8)$$

The local gray scale distribution can thus be approximately described with 2^P bin discrete distribution of LBP codes:

$$T \approx t\left(LBP(\text{given block of pattern})_{P,R}\ (x_c,y_c)\right) \quad (9)$$

In the calculation of the feature vector for given image (MxN), only the central part is considered, as large neighborhood cannot be used on the borders. The distribution of the codes is used as feature vector, denoted by s.

$$T = t\left(LBP(\text{given block of pattern})_{P,R}\ (x,y)\right)$$
$$x \in \{[R] \dots \dots N - 1 - [R]\}, Y \in \{[R] \dots \dots M - 1 - [R]\} \quad (10)$$

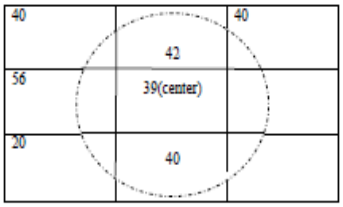
Thus CMBLBP (as from figure 5):

$$\sum_1^n s(T_i - T_c)2^n \quad (11)$$

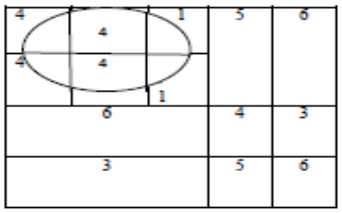
Where T_c is the average intensity of the central pixel, T_i=(1…….n),are those in the neighborhood rectangles, n is number of blocks, and the output is computed as

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases} \quad (12)$$

The detail description of such CMBLBP operator



(a)



(b)

Figure 5. CMBLBP Feature selection

can be found in the above figure 5. In this method, the scaling problem of circular LBP is reduced by the property of multi block, so in our system the detector can detect diverse image structure (as edge, lines, spots, flats, corner areas) also at different scale and location with varying image size (better than [20]) and varying illumination.

Comparing to local binary pattern and MB pattern alone, our system captures larger area with greater accuracy. Collectively we get 256 bin patterns.

Actually in net we are getting LBP at first stage, in second stage we calculate integral histogram $(I_K^H)_1^{n(Labels)}$ of the given test. The labels(n) depends on LBP operator used if n=4,so 2^4=16,integral histograms will be created. In the final stage integral histogram enables us to calculate the feature vector as CHLBP or CHMBLBP. So the CHLBP features are the binary features as normal Haar features. Region separation is defined as R^+ if q pixels of region R^+ have more label compared to b pixels of region R^- .

A. Algorithm

- Take input image as I(x,y)
- For given block b_n
- Compute LBP operator O_P
- Compute LBP(for given block of pattern)$_{P,R}(x_c,y_c)=\sum_{P=0}^{P-1} s(g_P - g_C)\,2^P$
- Calculate $\sum_1^n s(T_i - T_c)2^n$

Where T_i ,T_c are values of i^{th} block and central block ,n is number of blocks.

- Compute $(I_K^H)_1^{n(Labels)}$

For given n(label),by$(I_K^H)_1^{n(Labels)}$,where n depends on $\rightarrow LBP_{P,R}^u$

- Compute IH(the number of integral Haar histograms)
- Compute Feature vector from $s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$

Perform classification

- Start with weight W_i
- Repeat for WL(weak learner)=1……M

(a)Calculate least square error
(b)update additive model $F(x) \leftarrow F(x) + f_m(x)$
Here $f_m(x)$ is weak learner, $F(x)$ is strong learner.
(c)update $W_i \leftarrow W_i e^{-y_i f_m(x)_i}$
(d)Normalization

- Output of the classifier F(x)=Sign$[\sum_{m=1}^{M} f_m(x)]$

In each step the weak classifier $f_m(x)$ are chosen so as to minimize the weighted squared error.

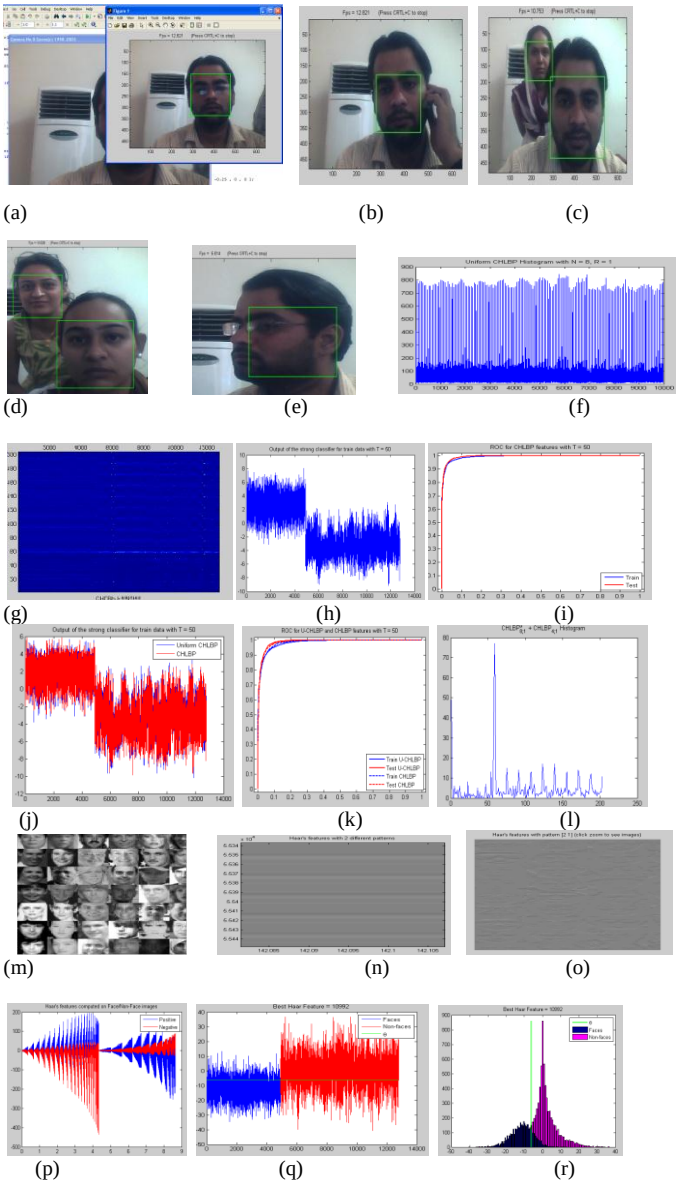
$$J_{wse}=\sum_{i=1}^{N} W_i(Y_i - f_m(x_i))^2$$

Our purpose of using gentle boosting is to minimize error as cost function using adaptive Newton step function as;

$$J=E[e^{-y_i F_x}]$$

VII. Database, Experiments and Results

In this section, we trained our system using two gray scale pixels frontal face database, Viola and Jones and Ole Jensen dataset. It consist of 4,916 images and Ole Jensen consist of 5,000 image. The negative sample was chosen from 10,000 images. LBP histogram features and are used for circular Haar like local binary operator as shown in the figure 6 below for full face description. of size 12x12 pixels and shifting of two pixels is used as scanning window.



(a) (b) (c)

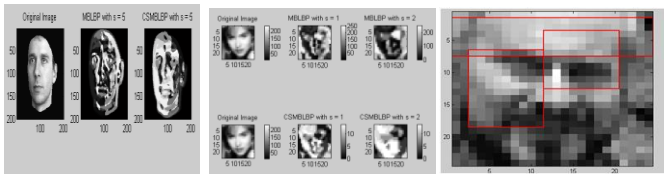
(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

(m) (n) (o)

(p) (q) (r)



(s) (t) (u)

Figure 6. (a-b) test on real world,(c-d) variable face size with different illumination,(e) at angle and with spectacles, (f-g)CHLBP features and histogram (h)output of strong classifier,(i)ROC for uniform CHLBP,(j)strong classifier output(k)Uniform CHLBP and CHLBP (l) $LBP^{u}_{P,R}$ for p=4 and 8 with R=1,(m)database,(n-r)Haar features(s-t)MBLBP & CSMBLBP,(u) Three Haar features

TABLE I. COMPARATIVE RESULTS

False Alarm	6	10	21	400	Net Accuracy
our	90%	92%	93%	98%	98%
Using MB alone[21]	80.1%	-	85.6%	-	92%
Viola and jones[12]	-	78.3%	-	93.7%	88%

Figure 6 above gives output from our detector. Part(a-b) shows performance on real world life clips.

The processing time of our detector for a 320x240 image is far less than 0.1s. The compared results are as shown in tabular form in the table 1 given above.

VIII. CONCLUSION

In this paper we use CSMBLBP, and applied to real world applications. It gives best performance in all respect like for varying face sizes, varying illumination, varying angle, different face expression, rotation, scaling like challenges. We perform these results on Dual core Intel processor with 3GB RAM.

REFERENCES

[1] B.Wu, H.Z. Ai, C. Huang, and S.H. Lao. Fast rotation invariant multi-view face detection based on real Adaboost. In FG, 2004.

[2] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1998.

[3] J. Friedman, T. Hastie, and R. Tibshirani. Additive logistic regression: A statistical view of boosting. Annals of Statistics, 2000

[4] B. Froba and A. Ernst. Face detection with the modified census transform. In AFGR, 2004.

[5] P. Y. Simard, Y. A. L. Cun, J. S. Denker, and B. Victorri. Transformation invariance in pattern recognition - tangent distance and tangent propagation. Neural Networks: Tricks of the Trade, 1998.

[6] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In IEEE Conference on Computer Vision and Pattern Recognition, 2001

[7] S. Zhao, Y. Gao, and B. Zhang, "Sobel-LBP," in Proceedings of IEEE International Conference on Image Processing (ICIP '08), pp. 2144–2147, October 2008.

[8] M. Heikkil¨a, M. Pietik¨ainen, and C. Schmid, "Description of interest regions with local binary patterns," Pattern Recognition, vol. 42, no. 3, pp. 425–436, 2009.

[9] R. Lienhart and J. Maydt. An extended set of Haar-like features for rapid object detection. In ICIP, 2002.

[10] C. Huang, H. Ai, Y. Li, and S. Lao. Learning sparse features in granular space for multi-view face detection. In IEEE International conference on Automatic Face and Gesture Recognition, April 2006.

[11] S. Z. Li, L. Zhu, and Z. Q. Zhang et al. Statistical learning of multi-view face detection. In ECCV, 2002.

[12] Viola, P., Jones, M.: Fast Multi View Face Detection. Technical Report TR2003-96, July 2003, Mitsubishi Electric Research Laboratories.

[13] C.J. Edward, C.J. Taylor, and T.F. Cootes, "Learning to Identify and Track Faces in an Image Sequence," Proc. Int'l Conf. Automatic Face and Gesture Recognition, pp. 260-265, 1998..

[14] T. Mita, T. Kaneko, and O. Hori. Joint Haar-like features for face detection. In ICCV, 2005.

[15] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. Pattern Recognition, January 1996.

[16] K. I. Chang, L. V. Tsap, and M. C. Shin, "Does color space transformation make any difference on skin detection?," IEEE Workshop on Applications of Computer Vision, Orlando, FL, pp. 275–279, December 2002.

[17] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in Proceedings of the International Conference on Computer Graphics and Vision, 2003.

[18] E. Osuna, R. Freund, and F. Girosi. Training support vector machines: an application to face detection. In CVPR, 1997.

[19] C. P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. Proc. of ICCV, pages 555-562, 1998.

[20] P.K Suri,Ekta Walia,Amit Verma, "Noval face detection using Gabor filter bank with varaible threshold" (accepted) in International conference on high performance architecture and grid computing in collaboration with springer germany, HPAGC-2011.

[21] Lun Zhang, Rufeng Chu, Shiming Xiang, Shengcai Liao, Stan Z. Li,"Face Detection Based on Multi-Block LBP Representation",Center for Biometrics and Security Research & National Laboratory of Pattern Recognition Institute of Automation, Chinese Academy of Sciences 95 Zhongguancun Donglu Beijing 100080, China, 2010

[22] K. Zhang, G. Lafruit, R. Lauwereins and L.V. Gool, "Joint Integral Histogram and its Application in Stereo Matching," Proceeding of 2010 IEEE 17th International Conference on Image Processing, Hong Kong, September 26-29, 2010.

[23] J. Trefny and J.Matas, "Extended Set of Local Binary Patterns For Rapid Object Dtection," Computer Vision Winter Workshop 2010, Libor ˜ Spa˜cek, Vojt˜ech Franc (eds.), Czech Republic, February 2010.

[24] S. Paisitkriangkrai, C. Shen and J.Zhang, "Face Detetcion with Effective Feature Extraction," CoRR, 2010.

[25] Jo Chang-yeon, "Face Dtection using LBP features," Final Project Report, December 2008.

[26] C. Zhang and Z. Zhang, "A Survey of Recent Advances in Face Detection," Technical Report, June 2010.

A new vehicle detection method

Zebbara Khalid
LabSIV, Department of Computer Science
Faculty of Science, Ibn Zohr University
Agadir, Morocco

Abdenbi Mazoul
LabSIV, Department of Computer Science
Faculty of Science, University of Ibn Zohr
Agadir, Morocco

Mohamed El Ansari
LabSIV, Department of Computer Science
Faculty of Science, University of Ibn Zohr
Agadir, Morocco

***Abstract*—This paper presents a new vehicle detection method from images acquired by cameras embedded in a moving vehicle. Given the sequence of images, the proposed algorithms should detect out all cars in realtime. Related to the driving direction, the cars can be classified into two types. Cars drive in the same direction as the intelligent vehicle (IV) and cars drive in the opposite direction. Due to the distinct features of these two types, we suggest to achieve this method in two main steps. The first one detects all obstacles from images using the so-called association combined with corner detector. The second step is applied to validate each vehicle using AdaBoost classifier. The new method has been applied to different images data and the experimental results validate the efficacy of our method.**

Keywords-component; intelligent vehicle; vehicle detection; Association; Optical Flow; AdaBoost; Haar filter.

I. INTRODUCTION

Detection of road obstacles [1] [2] [3] [4] [5] [6] is an important task in the intelligent transportation. A number of sensors embedded in IV to perform the vehicle detection task. These sensors can be classified into passive and active sensors. Known that active sensors are expensive and cause pollution to the environment, we propose to use passive sensors in our vehicle detection approach. The data we are going to process to achieve vehicle detection are images taken from a camera embedded in a moving car.

In the field of technical obstacle detected by vision system, two approaches existed: the first approach is unicameral approach that uses a single camera that consists of an image interpretation with former knowledge of information about these obstacles. This information can be texture information [7], color [8], [9]. The second one is the stereo or multi-camera approach which is based on the variation map after matching primitives between different views of the sensor [10], [11] and [12]. Vehicle detection algorithms have two basic step; Hypothesis Generation (HG) and Hypothesis Verification (HV) [13]. In the hypothesis Generation step, the algorithm hypothesizes the locations of vehicles in an image. In the Hypothesis Verification (HV) step, the algorithm verifies the presence of vehicle in an image. The methods in the HG step can be categorized into tree methods; Knowledge-based methods which use symmetry of object, color, corners and edges; Stereo-vision-based methods which use two cameras; Motion-based Methods which track the motion of pixels between the consecutive frames [14]. The methods in the HV step are Template-based methods and Appearance methods. Template-based methods use predefined patterns of the vehicle class. Appearance-based methods include pattern classification system between vehicle and non vehicle. There are a many works [15][16][17] tackling realtime on-road vehicle detection problem. All the papers used monocular cameras and have real-time constraints. [15] used horizontal and vertical edges (Knowledge-based methods) in HG step. The selected regions at HG step are matched with predefined template in HV step. [16] used horizontal and vertical edges in HG step. However, they use Haar Wavelet Transform and SVMs (Appearance-based methods) in HV step. [17] detected long-distance stationary obstacles including vehicles. They used an efficient optical flow algorithm [18] in HG step. They used Sum of squared differences (SSD) with a threshold value to verify their hypothesis.

This paper presents a new approach for vehicle detection. At each time, the decision of the presence of vehicles in the road scene is made based on the current frame and its preceding one. We use the association approach [20], which consists in finding the relationship between consecutive frames. This method exploits the displacement of edges in the frames. At each edge point in one frame we look for its associate one in the preceding frame if any. Obstacles can be detected on the basis of the analysis of association results. Adaboost classifier is used to verify is an obstacle is a vehicle.

II. METHOD VEHICLE DETECTION

This section details de main steps of the proposed method. We extract the edge points and corners of the consecutive images. We keep only the edge points belonging to curves containing corners. The association is performed between consecutive images. We analyze the association results to detect obstacles (objects). Finally, Adaboost is used to decide if a detected object is a vehicle or not.

A. Detecting Corner

We use Shi and Tomasi [19] corner detector that is modified from the Harris corner detector. Shi and Tomasi corner detector is based on the Harris corner detector. Affine transformation is used instead of a simple translation. Given the image patch over the area *(u,v).* Shi and Tomasi corner detector finds corner with applying Affine transformation A and shifting it by *(x,y)* (Eq. 1). .

$$S = \sum_{u}\sum_{v}(I(u,v) - I(A(u,v) - (x,y)))^2 \qquad (1)$$

After calculating the point's corners threshold was performed to remove small close point's corners, points corners

in a vehicle are much more compared to trees or features of the road.

B. Detecting Edge and filtring

Canny operator is used to detect edge points of the consecutive images. The edge curves are formed by grouping edge points using morphological operations. Among the resulting curves, we keep only the ones crossing at least one of the corners calculated in subsection A.

C. Association

The rest of this subsection describes the method we use to find association between edges of successive frames. Let C_{k-1} be a curve in the image $I_{K-1}$and C_k be its corresponding one in the image I_K. Consider two edges P_{k-1} and Q_{k-1} belonging to the curves C_{k-1} and their corresponding ones P_k and Q_k belonging to the curve C_k (see Fig. 1). We define the associate point of the point P_{k-1} as the point belonging to the curve C_k which has the same y coordinate as P_{k-1}. Note that the association is not correspondence neither motion. Two associate points are two points belonging to two corresponding curves of two successive images of the same sequence and having the same y-coordinate. From Fig. 1, we remark that the point Q_k meets these constraints. Consequently, Q_k constitutes the associate point of the point P_{k-1}.

In practice, we assume that the movement of the objects from one frame to the other is small. So, if x_1 and x_2 represent the x-coordinates of P_{k-1} and Q_k, respectively, x_2 should belongs to the interval $[x_1 - Dx, x_1 + Dx]$,where Dx is a threshold to be selected. This constraint allows the reduction of the number of associate candidates. The gradient magnitude is used to choose the best associate one. As a similarity criterion, the absolute difference between the gradient magnitudes of the edges is used. As we see in Fig. 1, the point P_k represents the match of the point P_{k-1}. However, the point Q_k constitutes the associate of the point P_{k-1}. We remark that the points P_k and Q_k are different because of the movement of the point P_k in the image I_k.

We could not find the association for all edges because of the different viewpoints and then objects movement. It is the same as in the matching algorithm, where some parts are visible in one image but occluded in the other one.

Association approach is a technique used to find the relationship between successive frames, this method exploit the displacement of edges in the frames. Let Q_k be an edge

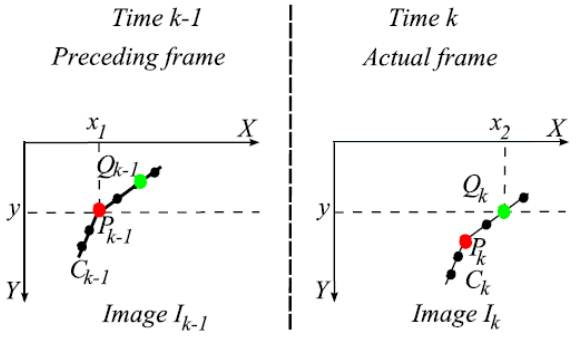


Figure 1. I_{k-1} and I_k represent successive images of the same sequence, e.g. left sequence. The point Q_k in the image I_k constitutes the associate point of the point P_{k-1} in the image I_{k-1}. The points P_k and P_{k-1} are in red color. The points Q_k and Q_{k-1} are in green color.

point belonging to the curves C_k in the image I_k. The associate point of Q_k can be found as a correspondent point P_{k-1} belonging to the curves C_{k-1} in the horizontal neighborhood of Q_k in previous image I_{k-1}.(More details about association method are described in [20]).

The associated points should belong to the same object contour and they should have similar or closer gradient magnitudes and orientation. In this work, we use an important cost function (Eq. 2) described below in this paper. This function computes the distance between two candidate associate points using gradient magnitudes. The edge with smaller cost will be considered as associated pairs of features. Because of vertical movement of scene, the association approach does not guarantee that each feature in the image have its associated point. But some good associates' points are enough to construct the vehicle objects.

$$F(d_x) = \min \sum_{x=u-w}^{u+w} \left(I(x,y) - I(x+d_x, y)\right) \quad (2)$$

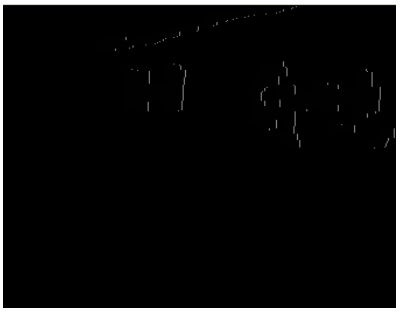
Where d_x is the distance that a contour moves between instant t0 and t1. Given point (u, v) in image I_t, the algorithm should find the point (if exist) $(u + dx, v)$ in image I_{t+1} that minimizes function of cost F (Fig. 2). And w is the neighbourhood window around (x,y).

D. Detection of Objects

Let us consider *Ass* the image association and M and N be the image width and height, respectively. At each pixel (x,y) in the current image, Ass(x,y) is the distance between the pixel (x,y) and its associate one in the preceding image (frame). The obstacles can be detected by using the following functions.

$$F_1(i) = \sum_{j=1}^{N} Ass(i,j) \quad (3)$$

$$F_2(j) = \sum_{i=1}^{M} Ass(i,j) \quad (4)$$



(a)

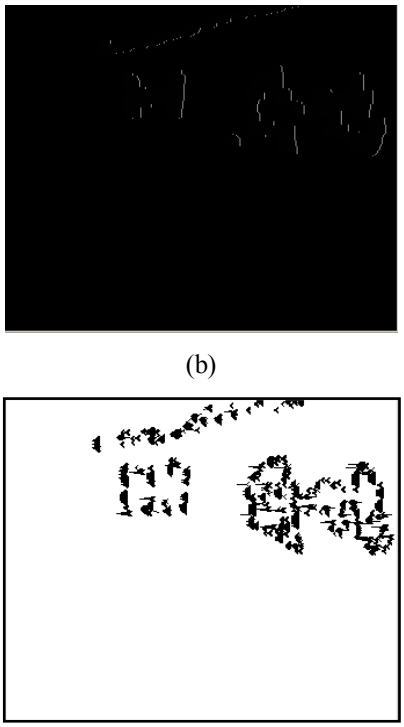


Figure 2. (a) Edge detection at instant t_0. (b) Edge detection at instant t_1 (c) Vector Association.

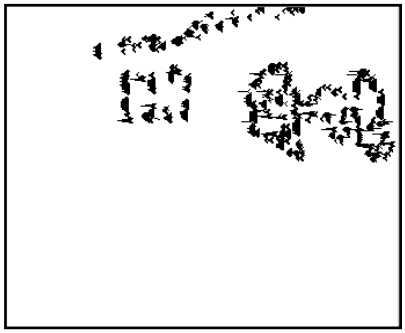
Where i = 1,..,M and j = 1,..,N.

The values of the function F_1 and F_2 should be maximum at the areas where there are obstacles. The function F_1 allows to determine the horizontal bounds of obstacles. The function F_2 allows to determine the vertical bounds of obstacles. The segmentation of the two functions helps to determine the horizontal and vertical bounds of obstacles. Fig. 3 illustrates an example of the computation by equations F_1 and F_2. Fig. 3 (a) depicts the image association, Fig. 3(b) the computed function F_1, and Fig. 3(c) the computed function F_2.

E. Validation using Adaboost

In the step of detecting and locating faces, we propose an approach for robust and fast algorithm based on the density of images, AdaBoost, which combines simple descriptors (Haar feature) for a strong classifier.

The concept of Boosting was proposed in 1995 by Freund [21]. The Boosting algorithm uses the weak hypothesis (error



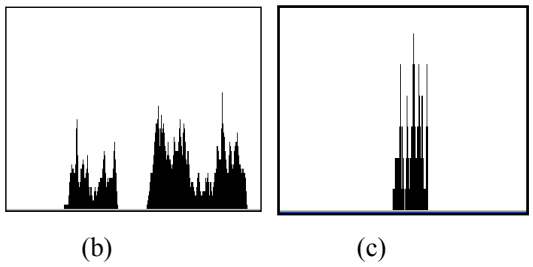


Figure 3. (a) Image Association (b) the computed function by equation F_1 (3). (c) the computed function by equation F_2 (4).

rate $\varepsilon < 0.5$) a priori knowledge to build a strong assumption. In 1996 Freund and Schapire proposed the AdaBoost algorithm which allowed automatic choosing weak hypothesis with adjusted weight. AdaBoost does not depend on a priori knowledge [22].

In 2001, Viola and Jones applied the AdaBoost algorithm in the detection of faces for the first time. With simple descriptors (Haar feature), the method of calculating value descriptors (full image), the cascade of classifiers, this method has become reference face detection for its qualities of speed and robustness.

In 2002, Lienhart et al. extended Haar descriptors, experienced in several of AdaBoost algorithms: Discrete Adaboost, Real Adaboost, Gentle Adaboost and Logitboost.These codes learning and detection algorithm AdaBoost are published in the function library OpenCV (Open Source Computer Vision) [23] [24]. [25] Using descriptors histograms of oriented gradients for human detection and bicycles.

In our work we applied the algorithm "Gentle AdaBoost" using the OpenCV library function using two waterfalls - "haarcascade_car_1" and "haarcascade_car_2" - to detect and locate most vehicles in the sequences images.

III. RESULTS

We have performed a number of experiments and comparisons to demonstrate the proposed Association approach in the context of vehicle detection.

The system was implemented on a Intel® Core ™ CPU 2.99 Ghz. We tested the system on different frames of images. The system is able to detect most vehicles in different images in 20 milliseconds it's fast more than algorithm of optical flow [26].

(a)

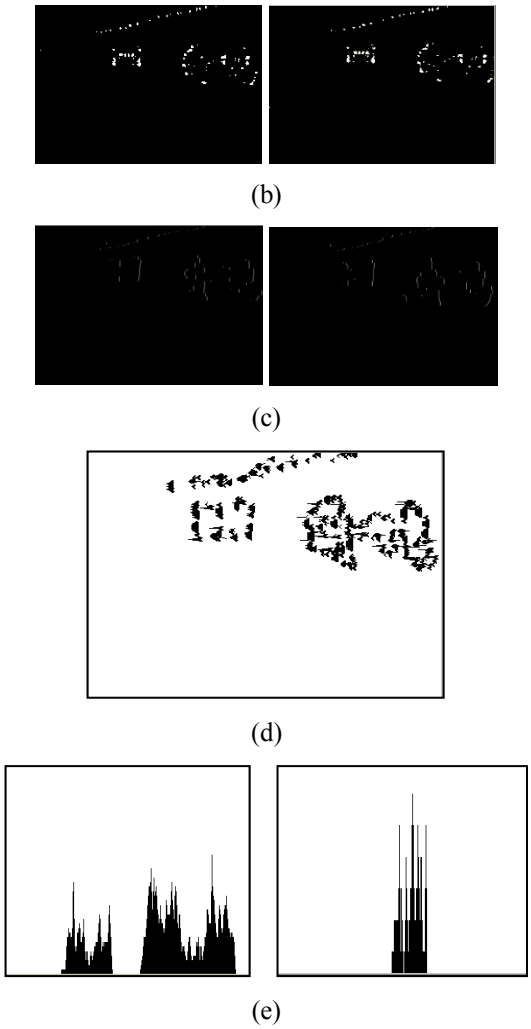


Figure 4. (a) at instant t0 and t1. (b) point's corners of images (a). (c) edges that cross points corners detected at (b). (d) Association vectors of edges from (c) calculated between instants t_0 and t_1. (e right) vertical and (e left) horizontal projection of the associated edges points.

Fig. 4 shows each step of our approach for detection of vehicles using Association and Adaboost. The results illustrate several Strong points of the proposed method. Fig. 4.a shows an image at instant t0 ant t1. In Fig. 4.b, the point's corners are calculated successfully after threshold to eliminate other obstacles (tree,), although we have only point's corners of vehicles. In Fig. 4.c, shows edges that cross points corners and we keep only edges of vehicles. In Fig. 4.d, shows associations vectors for each edge in the frame t0. In Fig. 4.e shows results calculate in section detection of objects by formulas (3) and (4) to determine abscises and ordinates of obstacles.

The proposed method has been tested on other real road images depicted in Fig. 6. The HG and HV results are shown in Fig. 6.b and Fig. 6.c respectively. It's clear that the results computed by our approach are very satisfactory.

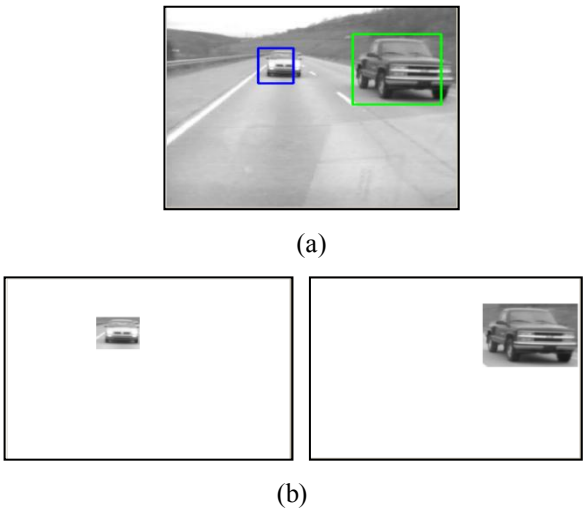


Figure 5. (a) Bounding box. (b) Validation of objects using AdaBoost.

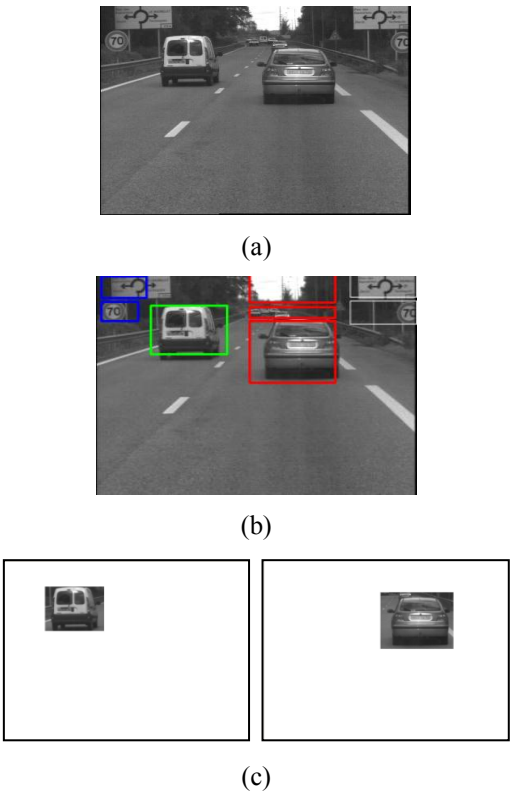


Figure 6. (a) Original image.(b) Hypothesis Generation (HG) and (c) Hypothesis Verification (HV).

IV. CONCLUSION

This paper presents a new vehicle detection method based on association notion descrebed above. In order to select more reliable features, the corner detector is used. Based on horizontal and vertical projection of the associated edge points, the focused sub-region is selected as region of interest.

The experiment results have validated the efficacy of our method, and they show that this method is capable to work in real time. In the future, we plan to improve our vehicle detection method, which will be tested to detect much more

complex obstacles (pedestrian, traffic light...) under different weather conditions.

REFERENCES

[1] R Manduchi, A Castano, A Talukder, L Matthies 2005. Obstacle Detection and Terrain Classification for Autonomous Off-Road Navigation.

[2] R. Labayrade, D. Aubert, J. P. Tarel, "Real Time Obstacle Detection on Non Flat Road Geometry through V-Disparity Representation", IEEE Intel- ligent Vehicules Symposium, Versailles, June 2002.

[3] M. Bertozzi, A. Broggi - "GOLD: A parallel real- time stereo vision system for generic obstacle and lane detection", IEEE Transaction on image pro- cessing, Vol. 7, N1, January 1998.

[4] T.A. Williamson - "A high-performance stereo vi-sion system for obstacle detection ", Phd, Carnegie Mellon University, September 1998.

[5] G. Toulminet, A. Bensrhair, S. Mousset, A. Broggi, P. Mich, "Systeme de stereovision pour la detec- tion d'obstacles et de vehicule temps reel". In Procs. 18th Symposium GRETSIIO1 on Signal and Image Processing, Toulouse, France, September 2001

[6] Tuo-Zhong Yao, Zhi-Yu Xiang, Ji-Lin Liu 2009 .Robust water hazard detection for autonomous off-road navigation in Journal of Zhejiang UniversityScience.

[7] T. Kalinke, C.Tzomakas, and W.Seelen (1998). A Texture-based object detection and an adaptive model-based classification.

[8] R.Aufrere, F.Marmoiton, R.Chapuis, F.Collange, and J.Derutin (2000).Road detection and vehicles tracking by vision for acc.

[9] R. Chapuis, F.Marmoiton and R.Aufrere (2000). Road detection and vehicles tracking by vision for acc system in the velac vehicle.

[10] U.Franke and A.Joos (2000). Real-time stereo vision for urban traffic scene understanding.

[11] D.Koller, T.Luong, and J.Malik (1994). Binocular stereopsis and lane marker flow for vehicle navigation: lateral and longitudinal control.

[12] R. Labayade, D.Aubert, and J. Tarel (2002). Real time obstacle detection in stereo vision on non flatroad geometry through v-disparity representation.

[13] R. Miller. On-road vehicle detection: A review. IEEE Trans. Pattern Anal. Mach. Intell., 28(5):694–711, 2006. Member- Zehang Sun and Member-George Bebis. 2

[14] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In Proc. of the 7th IJCAI, pages 674–679, Vancouver, Canada, 1981. 1, 2, 3

[15] M. Betke, E. Haritaoglu, and L. Davis. Real-time multiple vehicle detection and tracking from a moving vehicle. 12(2):69–83, 2000. 2

[16] G. B. Z. Sun, R. Miller and D. DiMeo. A real-time precrash vehicle detection system. In Proceedings of the 2002 IEEE Workshop on Applications of Computer Vision, Orlando, FL, Dec. 2002. 2.

[17] A. Wedel, U. Franke, J. Klappstein, T. Brox, and D. Cremers.Realtime depth estimation and obstacle detection from monocular video. In K. F. et al., editor, Pattern Recognition (Proc. DAGM), volume 4174 of LNCS, pages 475–484, Berlin, Germany, September 2006. Springer. 2

[18] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(10):1025–1039, 1998. 2

[19] J. Shi and C. Tomasi. Good features to track. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR'94), Seattle, June 1994. 1, 2, 3.

[20] M. El-Ansari, S. Mousset, and A. Bensrhair, "Temporal consistent real-time stereo for intelligent vehicles," Pattern Recognition Letters, vol. 31, no. 11, pp. 1226–1238, August 2010.

[21] Y. Freund, R.E. Schapire, Experiments with a New Boosting Algorithm, In Proc. 13th Int. Conf. on Machine Learning, pp. 148.-156, 1996.

[22] Y. Freund, Boosting a weak learning algorithm by majority, Information and Computation, 121(2):256–285, 1995.

[23] R. Lienhart, J. Maydt, An Extended Set of Haar-like Features for Rapid Object Detection, IEEE ICIP 2002, Vol. 1, pp. 900-903, Sep. 2002.

[24] R. Lienhart, A. Kuranov, V. Pisarevsky, Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection, MRL Technical Report, May 2002.

[25] R. Miller. On-road vehicle detection: A review. IEEE Trans. Pattern Anal. Mach. Intell., 28(5):694–711, 2006. Member-Zehang Sun and Member-George Bebis.

[26] Jaesik Choi. Realtime On-Road Vehicle Detection with Optical Flows and Haar-like feature detector.

Multimodal Biometric Person Authentication using Speech, Signature and Handwriting Features

Eshwarappa M.N.
Telecommunication Engineering Department
Sri Siddhartha Institute of Technology
Tumkur-572105, Karnataka, India

Dr. Mrityunjaya V. Latte
Principal and Professor
JSS Academy of Engineering and Technology
Bangalore-560060, Karnataka, India

***Abstract*—The objective of this work is to develop a multimodal biometric system using speech, signature and handwriting information. Unimodal biometric person authentication systems are initially developed for each of these biometric features. Methods are then explored for integrating them to obtain multimodal system. Apart from implementing state-of-the art systems, the major part of the work is on the new explorations at each level with the objective of improving performance and robustness. The latest research indicates multimodal person authentication system is more effective and more challenging. This work demonstrates that the fusion of multiple biometrics helps to minimize the system error rates. As a result, the identification performance is 100% and verification performances, False Acceptance Rate (FAR) is 0%, and False Rejection Rate (FRR) is 0%.**

Keywords- Biometrics; Speaker recognition; Signature recognition; Handwriting recognition; Multimodal system.

I. INTRODUCTION

In the present era of e-commerce more and more services are being offered over the electronic devices and internet. These include banking, credit card facility, e-shopping, etc. To ensure proper use of these facilities only by the authorized or genuine users and avoid any misuse by the unauthorized or imposter users, some person authentication scheme is embedded into these services. Currently, person authentication is done mostly using one or more of the following means: text passwords, personal identification numbers, barcodes and identity cards. The merit of these schemes is that they do not change their value with respect to time and also unaffected by the environment in which they are used. The main demerit of them is that they can be easily misused or forgotten. Also, with time more and more services are being offered over the electronic devices and internet. Hence it becomes unmanageable to keep track of the authentication secrets for different services. The alternative that provides relief from all these demerits is the use of biometric features for person authentication. Any physiological and/or behavioural characteristics of human can be used as biometric feature provided it possesses the following properties: universality, distinctiveness, permanence, collectability, circumvention, acceptability and performance [2].

Some of the commonly used biometric features include speech, face, signature, finger print, handwriting, iris, DNA, Gait, etc. In practice, no single biometric can satisfy all the desirable characteristics mentioned above for it to be used for person authentication. This is due to the problems associated with noisy data, intra-class variation, non-universality, spoof attacks and high error rates [2]. To overcome this limitation, multiple biometric features can be used for person authentication. This resulted in the development of multimodal biometric person authentication system [2]. Thus biometric system can be classified as unimodal system and multimodal system based on whether single or multiple biometric features are used for person authentication. Biometric security system becomes a powerful tool compared to electronics based security systems [1]. Biometrics is fast becoming applicable in various walks of life. Basically, it deals with the use of computer technology and signal processing to identify people based on their unique physical and behavioural characteristics such as fingerprints, voice scans, retinal patterns, facial characters and human DNA mapping. Typically, a biometric system comprises a sensor, interface and a signal processor with driver software. The various different biometric procedures fall into two categories: Static process relating to the identification of fingerprints, hand geometry, Iris or retina and face, and Dynamic processes relating to the recognition of handwriting, keyboard typing patterns, voice, lip movement and behaviour analysis.

A biometric sensor works on the inputs provided by any of the human characteristics and applies an algorithm on the scanned biometric data. This is then compared with, and matched to, a template that has already been created earlier and approved by the user. The most specific and reliable biometric data is obtained from the DNA sequencing of any subject. The matching and comparing process creates a 'score' based on how closely the sampled biometric matches with the template already obtained. A match score is known as genuine score if it is a result of matching two samples of a biometric trait of the same user. It is known as an imposter score if it is the result of matching two samples of a biometric trait originating from different users. An imposter score that exceeds the predefined threshold results in a false accept, while a genuine score that falls below the predefined threshold results in a false reject. The False Accept Rate (FAR) of a biometric system is the fraction of imposter scores exceeding the threshold. Similarly, the False Reject Rate (FRR) of a system is defined as the fraction of genuine scores falling below the threshold. Regulating the value of threshold changes the FRR and the FAR values, but for a given biometric system,

it is not possible to decrease both these errors simultaneously. In real-world biometric system, biometric measure is referred in terms of FAR and FRR. The FAR measures the percentage of invalid users who are incorrectly accepted of genuine users and the FRR measures the percentage of valid users rejected as imposters. The Equal Error Rate (EER) refers to the point where the FAR equals the FRR. Lower the value of EER, the more accurate the biometric system.

There are several multimodal biometric person authentication systems developed in the literature [2-12]. Person authentication based on speech and face features, is one of the first multimodal biometric system [3]. Two acoustic features from speech and three visual features from face are used to build a multimodal system. Multimodal system using face and fingerprint features is then proposed [4]. Finger print verification on top of the face recognition is used to improve the recognition accuracy. The use of clustering algorithms for the fusion of decisions from speech and face modalities are explored [5]. A practical multimodal system using face, voice and lip movement is then developed [6]. The focus is on improving security by considering a dynamic feature like lip movement. In 2004, A. K. Jain et.al., proposed the framework for multimodal biometric person authentication [2]. They discussed in detail about the significance of biometric person authentication and desirable characteristics for a physiological and/or behavioural characteristics of human to be useful as biometric feature. Several biometric features in use are described in terms of these characteristics to highlight the strength and weaknesses of each of them. Details about the different levels of fusion, security and privacy concerns are also discussed. In recent times much of the interest is in audio-visual multi biometric systems [12].

The objective of our work is to develop a multimodal biometric person authentication system using speech, signature and handwriting biometric features. The motivation for the same are explained as follows: (1) speech is both physiological and behavioural, and signature and handwriting are behavioural biometric features, (2) each of these biometric features can be collected using sensors which are cheap and provide reasonably good quality data. All these features are non-intrusive type, easy to collect and hence acceptability among users will be high, (3) there are several practical applications where these three modalities fit in very well, like banking transaction. To complete a financial transaction, you can write the required amount using an electronic pen on the cheque displayed onscreen and put your signature. This enables giving both handwriting and signature features. You can read the amount written and other details that provide speech data, (4) speech is one of the mostly explored biometric features by the speech processing community for the development of speaker recognition system. Speech biometric feature will immensely benefit from the developments available in the speaker recognition literature, (5) the recent trend in human-computer interaction is the electronic-pen based input to the computer that includes Personal Digital Assistant (PDA) and Tablet-Personal Computers (PCs).

With this integrated input device, we have an easy way of capturing signatures and handwriting information. Thus these features can also be integrated into the multimodal system along with speech, (6) most of the existing signature verification systems are online type that uses dynamic features like time and pressure information. However, development of offline signature verification system may benefit from the rich image processing techniques available. A hybrid system using online and offline features may then be developed for increased robustness, (7) most of the handwriting recognition system is meant for forensic investigation. However, handwriting may also have significant information for person authentication. A detailed exploration is required from this perspective, (8) a person authentication system using handwriting information may also benefit from the speaker recognition literature by drawing parallels between the two. It may also be possible to extract some synchronous features between speech and handwriting to reduce spoof attacks.

The present work mainly deals with the implementation of multimodal biometric system employing speech, signature and handwriting as the biometric modalities. This includes feature extraction techniques, modelling techniques and fusion strategy used in biometric system. The organization of the paper is as follows: Section II deals with speaker recognition system, signature recognition system and handwriting recognition system using different feature extraction and modelling techniques, and Section III deals with multimodal biometric person authentication system by combining speaker, signature and handwriting recognition systems using fusion strategy. Section IV provides conclusion and suggestion for future work.

II. Development of Unimodal Systems

A. Speaker Recognition System

Speaker recognition is the task of recognizing speakers using their speech signal. The unimodal biometric system using speech analyzes and extracts speaker-specific features from the speech signal. The extracted features are then separately modeled to obtain one reference model for each speaker. During testing same analysis and feature extraction are carried out to extract speaker-specific features. These features are compared with the reference models to decide on the speaker. The speaker of the reference model that matches closely with the test speech features is declared as the speaker. In person authentication case, claimed identity is given along with the test speech. Hence comparison is done only with the claimed identity reference model and the claim is accepted or rejected based on the comparison with a preset threshold.

The state of the art system builds a unimodal system by analyzing speech in blocks of 10-30 milli seconds with shift of half the block size. Mel Frequency Cepstral Coefficients (MFCCs) are the mostly used features extracted from each of the blocks [13]. The MFCCs from the training or enrollment data are modeled using Vector Quantization (VQ) technique [14]. The MFCCs from the testing or verification data are compared with the VQ to validate the identity claim of the speaker. The MFCCs represent mainly the vocal tract aspect of speaker information and hence take care of only physiological aspect of speech biometric feature. Another important physiological aspect contributing significantly to speaker characteristics is the excitation source [16]. The behavioral biometric aspect of speech is present at longer duration levels

that can be characterized using supra-segmental features like speaking rate, pitch contour, duration etc. In this work, apart from the development of conventional speaker recognition system using MFCC and VQ features termed as baseline system, methods will also be explored to model excitation source and supra-segmental speaker-specific features. These features are then integrated into the baseline system. This may result in an improved and robust speaker recognition system.

Speaker recognition is the task of recognizing the speakers using their voices [17]. Speaker recognition can be either identification or verification depending on whether the goal is to identity the speaker among the group of speaker or verify the identity claim of the speaker. Further, speech from the same text or arbitrary text may be used for recognizing the speakers and accordingly we have text dependent speaker identification and verification approaches. The present work approaches text dependent speaker identification and verification of a speaker through identification. In this work, two different feature extraction and modeling techniques are used for text dependent speaker recognition. The feature extraction techniques are: (1) Mel Frequency Cepstral Coefficients (MFCC) are derived from cepstral analysis of the speech signal, (2) a new feature set, named the Wavelet Octave Coefficients of Residues (WOCOR), is proposed to capture the spectro-temporal source excitation characteristics embedded in the linear predictive residual of speech signal [16]. The two modeling techniques are used for modeling the person information from the extracted features are: (1) Vector Quantization (VQ), (2) Gaussian Mixture Modeling (GMM).

1) Feature extraction phase.

The speaker information is present both in vocal tract and excitation parameters [18]. The MFCCs represent mainly the vocal tract aspect of speaker information and hence take care of only physiological aspect of speech biometric feature. The vocal tract system can be modeled as a time-varying all-pole filter using segmental analysis. The segmental corresponds to processing of speech as short 10 to 30 milliseconds overlapped 5 to 15 milliseconds windows.

The vocal tract system is assumed to be stationary within the window and is modeled as an all-pole filter of order P using linear prediction analysis. The feature vectors that are extracted from smooth spectral representations are cepstral coefficients. In the present work we are using MFCC as feature vectors. The cepstral analysis used for separating the vocal tract parameters and excitation parameters of speech signal s(n). This analysis uses the fundamental property of convolution. The cepstral coefficients (C) are derived by using Fast Fourier Transform (FFT) and Inverse FFT (IFFT) which is given by equation (1).

$$C = real(IFFT(\log|IFFT(s(n))|)) \qquad (1)$$

Human auditory system does not perceive the spectral components in linear scale, but it will perceive on a nonlinear scale. So we can use the nonlinear scale, Mel frequency scale, to extract the spectral information. The critical band filters are used to compute the MFCC feature vectors by mapping the linear spaced frequency spectrum (f_{HZ}) into nonlinearly spaced frequency spectrum (f_{Mel}) using equation (2).

$$f_{Mel} = 2595\log_{10}\left(1+\frac{f_{HZ}}{700}\right) \qquad (2)$$

When a speech signal is given as an input to the feature extractor, it will truncate entire speech signal into frames of length 10-30 ms to make it quasi-stationary. Hamming window is used for eliminating the Gibbs oscillations, which occur by truncating the speech signal. But, due to windowing, samples present at the verge of window are weighted with lower values. In order to compensate this, we will try to overlap the frame by 50%. After windowing, we compute the log magnitude spectrum of each frame and calculating the energy in each critical filter bank. After finding the energy coefficients, we find the feature vectors using Discrete Cosine Transform (DCT) analysis. Compute the MFCC feature vectors for the entire frame of the speech signal for the individual speaker. In order to avoid channel mismatch we used cepstral mean subtraction procedure for the entire utterance. Liftering is a procedure which is used to eliminate the effects of different roll off in various telephone channels on cepstral coefficients. In this work, the conventional speaker recognition system using MFCC feature will be the baseline system.

The new feature set used in our work is the Wavelet Octave Coefficients of Residues (WOCOR). A time-frequency vocal source feature extraction by pitch-synchronous wavelet transform, with which the pitch epochs, as well as their temporal variations within a pitch period and over consecutive periods can be effectively characterized [40]. The wavelet transform of time signal x(t) is given by equation (3).

$$w(a,\tau) = \frac{1}{\sqrt{|a|}}\int_t x(t)\psi^*\left(\frac{t-\tau}{a}\right) \qquad (3)$$

Where $\psi(t)$, a and τ are the mother wavelet function, scaling (or dilation) parameters and translation parameter respectively. Where $\psi\left(\frac{t-\tau}{a}\right)\frac{1}{\sqrt{a}}$ is named the baby wavelets. It is constructed from the mother wavelet by first, scaling $\psi(t)$ which means to compress or dilate $\psi(t)$ by parameter a and then moving the scaled wavelet to the time position of parameter τ . The compression or dilation of $\psi(t)$ will change the window length of wavelet function, thus changing the frequency resolution. Therefore, the ensemble of $\psi\left(\frac{t-\tau}{a}\right)\frac{1}{\sqrt{a}}$ constitutes the time-frequency building blocks of the wavelet transform [30]. The wavelet transform of discrete time signal x(n) is given by equation (4).

$$w(a,b) = \frac{1}{\sqrt{a}}\sum_n x(n)\psi^*\left(\frac{n-b}{a}\right) \qquad (4)$$

Where $a = \{2^k | k=1,2\ldots,K\}$ and $b = 1,2\ldots,N$, and N is the window length. $\psi^*(n)$ is the conjugate of the fourth-order Daubechies wavelet basis function $\psi(n)$. K=4 is selected such that the signal is decomposed into four sub-bands at different octave levels. At a specific sub-band, the time-varying characteristics within the analysis window are

measured as parameter b changes. To generate the feature parameters for pattern recognition, the wavelet coefficients with specific scaling parameters are grouped is given by equation (5).

$$W_k = \{ w(2k,b) | b = 1,2..N \} \quad (5)$$

where N is the window length. Each W_k is called an octave group. Then WOCOR parameters can be derived by using equation (6).

$$WOCOR_M = \left\{ \mathrm{P}W_k(m)\,\mathrm{P} \middle| \begin{matrix} m = 1,2..M \\ k = 1,2..4 \end{matrix} \right\} \quad (6)$$

where ||.|| denotes two-norm operation. Finally, for a given speech utterance, a sequence of $WOCOR_M$ feature vectors is obtained by pitch-synchronous analysis of the LP residual signal. Each feature vector consists of 4M components, which are expected to capture useful spectro-temporal characteristics of the residual signal. For each voiced speech portion, a sequence of LP residual signals of 30 ms long is obtained by inverse filtering the speech signal. The neighboring frames are concatenated to get the residual signal, and their amplitude normalized within (-1, 1) to reduce intra-speaker variation. Once the pitch periods estimated, pitch pulses in the residual signal are located. For each pitch pulse, pitch-synchronous wavelet analysis is applied with a Hamming window of two pitch periods long. For the windowed residual signal x(n) the wavelet transform is computed using equation (4).

2) Training Phase.

For speaker recognition, pattern generation is the process of generating speaker specific models with the collected data in the training stage. The mostly used modeling techniques for modeling include vector quantization [14] and Gaussian mixture modeling [15]. The VQ modeling involves clustering the feature vectors into several clusters and representing each cluster by its centroid vector for all the feature comparisons. The GMM modeling involves clustering the feature vectors into several clusters and representing all these clusters using a weighted mixture of several Gaussians. The parameters that include mean, variance and weight associated with each Gaussian are stored as models for all future comparisons. A GMM is similar to a VQ in that the mean of each Gaussian density can be regarded as a centroid among the codebook. However, unlike the VQ approach, which makes hard decision (only a single class is selected for feature vector) in pattern matching, the GMM makes a soft decision on mixture probability density function. This kind of soft decision is extremely useful for speech to cover the time variation.

In training phase, the first modeling technique we used in this work is Vector Quantization (VQ). After finding the MFCC feature vectors for the entire frame of the speech signal for the individual speaker, we have to find some of the code vectors for the entire training sequence with less number of code words and having the minimum mean square error. To find minimum mean square error with less number of code words by using VQ, we have two most popular methods namely K-means algorithm and Linde-Buzo and Gray (LBG) algorithms [23]. Vector quantization process is nothing but the idea of rounding towards the nearest integer.

The second modeling technique we used in our work, the Gaussian Mixture Modeling (GMM), which is most popular generative model in speaker recognition. The template models, VQ codebooks, can also be regarded as a generative model, although it does not model variations. The pattern matching can be formulated as measuring the probability density of an observation given the Gaussian. The likelihood of an input feature vectors given by a specific GMM is the weighted sum over the likelihoods of the M unimodal Gaussian densities [32], which is given by equation (7).

$$P(x_i | \lambda) = \sum_{j=1}^{M} w_j b(x_i | \lambda_j) \quad (7)$$

The likelihood of x_i given j^{th} Gaussian mixture is given by

$$b(x_i | \lambda_j) = \frac{1}{(2\pi)^{D/2} |\sum_j|} \exp\left\{ -\frac{1}{2} (xi - \mu j)^T \sum_j^{-1} (xi - \mu j) \right\} \quad (8)$$

Where D is the vector dimension, μ_j and Σ_j are the mean vectors and covariance matrices of the training vectors respectively. The mixture weights w_j are constructed to be positive and the sum to be one. The parameters of a GMM are: Mean (μ_j), Covariance (Σ_j) and Weights (w_j) can be estimated from the training feature vectors using the maximum likelihood criterion, via the iterative Expectation-Maximization (EM) algorithm (32). The next stage in the speaker recognition system will be the testing phase.

3) Testing Phase.

In this phase, feature vectors are generated from the input speech sample with same extraction techniques as in training phase. Pattern matching is the task of calculating the matching scores between the input feature vectors and the given models in recognition. The input features are compared with the claimed speaker pattern and a decision is made to accept or reject the claiming. Testing phase in the person authentication system includes matching and decision logic. The testing speech is also processed in a similar way and matched with the speaker models using Euclidean distance in case of VQ modeling and likelihood ratio in case of GMM modeling. Hence matching gives a score which represents how well the feature vectors are close to the claimed model. Decision will be taken on the basis of matching score, which depends on the threshold value.

The alternative is to employ verification through identification scheme. In this scheme the claimed identity model should give best match. The test speech compared with the claimed identity model, if it gives best match, then it is accepted as genuine speaker, otherwise, rejected as imposter.

For testing the performance of speaker recognition system, we have collected the speech database of students of SSIT at a sampling frequency of 8 kHz. Figure 1 shows speaker 1 sample speech signal of four sentences, which is collected by using microphone.

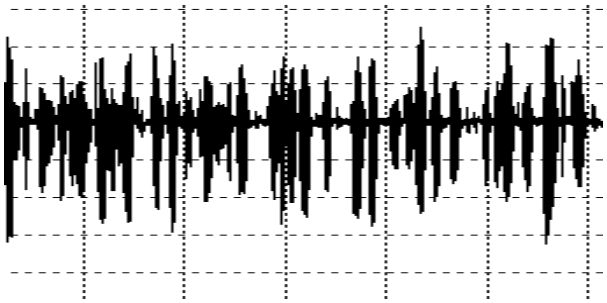


Figure 1. Sample of speech signals of speaker 1.

The SSIT database contains the speech data of 30 speakers, among them 20 were male and the remaining 10 were female. Four sentences are used for each speaker and 24 number of utterances of each sentence for each speaker. First 16 utterances are used for training and the remaining 8 utterances are used for testing. Table I shows the experimental results of different speaker identification and verification systems using SSIT speech database.

TABLE I. SPEAKER RECOGNITION SYTEM

Code book size	**MFCC-VQ based System**		
	Speaker Identification	***FAR***	***FRR***
32	98.75%	0%	0%
64	100%	0%	0%
Code book size	**WOCOR-VQ based System**		
	Speaker Identification	***FAR***	***FRR***
32	89.1667%	0.1293%	3.75%
64	96.25%	0.22%	1.624%
Gaussians	**MFCC-GMM based System**		
	Speaker Identification	***FAR***	***FRR***
32	100%	0%	0%
64	100%	0%	0%%
Gaussians	**WOCOR-GMM based System**		
	Speaker Identification	***FAR***	***FRR***
32	94.5833%	0.113%	3.333%
64	100%	0%	0%%

The performance of the conventional MFCC-VQ based speaker recognition system with code book size of 64 gives better result compared to WOCOR-VQ based system. The WOCOR-GMM and MFCC-GMM based speaker recognition systems with 64 Gaussians also gives better result. Finally, we combine the matching scores of MFCC-VQ and WOCOR-GMM based systems. These combined system become one of the baseline system, which is used for developing the multimodal system for person authentication. The reasons for the same are the different feature extraction and modeling techniques are used.

B. Signature Recognition System

Signature recognition is the task of recognizing signatories by using their signatures. Signature is a behavioral biometric, the features of signature are variant with respect to time and the forgers can easily fool the system by reproducing the signatures of the correct persons. Irrespective of the above limitations we can still use signature as our best biometric feature, since the signature is a unique identity of an individual and is being used extensively in practical systems. No two signatures can be identical, unless one of them is a forgery or copy of the other [35]. The signature recognition systems find applications in government, legal and commercial areas. Signature verification is the verification of given signature of claimed identity of a person. There are two types of signature verification systems in practice, namely, online and offline [17], [18]. Online signature verification uses information collected dynamically at the time of signature acquisition like timing, acceleration, velocity, pressure intensity and also termed as dynamic signature verification. Offline signature verification uses only the scanned image of signature and also termed as static signature verification.

In case of online signature verification during the training phase, the user supplies a set of reference signatures measured in terms of dynamic features mentioned above. These dynamic features along with signatures are stored as reference templates. When a test signature is input to the system in terms of these dynamic features, it is compared to each of the reference signatures of the claimed person. Based on the resulting comparison distance, the claimed identity is either accepted or rejected. Most of the existing signature verification systems are based on online approach. Not much importance has been given to the offline signature verification as it is relatively complex. The complexity may be due to the two dimensional nature of offline signature compared to one dimensional online signature. However, once we have signature images, then we can view signature verification as a pattern recognition problem. The online and offline approaches exploit different aspect of signature information for verification, namely, dynamic and static. The development of offline signature verification and integrating with existing online system may provide improved performance as well as robustness. It is therefore aimed to develop offline signature verification system static features like aspect ratio, horizontal projection profile, vertical projection profile and discrete cosine transform features.

1) Feature extraction phase.

Feature extraction plays a very important role in offline signature verification. Unlike our speaker recognition case, we are not going model the feature vectors up to some codebook level. Here feature vectors itself will give the training sequence. In this work the features of signature are extracted by using Discrete Cosine Transform (DCT) analysis, Vertical Projection Profile (VPP) analysis and Horizontal Projection Profile (HPP) analysis. The VPP and HPP are static features of a signature and DCT is a global feature of a signature image. Since our signature is an image, it will have the gray levels from 0 to 255 and to compute the maximum gray level the histograms of all images are used. VPP and HPP are the kind of histograms. VPP gives the horizontal starting and ending points and HPP gives the vertical starting and ending points of the image. The size of VPP and HPP is equal to the number of columns and the number of rows in the signature image respectively. Since, the size of signature regions are not

constant even for a single user, in this work we are taking average value of vertical projection profile as a feature.

$$vpp_{avg} = \frac{1}{N}\sum_{q=1}^{N}\sum_{p=1}^{M} A(p,q) \quad (9)$$

$$hpp_{avg} = \frac{1}{M}\sum_{p=1}^{M}\sum_{q=1}^{N} A(p,q) \quad (10)$$

The signature image intensity A (p, q) at p^{th} row and q^{th} column indices respectively. Where M is number of rows in an image and N is number of columns in image. Equation (11) gives the DCT coefficient corresponding to p^{th} row and q^{th} column of an input signature image.

$$B_{pq} = \alpha_p \alpha_q \sum_{m=0}^{M-1}\sum_{n=0}^{N-1} A_{mn} \cos\left(\frac{\pi(2m+1)p}{2M}\right)\cos\left(\frac{\pi(2n+1)q}{2N}\right) \quad (11)$$

where $\alpha_p = \left\langle \frac{1}{\sqrt{M}} \right.$ for p=0 and $\alpha_p = \left\langle \sqrt{\frac{2}{M}} \right.$ for $1 \le p \le (M-1)$

$\alpha_q = \left\langle \frac{1}{\sqrt{N}} \right.$ for q=0 and $\alpha_q = \left\langle \sqrt{\frac{2}{N}} \right.$ for $1 \le q \le (N-1)$

The performance of signature recognition system depends on the way in which the DCT coefficients are considered. The zonal coding of DCT coefficients of signature image are used for better performance, which gives concentration at low spatial frequencies.

2) Testing Phase.

For the identification or verification, same set of features which have been extracted during registration process are extracted from the input samples scanned or recorded using input devices like writing pads, to form the feature vectors. Verification is 1 to 1 matching while identification is 1 to n matching. In verification, the individual claims his/her identity which is verified by comparing these features vectors by the feature vectors of the individual which he/she claimed to be. If the matching score crosses the predefined threshold then the system verifies the individual as authentic user. In identification, the feature vectors of the individual are compared with the feature vectors of every individual stored in the database. If the highest matching score crosses the predefined threshold, then it identifies the individual as the person whose matching score is the highest otherwise the system suggest few top most matches. The matching algorithm is needed to compare the samples and computes the matching score and decide if two samples belong to the same individual or not by comparing the matching score against the acceptance threshold. However, it is possible that sometimes the output of a biometrics system may be wrong. Therefore, the performance of a biometrics system is measured in terms of two errors: FAR and FRR. In order to design the multimodal system using speech and signature features, we have collected the signature database from the same 30 students who had given their speech samples while collecting the speech database. For every writer we have taken 24 samples of signatures and scanned them by using HP Scan jet 5300C scanner at 300 digits per inch resolution and stored them in 'bmp' format. After scanning the signatures, we have cropped all the 24 signatures of individual writer by using Windows Picture manager. Figure 2 shows one of the sample signature of user 1.

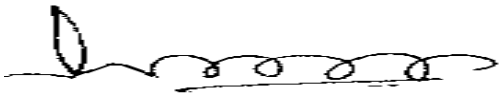


Figure 2. Sample signature of user 1.

During the training session, we considered the first 16 signatures of each writer and extract the features from those signatures by using VPP, HPP and DCT analysis. The three feature models are obtained for all 30 users. In testing phase, we have used the remaining 8 signatures for each writer. For the given test signature, we have to extracted the VPP, HPP values and DCT coefficients separately by VPP-HPP-DCT analysis. After getting these values, we found the minimum distance between the VPP-HPP-DCT values and the feature vectors of all the writers corresponding to each of the model. Table II shows the performance of different signature recognition systems using SSIT signature database. The VPP-HPP-DCT method gives highest performance (86.66%) compared to the other systems.

TABLE II. SIGNATURE RECOGNITION SYTEM

System	VPP-HPP-DCT based System
	Signature Identification
VPP	21.25%
HPP	30.4167%
VPP-HPP	56.25%
DCT	72.9167%
VPP-HPP-DCT	86.667%

To improve the performance of the signature recognition system, along with the baseline VPP-HPP system the DCT coefficients are used. A modified system uses VPP and HPP vectors with Dynamic Time Warping (DTW) for the optimal cost. DTW is a pattern matching technique which aims at finding the minimum cost path between the two sequences having different lengths [26]. A very general approach to find distance between two time series of different sizes is to resample one of the sequence and comparing the sample by sample. The drawback of this method is that there is a chance of comparing the samples that might not correspond well. This means that comparison of two signals correspond well when there is a matching between troughs and crests. DTW solves this method by considering the samples with optimum alignment. The DTW computation starts with the warping of the indices of two sequences. The two sequences are compared with some distance measures like Euclidean distance at each and every point, so as to obtain the distance matrix. These distances in the matrix are termed as local distances. Let the Matrix be D and the sequences are A, B with lengths M, N respectively. Then D is calculated using equation (12).

$$D(i, j) = \text{distance}(A(i), B(j)) \quad (12)$$

where i varies from 1 to M and j varies from 1 to N. The distance here considered is Euclidean distance. The modified

feature vectors obtained from the signature image A(i,j) of size MxN are given in the equations (13) and (14).

$$vpp_{(j)} = \sum_{i=1}^{M} A(i, j) \text{ where j=1,2,3……. N} \quad (13)$$

$$hpp_{(j)} = \sum_{j=1}^{N} A(i, j) \text{ where i=1,2,3….. M} \quad (14)$$

Calculate the DTW distance values separately for VPP and HPP vectors from all the users for all the training images to the testing image and obtain distances from each user using average distance method. Normalize each of the distance of a particular feature using one of the normalization methods and use sum rule for fusion of match scores obtained using each model. Assign the test signature to the user who produces least distance in fused sum vector. Table III shows the results of signature verification system using SSIT signature database.

TABLE III. SIGNATURE VERIFICATION SYTEM

System	VPP-HPP-DCT based System		
	FAR	***FRR***	***Average Error***
VPP-HPP-DCT	1.2931%	37.5%	19.396%
Modified VPP-HPP-DCT	0.1149%	3.333%	1.7241%

The VPP-HPP-DCT based signature identification gives better result compare to other systems. The modified VPP-HPP-DCT system gives even better in signature verification. These systems are used in multimodal biometric system for identification and verification of test signature respectively.

C. Handwriting Recognition System

Handwriting biometric feature can also be used for person authentication [25]. Most of the existing works on handwriting information is for forensic investigation. The scope includes identifying the author of the given handwritten script from the group of available large population. The end result may be a subgroup of most likely population. This subgroup may then be carefully analyzed by the human experts to identity the correct person who might have written the script. Thus using handwriting information in criminal investigation is an age old method. Handwriting biometric feature may also posses several characteristics to qualify it for use in person authentication. Relatively few works have been done in this direction [25]. With the integration of pen-based input devices in PDA and Tablet PCs strongly advocates the use of handwriting information for person authentication due to ease of collection. Handwriting verification can also be done either in online or offline mode as in signature verification. Online handwriting verification exploits similar dynamic features as in signature verification. Thus it is easy to extend the online signature verification approach to handwriting verification. Initially an online handwriting verification system will be developed. However, it should be noted that there is a significant difference between signature and handwriting. Signature is one pattern from hand, but it will not use any language specific information. Alternatively, handwriting exploits language information at various levels, starting from character set. Since this has been trained during initial days of learning stage of language, it is possible that, we may find more regular and reliable feature from handwriting for person authentication. The offline handwriting verification can be approached by using the information from the offline signature verification literature and also from the speaker recognition literature. An offline system is initially developed using the technique developed for offline signature verification. Later techniques available in speaker recognition literature can be mapped here to further improve the performance or develop a new technique for verification. For instance, well known text dependent speaker verification technique using Dynamic Time Warping (DTW) can be extended to offline handwriting verification in the text dependent mode using VPP features. Finally a hybrid handwriting verification system using offline and online approaches may be developed to provide improved performance and robustness.

The methods in handwriting biometrics account for both offline and online with verification and identification modes. The two recent approaches for handwriting recognition which have proved fruitful are based on a textural feature, whereas the second method zooms in on character shape elements [29]. The first method refers to angles and curvature in handwriting which are determined by the degrees of freedom in wrist and finger movement, which in turn depend on the pen grip attitude and the applied grip forces. The other approaches include use of Hidden Markov Models, Gray level distribution [28], Support vector machine, and connected component contours [27]. The Gray level distribution approach is used for handwriting recognition.

1) Feature extraction phase.

Mainly dynamic time warping in context of images is used for word matching which uses vectors like normalized upper word and lower profile, back ground ink transitions etc.,. The features from the handwriting image considered in our work are VPP vector and HPP vector. The VPP is an array that contains sum of gray levels of each column in a handwriting image. This feature signifies the variations of Gray level distribution along the length of the image. This VPP vector is a unique feature for a given user and will vary from user to user. Even the same user will have variations. The important and the uniqueness of the information present in the HPP vectors are equally important as that of VPP vectors. So along with the VPP vector extraction, another feature HPP vector is obtained from the handwriting image. This HPP vector gives the information about the variations of the handwriting along the lateral extent. The handwriting recognition system runs on the same lines as of the signature recognition system.

2) Testing Phase.

Writer recognition system is built using the individual words, segmented from the sentence considered for handwriting and combined later for better performance. In order to obtain a correct segmentation, a threshold is calculated that distinguish words and characters. After obtaining the threshold, words are segmented by obtaining the VPP vector and examining its intensity profile. The each word extracted from the sentence now act as the images to be tested. The algorithm proposed for a full sentence is applied for each word. First obtain the image from which words should be segmented out and let N numbers of words are segmented.

Consider one word and apply already proposed algorithm for all sentence. Obtain the DTW distances from each user by averaging method. Next, normalize the distances and repeat the same procedure for all the words. The normalized distances are fused using the fusion principle. Obtain minimum distance and its corresponding user there by identifying the user. At the fusion level, distances are fused using sum rule. The similar procedure is used for finding the HPP vectors.

For handwriting recognition system we created database of same 30 users of speech and signature recognition systems. The same sentence used for verification for all users. The sentence is written on A4 sheet with eight equal rectangular boxes. This sheet was scanned using HP scanner, at a resolution of 300 digits per inch. Then sentences are separated out and stored in bits mapping format. In our experiments, five samples are used for training and three samples are used for testing. Figure 3 shows the sample of handwriting of user 1.

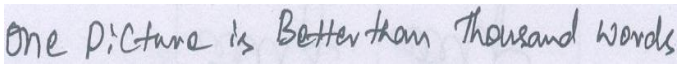


Figure 3. Sample of handwriting of user 1.

Table IV shows the performance of handwriting identification and verification system using SSIT database.

TABLE IV. HANDWRITTING RECOGNITION SYTEM

System	VPP-HPP-based System		
	Identification	***FAR***	***FRR***
HPP	80.63%	1.461%	26.6%
VPP	90.02%	0.062%	0.362%

The VPP based handwriting recognition system gives better result compared to HPP based system. The reason is that, the richest image information obtained from Gray level distribution along the length of the handwriting image. The HPP vectors are equally important as that of VPP vectors which gives the information about the variations of the handwriting along the lateral extent. The combined feature gives the complete behavior of handwriting image of a user, hence VPP-HPP based system is one the unimodal system in our multimodal biometric person authentication.

III. MULTIMODAL BIOMETRIC PERSON AUTHENTICATION SYSTEM

A. Development of Multimodal System

Multimodal or Multi-biometric systems, remove some of the drawbacks of the unimodal systems by grouping the multiple sources of information. These systems utilize more than one physiological or behavioral characteristic for enrollment and identification. Once the unimodal systems are developed, then the next step is to develop multimodal system by integrating them suitably. The unimodal systems using speech, signature and handwriting information are ranked according to their performance. Based on this, the best performing system is used as the baseline system to which other systems are integrated. The integration can be done at any of the following three levels: feature, measurement and score levels [2]. A tight integration is possible if it is done at the feature level. However, the difficulty associated is the different nature of features and also significant variation of person information. This difficulty can be overcome by integrating at the measurement level, but the level of person information present may be smoothened out to some extent due to modeling. To that extent the level of coupling will be loose or moderate. The integration of measurement values may also harm the combined system, if one of the systems provides poor performance. Under such condition, the safe way to integrate is at the score level. This level of fusion is immune to any poor performance, since already decision is made about the person. But the amount of improvement achieved after combination may be relatively low. For the selected baseline system, only the next best performing unimodal system is integrated at the feature level. One approach for the same is to extract same features say, Discrete Cosine Transform (DCT) values and normalize them suitably and combine them. Alternatively, the features can be applied to one more level of smoothing using feature modeling techniques to obtain modified features that are similar for both biometric features. These features are then used for modeling. As a result of this a bimodal person authentication system is developed by integrating at the feature level. At the next level, the best performing unimodal system is integrated with the feature level integrated bimodal system. The integration is done at the measurement level. The measurement scores from the two systems are suitably normalized combined and evaluated [36]. This results in tri-modal biometric person authentication system. At the final level, the unimodal system based on the fourth biometric feature is integrated into the tri-modal system at the score level. For this several combination techniques are explored to obtain maximum gain. This will result in the development of the multimodal biometric person authentication system using all the possible biometric features.

The particular biometric feature selection for developing unimodal system needs to satisfy different characteristics, as we mentioned above. However, some of the parameters are observed and studied during the design of unimodal systems. The following are the certain parameters to decide whether a biometric trait can be used for person authentication or not.

They are: (1) how common the trait is found in individual, (2) how much the trait varies from individual to individual, (3) how the trait varies with the age of the individual, (4) how easily the trait be collected, (5) how easily the trait can be processed and how is the accuracy and speed of the system built using the trait, (6) how people adopt the technology in their day to day life. The speech, signature and handwriting biometrics are fulfill the above requirements and characteristics of biometric person authentication. By combining the offline signature recognition, offline handwriting recognition and the text dependent speaker recognition systems, the challenge-response type of authentication can be facilitated.

With these factors, the three best performing unimodal systems are combined using score level fusion. Since we are using score level fusion, there are no special steps involved in the training of biometric system. In the score level fusion, scores obtained at the output of the classifier are fused using

some rules. The simple rules of fusion are Sum rule, Product rule, Min Rule, Max rule and Median rule. The Sum rule and Product rule assume the statistical independence of scores from the different representations [6]. The outputs on the individual matchers need not be on the same numerical scale. Due to these reasons, score normalization is essential to transform the scores of the individual matchers into a common domain prior to combining them. Score normalization is a critical part in the design of a combination scheme for matching score level fusion. Min-max and Z-Score normalization are the most popular techniques used for normalization. The present work uses the Z-Score normalization techniques for the individual matcher and the Sum rule for integrating the normalized scores. Using these two principle techniques, the multimodal biometric system is designed using three unimodal systems. Once the multimodal system is developed the next stage is performance and robustness evaluation.

B. Performance and Robustness Evaluation

There are standard databases for the individual evaluation of the unimodal biometric systems, like YOHO database, IITG database etc... However, such an evaluation is only for finding the performance of the particular unimodal system in an absolute sense. To have comparative study evaluate the strength of multimodal system on common platform, it is proposed to develop a multimodal database for these three biometric features. For this reason we have prepared our own SSIT database of 30 users.

The database consists of 24 samples of speech information, 24 samples of signature and 8 samples of handwriting for each user. Once the database is developed, then the performance is evaluated first for each of the unimodal systems. The performance is then evaluated for multimodal system using all the three features. Such evaluation provides a systematic comparison between unimodal and multimodal systems. The main features considered in developing a multimodal system are handwriting, signature and speech. The following are the steps involved in the implementation of multimodal biometric person authentication system based on unimodal system performance.

a) Collect the individual matching scores of the unimodal systems for every user.

b) Normalize the matching scores using normalization techniques and integrate the scores by using fusion rules.

c) Assign the multiple biometric to a particular person who produces the minimum score.

Table V shows results of different multimodal systems based on the combination of different unimodal systems.

TABLE V. PERFORMANCE OF MULTIMODAL BIOMETRIC SYSTEMS

System	VPP-HPP-based System	
	FAR	***FRR***
VQ-MFCC for Speech VPP-HPP-DCT for Signature VPP-HPP with DTW for Hadwriting	0%	0%
VQ-WOCOR for Speech VPP-HPP-DCT for Signature VPP-HPP with DTW for Hadwriting	0%	0%
GMM-MFCC for Speech VPP-HPP-DCT for Signature VPP-HPP with DTW for Hadwriting	0%	0%
GMM-WOCOR for Speech VPP-HPP-DCT for Signature VPP-HPP with DTW for Hadwriting	0%	0%

Table V proves the advantages of multimodal biometric system through its performance and robustness evaluation by using more number of biometrics for person authentication. The other major factors to be concentrated along with the development of multimodal system are the fusion rules and the normalization techniques. The score level fusion technique with Sum rule is employed in all the cases. The normalization techniques are used for the maintenance of the homogeneity among the scores obtained from different features. As a result, we are using the three possible combinations of unimodal systems to develop four multimodal systems. The each multimodal system identification performance is 100% and the verification performance is 0% error rates, even though there are some error rates in respective unimodal system.

IV. CONCLUSION AND FUTURE WORK

The trend of multimodal biometrics is spreading for the authentication process to maintain the interests regarding the security as strong as possible. The vital features that encourage the use of multimodal biometrics are the performance and accuracy along with the ability to overweigh the drawbacks of unimodal biometric systems. In this work we demonstrated multimodal biometric person authentication system using three biometric features. We generated our own database of 30 users and effectively using the principle of matching score fusion and normalization technique for developing multimodal system. Further, we combined the multimodal systems shown in Table V using normalization and fusion techniques. This system gives the identification performance is 100% and the verification performance is 0%, in terms of FAR 0% and FRR is 0%. As a result, we implemented multimodal biometric person authentication system using speech, signature and handwriting features which provides 0% error rates.

The future work may include integrate the biometric features at the feature level for improving the performance.

Further, to make the biometric person authentication system more practical: use more number of users, different sessions of collecting data of the same users, and multiple sensors for data collection. Also, by combining the offline system with online system may improve the performance.

ACKNOWLEDGMENT

I would like to thank my students for providing their speech, signature and handwriting information in generating the multimodal SSIT database.

REFERENCES

[1] L.Gorman, "Comparing passwords, tokens and biometrics for user authentication," IEEE Proc., vol. 91, no.12, Dec. 2003.

[2] A.K. Jain, A Ross and S. Prabhaker, "An introduction to biometric recognition," IEEE Trans. Circuits and Systems for Video Technology, vol. 14, no. 1, pp, 4-20, Jan. 2004.

[3] R. Bruneelli and D.Falavigna, "Person identificationusing multiple cues," IEEE Trans. PAMI, vol. 17, no.10, pp.955-966, oct.1995.

[4] L. Hong and A.K. Jain, "Integrating faces and fingerprints for person identification," IEEE PAMI, vol. 20, no. 12, pp. 1295-1307, Dec. 1998.

[5] V. Ghattis, A.G. Bors and I. Pitas, " Multimodal decision level fusion for person authentication," IEEE Trans. Systems, Man and Cybernatics, vol. 29, no. 6, pp, 674-680, Nov. 1999.

[6] R. W.. Frischholz and U. dieckmann, "Bioid: A multimodal biometric identification system," IEEE Computer Society, pp. 64-68, Feb.2000. Name Stand. Abbrev., in press.

[7] A. Kumar et. al., "Person verification using palmprint and handgeometry biometric," proc. Fourth Int. Conf. AVBPA, pp.668-678, 2003.

[8] S.Ribaric, D. Ribaric and N. Pavesic, "Multimodal biometric user identification system for network based applications," IEEE Proc. Vision, Image and Signal Processing, vol. 150, no.6, pp.409-416, 2003.

[9] A.K. Jain and Ross, "Learning user specific parameters in multibiometric system," Proc. Int. Conf. Image Processing (ICIP), pp. 57-60,2002.

[10] A. K.Jain, L.Hong and Y. Kulkarni, " A multimodal biometric system using fingerprint, face and speech," Proc. Second Int. Conf. AVBPA, pp.182-187, 1999.

[11] S.Ribaric, I. Fratic and K. Kris, " A biometric verification system based on the fusion of palmprint and face features," Proc. Fourth Int. Symposium Image and Signal Processing, pp. 12-17, 2005.

[12] B.Duc et. al., "Fusion of audio and video information for multimodal person authentication," Pattern Recogntion Letters, vol. 18, pp.835-845, 1997.

[13] S. Furu, "Cepstral analysis techniques for automatic speaker verification," IEEE Trans, Acoust., Speech Signal Processing, vol. 29(2), pp.254-272, 1981.

[14] F.K.Soong, A.E.Rosenberg, L.R. Rabiner and B.H. Jvang," A Vector quantization approach to speaker recognition," Proc., IEEE, Int., Conf., Acoust.,

Speech Signal Processing, vol. 10, pp.387-390, Apr.1985.

[15] D.A.Reynolds, "Speaker identification and verification using gaussian mixture speaker models," Speech Communication, vol. 17, no.1-2, pp.91-108, 1995.

[16] S.R.M. Prasana, C.S. Gupta, and B. Yegnanarayana, "Extraction of speaker-specific excitation information from linear prediction residual of speech," Speech Communication, vol. 48, no.10, pp.1243-1261, oct.2006.

[17] B.S.Atal, "automatic recogntion of speakers from their voices," IEEE Proc. vol. 64(4), pp.460-475, Apr. 1976.

[18] A.Eriksson and P.Wretling, "How flexible is the Human Voice? Acase study of Mimicry," Proc. European Conf. on Speech Tech. Rhodes, 1043-1046, 1997.

[19] V.S.Nalwa, "Automatic online signature verification," Proc. IEEE, vol. 85, no.2, pp.213-239, Feb. 1997.

[20] W.Hou, X.Ye and K.Wang, "A survey of offline signature verification," Proc.Int. Conf. Intelligent Mechatronics and Automation, pp. 536-541, Aug.2004.

[21] T. Scheidat, c. Vielhauer and J. dittmann, "Single-semantic multi-instance fusion of handwritting based biometric authentication systems," Proc.Int. Conf. IEEE-ICIP, pp. II-393-II-396, 2007.

[22] A.Rosenberg, "Automatic speaker verification: a review," Speech Communication, vol. 17, no.1-2, pp.91-108, 1995.

[23] Y.Linde, A.Buzo, and R.M.Gray, " An algorithm fot vector quantizer design," IEEE Trans. on Communications, vol. COM_28(1), pp.84-96, Jan. 1980.

[24] Chaur-Heh Hsieh, "DCT based code book design for vector quantization of images," IEEE Trans. Systems for Video Technology, vol. 2, no.4, pp.401-409, Dec 1992.

[25] F.Ramann C.Vielhuue, and R. Steinmetz, "Biometrics applications based on handwriting," IEEE Proc. Int. Conf. on Multimedia and Expo, ICME, vol2, pp.573-576, 2002.

[26] T.M. Mat and R. Manmatha, "Word image matching using dynamic time warping," Proc. IEEE, Computer Vision and Pattern Recgn., vol.2, pp.521-527, June 2003.

[27] L.Schomaker and M. Bulacu, "Automatic writer identification using connected component contours and edge-based features of uppercase western sript," IEEE Tran. Pattern Analysis and Machine Intelligence, vol. 26, pp.787-789, 2004.

[28] M.Wirotius, A. Seropian, and N. Vincent, "Writer identification from Gray level distribution," Proc. 7th Int. Conf. on Document Analysis and Recognition, ICDAR, pp.1168-11721, 2003.

[29] F. Nobard, "Handwritten signature verification: Global approach, Fundamentals in handwriting recognition," Springer-verlag, berlin series: Computer and System Science, 124, pp.445-495, 1991.

[30] Y.T. Chan. Wavelet Basics. Kluwer Academic Publishers Group, 1996.

[31] G. strang and T. Nguyen. Wavelets and Filter Banks. WellesleyCambridge Press, 1996.

[32] A. Sanker and C.H. Lee, "A maximum-Likelihood approach to stochastic matching for robust speech recognition," IEEE Trans. Speech-Audio Processing, 4(3): 190-202, 1996.

[33] L.E. Baum and T. Petie, "Stastical inference for probabilistic functions of finite state Markov chains," Ann. Mat. Stat., 37, pp.1554-1563, 1966.

[34] F.Leclerc and R. Plamodon, "Automatic Signature Verification," Int. Jr. of Patt. Recogn. and Art. Intelligence, vol. 18, no. 3, pp. 643-660, 1994.

[35] M.Ammar, Y.Yoshido, and T.Fukumura, "Structural description and classification of signature images," Patt. Recogn. vol. 23, no. 7, pp. 697-710, 1990.

[36] A. Jain, K.Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," Elsevier, Patt. Recogn. Journal, vol. 38, pp. 2270-2285, Jan. 2005.

[37] I.Daubechies, "Ten Lectures on Wavelets," Phaladelphia, PA: Siam, pp.36-106, 1992.

[38] Earl Gose, Richard Johnsonbaugh, and Steve Jost, Pattern Recognition and Image Analysis, PHI: Pentice Hall publisher, pp. 329-409, 1997.

[39] Meenakhsi, Sargur Srihari, and aihuxu, "Offline signature verification and identification using distance measures," Int. Journal, Patt. Recogn. and Art. Intelligence, vol. 18, no.7, pp. 1339-1360, 2004.

[40] Nengheng Zheng, Tan Lee and P.C. Ching, "Integration of complementary Acoustic Features for Speaker Recognition," IEEE Signal Processing Letters, vol. 14, no.3, March 1997.

A Fuzzy Decision Tree to Estimate Development Effort for Web Applications

Ali Idri
Department of Software Engineering
ENSIAS, Mohammed Vth –Souissi University
BP. 713, Madinat Al Irfane, Rabat, Morocco

Sanaa Elyassami
Department of Software Engineering
ENSIAS, Mohammed Vth –Souissi University
BP. 713, Madinat Al Irfane, Rabat, Morocco

***Abstract*— Web Effort Estimation is a process of predicting the efforts and cost in terms of money, schedule and staff for any software project system. Many estimation models have been proposed over the last three decades and it is believed that it is a must for the purpose of: Budgeting, risk analysis, project planning and control, and project improvement investment analysis. In this paper, we investigate the use of Fuzzy ID3 decision tree for software cost estimation, it is designed by integrating the principles of ID3 decision tree and the fuzzy set-theoretic concepts, enabling the model to handle uncertain and imprecise data when describing the software projects, which can improve greatly the accuracy of obtained estimates. MMRE and Pred are used, as measures of prediction accuracy, for this study. A series of experiments is reported using Tukutuku software projects dataset. The results are compared with those produced by three crisp versions of decision trees: ID3, C4.5 and CART.**

Keywords- Fuzzy Logic; Effort Estimation; Decision Tree; Fuzzy ID3; Software project.

I. INTRODUCTION

Estimation software project development effort remains a complex problem, and one which continues to attract considerable research attention. Improving the accuracy of the effort estimation models available to project managers would facilitate more effective control of time and budgets during software project development. Unfortunately, many software development estimates are quite inaccurate. Molokken and Jorgensen report in recent review of estimation studies that software projects expend on average 30-40% more effort than is estimated [13]. In order to make accurate estimates and avoid gross misestimations, several cost estimation techniques have been developed. These techniques may be grouped into two major categories: parametric models, which are derived from the statistical or numerical analysis of historical projects data [5], and non-parametric models, which are based on a set of artificial intelligence techniques such as artificial neural networks [9][4], case based reasoning [19], decision trees [20] and fuzzy logic [23][17]. In this paper, we are concerned with cost estimation models based on fuzzy decision trees especially Fuzzy Interactive Dichotomizer 3.

The decision tree method is widely used for inductive learning and has been demonstrating its superiority in terms of predictive accuracy in many fields [24][10]. The most widely used algorithms for building a decision tree are ID3 [11], C4.5 [12] and CART [14].

There are three major advantages when using estimation by decision trees (DT). First, decision trees approach may be considered as "white boxes", it is simple to understand and easy to explain its process to the users, contrary to other learning methods. Second, it allows the learning from previous situations and outcomes. The learning criterion is very important for cost estimation models because software development technology is supposed to be continuously evolving. Third, it may be used to feature subset selection to avoid the problem of cost driver selection in software cost estimation model.

On the other hand, fuzzy logic has been used in software effort estimation. It's based on fuzzy set theory, which was introduced by Zadeh in 1965 [15]. Attempts have been made to rehabilitate some of the existing models in order to handle uncertainties and imprecision problems. Idri et al. [3] investigated the application of fuzzy logic to the cost drivers of intermediate COCOMO model while Pedrycz et al. [25] presented a fuzzy set approach to effort estimation of software projects.

In two earlier works [1][2] we have empirically evaluated the use of crisp decision tree techniques for software cost estimation. More especially, the two used crisp decision tree techniques are the ID3 and the C4.5 algorithms. The two studies are based on the COCOMO' 81 and a web hypermedia dataset. We have found that the decision tree designed with the ID3 algorithm performs better, in terms of cost estimates accuracy, than the decision tree designed with C4.5 algorithm for the two datasets.

The aim of this study is to evaluate and to discuss the use of fuzzy decision trees, especially the fuzzy ID3 algorithm in designing DT for software cost estimation.

Instead of crisp DT, fuzzy DT may allow to exploit complementary advantages of fuzzy logic theory which is the ability to deal with inexact and uncertain information when describing the software projects.

The remainder of this paper is organised as follows: In section II, we present the fuzzy ID3 decision tree for software cost estimation. The description of dataset used to perform the empirical studies and the evaluation criteria adopted to measure the predictive accuracy of the designed models are given in section III. Section IV focuses on the experimental design. In Section V, we present and discuss the obtained results when the

fuzzy ID3 is used to estimate the software development effort. A comparison of the estimation results produced by means of the fuzzy ID3 model and three other crisp decision tree models is also provided in section V. A conclusion and an overview of future work conclude this paper.

II. Fuzzy ID3 For Software Cost Estimation

Based on the Concept Learning System algorithm, Quinlan proposed a decision tree called the Interactive Dichotomizer 3 (ID3). The ID3 technique is based on information theory and attempts to minimize the expected number of comparisons. The fuzzy ID3 is based on a fuzzy implementation of the ID3 algorithm [16][21]. It's formed of one root node, which is the tree top, or starting point, and a series of other nodes. Terminal nodes are leaves (effort). Each node corresponds to a split on the values of one input variable (cost drivers). This variable is chosen in order to reach a maximum of homogeneity amongst the examples that belong to the node, relatively to the output variable.

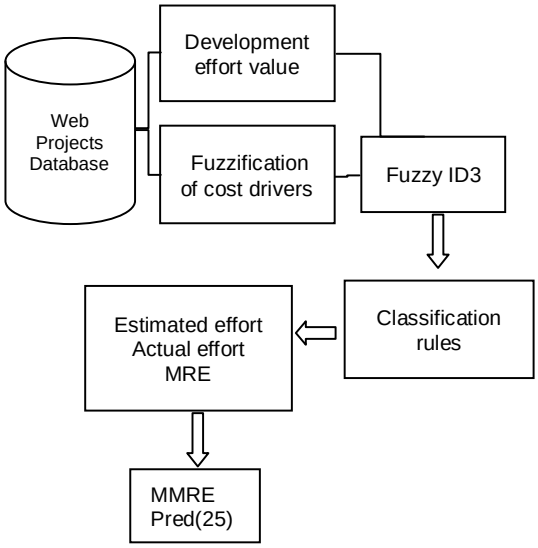


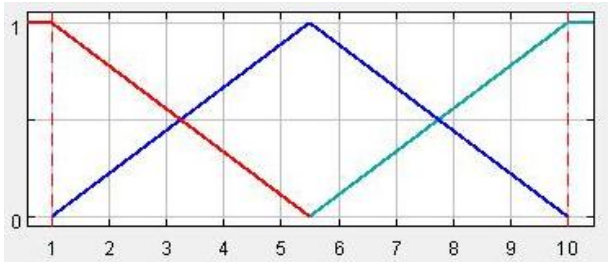
Figure 1. Fuzzy decision tree induction process

Fig. 1 illustrates the fuzzy decision tree induction process that consists on the fuzzification of the web cost drivers, the construction of the fuzzy decision tree, the prediction with the classification rules and the measure the accuracy of the estimates generated by the fuzzy ID3 decision tree.

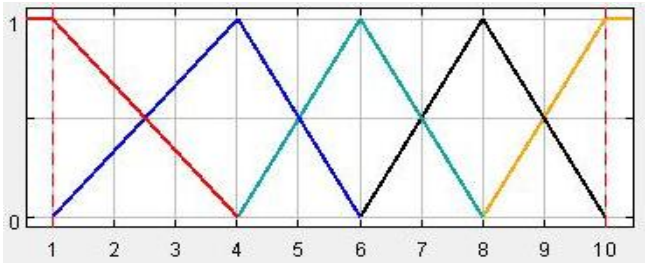
The fuzzification of the software cost drivers converts crisp cost drivers into membership degrees to the different fuzzy sets of the partition. Many algorithms can be found in the specialized literature for generating partitions from data, we chose the Hierarchical Fuzzy Partitioning (HFP) [22]. It corresponds to an ascending procedure. At each step, for each given variable, two fuzzy sets are merged. This method combines two different clustering techniques, hierarchical clustering and fuzzy clustering techniques.

The triangular membership functions are used to represent the fuzzy sets because of its simplicity, easy comprehension, and computational efficiency.

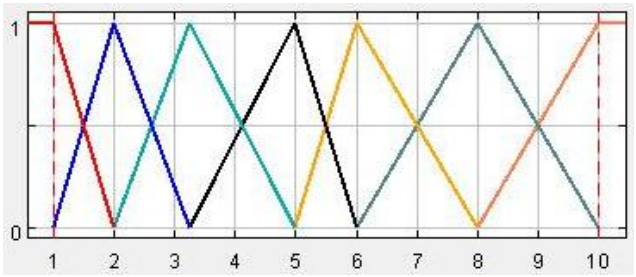
Figure 2 illustrates the membership functions associated to the fuzzy sets of the team experience attribute.



(a) Membership function of 3 fuzzy sets defined for the teamExp cost driver



(b) Membership function of 5 fuzzy sets defined for the teamExp cost driver



(c) Membership function of 7 fuzzy sets defined for the teamExp cost driver

Figure 2. Membership functions associated to the fuzzy sets of the teamExp attribute

The fuzzy decision tree is interpreted by rules, Each path of the branches from root to leaf can be converted into a rule with condition part represents the attributes on the passing branches from root to the leaf and the conclusion part represents the class at the leaf of the form: IF (condition 1 and condition 2 .. and condition n) THEN C, where the conditions are extracted from the nodes and C is the leaf.

Fig. 3 illustrates an example of fuzzy ID3 decision tree for software development effort where MF represents the membership function used to define fuzzy sets for each cost driver.

III. Data Description and Evaluation Criteria

This section describes the dataset used to perform this empirical study and the evaluation criteria adopted to measure the estimates accuracy of the designed software cost estimation model based on fuzzy ID3 method.

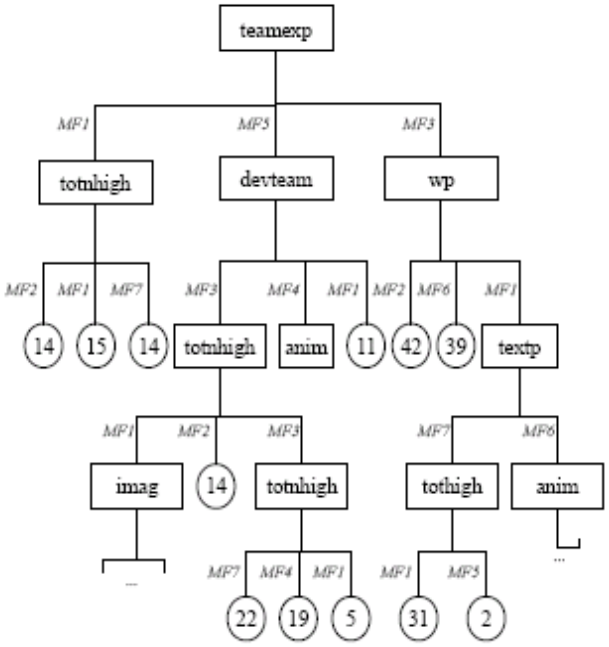


Figure 3. An example of fuzzy ID3 decision tree for software development effort

A. Data Descriptions

The Tukutuku dataset contains 53 web projects. [7] Each web application is described using 9 numerical attributes such as: the number of html or shtml files used, the number of media files and team experience (see Table I). However, each project volunteered to the Tukutuku database was initially characterized using more than 9 software attributes, but some of them were grouped together. For example, we grouped together the following three attributes: the number of new Web pages developed by the team, the number of Web pages provided by the customer and the number of Web pages developed by a third party (outsourced) in one attribute reflecting the total number of Web pages in the application (TotWP).

TABLE I. SOFTWARE ATTRIBUTES FOR THE TUKUTUKU DATASET

Attributes	*Description*
TeamExp	Average team experience with the development language(s) employed
DevTeam	Size of development team
TotWP	Total number of web pages
TextPages	Number text pages typed (~600 words)
TotImg	Total number of images
Anim	Number of animations
AV	Number of audio/video files
TotHigh	Total Number of high effort features/functions
TotNHigh	Total Number of low effort features/functions

B. Evaluation criteria

We employ the following criteria to measure the accuracy of the estimates generated by the fuzzy ID3. A common criterion for the evaluation of effort estimation models is the magnitude of relative error (MRE), which is defined as

$$MRE = \left| \frac{Effort_{actual} - Effort_{estimated}}{Effort_{actual}} \right| \quad (1)$$

where $Effort_{actual}$ is the actual effort of a project in the dataset, and $Effort_{estimated}$ is the estimated effort that was obtained using a model or a technique.

The *MRE* values are calculated for each project in the datasets, while mean magnitude of relative error (MMRE) computes the average over *N* projects.

$$MMRE = \frac{1}{N} \sum_{i=1}^{N} \left| \frac{Effort_{actual,i} - Effort_{estimated,i}}{Effort_{actual,i}} \right| \times 100 \quad (2)$$

The acceptable target values for *MMRE* are $MMRE \leq 25$. This indicates that on the average, the accuracy of the established estimation model would be less than 25%.

Another widely used criterion is the prediction *Pred(p)* witch represents the percentage of *MRE* that is less than or equal to the value p among all projects. This measure is often used in the literature and is the proportion of the projects for a given level accuracy [18]. The definition of *Pred(p)* is given as follows:

$$Pred(p) = \frac{k}{N} \quad (3)$$

Where *N* is the total number of observations and *k* is the number of observations whose *MRE* is less or equal to *p*. A common value for *p* is 25, witch also used in the present study. The prediction at 25%, *Pred(25)*, represents the percentage of projects whose *MRE* is less or equal to 25%. The acceptable values for *Pred(25)* are $\Pr ed(25) \geq 75$.

IV. EXPERIMENT DESIGN

This section describes the experiment design of the fuzzy ID3 decision tree on the Tukutuku dataset. The Hierarchical Fuzzy Partitioning method is chosen for generating the partitions.

The use of fuzzy ID3 to estimate software development effort requires the determination of the parameters, namely the number of input variables, the maximum number of fuzzy sets for each input variable and the significant level value. The last two parameters play an essential role in the generation of fuzzy decision trees. It greatly affects the calculation of fuzzy entropy and classification results of Fuzzy Decision trees.

The number of input variables is the number of the attributes describing the historical software projects in the used

dataset. Therefore, when applying fuzzy ID3 to Tukutuku dataset, the number of input variables is equal to 9. Concerning the significant level parameter, is the membership degree for an example to be considered as belonging to the node, is fixed to 0.2 for all experiments.

In the present paper we are interested in studying the impact of the number of fuzzy sets on the accuracy of fuzzy ID3. A series of experiments is conducted with the fuzzy ID3 algorithm each time using a different value of the fuzzy sets. The number of fuzzy sets is varied within the interval [3, 9].

V. Overview Of The Experimental Results

This section presents and discusses the results obtained when applying the fuzzy ID3 to the Tukutuku dataset. The calculations were made using Fispro software [8]. We conducted several experiments using different configurations of fuzzy ID3 obtained by varying the number of fuzzy sets. The aim is to determine which configuration improves the estimates.

The results for the different configurations have been compared. Fig. 4 and Fig. 5 show the accuracy of the fuzzy ID3 model, measured in terms of MMRE and Pred, on Tukutuku dataset.

Fig. 4 compares the accuracy of the model, in terms of MMRE, when varying the number of the fuzzy sets. We note that the fuzzy ID3 model generates a lower MMRE when increasing the number of fuzzy sets. For example, when setting the number of fuzzy sets at 9 the model produces a prediction error equal to 2.38 (MMRE=2.38) and when setting the number of fuzzy sets at 4 the model produces a prediction error equal to 52.19 (MMRE = 52.19).

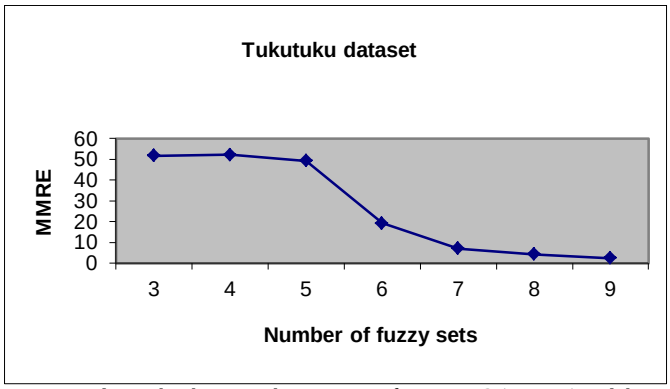


Fig. 4 Relationship between the accuracy of Fuzzy ID3 (MMRE) and the number of fuzzy sets

Fig. 5 shows the results of the model, in terms of Pred(25), when varying the number of the fuzzy sets. From this figure, we note that the accuracy of fuzzy ID3 model performs much better when increasing the number of the fuzzy sets and it's acceptable for the number of fuzzy sets greater than or equal to 6.

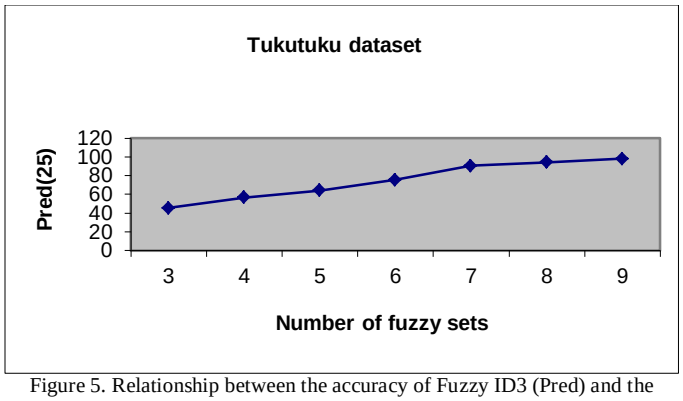


Figure 5. Relationship between the accuracy of Fuzzy ID3 (Pred) and the number of fuzzy sets

Table II summarizes the results obtained using different configurations of fuzzy ID3 for Tukutuku dataset. It shows the variation of the accuracy according to the number of fuzzy sets for Tukutuku dataset.

TABLE II. MMRE AND PRED RESULTS OF DIFFERENT FUZZY ID3 CONFIGURATIONS FOR TUKUTUKU DATASET

Number of fuzzy sets	MMRE	Pred(25)
3	51,68	45,28
4	52,19	56,6
5	49,3	64,15
6	19,27	75,47
7	7,09	90,57
8	4,3	94,34
9	2,38	98,11

The comparisons between the results produced by the fuzzy ID3 decision tree model and three other decision trees models: crisp ID3 decision tree model, C4.5 decision tree model [2] and CART model [6].

The best results obtained by means of the 4 models are compared in terms of MMRE and Pred(25). The comparison result is given in table III.

TABLE III. RESULT OF THE DIFFERENT MODELS USED ON TUKUTUKU DATASET

	Performance Criteria	
Decision tree models	Evaluation MMRE	Pred(25)
Crisp ID3	32	70
C4.5	28	70
CART	25	78
Fuzzy ID3	2,38	98,11

The experimental results show that the fuzzy ID3 model shows better estimation accuracy than the other crisp models in terms of MMRE and Pred(25).

For example, the improvement is 92.56% based on the fuzzy ID3 model MMRE and the crisp ID3 MMRE and is the 90.48% based on the fuzzy ID3 model MMRE and the CART model MMRE.

VI. CONCLUSION

In this paper, we have empirically studied a fuzzy ID3 model for software effort estimation. This fuzzy ID3 model is trained and tested using the tukutuku software projects dataset. The results show that the use of an optimal number of fuzzy sets improves greatly the estimates generated by fuzzy ID3 model. The comparison with the crisp decision tree models shows encouraging results.

To generalize this affirmation, we are looking currently in applying the fuzzy ID3 decision tree model on other historical software projects datasets.

REFERENCES

[1] A. Idri, S. Elyassami, Software Cost Estimation Using Decision Trees, In Proceeding of Sixième Conférence sur les Systèmes Intelligents: Théories et Applications (SITA'10), Rabat, Morocco, 4-5 Mai, 2010. pp. 120-125

[2] A. Idri, S. Elyassami, Web Effort Estimation Using Decision Trees, In Proceeding International Symposium on INnovations in Intelligent SysTems and Applications (INISTA 2010), Kayseri, Turkey, 21-24 Juin, 2010. pp. 224-228.

[3] A. Idri, L. Kjiri, and A. Abran, "COCOMO Cost Model Using Fuzzy Logic", 7th International Conference on Fuzzy Theory & Technology, Atlantic City, NJ, February, 2000. pp. 219-223.

[4] A. Idri, and A. Abran, and S. Mbarki, "An Experiment on the Design of Radial Basis Function Neural Networks for Software Cost Estimation", in 2nd IEEE International Conference on Information and Communication Technologies: from Theory to Applications, 2006, Vol. 1, pp. 230-235.

[5] B.W. Boehm, Software Engineering Economics, Place: Prentice-Hall, 1981.

[6] E. Mendes, "Cost Estimation techniques for web projects", 2008, pp. 203–239..

[7] B.A Kitchenham and E. Mendes, "A Comparison of Cross-company and Within-company Effort Estimation Models for Web Applications", Proceedings of EASE Conference, 2004, pp. 47-56.

[8] Guillaume, S., Charnomordic, B., Lablee, J.-L., 2002. FisPro: Logiciel open source pour les systemes d'inference floue. http://www.inra.fr/bia/M/fispro. INRA-Cemagref.

[9] G. R. Finnie, and G. Witting, and J.-M. Desharnais, "A Comparison of Software Effort Estimation Techniques: Using Function Points with Neural Networks, Case-Based Reasoning and Regression Models", Systems and Software, Vol. 39, No. 3, 1997, pp. 281-289.

[10] H. Berger, D. Merkl, and M. Dittenbach, "Exploiting Partial Decision Trees for Feature Subset Selection in e-Mail Categorization," in Proceedings of the 2006 ACM Symposium on Applied Computing (SAC 2006), Dijon, France, 2006, pp. 1105-1109.

[11] J. R. Quinlan, "Induction on decision tree," Machine Learning, Vol. 1, 1986, pp. 81-106.

[12] J. R. Quinlan, C4.5: Programs for Machine Learning, Morgan Kaufmann Publishers, San Mateo, CA, 1993.

[13] K. Molokken, and M. Jorgensen, "A Review of Surveys on Software Effort Estimation", in International Symposium on Empirical Software Engineering, 2003, pp. 223-231.

[14] L. Breiman, J.H. Friedman, R.A. Olsen & C.J. Stone. Classification and Regression Trees. Wadsworth, 1984.

[15] L. A. Zadeh, "Fuzzy sets", Information and Control, vol 8, 1965, pp. 338–353.

[16] M. Umano, H. Okamoto, I. Hatono, H. Tamura, F. Kawachi, S. Umedzu, J. Kinoshita, "Fuzzy Decision Trees by Fuzzy ID3 algorithm and Its Application to Diagnosis Systems", In Proceedings of the third IEEE Conference on Fuzzy Systems, vol. 3, Orlando, 1994, pp. 2113-2118.

[17] M. W. Nisar, and Y.-J. Wang, and M. Elahi, "Software Development Effort Estimation Using Fuzzy Logic – A Survey", in 5th International Conference on Fuzzy Systems and Knowledge Discovery, 2008, pp. 421-427.

[18] M. Korte and D. Port, "Confidence in Software Cost Estimation Results Based on MMRE and PRED", PROMISE'08, May 12-13, 2008, pp . 63-70.

[19] M. Shepperd and C. Schofield. "Estimating Software Project Effort Using Analogies." Transactions on Software Engineering, vol. 23, no. 12, 1997, pp. 736-747.

[20] R. W. Selby, and A.A. Porter, "Learning from examples: generation and evaluation of decision trees for software resource analysis", IEEE Transactions on Software Engineering, Vol. 14, No. 12, 1988, pp. 1743-1757.

[21] R. Weber, Fuzzy ID3: a class ofmethods for automatic knowledge acquisition, Proceedings ofthe 2nd International Conference on Fuzzy Logic and Neural Networks, Iizuka, Japan, July 17–22, 1992, pp. 265–268.

[22] S. Guillaume and B. Charnomordic, "Generating an interpretable family of fuzzy partitions", IEEE Transactions on Fuzzy Systems, 12 (3), June 2004, pp. 324– 335.

[23] V. Sharma, and H. K. Verma, "Optimized Fuzzy Logic Based Framework for Effort Estimation in Software Development", Computer Science Issues, Vol. 7, Issue 2, No. 2, 2010, pp. 30-38.

[24] W. Pedrycz and Z. A. Sosnowski, "The design of decision trees in the framework of granular data and their application to software quality models," Fuzzy Sets and Systems, Vol. 1234, 2001, pp. 271-290.

[25] W. Pedrycz, J.F. Peters, S. Ramanna, A Fuzzy Set Approach to Cost Estimation of Software Projects, Proceedings of the 1999 IEEE Canadian Conference on Electrical and Computer Engineering Shaw Conference Center , Edmonton Alberta, Canada. 1999, pp. 1068-1073.

AUTHORS PROFILE

A. Idri is a Professor at Computer Science and Systems Analysis School (ENSIAS, Rabat, Morocco). He received DEA (Master) (1994) and Doctorate of 3rd Cycle (1997) degrees in Computer Science, both from the University Mohamed V of Rabat. He has received his Ph.D. (2003) in Cognitive Computer Sciences from ETS, University of Quebec at Montreal. His research interests include software cost estimation, software metrics, fuzzy logic, neural networks, genetic algorithms and information sciences.

S. Elyassami received her engineering degree in Computer Science from the UTBM, Belfort-Montbeliard, France, in 2006. Currently, she is preparing her Ph.D. in computer science in ENSIAS. Her research interests include software cost estimation, soft

An Extended Performance Comparison of Colour to Grey and Back using the Haar, Walsh, and Kekre Wavelet Transforms

Dr. H. B. Kekre
Senior Professor
Computer Engineering Department
Mukesh Patel School of Technology, Management, and Engineering
NMIMS University
Mumbai, India

Dr. Sudeep D. Thepade
Associate Professor
Computer Engineering Department
Mukesh Patel School of Technology, Management, and Engineering
NMIMS University
Mumbai, India

Adib Parkar
Research Assistant
CSRE
Indian Institute of Technology, Bombay
Mumbai, India

***Abstract* – The storage of colour information in a greyscale image is not a new idea. Various techniques have been proposed using different colour spaces including the standard RGB colour space, the YUV colour space, and the YCbCr colour space. This paper extends the results described in [1] and [2]. While [1] describes the storage of colour information in a greyscale image using Haar wavelets, and [2] adds a comparison with Kekre's wavelets, this paper adds a third transform – the Walsh transform and presents a detailed comparison of the performance of all three transforms across the LUV, YCbCr, YCgCb, YIQ, and YUV colour spaces. The main aim remains the same as that in [1] and [2], which is the storage of colour information in a greyscale image known as the "matted" greyscale image.**

Keywords – Colouring; Colour to Grey; Matted Greyscale; Grey to Colour; LUV Colour Space; YCbCr Colour Space; YCgCb Colour Space; YIQ Colour Space; YUV Colour Space; Haar Wavelets; Kekre's Wavelets; Walsh Transform.

I. Introduction

The efficient storage of colour images in digital format has always been a dilemma faced when working with many, high resolution, digital images. With the massive improvements in memory technology these days, from traditional, slower hard disk drives to the newer solid state drives, however, this has become less of a problem due to the increases in available memory. Despite this, digital colour images can take up to three times the storage space of a greyscale image in uncompressed format, and for larger images, this is not a negligible increase. This is why various algorithms have been proposed, and many commercially implemented, that reduce the size of a stored digital image with little or no loss of image information and quality. The JPEG standard is a prime example of such an algorithm.

Another challenge faced when working with digital images today is to find a way to create (or recreate) a colour image from a greyscale image: in essence, to "re-colour" a monochrome image. Unlike the storage problem mentioned above, this one is much harder to solve since it entails the extraction of more information from a source that has less information, i.e. the retrieval of colour (multi-plane) information from a greyscale (single-plane) source. This is, by definition, not directly possible – how can one create information when none exists? Hence, this problem needs a more indirect solution – the missing information needs to be provided either externally (separated from the greyscale image to be coloured), or somehow embedded as "extra" information in the greyscale image itself. Examples of the former can be found in [3], [4], [5], and [6].

The technique described in this paper, on the other hand, is based on the latter principle – hiding of the colour information in the greyscale image itself. In short, the colour information is hidden in the transform domain – this transform is a wavelet transform – of the greyscale image and hence can be extracted from the transformed image. The technique was first described in [1], and this paper is an extension of the technique to more transforms and colour spaces. A similar technique can be found in [7].

An advantage of using this technique as compared to those based on retrieving colour information externally as in [3], [4], [5], and [6] is that this method also indirectly decreases image storage requirements. Having colour information stored in a greyscale image means that the greyscale image is now all that needs to be stored (or transmitted) and from it the colour image can be recreated. Hence, for example, the greyscale image can be transmitted via fax to a recipient and the recipient can recreate the colour image without needing any external information. Thus, multi-plane (colour) data is "compressed" into a single-plane (greyscale).

II. Overview of the Colour Information Embedding Technique

This section provides a simple, high level explanation of the technique used to "hide" the colour information in the greyscale image and is as found in [1].

The general procedure is as follows. First, the original colour image is converted to its greyscale equivalent in some colour space. This image is then transformed using an appropriate image transform into the transform domain. In the transform domain, the transformed image is modified by

embedding the colour components of the image. These components are embedded by replacing the lesser significant portions of the transformed image (high frequency region) by scaled down versions of the colour components. The most significant portion of the transformed image is, of course, left untouched. This modified transformed image is then inverse transformed back into the spatial domain. The greyscale image so obtained is known as the matted greyscale image. Reconstructing the colour image from this matted greyscale image is now straightforward. One simply has to apply the transform, extract the scaled down versions of the colour components, scale them up to their original size, and create a colour image using these scaled up components. Of course, a certain amount of information will be lost due to this scaling of the components.

In this paper, the above technique has been applied using three transforms – the Haar transform, Kekre's Wavalet Transform (KWT), and the Walsh transform. In addition, to provide a comprehensive comparison, the following colour spaces have been considered for all three transforms - LUV, YCbCr, YCgCb, YIQ, and YUV. The subsequent sections provide a quick reference for all of these colour spaces and the transforms used followed by a detailed description of the technique used.

III. YCbCr Colour Space

The YCbCr model defines a colour space in terms of one luminance (brightness) and two chrominance (colour) components. It is one of the most extensively used colour spaces and has been considered for many applications such as those described in [7], [8], and [9]. In the YCbCr colour space, the Y component gives luminance and the Cb and Cr components give the chromaticity values of the colour image. To get the YCbCr components, the conversion of the RGB components to YCbCr components must be known. The RGB to YCbCr conversion matrix is given below.

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 0.2989 & 0.5866 & 0.1145 \\ -0.1688 & -0.3312 & 0.5 \\ 0.5 & -0.4184 & -0.0816 \end{bmatrix} . \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

To get the RGB values from the YCbCr components, the following conversion matrix can be used.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & -0.001 & 1.402 \\ 1 & -0.3441 & -0.714 \\ 1 & 1.7718 & 0.001 \end{bmatrix} . \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} \quad (2)$$

IV. Kekre's LUV Colour Space

The Kekre's LUV colour space is a colour space generally used in techniques involving the colourization of images such as those described in [4], [10], [11], and [12]. In the Kekre's LUV colour space, the L component provides the luminance, while the U and V components contain the colour information. The RGB to Kekre's LUV conversion matrix is given below.

$$\begin{bmatrix} L \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.33333 & 0.33333 & 0.33333 \\ -0.3333 & 0.16667 & 0.16667 \\ 0 & -0.5 & 0.5 \end{bmatrix} . \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (3)$$

To get the RGB values from the Kekre's LUV components, the following conversion matrix can be used.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} . \begin{bmatrix} L \\ U \\ V \end{bmatrix} \quad (4)$$

A negative value for the U component in the Kekre's LUV colour space indicates prominence of the red component in the colour image. Similarly, a negative value for the V component indicates prominence of the green component over the blue component in the colour image.

V. YCgCb Colour Space

The YCgCb colour model [15], [18] is a newly proposed colour space similar to the LUV colour space described in the previous section. Since it is newer than the LUV colour space, it has not yet been used extensively.

In the YCgCb colour space, the Y component provides the luminance, while the Cg and Cb components contain the chromaticity values. The RGB to YCgCb conversion matrix is given below.

$$\begin{bmatrix} Y \\ Cg \\ Cb \end{bmatrix} = \begin{bmatrix} 0.33333 & 0.33333 & 0.33333 \\ 0.33333 & -0.3333 & 0 \\ 0.33333 & 0 & -0.3333 \end{bmatrix} . \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (5)$$

To get the RGB values from the YCgCb components, the following conversion matrix can be used.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} . \begin{bmatrix} Y \\ Cg \\ Cb \end{bmatrix} \quad (6)$$

A negative value for the Cg component in the YCgCb colour space indicates prominence of the green component over the red component in the colour image. Similarly, a negative value for the Cb component indicates prominence of the blue component over the red component.

VI. YIQ Colour Space

YIQ [19] is the colour space used by the NTSC colour TV system, employed mainly in North and Central America. 'I' stands for "in phase" and 'Q' stands for "quadrature," referring to the components used in quadrature amplitude modulation.

As in the YCbCr colour space, the Y component gives luminance and the I and Q components give the chromaticity values of the colour image. To get the YIQ components, the conversion of the RGB components to the YIQ components is defined by the following conversion matrix.

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ 0.596 & -0.275 & -0.321 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} . \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (7)$$

To get the RGB values from the YIQ components, the following conversion matrix can be used.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1 & 0.956 & 0.621 \\ 1 & -0.272 & -0.647 \\ 1 & -1.107 & 1.704 \end{bmatrix} . \begin{bmatrix} Y \\ I \\ Q \end{bmatrix} \quad (8)$$

VII. YUV Colour Space

YUV is a colour space [19] that encodes a colour image or video taking human perception into account, allowing reduced bandwidth for chrominance, thereby typically enabling transmission errors or compression artefacts to be more efficiently masked by human perception than using a direct RGB representation.

Similar to the other colour spaces discussed previously, the Y component gives luminance, and the U and V components provide the chrominance values. The RGB to YUV conversion matrix is shown below.

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.1471 & -0.2889 & 0.436 \\ 0.615 & -0.5149 & 0.10001 \end{bmatrix} . \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (9)$$

To get the RGB values from the YUV components, the following conversion matrix can be used.

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 0.74952 & -0.509 & 1.1398 \\ 1.0836 & -0.2247 & -0.5806 \\ 0.97086 & 1.9729 & 0.00001467 \end{bmatrix} . \begin{bmatrix} Y \\ U \\ V \end{bmatrix} \quad (10)$$

VIII. HAAR Transform

The Haar functions were proposed as a sequence in 1909 by Alfréd Haar [13]. Haar used these functions to give an example of a countable orthonormal system for the space of square-integrable functions on the real line. The study of wavelets, and even the term "wavelet", did not come until much later [14]. The Haar wavelet is also the simplest possible wavelet. As the Haar wavelet is not continuous, it is also not differentiable. This is a technical disadvantage of Haar wavelets.

The Haar wavelet's mother wavelet function ψ(t) can be described as follows.

$$\psi(t) = \begin{cases} 1, & 0 < t \leq \frac{1}{2} \\ -1, & \frac{1}{2} < t \leq 1 \\ 0, & otherwise \end{cases} \quad (11)$$

IX. KEKRE'S Wavelet Transform

The KWT matrix is a generic version of the LUV colour space matrix. Unlike most other transforms (wavelet or otherwise), the size of the KWT matrix need not be a power of two which is definitely an advantage of this transform.

The general form of an N x N KWT matrix [19], [20] is as follows.

$$\begin{bmatrix} 1 & 1 & 1 & \dots & 1 & 1 \\ -N+1 & 1 & 1 & \dots & 1 & 1 \\ 0 & -N+2 & 1 & \dots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & 1 \\ 0 & 0 & 0 & \dots & -N+(N-1) & 1 \end{bmatrix}$$

Figure 1 – General N x N KWT matrix

As can be seen in figure 1 above, in the KWT matrix, all values above the diagonal are one (including the diagonal itself). The diagonal just below the primary diagonal has specific values, and all remaining values in the matrix are zeroes. In general, the value of K_{xy} in the KWT matrix, where 'x' is the row number starting from 1 and 'y' is the column number also starting from one, is given by the following equation.

$$K_{xy} = \begin{cases} 1, & x \leq y \\ -N+(x-1), & x = y+1 \\ 0, & otherwise \end{cases} \quad (12)$$

The inverse KWT matrix is nothing but the transpose of the above general matrix as the KWT is orthogonal. Normalization is required to get back the identity matrix.

X. WALSH Transform

The Walsh transform was first defined in 1923 by Walsh [16], although in 1893 Hadamard [17] had achieved a similar result by the application of certain orthogonal matrices, generally called Hadamard matrices, which contain only the entries +1 and -1.

The image algebra formulation of the fast Walsh transform is identical to that of the fast Fourier formulation, with the exception that the template 't' used for the Walsh transform is as follows.

$$t(p)_{(u,v)}(x,y) = \begin{cases} 1, & \lfloor u/p \rfloor \; is \; even \; and \; (x,y) = (u,v) \\ 1, & \lfloor u/p \rfloor \; is \; even \; and \; (x,y) = (u+p,v) \\ -1, & \lfloor u/p \rfloor \; is \; odd \; and \; (x,y) = (u,v) \\ 1, & \lfloor u/p \rfloor \; is \; odd \; and \; (x,y) = (u-p,v) \\ 0, & otherwise \end{cases} \quad (13)$$

The Walsh transform shares the important property of separability with the Fourier transform. Thus, the two dimensional Walsh transform can also be computed by taking the one-dimensional Walsh transforms along each row of the image, followed by another one-dimensional Walsh transform along the columns.

XI. Colour to Matted Grey Conversion

The most trivial way to convert a colour image to greyscale for printing is to retain and use the luminance component of the colour image.

The problem with this approach is that regions that have contrasting colours with similar luminance components would be assigned the same output luminance level and would, therefore, look the same.

The other option is to map colours to textures [7]. One can control halftone dots or patterns as a function of the colours, for example, as a function of hue and saturation. Hence, regions of different colours with similar luminance will look different after mapping because they would have different textures [3]. The procedure proposed in [7] produces a continuum of textures using the DWT that naturally switch between patterns without causing visual artefacts. The DWT decomposes an image into several sub-bands [14] each representing different spatial frequency contents.

In this paper, as explained previously, is an extension of [1] and [2], and is a more detailed study of a technique based on the one described in [7]. Here, the wavelet transforms used are the basic Haar transform, the KWT, and the Walsh transform. Five different colour spaces are considered – LUV, YCbCr, YCgCb, YIQ, and YUV. The greyscale image obtained by following the procedure outlined below is known as the matted greyscale image and it contains the colour information about the image embedded within its transform.

As various colour spaces are used, the procedure is outlined for a general colour space 'ABC' where 'A' is the luminance component, and 'B' and 'C' are the chromaticity components. As two different transforms are used, even the transforms are generalized as just 'the transform.' The following procedure has been taken from [1].

The steps involved to create the matted greyscale image are as follows.

1. The original colour image is converted from the RGB colour space into the ABC colour space using the appropriate conversion matrix, that is, using equations (1), (3), (5), (7), or (9) for the YCbCr, LUV, YCgCb, YIQ, and YUV colour spaces respectively. The 'A' component is the luminance and is considered as the "original greyscale" image.
2. The 'A' component is transformed using the transform into the transform domain. Let this transformed image be known as T_o.
3. T_o can be divided into 4 regions as shown in figure 2 below.

LL	LH
HL	HH

Figure 2 – Structure of T_o

4. Most of the information in the image is found in the LL region of T_o which corresponds to the low frequency components of the image. This region is left untouched. The LH and HL regions of T_o are replaced by scaled down versions of the 'B' and 'C' components respectively. Thus we now have a modified transformed image that contains scaled down versions of the chromaticity components of the colour space being used. Let this be known as T_m.
5. The inverse transform is now applied to T_m to get a new greyscale image in the spatial domain. This greyscale image now contains colour information hidden within its transform and is known as the matted greyscale image.

XII. Colour Extraction From Matted Grey

Since the matted greyscale image already contains the colour information hidden within its transform, extracting the colour image from the matted greyscale image is a straightforward procedure. It consists of the following steps.

1. The matted greyscale image is read or scanned.
2. Once available in digital form, the matted greyscale image is transformed using the transform into the transform domain. The transformed image obtained will be T_m.
3. T_m can be represented as shown in figure 3 below.

LL	B'
C'	HH

Figure 3 – Structure of T_m

4. The B' and C' regions of T_m are the scaled down versions of the 'B' and 'C' components of the original colour image. Thus by extracting these two regions from T_m and scaling them up back to their original size we get B_{approx} and C_{approx} which are approximations of the original 'B' and 'C' components of the original colour image.
5. To retrieve an approximation for the 'A' component, we replace regions B' and C' in T_m by zeroes and perform an inverse transformation. The image obtained in the spatial domain is an approximation of the original 'A' component of the image, A_{approx}.

Now the approximations for the ABC components are used to convert the image back to the RGB colour space using the appropriate conversion matrices, that is, equations (2), (4), (6), (8), and (10) for the YCbCr, LUV, YCgCb, YIQ, and YUV colour spaces respectively.

XIII. Implementation and Results

The implementation of the technique described in the previous section was an extension of the implementation described in [1] and [2]. The technique was broadened to include the Walsh transform and the results obtained using the Walsh transform were also added.

The technique was applied to a large number of images. All the images were of the size 256 x 256 pixels, and belonged to various categories such as people, objects, vehicles, animals, cartoons, and nature.

On applying the technique proposed, the following set of images was derived from the each original image – original colour, original greyscale, matted greyscale, and reconstructed colour.

Performance was measured by calculating the mean square error (MSE) between the original greyscale image and the matted greyscale image as well as between the original colour image and the reconstructed colour image. The MSE is the mean of the square of the Euclidean distances between each pixel value. The Euclidean distance 'd' between two images I_m and I_o is defined as follows.

$$d = \sqrt{\sum (z_m - z_o)^2} \qquad (14)$$

Where

z_m is the value of the pixel in I_m,

z_o is the value of the corresponding pixel in I_o,

and the summation is over all pixels in the images.

It must be noted that for a colour image, each pixel will have 3 values, one for the red (R) plane, one for the green (G) plane, and one for the blue (B) plane.

The MSE thus provides an objective criterion that can be used as a measure of the degree of similarity between two given images. The greater the similarity between the two images at the pixel level, the lower the MSE.

Table 1 shows the MSE values between the matted greyscale images and the original greyscale images averaged across all images for each of the five colour spaces and all transforms.

TABLE 1 – GREYSCALE MSE ACROSS ALL IMAGES (MSE BETWEEN ORIGINAL GREYSCALE AND MATTED GREYSCALE IMAGE)

	MSE (Haar)	MSE (KWT)	MSE (Walsh)
YCbCr	411.221335	401.618715	230.4947
LUV	341.32406	335.38614	207.42726
YCgCb	317.00611	309.915005	204.9934
YIQ	432.70043	423.13666	236.232875
YUV	643.38025	603.0701	237.407555

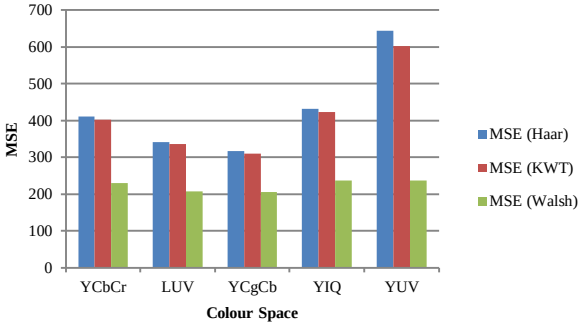


Figure 4 – Greyscale MSE across all images

Figure 4 illustrates the results graphically. As can be clearly seen from figure 4, the YCgCb colour space provided the best results in terms of the MSE across all images for the matted greyscale image. Also, the KWT consistently outperformed the Haar transform across all colour spaces, but not by an extremely significant amount. On the other hand, the Walsh transform performed spectacularly well and thoroughly defeated the other two transforms across all colour spaces.

The results for the reconstructed colour images, however, were incredibly different. The table below shows the MSE values between the reconstructed colour images and the original colour images.

TABLE 2 – COLOUR MSE ACROSS ALL IMAGES (MSE BETWEEN ORIGINAL COLOUR AND RECONSTRUCTED COLOUR IMAGE)

	MSE (Haar)	MSE (KWT)	MSE (Walsh)
YCbCr	245.36608	268.69903	804.9029
LUV	233.00122	249.32088	724.70764
YCgCb	245.31157	263.90563	684.758285
YIQ	249.90365	265.723	811.5984
YUV	417.922255	498.013915	1554.5535

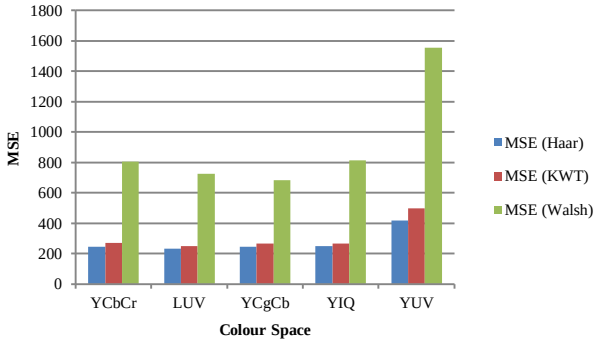


Figure 5 – Colour MSE across all images

Figure 5 illustrates the results graphically. It can be noticed that in the case of the reconstructed colour images, there is no clear winner across four colour spaces when the MSE is averaged across all images: YCbCr, LUV, YCgCb, and YIQ. The YUV colour space clearly underperforms using this technique for both greyscale and colour images and hence is not recommended. The colour space that performs the best is the LUV colour space, but not by a large amount. The YCgCb colour space, the YIQ colour space, and the YCbCr colour space give almost identical performance results based on the MSE, and in the case of the Walsh transform, the YCgCb space performs the best.

However, when we study the results with respect to the transforms used, the Walsh transform is literally blown away and produces extremely poor results. Across all colour spaces, the colour image obtained from the matted greyscale image when using the Walsh transform is consistently and significantly poorer than those obtained by the Haar transform

and the KWT. For the reconstructed colour images, the Haar transform performs slightly better than the KWT.

Through these results we can hazard a hypothesis that the better a transform performs when creating a matted greyscale image, the worse its resultant reconstructed colour image is. Ideally, we would like a transform that would minimize both the greyscale and colour MSEs – none of these appear to have that property.

Some of the images used when implementing this procedure are shown below.

6a – Original Colour

6b – Original Greyscale

6c – Matted Greyscale (Haar)

6d – Matted Greyscale (KWT)

6e – Matted Greyscale (Walsh)

6f – Reconstructed Colour (Haar)

6g – Reconstructed Colour (KWT)

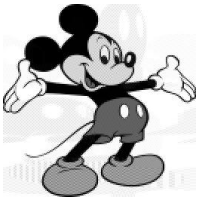
6h – Reconstructed Colour (Walsh)

Figure 6 – Colour to grey and back of Book image

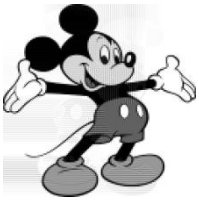
In figure 6, we see a simple image of a book. While the cover of the book is textured, on the whole there are not too many colours in this picture. Even so, one can clearly see the reconstructed colour image when using the Walsh transform is nowhere near as good as those using the Haar and KWT transform.

7a – Original Colour

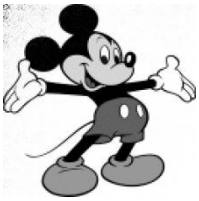
7b – Original Greyscale



7c – Matted Greyscale (Haar)



7d – Matted Greyscale (KWT)



7e – Matted Greyscale (Walsh)

7f – Reconstructed Colour (Haar)

7g – Reconstructed Colour (KWT)

7h – Reconstructed Colour (Walsh)

Figure 7 – Colour to grey and back of Cartoon image

Figure 7 shows a cartoon image. Colours in the original are clear cut with prominent boundaries and also have deep saturation. For this image, the Haar transform appears to create a good matted greyscale image (figure 7c), better than those by the KWT and the Walsh transform (figures 7d and 7e). The Walsh transform's matted greyscale image displays a discolouration at the top left corner while the KWT shows discolouration at near the bottom left.

These spots of discolouration persist in the reconstructed colour images seen in figures 7g and 7h for the KWT and the Walsh transform respectively. The Haar transform does not perform too well in reconstructing the colour image either, but once again, the Walsh transform clearly performs the worst.

This image leads to a possibility that this technique does not work very well for "cartoon-like" images – images with clearly defined stretches of single colours – especially reds and yellows.

8a – Original Colour

8b – Original Greyscale

8c – Matted Greyscale (Haar)

8d – Matted Greyscale (KWT)

8e – Matted Greyscale (Walsh)

8f – Reconstructed Colour (Haar)

8g – Reconstructed Colour (KWT)

8h – Reconstructed Colour (Walsh)

Figure 8 – Colour to grey and back of Deer image

Next, figure 8 depicts a more complex image than those in figures 6 and 7. There is a clear background and foreground; however, both the background and foreground are textured. For this image, the technique works much better when using the Haar and KWT transforms. When considering only the matted greyscale images, all three transforms perform adequately well; when also taking into consideration the reconstructed colour image, as has been the norm, the Walsh transform's reconstructed image is significantly inferior to those using the Haar transform and the KWT.

XIV. Conclusion

After applying the technique to embed colour information in a greyscale image using various colour spaces and three transforms and studying the results obtained, it can be safely concluded that the Walsh transform should not be considered as an appropriate transform to implement this algorithm. While its results when creating a matted greyscale image are acceptable, its deplorable results on reconstructing the colour image take it completely out of consideration.

The choice between the Haar transform and the KWT is a harder one to make. Not only do neither of them perform spectacularly well across all images, but also neither consistently outperforms the other. The MSE measure itself can be misleading at times as seen in the example images in the previous section (specifically figure 7) – while the KWT generally has a lower MSE for the matted greyscale image it does not guarantee the image will be perceived as better than the one created using the Haar transform. Moving on to the reconstructed colour images, the Haar transform slightly edges out the KWT in most cases, but the results are very close for this to be deemed significant.

Finally, when considering colour spaces, except for the YUV colour space, all the other colour spaces considered seem to be as good as each other. In general, the YCgCb colour space slightly outperforms the others.

Hence, in conclusion, it may be stated that while there is no "perfect" combination of colour space and image transform out of those considered here that outperforms all others when applying this procedure to hide colour information in a greyscale image, it can be confidently deduced that the Walsh transform and the YUV colour space must not be used.

References

[1] H. B. Kekre, Sudeep D. Thepade, Adib Parkar, "Storage of Colour Information in a Greyscale Image using Haar Wavelets and Various Colour Spaces", International Journal of Computer Applications (IJCA) September 2010.

[2] H.B. Kekre, Sudeep D. Thepade, Adib Parkar, "A Comparison of Haar Wavelets and Kekre's Wavelets for Storing Colour Information in a Greyscale Image", 2010.

[3] E. Reinhard, M. Ashikhmin, B. Gooch, and P. Shirley, "Color transfer between images," IEEE Computer graphics and applications, vol. 21, no. 5, pp. 34–41, September/October 2001.

[4] H. B. Kekre, Sudeep D. Thepade, Archana Athawale, Adib Parkar, "Using Assorted Color Spaces and Pixel Window Sizes for Colorization of Grayscale Images", ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010), Thakur College of Engg. And Tech., Mumbai, 26-27 Feb 2010.

[5] H. B. Kekre, Sudeep D. Thepade, Adib Parkar, "A Comparison of Kekre's Fast Search and Exhaustive Search for various Grid Sizes used for Colouring a Greyscale Image", 2nd International Conference on Signal Acquisition and Processing (ICSAP 2010), IACSIT, Bangalore, pp. 53-57, 9-10 Feb 2010.

[6] H. B. Kekre, Sudeep D. Thepade, Adib Parkar, "Performance Analysis of Kekre's Median Fast Search, Kekre's Centroid Fast Search and Exhaustive Search Used for Colouring a Greyscale Image", International Journal of Computer Theory and Engineering, Vol. 2, No. 4, August, 2010 1793-8201.

[7] Ricardo L. de Queiroz, Karen M. Braun, "Color to Gray and Back: Color Embedding into Textured Gray Images", 1464 IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 6, JUNE 2006.

[8] Son Lam Fung, A. Bouzerdoum, D. Chai, "A novel skin color model in YCbCr color space and its application to human face detection", In Proc. of International Conference on Image Processing (ICIP-2002), Vol.1, pp. I289-I292.

[9] Hideki Noda, Michiharu Niimi, "Colorization in YCbCr color space and its application to JPEG images", Pattern Recognition Society Published by Elsevier B.V., Vol.40, number 12, pp.3714-3720, December, 2007.

[10] H. B. Kekre, Sudeep D. Thepade, "Image Blending in Vista Creation using Kekre's LUV Color Space", In SPIT-IEEE Colloquium, SPIT Mumbai, INDIA, Feb 4-5,2008.

[11] H. B. Kekre, Sudeep D. Thepade, "Improving 'Color to Gray and Back' using Kekre's LUV Color Space", IEEE International Advanced Computing Conference 2009 (IACC '09), Thapar University, Patiala, INDIA, 6-7 March 2009.

[12] H. B. Kekre, Sudeep D. Thepade, "Boosting Block Truncation Coding using Kekre's LUV Color Space for Image Retrieval", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, No.3, Summer 2008.

[13] Haar, Alfred, "Zur Theorie der orthogonalen Funktionen systeme". (German), Mathematische Annalen, volume 69, No. 3, 1910, pp. 331–371.

[14] Charles K. Chui, "An Introduction to Wavelets", Academic Press, 1992, San Diego, ISBN 0585470901.

[15] H. B. Kekre, Sudeep D. Thepade, Nikita Bhandari, " Colorization of Greyscale Images using Kekre's Biorthogonal Color Spaces and Kekre's Fast Codebook Generation", Advances in Multimedia-An International Journal (AMIJ), Volume I, Issue 3, 2011.

[16] J. Walsh, "A closed set of normal orthogonal functions," American Journal of Mathematics, vol. 45,no. 1, pp. 5-24, 1923.

[17] M. J. Hadamard, "Resolution d'une question relative aux determinants," Bulletin des Sciences Mathematiques, vol. A17, pp. 240-246, 1893

[18] Sudeep Thepade, Ph.D. Thesis, "New Approached of Feature Vector Extraction for Content Based Image Retrieval", pp. C3-24 to C3-27, Supervisor Dr.H.B.Kekre, MPSTME, SVKM's NMIMS (deemed to be University), Mumbai, 2011.

[19] H. B. Kekre, Sudeep D. Thepade, Avni Agrawal, Naman Agrawal, "Performance Comparison of IRIS Recognition Techniques using Wavelet Pyramids of Walsh, Haar and Kekre Wavelet Transforms", International Journal of Computer Applications (IJCA), Number 2, Article 4, March 2011.

AUTHORS PROFILE

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg. from Jabalpur University in 1958, M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S.Engg. (Electrical Engg.) from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. at Thadomal Shahani Engineering College, Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. He has guided 17 Ph.D.s, 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing and Image Processing. He has more than 350 papers in National / International Conferences / Journals to his credit. Recently nine students working under his guidance have received best paper awards. Currently he is guiding ten Ph.D. students.

Dr. Sudeep D. Thepade has Received B.E.(Computer) degree from North Maharashtra University with Distinction in 2003, M.E. in Computer Engineering from University of Mumbai in 2008 with Distinction, Ph.D. from SVKM's NMIMS (Deemed to be University) in July 2011, Mumbai. He has more than 08 years of experience in teaching and industry. He was Lecturer in Dept. of Information Technology at Thadomal Shahani Engineering College, Bandra(W), Mumbai for nearly 04 years. Currently working as Associate Professor in Computer Engineering at Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS (Deemed to be University), Vile Parle(W), Mumbai, INDIA. He is member of International Advisory Committee for many International Conferences. His areas of interest are Image Processing and Biometric Identification. He has guided two M.Tech. projects and several B.Tech projects. He more than 115 papers in National/International Conferences/Journals to his credit with a Best Paper Award at International Conference SSPCCIN-2008, Second Best Paper Award at ThinkQuest-2009 National Level paper presentation competition for faculty, Best Paper Award for paper published in June 2011 issue of IJCSIS (USA), Editor's Choice Awards for IJCA (USA) in 2010 and 2011.

Adib Parkar is currently a research assistant at the Indian Institute of Technology (IIT), Bombay, India. He has received a Bachelors (B.E.) degree in Computer Science with Distinction from Mumbai University, India in 2010. He has also worked for Accenture as a software developer from 2010 - 2011. He has been an active IEEE Student Member for 4 years and was also a member of the Computer Society of India. His areas of interest lie in the fields of Image Processing and Artificial Intelligence.

A Prototype Student Advising Expert System Supported with an Object-Oriented Database

M. Ayman Al Ahmar
Deputy Dean, College of Information Technology
Ajman University of Science and Technology (AUST)
United Arab Emirates (UAE)

***Abstract*— Using intelligent computer systems technology to support the academic advising process offers many advantages over the traditional student advising. The objective of this research is to develop a prototype student advising expert system that assists the students of Information Systems (IS) major in selecting their courses for each semester towards the academic degree. The system can also be used by academic advisors in their academic planning for students. The expert system is capable of advising students using prescriptive advising model and developmental advising model. The system is supported with an object-oriented database and provides a friendly graphical user interface. Academic advising cases tested using the system showed high matching (93%) between the automated advising provided by the expert system and the advising performed by human advisors. This proves that the developed prototype expert system is successful and promising.**

Keywords- academic advising; expert system; object-oriented database.

I. INTRODUCTION

Student academic advising is an essential task in educational institutions. Traditionally a university student plans the courses semester-by-semester towards a degree through lengthy meetings with the human academic advisor. Advising meetings are usually held during the beginning of each academic semester. Since student advising is a time-consuming effort, there is a need for computerization of some parts of the advising process. Utilizing a computerized advising system, students can save the software consultation results and can then meet with the human advisor for further consultation (if there is still a need for the traditional face-to-face meeting). This hopefully will save valuable time for academic advisors and for students.

The objective of this research is to develop a prototype rule based Expert System (ES) for the academic advising of the students of the Information Systems (IS) Department, College of Information Technology, Ajman University of Science and Technology (AUST), UAE. The system is called "IS-Advisor" and helps students in course selection for each academic semester.

Literature reveals many endeavors in the field of automating academic advising activities including the application of expert systems [1-9]. There can never be a 'global' expert student advising system applicable to all academic institutions and departments because of the existence of academic regulations and expert advising knowledge and reasoning specific to each academic unit.

As an example to illustrate this concept, AUST regulations allow each student to register from three to six courses per semester. However, the accumulated advising experience and grade statistics related to the IS department show that students who can be advised to register six courses without difficulty are students with AGPA 3.00 or above (out of 4.00), whereas students with AGPA greater than 2.00 and less than 2.25 are better advised to register 4 courses only the next semester to give them a chance to increase their AGPA.

The proposed ES (IS-Advisor) represents such specific advising knowledge and reasoning as rules in its knowledge base component and reasoning strategies in its inference engine component. Thus, the ES developed in this research is unique in its specific knowledge base and reasoning strategies and is intended to be of great help to the department of IS. Another particular feature of IS-Advisor is the Object-Oriented (OO) architecture of its database as will be addressed in subsequent sections.

II. MODELS OF ACADEMIC ADVISING

From the literature we select two models of academic advising adopted in the proposed expert system: Prescriptive advising model and developmental advising model. The prescriptive advising model is characterized by an advisor-student relationship in which students follow the prescriptive procedure of their advisors without assuming responsibility for decision making [10]. The developmental advising models rely on a shared responsibility between the student and the advisor in which the advisor directs the student to proper resources [11].

Literature studies show findings that support both models. For example, Fielstein L. in the research paper titled "Developmental versus prescriptive advising: Must it be one or the other?" stated that: "…intuitive students appeared to endorse the developmental approach to advising. On the other hand, the more 'thinking' students did not value a collaborative relationship and seemed more content with the criteria associated with prescriptive advising" [12]. In general advisors need to look at each student as an individual with individual characteristics.

III. EXPERT SYSTEMS

"Expert Systems (ES) are programs that attempt to emulate the behavior of human experts, usually confined to a specific field" [13]. Regarding the domain of academic advising, ES technology seems to be the most successful method of computerization because the dialogue between human advisor and the student can be conveniently emulated by the dialogue between the ES and the student, and the reasoning of the academic advisor can be successfully automated by the reasoning power of ES; particularly the rule-based ES. A rule based ES captures human knowledge using If-Then rules in a rule-based knowledge base. Academic advising process can be successfully modeled in computers as a rule-based expert system since most advising regulations are based on academic 'rules' such as "if you pass course A, then you can register course B" and so on. The proposed system in this research (IS-Advisor) is modeled as an ES with an object-oriented database, thus the main components of IS-Advisor are: The OO database, the rule-based knowledge base, the inference engine, and the user interface. The ES can explain its results by tracing the If-Then rules used to reach the conclusions through a component called the explanation subsystem. The following sections explain the details of the developed ES by explaining its components. The system is developed using Kappa-PC expert system shell [14]. Kappa-PC supports OO modeling which is adopted in this system since OO database allows each student and each course to be modeled as a single object. The ES knowledge was compiled from the university and college regulations, the long term experience of the author as an academic advisor and the Deputy Dean of the College of IT, and discussing the knowledge with students and advisors for their feedback.

IV. THE DEVELOPED EXPERT SYSTEM

A. The OO Database (OODB)

An important objective in database design is to develop an efficient database structure so that data can be stored, accessed, and modified easily. Much of the work in creating an effective database is in the modeling. It is the application domain that determines how the database should be modeled in order to be successful. The nature of university subjects' and students' records (the domain of this research) reveals that the OO model is the most appropriate database modeling method. OO structure allows each course and each student to be constructed as a different object, and the database modeled as a collection of these objects. This structure gives more flexibility to each object to have whatever features (i.e. attributes or fields) required to identify it while maintaining the integrity of the whole system. The database of IS-Advisor consists of the main classes: Courses and Students. Fig. 1 presents a portion of the object hierarchy of IS-Advisor which is the Kappa-PC's graphical representation of the OO database structure. Each study plan course in the database includes the following data: Title, ID, plan semester number (1 to 8), number of pre-requisite courses, List of pre-requisite courses (if any), pre-requisite hours (Some courses have a specified number of hours as their pre-requisite), type of course (There are three types of courses: Compulsory courses, major elective courses, and university elective courses), keywords describing course contents (e.g. mathematics, programming, algorithm, management, marketing, etc.; these keywords are used to assist students in selecting courses based on their preferences as will be addressed later), course components (theory, lab, and/or tutorial), and course status (offered or not offered; note that fall -or odd- semester courses are offered in fall semester and spring -or even- semester courses are offered in spring semester). Each student object includes the following fields: ID, name, AGPA, passed compulsory courses, passed major elective courses, passed university elective courses, course grades semester-by-semester, earned credit hours, allowable courses, registered courses, course keyword preferences, and load preferences. Note that some data listed above are known and saved in the database (example: offered courses in a particular semester or AGPA of a student) and some data are inferred by the ES (example: lists of allowable and registered courses of a student). It is important to note that the proposed ES is intended to be used for course selection only, and based on courses selected by all students the timing of lectures will be determined manually by the timetabling committee in order to prevent the time conflict between courses. Thus the ES's recommended courses for students will be used as the input for the college timetabling committee. Therefore course timing is not a factor in the current version of the system and a component to automate the determination of lecture timings can be added to the system as a future work.

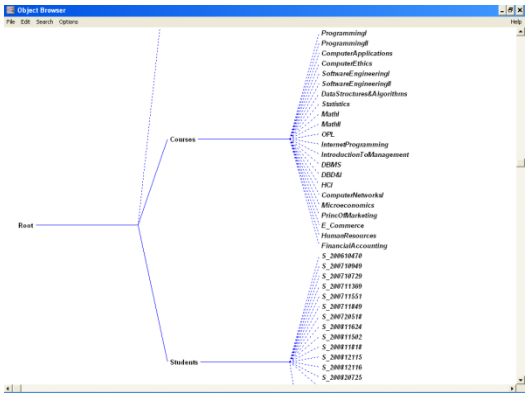


Figure 1. The Object hierarchy of the OO databse of IS-Advisor.

B. The Rule-Based Knowledge Base (RBKB)

The rules of the rule base can be classified into two categories: Academic rules and student-preference rules. Academic rules are rules that are concerned with academic regulation like pre-requisites, the minimum and maximum number of courses that can be registered by a student (usually: minimum 3 courses and maximum 6 courses), etc. As an example of this rule category, consider the following rules written in English:

Rule1:

If: The student passed Programming I AND Programming II is offered

Then: Add Programming II to the student's allowable courses list.

Rule2:

If: The student's passed hours are greater than or equal to 45AND Computer Ethics is offered

Then: Add Computer Ethics to the student's allowable courses list.

Rule3:

If: The student passed Computer Applications AND The student's passed hours are greater than or equal to 40 AND Computer Networks I is offered

Then: Add Computer Networks I to the student's allowable courses list.

Rule4:

If: The number of courses in the student's recommended courses list is less than 3

Then: Show the message: Students should register minimum 3 courses and maximum 6 courses. If your case is an exception, please contact your academic advisor.

Student-preference rules are If-Then rules related to preferences input by the student like preferred courses and preferred number of courses that the student is willing to register in a particular semester. As an example of this rule category, consider the following rule:

Rule5:

If: The student's course preference keyword is Management

Then: Mark all allowable courses having Management as a course keyword.

C. The Inference Engine (IE)

Kappa-PC ES shell supports both rule-based reasoning (forward- and backward-chaining) as well as the micro-managing of the reasoning using classical programming techniques (particularly list processing). IS-Advisor's inference engine uses both If-Then rules processing and list processing techniques. The overall reasoning procedure of the IS-Advisor is unique and different than other academic advising expert systems available in literature since it is based on the accumulated academic advising knowledge within the IS department at AUST.

There are three main steps performed in the process of determining the recommended courses for a particular IS student. In Step 1 all courses that are offered and can be registered by the student are stored in a list called Allowable Courses. Step 2 performs the ranking process for the courses contained in Allowable Courses list. The courses are ranked in a descending order as following: (1) Courses that are pre-requisite for subsequent courses (have the highest priority), (2) Courses matching student preferences (in case preferences are given), (3) Courses officially in the current student's registration semester (fall or spring) according to the study plan, (4) Courses whose pre-requisites were passed in the previous semester (in order not to leave a long time gap between a course and its pre-requisite), and (5) Remaining 'equal' allowable courses (if any) are displayed to the user in order to rank them as preferred. The list resulted from this step is called Ordered Allowable Courses. Step 3 is the filtering step that generates the ordered list of Recommended Courses based on the contents of the list Ordered Allowable Courses. This step follows one of the two advising models: Perspective advising (option 'One-Step Advising' in Fig. 4) or developmental advising (option 'Student's Preferences' in Fig. 4). In 'One-Step Advising' option the list of Recommended Courses is generated as following: (a) Students with AGPA greater than or equal to 3.00 are given the courses ranked from 1 to 6 (from the Ordered Allowable Courses list). (b) Students with AGPA greater than 2.24 and less than 3.00 are given the courses ranked from 1 to 5. (c) Students with AGPA greater than or equal to 2.00 and less than 2.25 are given the courses ranked from 1 to 4. Note that if the remaining number of courses for a student towards graduation is less than the number of courses that can be suggested by the system, then the students is recommended to take the remaining courses only. In "Students' Preferences" option the student is asked to select the number of courses he/she is willing to register (3 to 6 courses) and course keyword preferences. Consequently the list Recommended Courses is prepared as explained in 'One-Step Advising' option above however here level 2 of ranking (courses matching student's preferences) is activated and the number of courses is equal to the number of courses selected by the student (if possible). In addition, more system messages are given here during the user-system interaction in order to guide the student to consider a 'more' suitable course selection.

D. The User Interface and Sample Consultation

Interactions between the users and the system are supported through a friendly graphical user interface running under Windows environment. Fig. 2 shows the main screen of the system where various options are displayed. The Button "Offered Courses" presents all currently offered courses and the button "Study Plan Structure" displays a semester-by-semester structure plan of the IS program. The user can enter the user manual and get more help on using the system by selecting the "Help" button, or exit the system by clicking "Exit". The main option here is "Academic Advising" from which the user is directed to a screen asking for the student ID number (Fig. 3).

After a welcoming message and displaying the currently available student transcript data, the student is given two options as shown in Fig. 4. These two options work as explained while discussing the inference engine above. Fig. 5 shows the result of selecting the option One-Step Advising for a particular student whose AGPA is between 2.25 and 2.99. Note that the Recommended Courses list generated here is ranked by priority from 1 (highest priority) to 5 (lowest priority). The student can click here on the option "Explain!" and get the explanation screen shown in Fig. 6. This explanation screen (related to this 'consultation 1' example) gives the academic reasons for suggesting these courses and for ranking them this way.

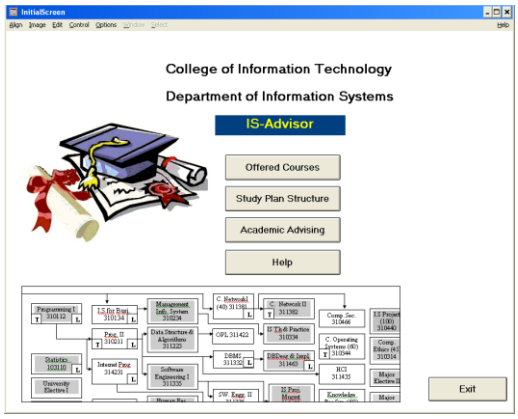


Figure 2. The main screen of IS-Advisor.

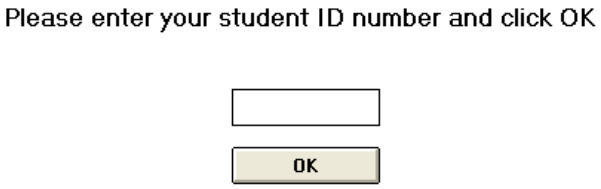


Figure 3. Entering student ID number.

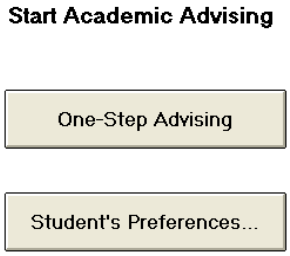


Figure 4. Academic advising models.

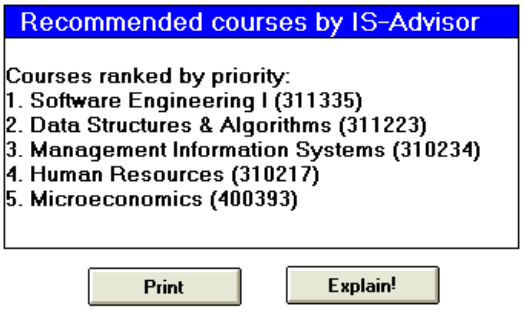


Figure 5. Recommended courses by the system.

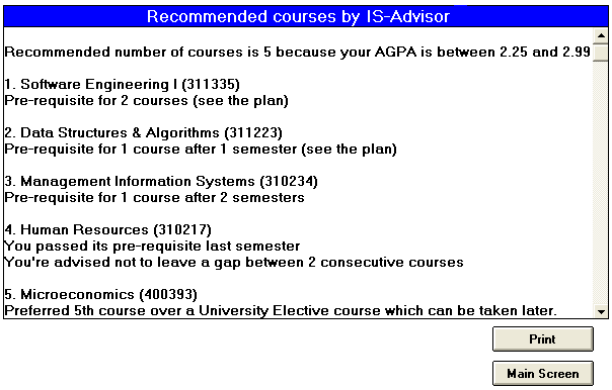


Figure 6. A Sample Explanation Screen (consultation 1).

As a second consultation (consultation 2) example for a student with AGPA greater than 2.99, the option 'Student's Preferences" on the screen of Fig. 4 results in various query screens as shown in Fig. 7, and Fig. 8.

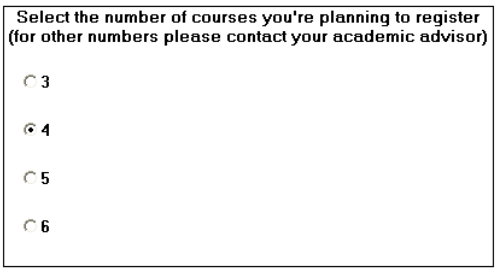


Figure 7. Specifying the number of courses.

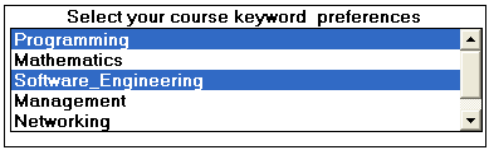


Figure 8. Selecting course keyword preferences.

The result of this consultation is displayed in Fig. 9 below.

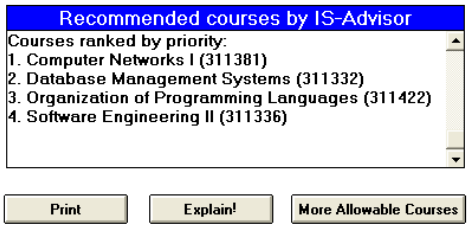


Figure 9. Recommended courses by the system.

The options Explain! And More Allowable Courses give the screens of Fig. 10 and Fig. 11 respectivily. Note that for this particular sample advising consultation, the student's prefered number of courses is 4 (see Fig. 7 and Fig. 10). Since the student's AGPA is above 2.99 in this example, the student can register up to 6 courses, this fact is given to the student in Fig. 10 and clicking on 'More Allowable Courses' will suggest to the student the list of allowable courses (Fig. 11). In case the student agrees to increase the number of courses, he/she can increase his prefered number of courses in the screen of Fig. 7.

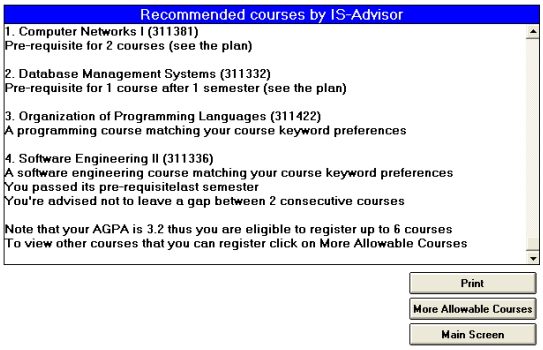


Figure 10. A Sample Explanation Screen (consultation 2).

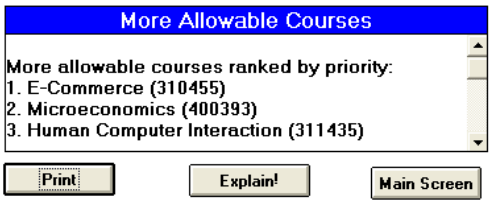


Figure 11. A Sample List of More Allowable Courses.

V. SYSTEM TESTING

The system was tested by comparing its results with randomly selected actual (i.e. human advisor guided) student registration files for three academic semesters. The results of comparing 130 registration processes show that 93% of the actual cases matched with the system's recommendations. This number shows that the prototype system is successful and going in the correct direction. The reasons for the 7% unmatched results include:

- Warned students: Warned students are students with AGPA less than 2.00 and the regulations require them to repeat some low grade courses to increase their AGPA. Such category of students is outside the scope of the current version of the prototype system. (Planned to be modeled in a modified version).
- Exceptional cases: The current version of IS-Advisor follows strictly the registration rules; however, there are few exceptions which are performed under some special conditions in human-guided advising situations and not counted for in IS-Advisor. To give an example: Some students (because of official medical reports) request to register less than 3 courses in a semester. Such exceptional cases cannot be handled by the current version of the system, and such students are directed by the system to contact their academic advisors (see Fig. 7).

VI. CONCLUSION AND FUTURE WORK

In this research a prototype expert system with an object-oriented database for student academic advising has been designed and developed. By implementing prescriptive advising model and developmental advising model, the system provides the students and advisors with a useful tool for quick and easy course selection and evaluation of various alternatives. The system has a graphical user interface and simple menus; information is displayed in a way that is familiar for both advisors and students. The present state of the system was discussed and illustrated with sample consultations. The system is successful and efficient. System testing revealed that 93% of academic advising test cases show an agreement between the system advising in course selection and human advising. Enriching the system by adding more data and knowledge rules is a continuous process. Many parts of the system can be improved further and some issues deserve future work, among them:

1. Currently the system operates as a stand-alone system. It would be better to connect IS-Advisor with the university's student information system. This will automate the process of importing students' data.
2. Advising of students with exceptional cases and warned students is outside the scope of the current system. This feature can be added in future developments of the system.
3. The system can be improved so that it automates the determination of lecture timings for courses based on the courses recommended for all students so that time conflict between lectures is prevented.

REFERENCES

[1] T. Feghali, I. Zbib, and S. Hallal, "A web-based decision support tool for academic advising", Educational Technology & Society, Vol. 14, No. 1, pp. 82–94, 2011.
[2] A. N. Nambiar and A. K. Dutta, "Expert system for student advising using JESS", International Conference on Educational and Information Technology (ICEIT), China, September 17-19, 2010.
[3] F. Albalooshi and S. Shatnawi, "HE-Advisor: A multidisciplinary web-based higher education advisory system", Global Journal of Computer Science & Technology, Vol. 10, No. 7, pp. 37-49, September 2010.
[4] R. Zucker, "ViCurriAS: A curriculum visualization tool for faculty, advisors, and students", Journal of Computing Sciences in Colleges, Vol. 25, No. 2, pp. 138-145, December 2009.
[5] D. Pokrajac and M. Rasamny, "Interactive virtual expert system for advising (InVEStA)", 36th Annual ASEE/IEEE Frontiers in Education Conference, San Diego, CA, USA, October 27-31, 2006.
[6] K. Kowalski, "On-line advising with JavaScript rule-based system", Proceedings of Society for Information Technology & Teacher Education International Conference, Chesapeake, VA, USA, pp. 2922-2927, 2004.
[7] R. M. Siegfried, A. M. Wittenstein, and T. Sharma, "An automated advising system for course selection and scheduling", Journal of Computing Sciences in Colleges, Vol. 18, No. 3, pp. 17-25, February 2003.
[8] A. M. Wittenstein and T. Sharma, "FROSH2: An expert system for freshman advisement", Proceeding of the National Conference on Undergraduate Research (NCUR), University of Wisconsin, Whitewater, Wisconsin, USA, April 25-27, 2002.

[9] M. Patankar, "A rule-based expert system approach to academic advising", Innovations in Education and Training International, Vol. 35, No. 1, pp. 49-58, February 1998.
[10] B. B. Crookston, "A developmental view of academic advising as teaching", Journal of College Student Personnel, Vol. 13, No. 1, pp. 12-17, 1972.
[11] C. M. Chando, "Predicting advising style preference from student characteristics", Doctoral dissertation, University of Memphis, UAS, 1997.
[12] L. L. Fielstein, "Developmental versus prescriptive advising: Must it be one or the other?", The National Academic Advising Association (NACADA) Journal, Vol. 14, No. 2, pp. 76-79, 1994.
[13] R. J. Schalkoff, "Intelligent systems: Principles, paradigms and pragmatics", Jones & Bartlett Publishers, 2009.
[14] Intellicorp, Kappa-PC 2.4 ES shell manuals, Intellicorp, Inc., USA, 1997.

AUTHORS PROFILE

Dr. Ayman Al Ahmar is Assistant Professor and the Deputy Dean of the College of Information Technology, Ajman University of Science and Technology, UAE. He received his B.Sc. (1994), M.Sc. (1997), and Ph.D. (2001) degrees from Middle East Technical University (METU), Ankara, Turkey. His current research interests include Artificial Intelligence, Software Engineering, and Engineering Information Systems. He is a member of IEEE and IEEE Computer Society.

Face Recognition Using Bacteria Foraging Optimization-Based Selected Features

Rasleen Jakhar
M.Tech Student
Lovely Professional University
Punjab, India

Navdeep Kaur
Ex-Lecturer
RIEIT
Railmajra, Punjab, India

Ramandeep Singh
Assistant professor
Lovely Professional University
Punjab, India

***Abstract*— Feature selection (FS) is a global optimization problem in machine learning, which reduces the number of features, removes irrelevant, noisy and redundant data, and results in acceptable recognition accuracy. This paper presents a novel feature selection algorithm based on Bacteria Foraging Optimization (BFO). The algorithm is applied to coefficients extracted by discrete cosine transforms (DCT). Evolution is driven by a fitness function defined in terms of maximizing the class separation (scatter index). Performance is evaluated using the ORL face database.**

Keywords- Face Recognition; Bacteria Foraging Optimization; DCT; Feature Selection.

I. INTRODUCTION

A. Face Recognition

Face Recognition (FR) is a matching process between a query face's features and target face's features. Face recognition (FR) has emerged as one of the most extensively studied research topics that spans multiple disciplines such as pattern recognition, signal processing and computer vision. [1]. The block diagram of Face Recognition system is shown in figure 1.

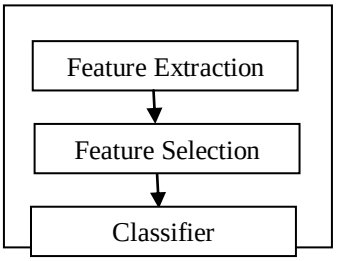


Figure 1: Face Recognition System

1) Feature Extraction

It is known that a good feature extractor for a face recognition system is claimed to select as more as possible the best discriminate features which are not sensitive to arbitrary environmental variations such as variations in pose, scale, illumination, and facial expressions [2].

Feature extraction algorithms mainly fall into two categories: geometrical features extraction and, statistical (algebraic) features extraction [1], [3], [4]. The geometrical approach represents the face in terms of structural measurements and distinctive facial features. These features are used to recognize an unknown face by matching it to the nearest neighbor in the stored database. Statistical features extraction is usually driven by algebraic methods such as principal component analysis (PCA) [5], and independent component analysis (ICA) [5], [6], [7], [8], [9], [10], [11].

a) Discrete Cosine Transform

DCT has emerged as a popular transformation technique widely used in signal and image processing. This is due to its strong "energy compaction" property: most of the signal information tends to be concentrated in a few low-frequency components of the DCT. The use of DCT for feature extraction in FR has been described by several research groups [10], [11], [12], [13], [14], [15] and [16]. DCT transforms the input into a linear combination of weighted basis functions. These basis functions are the frequency components of the input data.

The general equation for the DCT of an NXM image $f(x, y)$ is defined by the following equation:

$$F(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \cos\left[\frac{\pi.u}{2.N}(2x+1)\right] \cos\left[\frac{\pi.v}{2.M}(2y+1)\right] f(x,y) \quad \text{... (i)}$$

Where $f(x, y)$ is the intensity of the pixel in row x and column y; $u = 0, 1 \ldots N-1$ and $v = 0, 1 \ldots M-1$ and the functions $\alpha(u)$, $\alpha(v)$ are defined as:

$$\alpha(u),\alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & for\ u,v=0 \\ \sqrt{\frac{2}{N}} & for\ u,v\neq 0 \end{cases} \quad \ldots \text{(ii)}$$

2) Feature Selection

The feature selection seeks for the optimal set of d features out of m [17], [18] and [19]. Several methods have been previously used to perform feature selection on training and testing data. Among the various methods proposed for FS, population-based optimization algorithms such as Genetic Algorithm (GA)-based method [20], [21], [22] and Ant Colony Optimization (ACO)-based method have attracted a lot of attention [23]. In the proposed FR system we utilized an evolutionary feature selection algorithm based on swarm intelligence called the Bacteria Foraging Optimization.

B. Bacteria Foraging Optimization

Bacterial Foraging Optimization (BFO) is a novel optimization algorithm based on the social foraging behavior of

E. coli bacteria. The motile bacteria such as *E. coli* and salmonella propel themselves by rotating their flagella. To move forward, the flagella counterclockwise rotate and the organism "swims" (or "runs").While a clockwise rotation of the flagellum causes the bacterium randomly "tumble" itself in a new direction and then swims again [24], [25].

1) Classical BFO Algorithm

The original Bacterial Foraging Optimization system consists of three principal mechanisms, namely, chemo taxis, reproduction, and elimination-dispersal [25]:

a) Chemo taxis

Suppose $\theta^i(j,k,l)$ represents the bacterium at *j*th chemo tactic, *k*th reproductive, and *l*th elimination-dispersal step. *C*(*i*) is the chemo tactic step size during each run or tumble (i.e., run-length unit). Then in each computational chemo tactic step, the movement of the *i*th bacterium can be represented as

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\,\Delta(i)}}, \qquad \ldots \text{(iii)}$$

Where Δ(*i*) is the direction vector of the *j*th chemo tactic step. When the bacterial movement is *run*, Δ(*i)* is the same with the last chemo tactic step; otherwise, Δ(*i*) is a random vector whose elements lie in [−1, 1]. With the activity of run or tumble taken at each step of the chemo taxis process, a step fitness, denoted as *J*(*i, j, k, l*), will be evaluated.

b) Reproduction

The health status of each bacterium is calculated as the sum of the step fitness during its life, that is, $\sum_{j=1}^{Nc} J(i,j,k,l)$, where *Nc* is the maximum step in a chemo taxis process. All bacteria are sorted in reverse order according to health status. In the reproduction step, only the first half of population survives and a surviving bacterium splits into two identical ones, which are then placed in the same locations. Thus, the population of bacteria keeps constant.

c) Elimination and Dispersal

The chemo taxis provides a basis for local search, and the reproduction process speeds up the convergence which has been simulated by the classical BFO. While to a large extent, only chemo taxis and reproduction are not enough for global optima searching. Since bacteria may get stuck around the initial positions or local optima, it is possible for the diversity of BFO to change position to eliminate the accidents of being trapped into the local optima. Then some bacteria are chosen, according to a preset probability *Ped*, to be killed and moved to another position within the environment.

The original BFO algorithm is briefly outlined step by step as follows.

Step1. Initialize parameters n, S, Nc, Ns, Nre, Ned , Ped, C(i) (i = 1, 2, . . . , S), θi where

n: dimension of the search space,
S: the number of bacteria in the colony,
Nc: chemo tactic steps,
Ns: swim steps,
Nre: reproductive steps,
Ned: elimination and dispersal steps,
Ped: probability of elimination,
C(*i*): the run-length unit (i.e., the size of the step taken in each run or tumble).
θ^i: position of ith bacteria

Step2. Elimination-dispersal loop: $l = l + 1$.

Step3. Reproduction loop: $k = k + 1$.

Step4. Chemo taxis loop: $j = j + 1$.

Sub step 4.1. For $i = 1, 2 \ldots S$, take a chemo tactic step for bacterium *i* as follows:

Sub step 4.2. Compute fitness function, *J*(*i, j, k, l*).

Sub step 4.3. Let *J*last = *J*(*i, j, k, l*) to save this value since we may find better value via a run.

Sub step 4.4. Tumble. Generate a random vector $\Delta(i) \in \mathrm{R}^n$ with each element Δ*m*(*i*), $m = 1, 2 \ldots n$, a random number on [−1, 1].

Sub step 4.5. Move. Let

$$\theta^i(j+1,k,l) = \theta^i(j,k,l) + C(i)\frac{\Delta(i)}{\sqrt{\Delta^T(i)\,\Delta(i)}}, \qquad \ldots \text{(iv)}$$

This results in a step of size *C*(*i*) in the direction of the tumble for bacterium *i.*

Sub step 4.6. Compute *J*(*i, j* +1, *k, l*) with $\theta^i(j+1, k, l)$.

Sub step 4.7. Swimming

(i) Let *m* = 0 (counter for swim length).

(ii) While *m* < *Ns* (if has not climbed down too long), the following hold.

- Let *m* = *m* + 1.
- If *J*(*i, j* +1, *k, l*)< *J*last, let *J*last = *J*(*i, j* +1, *k, l*), then another step of size *C*(*i*) in this same direction will be taken as (iv) and use the new generated. $\theta^i(j+1, k, l)$ to compute the new *J*(*i, j* +1, *k, l*).
- Else let *m* = *Ns.*

Sub step 4.8. Go to next bacterium (*i* +1). if *i*≠*S,* go to Sub step 4.2 to process the next bacterium.

Step5. If *j* < *Nc* , go to Step 3. In this case, continue chemo taxis since the life of the bacteria is not over.

Step6. Reproduction

Sub step 6.1. For the given *k* and *l*, and for each $i = 1, 2 \ldots S$, let

$$J^i_{health} = \sum_{j=1}^{Nc+1} J(i,j,k,l) \qquad \ldots \text{(v)}$$

be the health of the bacteria. Sort bacteria in order of ascending values (J_{health})

Sub step 6.2. The *Sr* bacteria with the highest *J*health values die and the other *Sr* bacteria with the best values split and the copies that are made are placed at the same location as their parent.

Step7. If *k* <*Nre,* go to Step 2. In this case the number of specified reproduction steps is not reached and start the next generation in the chemotactic loop.

Step8. Elimination-dispersal: for *i* = 1, 2, . . . , *S*, with probability P*ed*, eliminate and disperse each bacterium, which results in keeping the number of bacteria in the population constant. To do this, if a bacterium is eliminated, simply disperse one to a random location on the optimization domain. If *l* <*Ned*, then go to Step 2; otherwise end.

II. BFO- BASED FEATURE SELECTION

In this proposed work, features of image are extracted using DCT technique. The extracted features are reduced further by using Bacteria Foraging Optimization to remove redundancy and irrelevant features. The resulting feature subset (obtained by BFO) is the most representative subset and is used to recognize the face from face gallery.

A. Bacteria Representation

Each bacteria's position represent one possible solution (feature subset) required for face recognition. The number of dimensions of search space is m where m is the length of feature vector extracted by DCT. In each dimension of search space, bacteria position is 1 or 0, where 1 or 0 indicates that this feature is selected or not selected, respectively, as required feature for next generation. In the each iteration of chemo taxis step, each bacteria tumbles to the new random position. Position of ith bacteria in jth chemo taxis and kth reproduction step is defined as:

$$\Theta^{i}(j, k) = F_1 F_2 \ldots F_m \quad \ldots(vi)$$

Where, m is the length of feature vector extracted by DCT. Each F_z = 1 or 0 (z=1,2,..m) Depending upon whether zth feature is selected or not for the next iteration.

1) Fitness Function

In each generation, each bacterium is evaluated, and a value of *goodness* or *fitness* is returned by a fitness function. This evolution is driven by the fitness function *F* [26]. Let w_1, w_2 ..., w_L *and* N_1, N_2... N_L denote the classes and number of images within each class, respectively. Let *M1*, *M2*, M_L *and* M_0 be the means of corresponding classes and the grand mean in the feature space, M_i can be calculated as:

$$M_i = \frac{1}{N_i} \sum_{j=1}^{Ni} W_j^{(i)}, \quad i = 1,2,....,L \quad \ldots (vii)$$

Where $W_j^{(i)}$, j=1,2,…,N_i , represents the sample image from class w_i and grand mean M_0 is:

$$M_0 = \frac{1}{N} \sum_{i=1}^{L} N_i M_i \quad \ldots (viii)$$

Where N is the total number of images of all the classes. Thus the between class scatter fitness function F is computed as follows:

$$F = \sqrt{\sum_{i=1}^{L} (M_i - M_0)^t (M_i - M_0)} \quad \ldots (ix)$$

2) Classifier

After the training phase, a typical and popular Euclidean distance is employed to measure the similarity between the test vector and the reference vectors in the gallery. Euclidean distance is defined as the straight-line distance between two points. For *N*-dimensional space, the Euclidean distance between two any points' *pi* and *qi* is given by:

$$D = \sqrt{\sum_{i=1}^{N} (p_i - q_i)^2} \quad \ldots(x)$$

Where p_i (or q_i) is the coordinate of *p* (or *q*) in dimension *i*.

III. PROPOSED BFO-BASED FEATURE SELECTION ALGORITHM

The algorithm proposed for feature extraction using BFO is discussed in figure 2. There are certain variations in BFO algorithm used in this work. Firstly, step 6.6 of the proposed algorithm moves the bacteria back to its previous position if current position is less suitable (checked using fitness function). So in this algorithm, bacteria have "memory" as they remember their previous position. Secondly, as there are chances that bacteria may get struck in local optima, elimination dispersal removes bacteria from its current position and moves it to "random" new position. In the proposed algorithm, position of bacteria is decided randomly in the each iteration. There is no need of using Elimination Dispersal.

IV. EXPERIMENTAL RESULTS

The performance of the proposed feature selection algorithm is evaluated using the standard Cambridge ORL gray-scale face database. The ORL database of faces contains a set of face images taken between April 1992 and April 1994 at the AT&T Laboratories (by the Oliver Research Laboratory in Cambridge, UK) [13] and [23].

The database is composed of 400 images corresponding to 40 distinct persons. The original size of each image is 92x112 pixels, with 256 grey levels per pixel. Each subject has 10 different images taken in various sessions varying the lighting, facial expressions (open/ closed eyes, smiling/ not smiling) and facial details (glasses/ no glasses). Four images per person were used in the training set and the remaining six images were used for testing.

The parameters used for BFO-based Feature Selection is shown in table 1.

TABLE I. BFO PARAMETER SETTING

Bacteria Size(S)	30
Number of chemo taxis steps (Nc)	10
Number of reproduction steps (Nre)	10

In this work, we test the BFO-based feature selection algorithm with feature vectors based on various sizes of DCT coefficient. The 2-dimentional DCT is applied to the input image and only a subset of the DCT coefficients corresponding to the upper left corner of the DCT array is retained. Subset sizes of 50x50, 40x40, 30x30 and 20x20 of the original 92x112 DCT array are used in this work.

1. Feature Extraction: Obtain the DCT array by applying Discrete Cosine Transformation to image.
2. Take the most representative features of size nxn from upper left corner of DCT Array.
3. Feature Selection: Define the BFO parameters: S,Nc,Nre,n
   Where S : Number of bacteria in the colony
   Nc: chemo tactic steps
   Nre: reproductive steps
   n: dimension of the search space
4. Place each bacteria at random position.
5. (Reproduction step k+1) For k = k + 1
6. (Chemo taxis step j+1) For j = j + 1
   6.1 For i = 1,2,…S, take chemotaxis step for bacteria i as follows:
   6.2 Compute the fitness function of the previous (j^{th}) chemo taxis step as defined in formula (vii), (viii), (ix).
   6.3 (Tumble)Tumble to random new position, $\theta^i(j+1,k)$, as defined in formula (vi).
   6.4 (Move) Move the bacteria to new position
   $$\theta^i(j+1,k).$$
   6.5 Compute $J(i,j+1,k)$ using $\theta^i(j+1,k)$ as in step 6.2.
   6.6 If J(i,j+1,k) < J(i,j,k) then:
   (i) (Move Back) Move bacteria back to its previous position:
   $$\theta^i(j+1,k) = \theta^i(j,k)$$
   (ii) Update Fitness function
   J(i, j+1, k) = J(i ,j, k)
   end if
   6.7 Go to next bacteria i +1 (for of step 6.1 ends)
   6.8 Store the current fitness of i^{th} bacteria in $J_c(i)$.
   (chemo taxis loop of step 6 ends).
7. (reproduction step) for given k and for each i = 1,2..S, Let
   $$J^i_{health} = J_c(i)$$
   be the health of bacteria. Sort the bacteria in descending order of J_{health}.
8. The bacteria with S_r lowest J_{health} values die and other bacteria with S_r best J_{health} values are split and copies that are made are placed at the same location as their parents.
   (reproduction loop of step 5 ends)
9. Pick up the position of bacteria B with *max* (J_{health}) value. This position represents the best feature subset of the features defined in step 2.
   (Feature Selection Ends)
10. Classification: calculate the difference between the feature subset (obtained in step 9) of each image of facial gallery and the test image with the help of Euclidean Distance defined in Formula (v).The index of the image which has the smallest distance with the image under test is considered to be the required index.

Figure 2: Face Recognition using BFO based Feature Selection

Each of 2- dimensional subset DCT array is converted to a 1-dimensional array using raster scan. This is achieved by processing the image row by row concatenating the consecutive rows into a column vector. This column vector is the input to the subsequent feature selection algorithm.

To calculate average recognition rate for each problem instance, 5 test images are randomly chosen from 40 classes. Average recognition is measured by knowing how many times correct faces were identified out of 5 trials. The average recognition rate is measured together with the CPU training time and the average number of selected features for each problem instance. The algorithm has been implemented in Mat lab 7 and the result for each problem instance (20X20, 30X30, 40X40, and 50X50 DCT Array) is shown in table 2

TABLE II. RESULTS OF BFO-FS ALGORITHM

DCT Feature Vector Size	Number of Features input to BFO-FS	Average no. of features selected by BFO-FS	Training time (in seconds)	Average Recognition Rate
20X20	400	215	102.610	100%
30X30	900	454	132.531	100%
40X40	1600	807	209.969	100%
50X50	2500	1281	272.110	100%

Following are the faces recognized by the proposed Algorithm for various number of features input to BFO-FS.

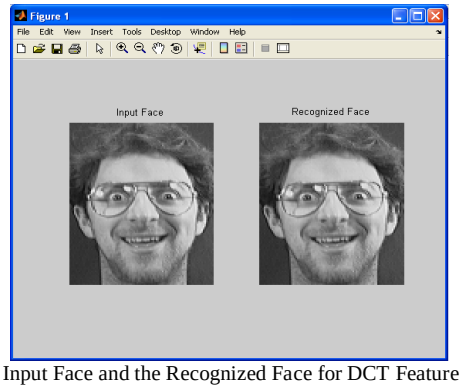


Figure: 3 Input Face and the Recognized Face for DCT Feature vector of 20X20

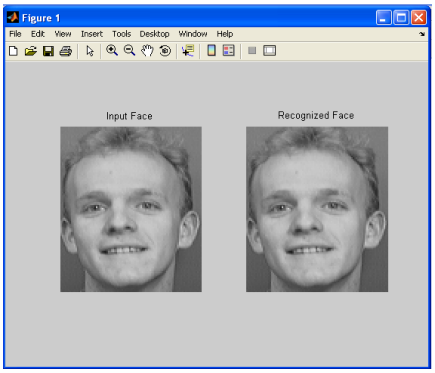


Figure: 4 Input Face and the Recognized Face for DCT Feature vector of 30X30

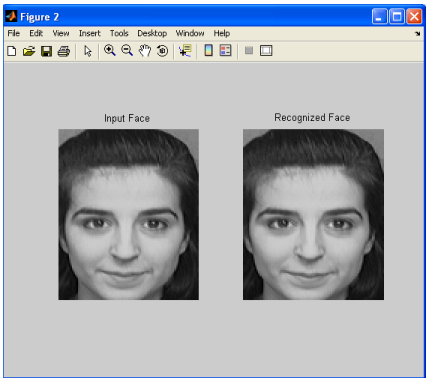


Figure: 5 Input Face and the Recognized Face for DCT Feature vector of 40X40

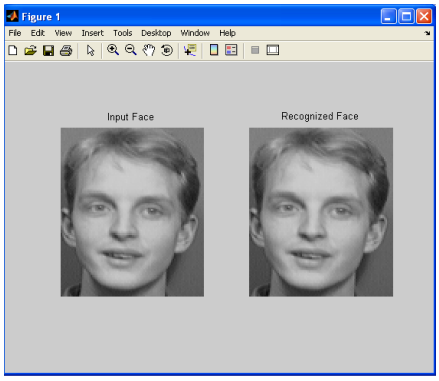
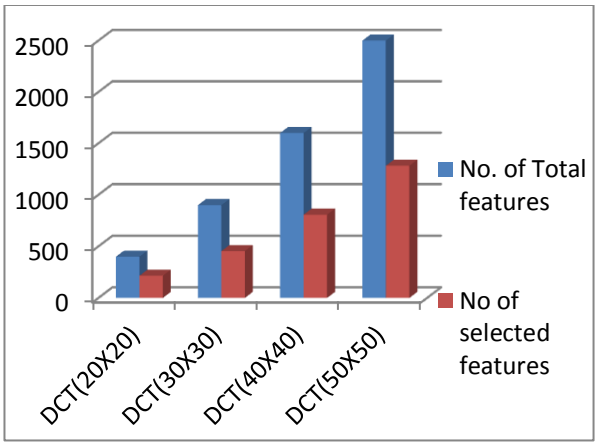


Figure: 6 Input Face and the Recognized Face for DCT Feature vector of 50X50

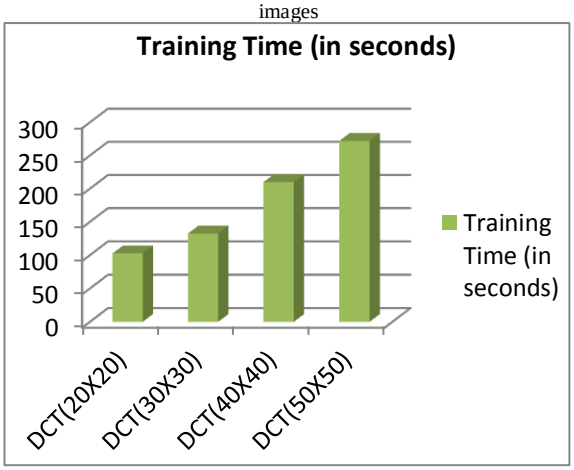
For each of the problem instance (20X20, 30X30, 40X40, and 50X50), algorithm is run 5 times and each time, random test image is chosen to be matched with face gallery. The test face matches with image in face gallery in each trial and average recognition rate is 100 % for each problem instance. The BFO-selection algorithm reduces the size of original feature vector to 53.7%, 50%, 50%,51% for problem instance of 20X20, 30X30, 40X40, and 50X50 respectively. For example, if the DCT of an image is calculated and 20X20 DCT subset is taken from upper left of DCT array, there are total 400 features which are given as an input to BFO-FS algorithm. BFO-FS reduces the 400 features to 215 which means only 215 features are required to recognize the face from facial gallery.

A. *Comparison of BFO with PSO*

If the proposed algorithm is compared with PSO-based feature selection described in [2], the average recognition rate of the proposed algorithm is better than that of PSO-based feature selection. The number of selected features by proposed algorithm is comparable to those selected by PSO-based feature selection. On the other hand, in terms of computational time, PSO-based selection algorithm takes less training time than the BFO-based selection algorithm in all tested instances which indicates that BFO is computationally expensive than PSO but the effectiveness' of BFO in finding the optimal feature subset compared to PSO compensates its computational inefficiency.



Graph: 1 showing the total no of features and the selected features for various images



Graph2: Showing the training time for different images

V. CONCLUSION

In this paper, a novel BFO-based feature selection algorithm for FR is proposed. The algorithm is applied to feature vectors extracted by Discrete Cosine Transform. The algorithm is utilized to search the feature space for the optimal feature subset. Evolution is driven by a fitness function defined in terms of class separation. The classifier performance and the length of selected feature vector were considered for performance evaluation using the ORL face database. Experimental results show the superiority of the BFO-based feature selection algorithm in generating excellent recognition accuracy with the minimal set of selected features.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A.Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399-458, 2003.

[2]. R.M. Ramadan and R. F. Abdel – Kader, "Face Recognition Using Particle Swarm Optimization – Based Selected Features", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 2, No. 2, June 2009.

[3] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 10, pp. 1042-1052, 1993.

[4] X. Yi-qiong, L. Bi-cheng and W. Bo, "Face Recognition by Fast Independent Component Analysis and Genetic Algorithm," *Proc. Of the 4th International Conference on Computer and Information Technology (CIT'04)*, pp. 194-198, Sept. 2004.

[5] M. A. Turk and A. P. Pentland, "Face Recognition using Eigenfaces," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 586-591, June 1991.

[6] H. R. Wilson, D. Levi, L. Maffei, J. Rovamo, and R. DeValois, "The Perception of Form: Retina to Striate Cortex", *Visual Perception: The Neurophisiologcal Foundations*, Academic Press, 1990.

[7] S. Chien, and I. Choi, "Face and Facial Landmarks location based on Log-Polar Mapping", *Lecture Notes in Computer Science – LNCS* 1811, pp. 379-386, 2000.

[8] S. Minut, S. Mahadevan, J. Henderson, and F. Dyer, "Face Recognition using Foveal Vision", *Lecture Notes in Computer Science – LNCS* 1811, pp. 424-433, 2000.

[9] M. Tistarelli, and E. Grosso, "Active Vision-Based Face Authentication", *Image and Vision Computing*, no. 18, pp. 299-314, 2000.

[10] A. S. Samra, S. E. Gad Allah, R. M. Ibrahim, "Face Recognition Using Wavelet Transform, Fast Fourier Transform and Discrete Cosine Transform," *Proc*. 46th *IEEE International Midwest Symp. Circuits and Systems (MWSCAS'03*), vol. 1, pp. 272- 275, 2003.

[11] Z. Yankun and L. Chongqing, "Efficient Face Recognition Method based on DCT and LDA", *Journal of Systems Engineering and Electronics*, vol. 15, no. 2, pp. 211-216, 2004.

[12] C. Podilchuk and X. Zhang, "Face Recognition Using DCT-Based Feature Vectors," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96),* vol. 4, pp. 2144-2147, May 1996.

[13] F. M. Matos, L. V. Batista, and J. Poel, "Face Recognition Using DCT Coefficients Selection," *Proc. of the 2008 ACM Symposium on Applied Computing, (SAC'08)*,pp. 1753-1757, March 2008.

[14] M. Yu, G. Yan, and Q.-W. Zhu, "New Face recognition Method Based on DWT/DCT Combined Feature Selection," *Proc. 5th International Conference on Machine Learning and Cybernetics*, pp. 3233-3236, August 2006.

[15] Z. Pan and H. Bolouri, "High Speed Face Recognition Based on Discrete Cosine Transform and Neural Networks," Technical Report, Science and Technology Research Center (STRC), University of Hertfordshire.

[16] Z. M. Hafed and M. D. Levine, "Face Recognition Using Discrete Cosine Transform", *International Journal of Computer Vision*, vol. 43, no. 3, pp. 167-188. 2001

.[17] C.-J. Tu, L.-Y. Chuang, *J.-Y. Chang*, and C.-H. Yang, "Feature Selection using PSO-SVM," *International Journal of Computer Science* (*IAENG)*, vol. 33, no. 1, IJCS_33_1_18.

[18] E. Kokiopoulou and P. Frossard, "Classification-Specific Feature Sampling for Face Recognition," *Proc IEEE 8th Workshop on Multimedia Signal Processing*, pp. 20-23, 2006.

[19] A. Y. Yang, J. Wright,Y. Ma, and S. S. Sastry, " Feature Selection in Face Recognition: A Sparse Representation Perspective," submitted for publication, 2007.

[20] X. Fan and B. Verma, "Face recognition: a new feature selection and classification technique," *Proc. 7th Asia-Pacific Conference on Complex Systems*, December 2004.

[21] D.-S. Kim, I.-J. Jeon, S.-Y. Lee, P.-K. Rhee, and D.-J. Chung, "Embedded Face Recognition based on Fast Genetic Algorithm for Intelligent Digital Photography," *IEEE Trans. Consumer Electronics*, vol. 52, no. 3, pp. 726-734, August 2006.

[22] M. L. Raymer, W. F. Punch, E. D. Goodman, L.A. Kuhn, and A. K Jain, "Dimensionality Reduction Using Genetic Algorithms," *IEEE Trans. Evolutionary Computation*, vol. 4, no. 2, pp. 164-171, July 2000.

[23] H. R. Kanan, K. Faez, and M. Hosseinzadeh, "Face Recognition System Using Ant Colony Optimization-Based Selected Features," *Proc. IEEE Symp. Computational Intelligence in Security and Defense Applications (CISDA 2007)*, pp 57-62, April 2007.

[24]. J. Adler (1966), "*Chemotaxis in bacteria*," Science, vol. 153, pp. 708–716.

[25] H. Chen, Y.Zhu and R. Hu, " Cooperative Bacterial Foraging Optimization", Research Article, Key Laboratory of Industrial Informatics, Shenyang Institute of Automation, Chinese Academy of Sciences, China,,2009.

[26] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 6, pp. 570-582, 2000.

Instant Human Face Attributes Recognition System

N.Bellustin
[1]Research Institute of Radio
Physics, Nizhny Novgorod, Russia

Y. Kalafati[2]
[2]Institute of Radio Engineering and Electronics RAS,
Moscow, Russia,

Kovalchuck, A. Telnykh, O. Shemagina and
V.Yakhno[3]
[3]Institute of Applied Physics, Nizhny Novgorod, Russia,

Abhishek Vaish, Pinki Sharma, Shirshu Verma[4]
[4]Indian Institute of Information Technology, Allahabad,
INDIA

***Abstract*— The objective of this work is to provide a simple and yet efficient tool for human attributes like gender, age and ethnicity by the human facial image in the real time image as we all aware this term that "Real-Time frame rate is a vital factor for practical deployment of computer vision system". In this particular paper we are trying to presents the progress towards face detection and human attributes classification system. We have developed an algorithm for the classification of gender, age and race from human frontal facial image As the basis of the classifier proposed algorithm uses training set neuron receptors that process visual information a study of the several variants of these classifiers and shows the principal possibility of sex determination, assessment of a person's age on a scale (adult - children) and recognition of race by using the neuron-like receptors.**

Keywords- Gender recognition; Age recognition; Ethnicity recognition; MCT; AdaBoost; attributes classifier.

I. INTRODUCTION

The human brain is a dynamical system whose state evolves with time. Stimuli from the outside world are by cortex neuron processed and transformed, with inner human brain models compared and then fed-back throughout the brain. The complex dynamical interaction with these inner models, which human foregoing experience accumulate, provide right making decision and true classify objects of different types. The human brain was million years constructed and tuned, and now it is arguably the best known classifier – it can learn complex

Classification tasks fast and surely. This high efficiency of natural biological system of image processing have long motivated researchers to seek to best understanding of the brain operation and to construct artificial electronic systems with these high characteristics.

One the main step in this direction is to construct an "Attribute Classifier" which classifies the human attributes like gender, age, race and emotional state from the facial image with the efficacy comparable – the present paper is devoted to this interesting and complicated problem. There is well known that human faces convey lots of information and the high-level semantic information about the identification of such attributes. Automatic human attributes classification based on human face image aims at recognizing the attributes according to face appearance in images. In this paper we present a system for face detection and recognition of the attributes of a person, which is based on a unified approach to the recognition of some types of attributes that can be associated with the face of the person – gender, age, race, presence of beard and mustache, glasses and so on. From the experiments that our approach can be used for attributes recognition by the human face at real-time.

II. RELATED WORK

In order to interact socially, we must be able to process human faces in diversity of ways. There is vast amount of literature on cognitive psychology attesting capabilities of human at identifying faces .Most of work to date has been on identity verification like gender, age and emotional state of human face , and only few work have been concerned with combination of human attributes (age, gender and ethnicity) classification in real-time .

A. Age Classification-

Age classification of person from the digital image is still challenging task of the image analysis. Numerous authors have highlighted this research question. However, authors have not achieved much accuracy like human beings.

Yi-Wen Chen, Meng-Ju Han,Kai-Tai Song and Yu-Lun Ho[1] presented experimental analysis of Classification of human- age using facial features. 52 feature point extracted by using Lucas- Kanade image alignment method. These feature points and corresponding located facial area are used to build an active appearance model (AAM). (AMM based feature points).In this paper they used different image processing skill to express features like gray level image ,edge image ,gray image with edge image and horizontal image .These features are sent into a support vector machine (SVM) to estimate the level of age group and achieved 87% accuracy. Experiment on FERET and FG-NET Database.

Asuman GÜNAY [2],presented Estimation of the age exactly for the security system.LBP (Local Binary pattern and histogram of these pattern , LBP histogram are extracted and concatenated into a feature vector. In the classification phased , minimum distance, nearest neighbor and k-nearest neighbor classifier were used for classification and achieved 80% system performance for age estimation . Experiment on FERET and FG-NET Database. Jian-Gang Wang[3],introduced novel Age

categorization method to classify face images into several Age-group. They considered age estimation as a multiclass problem. Aging feature extracted using Gabor feature, LBP, Feature fusion and Adaboost is used for the boosting of ECOC (Error correcting output coding) codes then after combine this code with SVM for the classification of age .when they applied Gabor feature, LBP, Feature fusion to ECOC_SVM then faced memory problem .For reducing these problem adopt PCA to reduce the dimension of LBP and Gabor features respectively before combing into single vector and achieved average accuracy is 85%. Experiment on FERET and FG-NET Database.

Guodong Guo [4], presented that Is Gender recognition affected by human age? For solving this problem they used empirical studies on huge face database of more 8000 images with the ages. Based on these affects on human faces, LBP and Histogram oriented gradients (HOG) methods are used to evaluate for sex categorization with age variation. This paper also used bio-inspired feature for sex recognition.SVM classifier used for the classification of SEX. Accuracies for gender classification in age variations are 86.55% (Young), 95.03% (Adult) and 89.04% (senior).

Ryotatsu Iga[5], developed an algorithm to estimate Gender and Age using (SVM) based on features like geometric arrangement and luminosity of facial images. The graph matching method with GWT method is used to detect the position of the face. GWT features, such as geometric arrangement color, hair and mustache are used for gender estimation. GWT features viz., texture spots, wrinkles, and flabs are used for age estimation.

Young H Kwon[6], presented classification from facial images. The primary features of the face are eyes, nose, mouth, chin, virtual top of the head and sides of the face are computed using ratios to identify young adults and seniors. In secondary feature analysis the wrinkle index computation is used to distinguish seniors from young adults and babies. A combination of primary features and secondary features determine face into one of the three classes viz., babies, young, adults and seniors.

B. Gender classification: -

In terms of gender classification, it is an easy task for humans but it is a challenging task in computer vision. Mostly, to classify the gender, face images are used, because people can easily cover their bodies with clothes.

Len Bui, Dat Tran[7],presented novel method for gender recognition by using 2D principal component analysis for extracting the feature vector and combine these 2D PCA with Support vector machine (SVM) discriminative method for classification. Experiments for this approach have been conducted on FERET data set. Average error rate 4.51%. Huchuan Lu and Hui Lin[8], presented combination of the ellipse face images, Gabor filters, Adaboost learning and SVM classifier. Harr-like feature Gabor feature or ICA method are used to extract facial appearance information and achieved better performance.

Srinivas Gutta [9],presented improved classifier capability by using hybrid approach The hybrid approach consist of an ensemble of RBF networks and inductive decision trees. Experiment of this approach conduct FERET Database and achieved accuracy are 96% on the gender classification task and 94% on the ethnic classification task. C.F Lin [10], presented an approach based on fuzzy support vector machine with good generalization ability. The fuzzy membership function assigned to each input face feature data the degree of one human face is belonging to male or female face The aim of the fuzzification in FSVM is that different contributions to the learning of the decision surface.

B.Moghaddam[11],Used support vector machine(SVM) with radial basis function kernels to classify from low resolution 12*12 “Thumbnail” faces .They used 1755 faces from FERET face database to evaluate the classifier and achieved Classification accuracy of 96%. R.Brunelli[12], used a set of 16 geometric features per image to train two competing networks with the radial basis function,one network for male and other for female and the classification rate was 79% on 168 training image show an error rate 21%. Hui-chang lain[13],presented Multi-view gender classification considering both shape and texture information to represent facial images. The face area is divided into small regions, from which local binary pattern (LBP) histograms are extracted and concatenated into a single vector efficiently representing the facial image. Support Vector Machine (SVM) classifier is used for classification. Guillaume Heusch [14], proposed LBP as an image Pre-processing face authentication for illumination vari-ations.LDA and HMM techniques are used for face authentication. Zehang Sun[15], Proposed gender classification from frontal facial images using genetic feature subset selection. Principal Component Analysis (PCA) is used to represent each image as a feature vector in a low-dimensional space. Genetic algorithms select a subset of features from the low-dimensional representation by disregarding certain eigenvectors that do not seem to encode important gender information. Bayes, Neural Network, SVM, and LDA classifiers are used and compared using Genetic Algorithm feature subset selection.

Ramesha K , K B Raja , Venugopal K R and L M Patnaik[16], proposed new feature based FEBFRGAC algorithm .The geometric features from a facial images are obtained based on the symmetry of the human faces and variation of gray levels, position of eye nose and mouth are located by applying the Canny edge operator. The gender is classified based on posteriori class probability and age is classified based on the shape and texture information using Artificial Neural Network. Gutta [17], Proposed an approach for recognizing the gender, ethnicity and age with facial images. Combination of Gabor filter, Adaboost learning and SVM classifier. Gabor filter banks and Adaboost learning are combined to extract key facial features of each pattern. Then used the Gabor+Adaboost features based SVM classifier to recognize the face image of each pattern.. Christian Ku¨blbeck [18], proposed the new feature set local structure features computed from a 3x3 pixel neighborhood using a modified version of the census transform with four stage classifier .

Each classifier consists of a set of lookup-tables for feature weights. The training of a stage classifier is done using a version of the boosting algorithm.Yoav Freund [19], introduced

the boosting algorithm AdaBoost, and explains the underlying theory of boosting and has also explained how to reduced generalization error, training error and relation to support vector machine

C. Ethnicity Classification-

Ethnicity is an important demographic attribute of human beings, and automatic face-based classification of ethnicity has promising applications in various fields and its identification present yet another challenge in face processing. Several authors have attempted this research problem as a challenge.

Hosoi, S. Takikawa, E. and Kawade, M [20] presented the Gabor wavelets transformation and retina sampling are combined to extract key facial features, and support vector machines that are used for ethnicity classification. Hui Lin, Huchuan Lu and Lihe Zhang[21], presented an MM-LBP (Multi-scale Multi-ratio LBP) method, which is a multimodal method for ethnicity classification. LBP (Local Binary Pattern) histograms are extracted from multi-scale, multi-ratio rectangular regions over both texture and range images, and Adaboost is utilized to construct a strong classifier from a large amount of weak classifiers. Yongsheng Ou, Xinyu Wu, Huihuan Qian and Yangsheng Xu [22], presented the real time race classification system using PCA for feature generation and IDA for feature extraction. For classification they have been used new classifier (combination of SVM classifier).Experiment on FERET database with 750 face images and achieved 82.5%.

In the light of the above literature review it can be concluded that classification of human attributes (age, gender and ethnicity) is not yet present with single classifier. This research is an attempt to build classification engine which will achieve the different goal like human gender, age and ethnicity.

III. PROPOSED METHODOLOGY

In the current work, we have presented an efficient system for human face detection and its attributes recognition like gender, age, ethnicity by single classifier using strong classifier which is collection of weak classifier.

The central contribution of this paper is the classification of attributes of the image search function that takes an arbitrary portion of the image in the space of states {-1, 0, 1}.

$$H : u \to \{-1,0,1\} \quad (1),$$

Where -1 - corresponds to one decision attribute label -1 (e.g., "women"), 1 - corresponds to another decision attribute label 1 (e.g. "man") and 0 - corresponds to abstaining from solutions (unknown).

Our approach is to maximize the use of "bionic" principles of construction of such systems, which have an analogy with the retinal receptive fields and visual cortex. In particular, we use simple tests derived from a nonlocal field of images that can be interpreted as "receptive fields", issuing a code description of the three segments the image and activate the "weak classifiers" and the ensemble of weak classifiers, combined in a committee or a strong classifier can be regarded as an analog artificial neurons located in the visual cortex. At the same time, developed and implemented system has a detection mode online recognition, and learning mode based on "past experience" of the system.

A. Description of the prepared Embodiment.

In the view of previously accomplished work in this direction, nobody has classified human attributes such as Gender, Age and Ethnicity with a single classifier. The overall architecture of the system is described below. We have divided our system in two subsystems. The first is the Facial attribute recognition system and the second is the Attribute classifier training system.

The Attribute recognition system includes a source of data (which may be a video camera or image viewer) as well as a face detection module for cropping face, an attribute recognition module for classifying attributes of the cropped face and a database for simultaneous storing processed images by the upper part of the architecture, i.e. Figure1 (a). In this mode Function (1) is used for processing the image fragment which is being reviewed, obtaining code description and decision-making. The Training subsystem Fig1 (b) includes a source database (which is united with the recognition system database) as well as tools for interaction with the operator, a teaching database and a training module. The overall objective of this subsystem is to build or upgrade Function (1) with further objective of achieving more accurate results in the future. The training system works with "previously acquired experience of the recognition system" which is stored in the image database and which forms a prototype of "short-term memory" in human brain while the "teacher" as an external power in relation to the system corrects the data so that the system is able to build an upgraded version of the classifier which is a prototype of "long-term memory" in human brain. An important feature of the described system is that it constantly updates its classifier. This fact allows the system to adapt to changing environmental conditions within the time and accumulate "acquired experience" in the structure of the strong classifier (1).

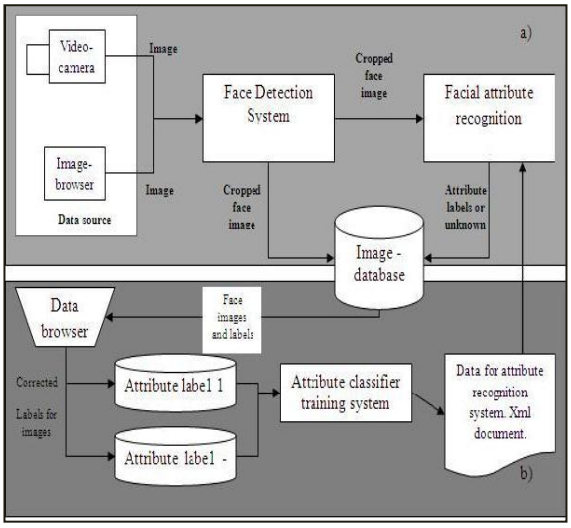


Figure1. Architecture of the recognition system attributes on the facial images a) and learning classifier system attributes b).

An original image which is obtained from a video camera or a viewer enters the face detection system. If a face is

detected on the image, the detector cuts it out and stores it in a specialized image database. Then the facial image is sent to the attribute detector which assigns it a corresponding tag: attribute name or the label "unknown". Later on, using a special program (data browser), the user can check the tags which have been assigned by the attribute recognition system and, if necessary, correct them. Thus, a database of training and updating for the attribute classifier is formed. The attributes are formed using the training procedure and are used by the attribute classifier while the system is working in the recognition mode.

B. Description of Attribute Recognition System

Attributes recognition system is a modified strong classifier in the notation of Viola and Jones [21], consisting of a set of weak classifiers that work with data that are obtained from simple attributes derived from the original image fragment .In Figure 2 ,we have shown the operation of recognition system, we set the size of the found face image in 30x30 pixels, and then convert the image to the integral form [20], to accelerate the computation of characteristic values obtained in this way the image is input to the strong classifier, which analyzes the image and takes one of three solutions.

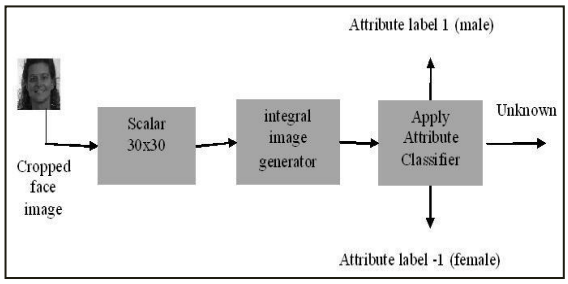


Figure2. Flowchart of Facial attributes recognition system

As shown in the figure, the strong classifier has three possible solutions: attribute label 1, attribute label -1 and the unknown. Solution is defined in DSS (Decision Support System). Thus, it has a strong classifier committee structure, shown in the above figure.

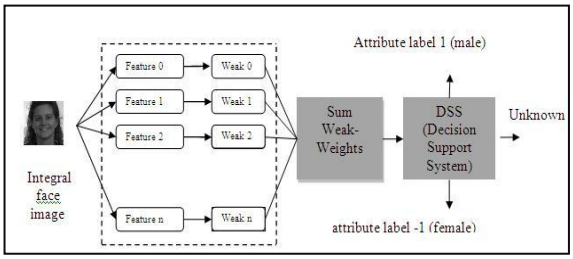


Figure3. The structure of the strong classifier

Simple tests, which are calculated from different parts of the image. Form the coded descriptions of these fragments, the code description to the inputs of weak classifiers, each of which shall decide exactly corresponding to the solution of a strong classifier or committee i.e. an attribute label, attribute label -1 and unknown. In addition, each weak classifier has its own weight, which is summed in the accumulator; the value goes to decision Support System, which makes the final decision on the recognition of the input image.

We consider all the elements of a strong classifier constructing, which includes forming code description of image fragments, the weak classifiers selecting and training.

C. Non-local features based on the Modified Census Transform.

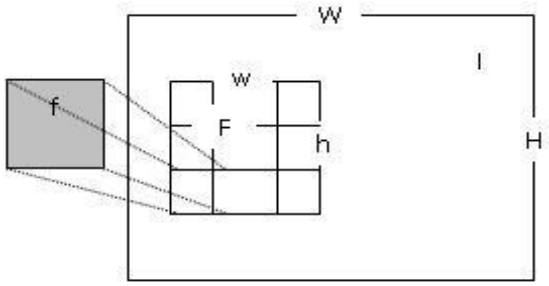


Figure4. Location attribute of the image.

Let there be F (w, h) - a rectangular region image **I** of width w and height **h** pixels. We divide this Image land into 9 equal parts as shown in Figure4 and we analyze the average brightness of each area of the fragment f u, comparing it with the average brightness of the fragment F.

Determine the average brightness of the fragment F as follows:

$$< I_F >= \frac{1}{w \cdot h} \sum_{x=x_o}^{x=x_0+w} \sum_{y=y}^{y=y_0+h} I(x,y), \quad (2)$$

$I(x,y)$ - Brightness of the pixel with coordinates x, y.

Determine the average brightness $< I_{f^n} >$ area f^n area owned fragment F followed as follows:

$$< I_{f^n} >= \frac{9}{w \cdot h} \sum_{x=x_i}^{x=x_i+w/3} \sum_{y=y_j}^{y=y_j+h/3} I(x,y), \quad (3)$$

For each area f^n we encode the brightness of the following rule:

$$c_{f^n} = \begin{cases} 1, & if \ < I_{f^n} > \ \geq \ < I_F > \\ 0, & otherwise \end{cases}, \quad (4)$$

Where c_{f^n} - binary code of brightness for each area of the image fragment F.

In accordance with (4) for each piece of the image we have a set of nine coded bits.

$c_{f^0} \ldots\ldots c_{f^8}$ Sequence which can be regarded as any C code image fragment 'I'. Thus, any rectangular portion of the image, we can describe an integer C, having a 9-bit word length in the decimal system $0 \leq C < 512$.

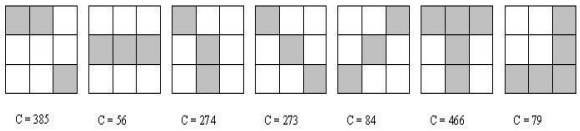


Figure5. Pattern of the codes C, Where the white area of the image coded with 1 and gray coded with 0.

D. Training a weak classifier

We shall consider a weak classifier as a function that takes the code C, the resulting non-local Modified Census Transform in the solution space:

$$h : C \to \{-1,0,1\}, \quad (5)$$

Since the feature value MCT has an integral nature, the function of the decision for the weak classifier (5) is simple lookup table, which contains 512 items that match the codes C.

E. To obtain such a lookup table, we have used the following procedure to study the weak classifier:

1) Suppose you have a database of images each of which is marked attribute label (-1, 1).

2) Let there be two histograms h1 [512] and h2 [512]. Each contains the distribution of the codes featured image database. Where h1 corresponds to the distribution of the codes for a subset of data from the attribute label = 1, and h2 corresponds to the distribution of codes for a subset of data from the attribute label = -1.

3) Then the formation of the decision rule is as follows (6):

$$h[C] = \begin{cases} 1, & h1[C] > h2[C] \\ -1, & h1[C] < h2[C] \\ 0, & h1[C] = h2[C] \end{cases}, \quad (6)$$

Thus, obtained is a weak classifier is fully compliant with our system and its properties are determined exclusively by the input data set.

F. The learning procedure is a strong classifier

We seek a strong classifier (1) as follows:

$$H(x) = F[\sum_{i=1}^{n} \alpha_i h_i(x) + T], \quad (7)$$

Where x- considered fragment, F [v] - a nonlinear function of making solutions, h_i - weak classifier (6), α_i - weight of a weak classifier, T - the threshold of decision-making and n - the number of weak classifiers in the committee. Here we use the following form of the nonlinear function F [v], which is implemented in the DSS (Decision Support System)

$$F[v] = \begin{cases} 1, v > 0 \\ -1, v < 0, \\ 0, v = 0 \end{cases} \quad (8)$$

r de 1 corresponds to the address "attribute label 1", -1, consistent with the decision "attribute label -1" and 0 corresponds to the solution of "unknown".

As we all aware with the well-known amplification method is one way to implement a class of associative machines. Classifiers is operating on the basis of the gain, study the examples belonging to very different distributions. To build a strong classifier (7), we are using the quite famous procedure AdaBoost, is widely known in the literature.

The purpose of using this algorithm is to find the final mapping function or hypothesis H, which will have low or acceptable level of error in this subset of labeled examples of learning. From other algorithms enhance AdaBoost differs as follows:

- AdaBoost adaptively adjusts the error of weak hypotheses returned by the weak learning model. In our case, the decision-making functions (6). Due to this property the algorithm got its name.

- Limiting performance AdaBoost depends on the performance of weak learning model (6) and only those distributions that are actually formed in the learning process.

A typical graph of the learning process of a strong classifier using AdaBoost procedure is shown in Figure 6. The initial data used in a database of men (attribute label 1) and women (attribute label -1), containing 1200 images.

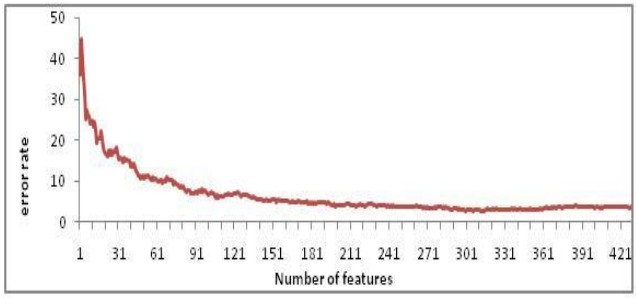


Figure6. Dependence of the overall recognition error of a strong classifier the number of features selected by AdaBoost procedure in the learning process based on data from men-women, containing 1200 images.

IV. EXPERIMENTAL PROCEDURE

In order to estimate the performance of the proposed system, we collected a set of images of human faces from the World Wide Web, FERET and FG-NET.

To study the recognition of attributes, built on a strong classifier (7), we chose several databases.

FERET database to conduct research on the recognition of men and women, database FG-NET for experiments on the recognition of age and a database l fw_funneled to conduct experiments on the recognition of race. Database Features in the following table:

TABLE1. DATABASES FEATURE.

Database	Attribute label 1	No. of Images	Attribute label -1	No. of Images
FERET	Male	785	Female	461
FG-NET	Adult	225	Child	390
RUS-IND	Europeans	910	Asians	646

For each image, face detection system finding the face and its location is stored in an internal database. Figure 1. No manual adjustments of the detector shall not be satisfied, except for removing artifacts search, i.e. sites background, taken for a person. In addition to these databases, for the age recognition experiment we have used our own database of photos which has been collected from World Wide Web contains 800 images of adult and 600 face images of children.

A. The choice of size and scale the face image analysis

Figure7. The characteristic image of the face, resulting detector

The detector has used in this investigation cuts the faces for analyzing from initial picture without hair style and other face features, which may be important for face classification. There is very interesting to investigate by use of computer simulations how the results of face classification depend on these circumstances. For this aim we realize experiments series for changing sizes of images rectangle and the size of the human face in the rectangle considered.

We experimented with nine different sizes of images 24, 27, 30, 33, 36, 39, 42, 45 and 48 pixels. By increasing the size of pixels we have not any significant in the accuracy of the classifier At the same time, learning increases significantly with increasing size of the images, since greatly increased the number of features and weaknesses of the classifiers involved in the training process. On this basis, we recorded the size of the training images by thirty pixels. This changes the scale of the face, increasing the size of the rectangle cut by 10%, 20% and 30%, respectively, we get the following scale: 100% - the initial size of human obtained by the detector are 90%, 80% and 70%. Further size reduction leads to significant areas of the background of the image.

Figure 8. Illustrates the changes in the scale of a person in the picture, depending on the cut size of the rectangle.

100% 90% 80% 70%

Figure8. Zoom in person, depending on the degree of increase found detector rectangle.

To test the dependency of the sharpness of the classification on the scale of the person we used a database FERET (Table 1). Testing results are shown in Figure

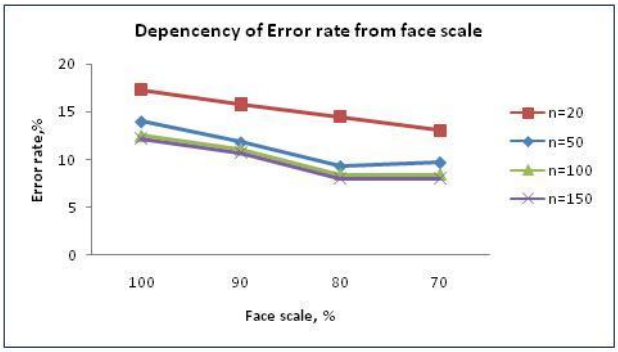


Figure9. Dependency of recognition errors on the scale person in the picture.

The above graph shows that, the face image with the different scale improves the accuracy of the system. The results are shown for different numbers of selected features. Thus in all further experiments, we used an image size of 30x30 pixels, and set the scale of human equal to 80% on the scale found by the detector.

B. Education and research Classifications.

All classifiers were trained and constructed as described above. For their building to used databases that are described in Table 1. Performed several dozen training cycles classifiers on random samples taken from these databases. 70% of the data used for training, the remaining 30% of the data used for testing. We investigated the following properties of classifiers:

- Dependency of recognition errors on the number of the selected algorithm AdaBoost weak classifiers
- Dependency of the ROC for different values of T (7). With regard to the test subset of the original data. ROC curves were constructed for each class individually recognizable.

In addition to these values fixed minimum error values for each of the data sets as learning and testing. A full description of the experiments can be found in the section Results.

V. EXPERIMENTAL RESULTS

A. Classification Performance of the attributes classifier

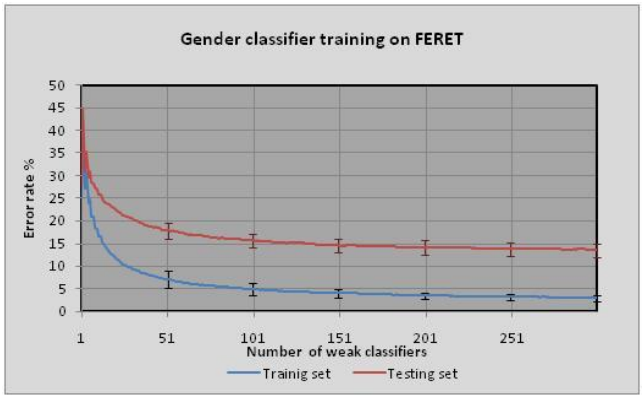
1) GENDER Classifier.

For training and testing database were used FERET, comprising 785 men and 461 images of a woman. 58 cycles were carried out for constructing a classifier on different subsets selected from the initial data at random. As the proportion of 70% of training subset, 30% of the testing subset. The average value of recognition errors, as in the test subset, and in the tutorial can be seen in

We carried out 300 cycles of updating the classifier, for each of the random data sets. It was found that the value of recognition errors on a test set of changes little after 250 weak

classifiers, while on the training dataset, the value of recognition errors continued to fall.

Averaged dependence of the FRR (FAR) for different values of T (7), shown in Figure 11.



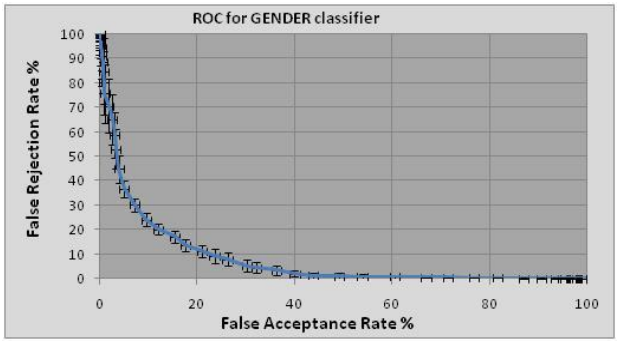


Figure11. Average ROC for GEDNDER classifier, based on FERET.

Figure12. Result of Gender Classifier
Blue- MALE, Red-FEMALE

2) AGE Classifier.

For training and testing database ware used FG-NET and the database collected during the study. The database includes 905 images of children and 722 adults image. The database is significantly more complex compared to a database FERET because the quality of images used in the studies of a lower light conditions and angle are not controlled. As in the previous case was carried out 58 cycles of building a classifier on different subsets selected from the initial data at random.

The proportion of 70% of training l subset, 30% of the test subset. The average value of recognition errors, as in the test subset, and in the tutorial can be seen in Figure 13.

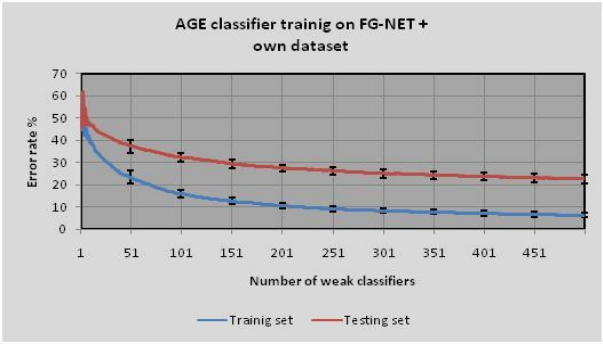


Figure13. Training process for AGE classifier.

We carried out 500 cycles of updating the classifier for each of the random data sets. It was found that the results of the classifier is somewhat worse than on the database FERET, but no trend for an end to learning have been identified. Averaged dependence of the FRR (FAR) for different values of T (7), shown in Figure 14.

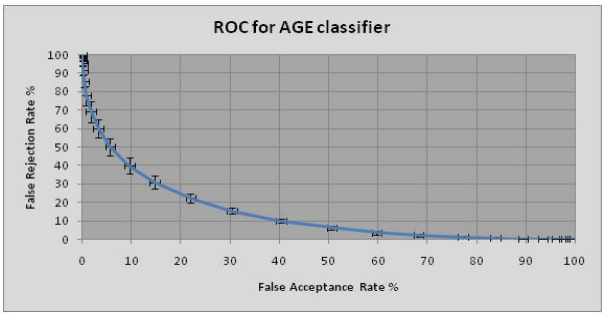


Figure14. Average ROC for AGE classifier, based on FG-NET

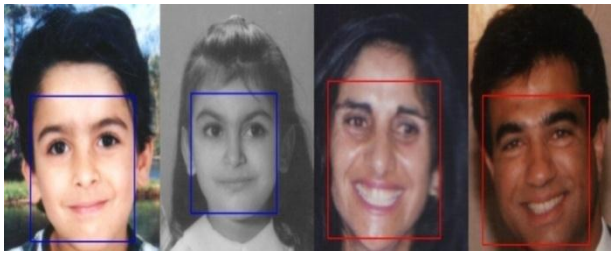


Figure15. Classification of age attributes on FG-NET.
Blue-CHILD, Red -ADULT

3) Ethnicity Classifier

To train our classifier for recognition of ethnicity we have used Indian Face Database [22] and the database collected during the study in Applied Physics Institute RAS. The database includes 910 images of Europeans and 722 Indians. As in the previous case was carried out 20 cycles of building a classifier on different subsets selected from the initial data at random. The ratio of 70% of training subset, 30% of the test subset. The average value of recognition errors, as in the test subset, and in the tutorial can be seen in Figure 16.

We carried out 200 cycles of updating the classifier for each of the random data sets. It was found that the results of the classifier are somewhat better than on the database FERET. Averaged dependence of the FRR (FAR) for different values of T (7), shown in Figure 17.

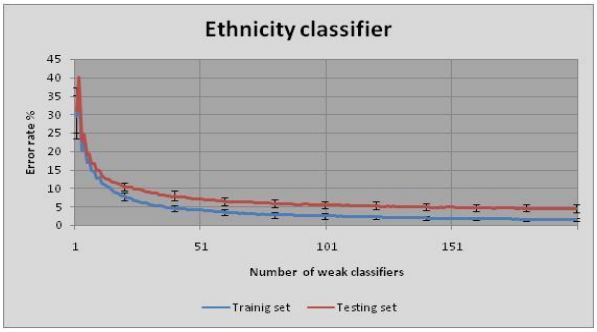


Figure 16.Training process for Ethnicity classifier.

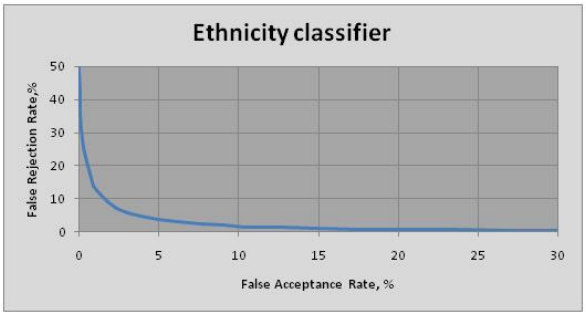


Figure 17. ROC for Ethnicity classifier.

As we all know that, a system with artificial intelligence cannot do work like human natural intelligence but still researches are moving towards the automation or minimize the human efforts. Therefore face recognition and computer vision are the active area of the research.

TABLE 2- RECOGNITION ACCURACY ACHIEVED

Human Attributes	Achieved Accuracy
Gender(Male, Female)	89%
Age(Adult, Child)	80%
Ethnicity	96%

VI. FINDINGS & DISCUSSION

In the present study was formulated and investigated neuron algorithm to automatically construct a classifier that separates the two-dimensional visual image into two classes according to some binary characters. Algorithm is formed from the available data and is quite universal because it uses a number of general principles of teaching:

a) Partitioning an existing database on two samples - a training and test,

b) Creation of a "strong" classifier in the form of a committee of the primary "weak" classifiers, whose voices in the decision on the committee are recorded in accordance with rating scales for each,

c) Analysis of errors of the classifier on the number of members on the committee,

d) Control of the invariance result with respect to partition the database into training and test samples.

Established authors program allows an effective and unique tool for computational and experimental determination of optimal values of those parameters, to find that speculative way is not possible. The result is determined by the brute force of several options, the calculation of total errors for each of them, and selection of options with the least error. As examples, which are important for practical applications in this paper, the study addressed the question of optimization of two parameters of the problem:

a) The scale of images used to create a classifier

b) The size of the face image on the compared fragment.

The results are shown in Fig. 9.

As shown in Fig. 10.12, after the number of items the committee is more than 50 error and the learning and test sample subside slowly, and the error on the test sample is 3-5 times less than on learning. Such behaviors of the curves are not exactly conventional ideas and are a new result that requires further study. It is not clear as to why the error curve for training database does not tend to zero.

VII. CONCLUSION & FUTURE WORK

The study showed versatility and promise of the proposed approach to the description of such complex objects as the human face, but at the same time allowed to see the boundaries of the new method. First, a simple increase in the number of elements in does not automatically lead to decay to zero errors, not only for the test sample, but also the learning part of the database. This is indicative of the fact that the quality of the classifier, constructed by the proposed method can and should be improved.

The above findings allow us to identify ways of further improving the method of Classification and enhance its effectiveness.

1. Firstly, we should construct a two-level classifier, which will be a committee of "strong" classifiers trained on different data. Such a qualifier in the second level can be built from first-level classifier for the scheme, similar to how the first-level classifiers ("strong" classifiers) are constructed from the original zero-level classifiers (weak classifiers).

2. Secondly, to improve the classification results should lead to introduce an additional threshold and extending the range of values of the internal variable, where the classifier says "I do not know.

Application of the elements with two thresholds which have three positions in voting (yes, no, I do not know), allows to build consistent and achieve better results in classification, if we can effectively address their learning according to a reasonable time.

Third, a wide enough field to improve the level of classification should be receptive to the use of elements of another type, than used here type"census". It is possible use elements with Gabor functions receptive structures or Haar functions receptive structure. Created soft tools allow make computer experiments for these variants comparing and choosing the best.

VIII. ACKNOWLEDGEMENT

The authors would like to thank Department of Science and Technology, Government of India and Russian Foundation of

Basic research for supporting the research under Indo-Russian bilateral arrangement through the project sanction number **DST/RFBR/P-69**

REFERENCES

[1] Yi-Wen Chen ; Meng-Ju Han ; Kai-Tai Song ; Yu-Lun Ho.: Image-based age-group classification design using facial features. In: System Science and Engineering (ICSSE), 2010 International Conference on 2010 , Page(s): 548 – 552.

[2] Asuman GÜNAY, Nabiyev, V.V.: Automatic Age Classification with LBP. In: Computer and Information Sciences, ISCIS '08. 23rd International Symposium 2008 , Page(s): 1 – 4

[3] Jian-Gang Wang; Wei-Yun Yau; Hee Lin Wang.: Age Categorization via ECOC with Fused Gabor and LBP Features. In: Applications of Computer Vision (WACV), 2009 Workshop 2009, Page(s): 1 - 6

[4] C.F. Lin, S.D Wang .:Fuzzy Support vector Machines. In: IEEE Transactions on Neural network, Vol.13,No.2,pp.464-471,Mar.2002

[5] Young H. Kwon and Niels Da Vitoria Lobo.: Age Classification from Facial Images. In: Journal of Computer Vision and Image Understanding, vol. 74, no. 1, pp. 1-21, April 1999

[6] Fukai, H.; Takimoto, H.; Mitsukura, Y.; Fukumi.: Age and gender estimation by using facial image M. In: Advanced Motion Control, 2010 11th IEEE International Workshop 2010 , Page(s): 179 - 184

[7] Guillaume Heusch,Yann Rodriguez and Sebasttien Marcel.: Local Binary Patterns as an Image processing for face Authentication. In: Proceedings of seventh International conference on Autometic Face and Gesture Recognition,pp.6-14.April 2006.

[8] Len Bui; Dat Tran; Xu Huang; Chetty, G.: Face Gender Recognition Based on 2D Principal Component Analysis and Support Vector Machine. In: Network and System Security (NSS), 2010 4th International Conference 2010 , Page(s): 579 - 582

[9] Huchuan Lu; Hui Lin.:Gender Recognition using Adaboosted Feature. In : Natural Computation, 2007. ICNC 2007. Third International Conference on Volume: 2 2007 , Page(s): 646 - 650

[10] B.Moghaddam and M.H Yang.:Learning Gender with support faces. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.24, No.5,pp. 707-711,2002.

[11] R.Brunelli and T.Poggio.:HyperBF Networks for gender classification.In: Proceedings of the DARPA Image Understanding Workshop,Pages 311-314.1992.

[12] Hui-Cheng Lian and Bao-Liang L.: Multi-view Gender Classification Using Local Binary Patterns and Support Vector Machines. In: Advances in Neural Networks - ISNN 2006.

[13] Hui Lin; Huchuan Lu; Lihe Zhang.: A New Automatic Recognition System of Gender, Age and Ethnicity.In: Intelligent Control and Automation, 2006. WCICA 2006. The Sixth World Congress on Volume: 2 ,2006 , Page(s): 9988 - 9991

[14] Christian Ku¨blbeck *, Andreas Ernst.: Face detection and tracking in video sequences using the modified census transformation. In: Image and Vision Computing 24 (2006) 564–572

[15] Ramesha K et al.: Feature Extraction based Face Recognition, Gender and Age Classification. In: (IJCSE) International Journal on Computer Science and Engineering,Vol. 02, No.01S, 2010, 14-23

[16] Hosoi, S.; Takikawa, E.; Kawade, M.: Ethnicity estimation with facial images Automatic Face and Gesture Recognition. In: Proceedings. Sixth IEEE international Conference 2004, Page(s): 195 - 200

[17] Gutta, S.; Wechsler, H.:Gender and ethnic classification of human faces using hybrid classifiers Neural Networks. In: International Joint Conference on Volume: 6 ,1999 , Page(s): 4084 - 4089 vol.6

[18] Zehang Sun, George Bebis, Xiaojing Yuan, and Sushil J. Louis.: Genetic Feature Subset Selection for Gender Classification: A Comparison Study.In: IEEE Workshop on Applications of Computer Vision, pp.165-170, 2002.

[19] Guodong Guo; Dyer, C.R.; Yun Fu; Huang, T.S.: Is gender recognition affected by age? In: Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International CONFERENCE 2009, Page(s): 2032 - 2039

[20] Y. Freund and R. E. Shapire.:A short introduction to boosting. In: Journal of Japanese Society for Artificial Intelligence, pages 771–780, 1999

[21] Paul Viola, Michael Jones.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2001.

[22] Indian Face Database. http://vis-www.cs.umass.edu/~vidit/IndianFaceDatabase/

AUTHORS PROFILE

Nikolai Sergeevich Bellustin received his graduate degree from radio physics faculty the University of Nizhny Novgorod. He defended the candidate thesis on the different theme. He defended his doctoral thesis. At present time he is a professor of Nizhny Novgorod technical university. The author has worked on various areas of scientific interests:

Yuri Dmitrievich Kalafati graduated from the Moscow physico-technical institute He defended the candidate thesis on the different theme. He defended his doctoral thesis. The author has worked on various areas of scientific interests:

Olga Vladimirovna Shemagina received his graduate degree from radio physics Nizhny Novgorod state university in 1996. He is the author of 15 printed works. Area of scientific interests: image processing and Computer vision.

G. Yakhno graduated from radio physics faculty Nizhny Novgorod state university in 1969. He defended the candidate thesis in 1977, on the theme "the Structure of collective activity in the excitable media". He defended his doctoral thesis in 1999 on the theme of "Transformation of the information in neuron systems". At present time he is a professor of Nizhny Novgorod state university. The author has work more than 150 areas of scientific interests: nonlinear dynamics in cognitive research.

Alexander Telnykh received his graduate degree from radio physics Nizhny Novgorod state university. He defended the candidate thesis in 2009 on the theme "Mathematical models neuron-like environments for the development of systems for the detection and recognition of objects set of classes". The author has work more than 30 areas of scientific interests: image processing, computer vision etc

Andrey V. Kovalchuck graduated from radio physics faculty the University of Nizhny Novgorod in 2006 is a graduate student. The author has worked on more than 18 areas of scientific interests: image processing, computer vision.

Abhishek Vaish received his PhD from Indian Institute of Information Technology, Allahabad India. He has done various researches on security domain.

Pinki Sharma received her M.Tech (CSE) degree from Bansathali University, India and she is currently Research Scholar at Indian Institute of Information Technology, Allahabad.

Mining Volunteered Geographic Information datasets with heterogeneous spatial reference

Sadiq Hussain
System Administrator, Examination Branch
Dibrugarh University, Dibrugarh, India

Prof. G.C. Hazarika
Director i/c, Centre for Computer Studies
Dibrugarh University, Dibrugarh, India

***Abstract*—When the information created online by users has a spatial reference, it is known as Volunteered Geographic Information (VGI). The increased availability of spatiotemporal data collected from satellite imagery and other remote sensors provides opportunities for enhanced analysis of Spatiotemporal Patterns. This area can be defined as efficiently discovering interesting patterns from large data sets. The discovery of hidden periodic patterns in spatiotemporal data could provide unveiling important information to the data analyst. In many applications that track and analyze spatiotemporal data, movements obey periodic patterns; the objects follow the same routes (approximately) over regular time intervals. However, these methods cannot directly be applied to a spatiotemporal sequence because of the fuzziness of spatial locations in the sequence. In this paper, we define the problem of mining VGI datasets with our already established bottom up algorithm for spatiotemporal data.**

***Keywords- data mining; periodic patterns; spatiotemporal data; Volunteered Geographic Information*.**

I. INTRODUCTION

There is an explosion of geographic information generated by individuals on the Web. Users provide geotagged photos and tweets, geotag Wikipedia articles, create gazetteer entries, update geographic databases like OpenStreetMap (OSM) and much more. Such user-generated geodata, also called Volunteered Geographic Information, VGI [1], is becoming an important source for geo-services like map generation, routing, search, spatial analysis and mashups. Different from traditional geodata, VGI often has no distinct classifying attributes or explicit taxonomy. Users are free to create new tagging schemas or add new properties or text. Although some schema checks may exist on the editor level through auto-completion or templates, these checks are not strict and can be ignored by the user. Analyzing the dynamic and heterogeneous schemas of VGI to find common conceptualizations is an important and complex task. For example, Deng et al. [2] use density based clustering and a document term matrix to find conceptualizations in geotagged Flickr images. Edwardes and Purves [3] explore the potential to develop a hierarchy of place concepts based on co-occurring characteristic terms in the description of geotagged photos of the British Isles. Extracting and exploring concepts is an important prerequisite to analyze the quality and consistency of a dataset and to evaluate its "fitness for use" [4]. We describe our work on using frequent pattern mining to extract and explore conceptualizations of VGI. Frequent pattern mining is used for effective classification in association rule mining [5]. Afrati et al. [1] use frequent sets to find approximate patterns, which is a promising technique for concept extraction and exploration. For geospatial data, frequent pattern mining is used to determine spatial association rules [6] and to perform co-occurrence analysis [5]. In our approach we transform VGI into a flat model of transaction objects, which can be input to our mining algorithms. Different from transactions of market basket data, which are the typical input to frequent pattern mining, geospatial patterns may occur rarely in a dataset but are nevertheless interesting. Ding et al. (2006) introduce a framework to mine regional association rules based on prior clustering to find patterns in sub regions. However, to extract concepts, mining sub regions is not an option. We explain what extensions to frequent pattern mining are needed to deal with the scale-dependency and introduce a bottom-up mining approach based on quadtrees. We developed a prototype framework to mine the frequent patterns apriori, which then can be efficiently accessed by clients. For this, we describe the OSM Explorer, which visualizes frequent patterns in the OSM dataset and performs data consistency and quality checks.

II. TRANSACTION MODEL

To employ frequent pattern mining to extract concepts from VGI, the heterogeneous geographic information needs to be transformed into transactions. A transaction has an associated set of items and is input record frequent pattern mining. We view each geoobject as a transaction having geometry and a set of attributes. Attributes can be key-value pairs (representing an attribute name and value) or just keys (like tags). Text has to be itemized first. For example, by using frequency term vectors or by extracting named entities, a text describing geographic information can be transformed into a set of attributes. In general, a geoobject is represented as a transaction as follows:

Transaction (ObjID, Geometry, List((Key, [Value])))

By determining frequent itemsets from such transactions one obtains frequent patterns of attribute names (if key is the name of a property), tags (if key is a tagname) or words (if key is the word of a frequency term vector). These frequent patterns cannot yet be seen as concepts, but they are good candidates for building concept hierarchies and classification models in a subsequent step. The result of some frequent

patterns in the OSM data, which can be interpreted as collaborative generated schemas for geographic concepts, is illustrated in Figure 1. The above process is discussed in more detail in [7].

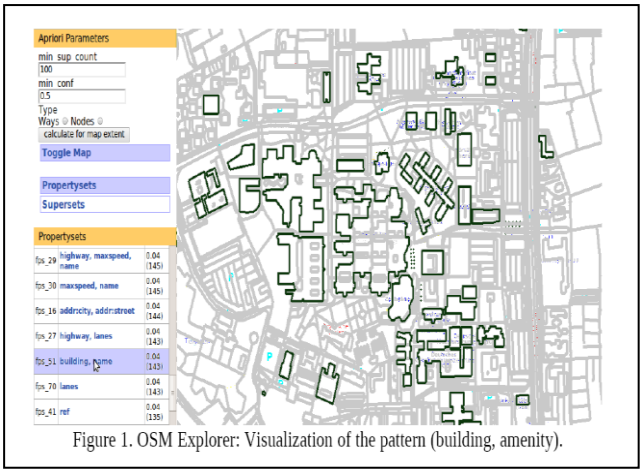


Figure 1. OSM Explorer: Visualization of the pattern (building, amenity).

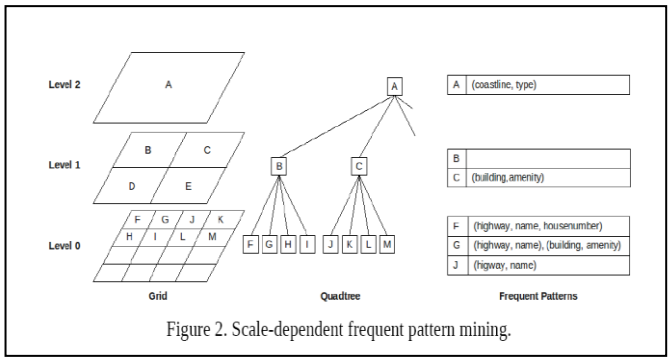


Figure 2. Scale-dependent frequent pattern mining.

III. Scale-Dependent Mining

A pattern is called frequent if it has a minimum support, that is, it occurs a minimum number of times in a given dataset. Patterns that only occur rarely are not considered. However, geospatial patterns occurring with a low frequency in a dataset can still be interesting for concept extraction. This is either because they are 1) densely clustered (and thus may represent a local/regional pattern on a large scale) or 2) they are widely distributed (and thus may represent a pattern on a small scale). We use a bottom-up approach based on a quadtree data structure to determine which items are candidates for itemset generation on a certain scale, as shown in Figure 2.

The items of every transaction in each leaf node of the quadtree (which constitutes a grid of cells over the input data space) are counted. The items that occur in a cell with at least a minimum frequency are used to generate frequent itemsets over the transactions within this cell. On the next higher level all items that have not been used so far are summarized. If they reach the minimum frequency they are input to frequent itemset mining at this level. This step is repeated until the root node is reached. The determined itemsets are linked to the according nodes in the quadtree, which then also allows for fast exploration, for example, via a map interface.

IV. Periodic Patterns in Object Trajectories

This section defines the problem of mining periodic patterns in spatiotemporal data. First, we motivate our research by discussing why previous work on event sequences is not expected to perform well when applied on object trajectories. We then proceed to a formal definition of the problem.

In our model, we assume that the locations of objects are sampled over a long history. In other words, the movement of an object is tracked as an n-length sequence S of spatial locations, one for each timestamp in the history, of the form $\{(l_0, t_0), (l_1, t_1), \ldots, (l_{n-1}, t_{n-1})\}$, where li is the object's location at time ti. If the difference between consecutive timestamps is fixed (locations are sampled every regular time interval), we can represent the movement by a simple sequence of locations l_i (i.e., by dropping the timestamps t_i, since they can be implied). Each location l_i is expressed in terms of spatial coordinates. Figure 3a, for example, illustrates the movement of an object in three consecutive days (assuming that it is tracked only during specific hours, e.g., (working hours). We can model it with sequence S = {<4, 9>, <3.5, 8>, . . . ,<6.5, 3.9>, <4.1, 9>, . . . }. Given such a sequence, a minimum support min_sup ($0 < \text{min\_sup} \leq 1$), and an integer T, called period, our problem is to discover movement patterns that repeat themselves every T timestamps. A discovered pattern P is a T-length sequence of the form $r_0 r_1 \ldots r_{T-1}$, where ri is a spatial region or the special character *, indicating the whole spatial universe. For instance, pattern AB*C** implies that at the beginning of the cycle the object is in region A, at the next timestamp it is found in region B, then it moves irregularly (it can be anywhere), then it goes to region C, and after that it can go anywhere, until the beginning of the next cycle, when it can be found again in region A. The patterns are required to be followed by the object in at least α (α = min_sup · [n/T]) periodic intervals in S.

Problem Definition

Let S be a sequence of n spatial locations $\{l_0, l_1, \ldots, l_{n-1}\}$, representing the movement of an object over a long history. Let T << n be a user specified integer called period (e.g., day, week, month). A periodic segment s is defined by a subsequence $l_i\, l_{i+1} \ldots l_{i+T-1}$ of S, such that i modulo T = 0. Thus, segments start at positions 0, T, . . . , ([n/T] − 1) · T, and there are exactly m = [n/T] periodic segments in S^1. Let s^j denote the segment starting at position lj·T of S, for $0 \leq j < m$, and let $s_i{}^j = l_j \cdot T + i$, for $0 \leq i < T$.

Definition The mining periodic patterns problem searches for all valid periodic patterns *P* in *S*, which are frequent and non-redundant with respect to a minimum support min_ sup. For simplicity, we will use "frequent pattern" to refer to a valid, non-redundant frequent pattern.

V. Mining Periodic Patterns

In this section, we present techniques for mining frequent periodic patterns and their associated regions in a long history of object trajectories. We first address the problem of finding frequent 1-patterns (i.e., of length 1). Then, we propose one method to find longer patterns; a bottom-up, level-wise technique.

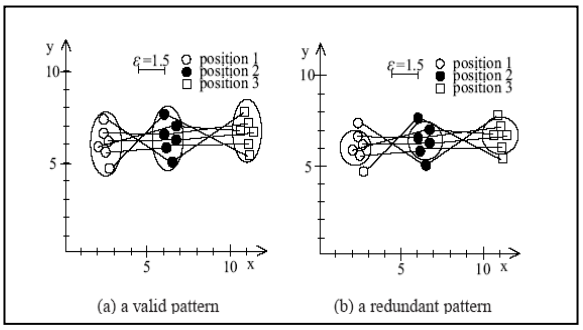


Figure 3: Redundancy of patterns

A. Obtaining frequent 1-patterns

Including automatic discovery of regions in the mining task does not allow for the direct application of techniques that find patterns in sequences (e.g., [8]), as discussed. In order to tackle this problem, we propose the following methodology. We divide the sequence S of locations into T spatial datasets, one for each offset of the period T. In other words, locations $\{l_i, l_{i+T}, \ldots, l_{i+(m-1)\cdot T}\}$ go to set R_i, for each $0 \le i < T$. Each location is tagged by the id $j \in [0, \ldots, m-1]$ of the segment that contains it. Figure 4a shows the spatial datasets obtained after decomposing the object trajectory. We use a different symbol to denote locations that correspond to different periodic offsets and different colors for different segment-ids. Observe that a dense cluster r in dataset R_i corresponds to a frequent pattern, having $*$ at all positions and r at position i. Figure 4b shows examples of five clusters discovered in datasets R_1, R_2, R_3, R_4, and R_6. These correspond to five 1-patterns (i.e., $r_{11}*****$, $*r_{21}****$, etc.). In order to identify the dense clusters for each R_i, we can apply a density-based clustering algorithm like DBSCAN [9]. Clusters with less than α ($\alpha = min\_sup \cdot m$) points are discarded, since they are not frequent 1-patterns according to our definition. Clustering is quite expensive and it is a frequently used module of the mining algorithms, as we will see later. DBSCAN [9] has quadratic cost to the number of clustered points, unless an index (e.g., R–tree) is available. Since R–trees are not available for every arbitrary set of points to be clustered, we use an efficient hash-based method.

B. A level-wise, bottom-up approach

Starting from the discovered 1-patterns (i.e., clusters for each R_i), we can apply a variant of the level-wise Apriori-TID algorithm [10] to discover longer ones. The input of our algorithm is a collection L_1 of frequent 1-patterns, discovered as described in the previous paragraph; for each R_i, $0 \le i < T$, and each dense region $r \in R_i$, there is a 1-pattern in L_1.Pairs $\langle P_1, P_2 \rangle$ of $(k-1)$-patterns in L_{k-1}, with their first $k-2$ non-$*$ regions in the same position and different $(k-1)$-th non-$*$ position create candidate k-patterns. For each candidate pattern P_{cand}, we then perform a segment-id join between P_1 and P_2, and if the number of segments that comply with both patterns is at least $min\_sup \cdot m$, we run a pattern validation function to check whether the regions of P_{cand} are still clusters. After the patterns of length k have been discovered, we find the patterns at the next level, until there are no more patterns at the current level, or there are no more levels.

C. Algorithm Level-wise Pattern Mining (L1, T, min sup);

k:=2;
while $(\mathcal{L}_{k-1} \neq \emptyset \wedge k < T)$
$\mathcal{L}_k$:=$\emptyset$;
Generation of quadtree
Dynamic Time Warping
for each pair of patterns $(P_1, P_2) \in L_{k-1}$
such that P_1 and P_2 agree on the first $k-2$
and have different $(k-1)$-th non-$*$ position
P_{cand}:=**candidate gen**(P_1, P_2);
if ($P_{cand} \neq$ *null*) **then**
P_{cand}:=$P_1 \bowtie_{P_1.sid=P_2.sid} P_2$; //segment-id join
if $(|P_{cand}| \ge min\_sup \cdot m)$ **then**
validate_pattern$(P_{cand}, \mathcal{L}_k, min\_sup)$;
k:=$k+1$;
return $\mathcal{P}$:=$\bigcup \mathcal{L}_k, \forall 1 \le k < T$;

function validate_pattern$(P_{cand}, \mathcal{L}_k, min\_sup)$;

1). $split$:=false; $prev\_size$:=$|P_{cand}|$
2). **for each** non-$*$ position i of P_{cand}
3). cluster points of R_i with $sid \in P_{cand}$;
4). **if** (more than one clusters with size $\ge min\_sup \cdot m$) **then**
5). $split$:=true;
6). **for each** cluster r with size $\ge min\_sup \cdot m$
7). P'_{new}:=$\{sid \mid sid \in r\}$;
8). **validate_pattern**$(P'_{cand}, \mathcal{L}_k, min\_sup)$;
9). **else** P_{cand}:=segment-ids in updated cluster r;
10). **if** $(\neg split)$**then**
11). **if** $(|P_{cand}| \ge min\_sup \cdot m)$ **then**
12). **validate_pattern**$(P_{cand}, \mathcal{L}_k, min\_sup)$;
13). **else** $\mathcal{L}_k$:=$\mathcal{L}_k \cup P_{cand}$;

14) Use Z test

To illustrate the algorithm, consider the 2-patterns $P_1 = r_{1x}r_{2y}*$ and $P_2 = r_{1w}*r_{3z}$ of Figure 3a. Assume that *MinPts* = 4 and ε = 1.5. The two patterns have common first non-$*$position and *Minimum Bounding Rectangles*(r_{1x}) overlaps *Minimum Bounding Rectangles*(r_{1w}). Therefore, a candidate 3-pattern P_{cand} is generated. During candidate pruning, we verify that there is a 2-pattern with non-$*$ positions 2 and 3 which is in L_2.Indeed, such a pattern can be spotted at the figure (see the dashed lines). After joining the segmentids in P_1 and P_2 , P_{cand} contains the trajectories shown in Figure 5b. Notice that the locations of the segment-ids in the intersection may not form clusters any more at some positions of P_{cand}. This is why we have to call **validate pattern**, in order to identify the valid patterns included in P_{cand}. Observe that, the segment-id corresponding to the lowermost location of the first position is eliminated from the cluster as an outlier. Then,

while clustering at position 2, we identify two dense clusters, which define the final patterns $r_{1a}\, r_{2b}\, r_{3c}$ and $r_{1d}\, r_{2e}\, r_{3f}$.

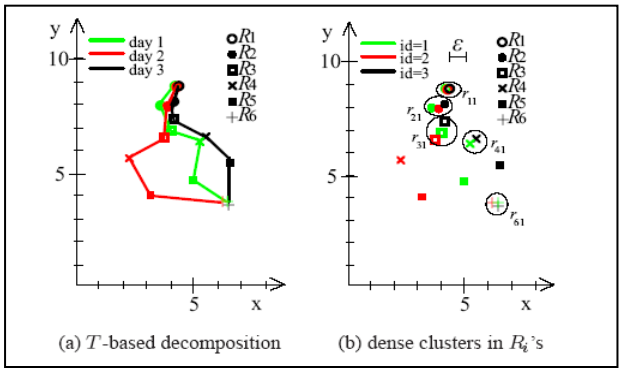


Figure 4: locations and regions per periodic offset

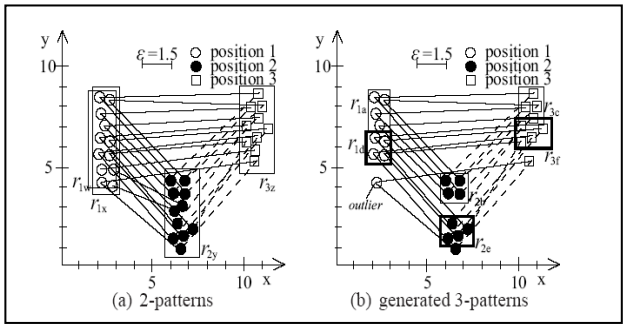


Figure 5: Example of the proposed Algorithm

VI. Conclusion and Future Work

Topics for future work include the automatic discovery of the period T related to frequent Periodic patterns and the discovery of patterns with distorted period lengths. For instance, the movement of an object may exhibit periodicity; however, the temporal length of the period may not be fixed but could vary between pattern instances. Public transportation vehicles may have this type of periodicity, since during heavy traffic hours, a cycle can be longer that usual. Building indexes based on distorted and shifted patterns is also an interesting direction for future work.

Secondly, a motivation for our work is the fast and efficient integration of heterogeneous user-generated geodata and to merge all information available for certain geographic objects. Another motivation is to help users using VGI based on automatically generated quality measures and extracted concepts. A lot work needs to be done regarding the transformation of textual descriptions into transaction objects, and an evaluation of discovered patterns for several data sources needs to be conducted.

References

[1] Afrati F, Gionis A, Mannila H, 2004, Approximating a collection of frequent sets. Proceedings of KDD '04, 12-19

[2] Deng D P, Chuang T R, Lemmens R, 2009, Conceptualization of Place via Spatial Clustering and Cooccurrence Analysis. Proceedings of the International Workshop on Location Based Social Networks, 49-55

[3] Edwardes A J and Purves S, 2007, A theoretical grounding for semantic descriptions of place. Proc. 7th Intern. Symp. on Web and Wireless Geographical Information Systems, LNCS 4857, 106-121

[4] Gervais M, Bédard Y et al., 2009, Data Quality Issues and Geographic Knowledge Discovery. In: Miller H J, Han J (eds), Geographic Data Mining and Knowledge Discovery, 99-115 Goodchild M, 2007, Citizens as sensors: the world of volunteered geography. GeoJournal 69(4):211-221

[5]. Han J, Gao J, 2009, Research challenges for Data Mining in Science and Engineering. In: Kargupta H, Han J et al. (eds), Next Generation of Data Mining, 3-27

[6] Koperski K and Han J, 1995, Discovery of Spatial Association Rules in Geographic Information Databases. Proc. 4th Intern. Symp. on Advances in Spatial Databases, LNCS 951, 47-66 Liu B, Hsu W, Ma Y, 1998, Integration of classification and association rule mining. Proc. KDD '98: 80-86

[7] Sengstock C, Gertz M, 2010, Anwendung von Frequent Itemset Mining auf nutzergenerierte Geodaten. Geoinformatik 2010, Kiel, 28-36

[8] J. Han, G. Dong, and Y. Yin. Efficient mining of partial periodic patterns in time series database. In Proc. of International Conference on Data Engineering, pages 106–115, 1999.

[9] M. Ester, H. P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In Proc. of ACM Knowledge Discovery andData Mining, pages 226–231, 1996.

[10] R. Agrawal and R. Srikant. Fast algorithms for mining association rules. In Proc. of VeryLarge Data Bases, pages 487–499, 1994.

Authors Profile

Sadiq Hussain MCA from Tezpur University, Assam,India in the year 2000 with CGPA 7.85. Currently, he is working as System Administrator of Dibrugarh University. He is in this position since December, 2008. He is in the charge of Computerization of Examination System and MIS of Dibrugarh University.

Prof. G.C. Hazarika
Date of birth : 01-01-1954
Academic Qualification: M.Sc. (Math.), Ph.D. (Math).
Positions held :
Director i/c, Centre for Computer Studies, Dibrugarh University, and Professor, Department of Mathematics, Dibrugarh University
Academic Positions held:
a. Computer Programmer: Joined as Computer Programmer, Dibrugarh University Computer Centre in Dec, 1977 and served till April, 1985.
b. Lecturer: Joined as Lecturer in the Department of Mathematics, Dibrugarh University in April, 1985.
c. Reader: Joined as Reader in a regular post in June, 1990.
d. Professor: Joined as Professor in a regular post in August, 1998.
Publications (a few)
1.Magnatic effect on flow through circular tube of non-uniform cross section with permeable walls
- Applied Science Periodical Vol. V. No.1, February, 2003
Jointly with B.C. Bhuyan.
2.Influence of Magnetic filed on Separation of a Binary Fluid Mixture in Free Convection flow Considering Soret Effect
- J. Nat. Acad. Math. Vol. 20 (2006), pp. 1-20
Jointly with B.R. Sharma and R.N. Singh
3. Effects of Variable viscosity and Thermal Conductivity on flow and heat transfer of a Stretching Surface of a rotating micropolar fluid with suction and blowing
- Bull. Pure and Appl. Sc. – Vol.-25 E No. 2, PP-361-370, 2006.
Jointly with P.J. Borthakur.
4. Effects of Variable viscosity and Thermal Conductivity on boundary Layer flow and heat transfer of micropolar fluid near an axisyusmetric Stagnation point on a moving cylinder- Proc. 51st .cong. of ISTAM, Dec-2006.
Research experiment:
Have guiles 11 Ph. D students and 9 M Phil students

Method for Extracting Product Information from TV Commercial

Kohei Arai
Information Science Department
Saga University
Saga, Japan

Herman Tolle
Software Engineering Department
Brawijaya University
Malang, Indonesia

***Abstract* — Television (TV) Commercial program contains important product information that displayed only in seconds. People who need that information has no insufficient time for noted it, even just for reading that information. This research work focus on automatically detect text and extract important information from a TV commercial to provide information in real time and for video indexing. We propose method for product information extraction from TV commercial using knowledge based system with pattern matching rule based method. Implementation and experiments on 50 commercial screenshot images achieved a high accuracy result on text extraction and information recognition.**

Keywords - text detection; information extraction; rule based classifying; patern matching;

I. INTRODUCTION

Nowadays, people use the information from Television (TV) commercial program as a reference before they buy the product. Since TV commercial only displays the important information in seconds, people who need such information has no insufficient time for noted it or even just for reading that information. In Japan, there is a company provide services for reading, note and distribute such information from TV commercials, but they still did it in manually. If these text occurrences could be detected, segmented and recognized automatically, it would be a valuable source for information extraction, indexing or retrieval.

The text information extraction (TIE) problem can be divided into the following sub-problems: (i) detection, (ii) localization, (iii) tracking, (iv) extraction and enhancement, and (v) recognition [1]. A TIE system receives an input in the form of a still image or a sequence of images. The images can be in gray scale or color, compressed or un-compressed, and the text in the images may or may not move. Text detection refers to the determination of the presence of text in a given frame (normally text detection is used for a sequence of images). Text localization is the process of determining the location of text in the image and generating bounding boxes around the text. Text tracking is performed to reduce the processing time for text localization and to maintain the integrity of position across adjacent frames. Although the precise location of text in an image can be indicated by bounding boxes, the text still needs to be segmented from the background to facilitate its recognition. This means that the extracted text image has to be converted to a binary image and enhanced before it is fed into an optical character recognition (OCR) engine. Text extraction is the stage where the text components are segmented from the background. Enhancement of the extracted text components is required because the text region usually has low-resolution and is prone to noise. Thereafter, the extracted text images can be transformed into plain text using OCR technology. Text in videos is usually not easily extracted especially when it is embedded in complex background scenes and suffers from poor visual quality due to the effects of motion blur and compression artifacts.

After extracting and recognition text from TV commercial screenshot, we should identify and classify what type of information from extracted text. Important product information from TV commercials is product name, product price, URL information, and phone number. In this paper we propose a novel method for extracting product information from TV commercials by using knowledge based method with pattern matching and classification rules for recognizing and classifying the information.

II. RELATED WORK

In text extraction from image, text localization methods can be categorized into three types: region-based, texture-based and hybrid approaches. Region-based schemes use the properties of color or grayscale in a text region or their differences with corresponding properties of background. These methods can be divided further into two sub-approaches: connected component (CC) and edge-based [2].

CC-based methods apply a bottom-up approach by grouping small components into successively larger ones until all regions are identified in the image. A geometrical analysis is required to merge the text components using the spatial arrangement of the components so as to filter out non text components and mark the boundaries of text regions. Among the several textual properties in an image, edge-based methods focus on the 'high contrast between the text and the background'. The edges of the text boundary are identified and merged, and then several heuristics are performed to filter out the non text regions.

Texture-based methods use the observation that text in images has distinct textural properties that distinguish them from the background. The techniques based on Gabor filters, Wavelet, FFT, spatial variance, etc. can be performed to detect the textural properties of a text region in an image [1]. First

class of localization methods can be found in some works [3] – [6]. Gllavata et al. [3] have presented a method to localize and extract text automatically from color images. First, they transform the color image into grayscale image and then only the Y component is used. The text candidates are found by analyzing the projection profile of the edge image. Finally, a *binarized* text image is generated using a simple binarization algorithm based on a global threshold. They [4] also have applied the same idea of the previously mentioned paper to localize text; in addition the algorithm has been extended with a local *thresholding* technique.

Cai et al. [5] have presented a text detection approach which uses character features like edge strength, edge density and horizontal alignment. First, they apply a color edge detection algorithm in YUV color space and filter out non text edges using a low threshold. Then, a local *thresholding* technique is employed to keep low-contrast text and further simplify the background. An image enhancement process using two different convolution kernels follows. Finally, projection profiles are analyzed to localize the text regions. Jain and Yu [6] first employ color reduction by bit dropping and color clustering quantization, and afterwards a multi-value image decomposition algorithm is applied to decompose the input image into multiple foreground and background images. Then, CC analysis is performed on each of them to localize text candidates. From all method described here, there is no method specific for TV commercial except Leinhart [9] and Gllavata [3].

III. The Proposed Method For Text Extraction from TV Commercials

A. Overview of Information Extraction from TV Commercial System

Information extraction from TV commercial is useful for help people recognize important information from short and fast commercials video. The block diagram of TV commercial Information extraction system is shown in Figure 1 consist five processes as follow: Video frame detection, Text detection, Text Localization, Text Extraction, Text Recognition and Text Identification. Text detection in TV commercial video should perform based on the characteristic of displayed text object in commercials. After the investigation on common TV commercials video, we found some typical pattern of the commercials that accepted as assumption in our research work, as follow:

1) Text information usually in bright or contrast color like: white, red, or yellow in dark background; or blue or black in bright background
2) Important information using bigger font size.
3) Important information usually appears in near end part of commercials.

For our text extraction process, we propose a new text extraction method using combination of edge-based method and CC-based method. We use edge-based method for text detection because the characteristic of commercial video which is contain contrast color of text with background. We use CC-based method for text localization because we need the information about text position and size that important in information recognition. Also both of two method has less complexity in algorithm for real time application,

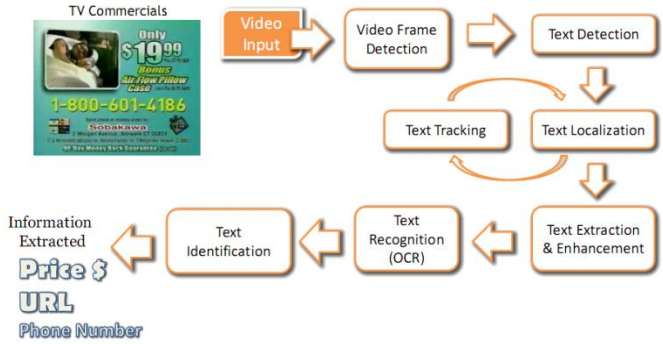


Figure 1. Block Diagram of Automatic TV Commercial Information Extraction

B. Video Frame Detection

By investigation on common TV commercial video samples, we found some typical pattern of the commercials that accepted as assumption in our research work, as follows:

1. Video length in 15 seconds, 30 seconds, 45 seconds or 60 seconds
2. Text information usually in bright or contrast color: Some pattern is: white, orange, or yellow in dark background, in the other hand, black, red or blue in bright background.
3. Important information is displayed more than once among a TVCM.
4. Important information using bigger font size.
5. Important information appears in the near of the end part of a TVCM.
6. Some important information displayed with product image.

Information extraction from TV commercial video is based on text extraction from screenshot image of commercials video. Commercial video was recorded from television and classify based on the video length. Then in determined time, we capture the screen shoot of running CM In automatic way, we can implement other research work on commercial detection like [14] for detecting the occurrence of commercials within a TV programs. In this paper, we assume that commercials are already separated from other TV programs. Using assumption that important information usually appears in near end part of commercials we generate screenshot of commercial video repeated in every 3 second but vary depend on commercials long. We use 15 seconds and 30 seconds type for TVCM. The screenshot number and its relationship with video frame are shown in Table 1.

TABLE I. Timing for Screenshot Taking and Number of Screenshot

CM Duration	Seconds of Screenshot	Total
15 seconds	5, 8, 11, 14	4
30 seconds	14, 17, 20, 23, 26, 29	6

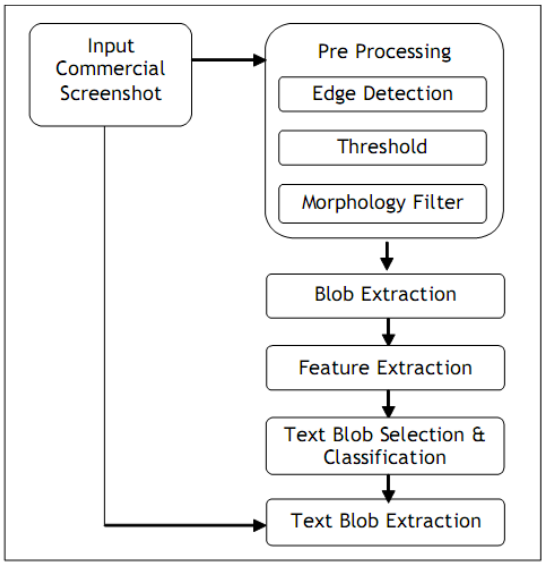


Figure 2. Text Extraction Algorithm

C. Text Detection Algorithm

The text detection stage seeks to detect the presence of text in a given image. The text detection method comprises of four main phases; pre processing, blob extraction, blob classification, and text blob extraction. The block schematic of the text extraction methodology is given in Figure 2. The detailed description of each phase is presented in the following subsections. This method also published in [12].

B. Pre-Processing

The objective of pre-processing is to convert image as binary picture that separates text object from background or non text object. We use combination of filter to process original image into binary image. Processes within pre-processing process as follows:

1. Extract Red channel.
2. Convert image into Grayscale
3. Homogeneity Edge Detection Filter
4. Binarization with threshold
5. Morphology Erosion Filter
6. Dilatation Filter

First, we extract red channel from image to get bright image information. After image is converted to grayscale and extract red channel, we implement homogeneity edge detection filter to get edge pattern of the object from image. The next process is binarization –after inverting image color- with an appropriate threshold number to produce black and white image. The heuristic value of threshold is 225 (for the images with quantization bits of 8 bits) that chosen empirically based on experiments. Combination of edge detection filter and appropriate threshold number will separates text from relatively complex background. In the end of pre-processing, we implement morphology erosion filter in 5x5 horizontal matrixes to combine small blob with left or right nearest blob. Figure 3 show step-by-step results from process in pre-processing process.

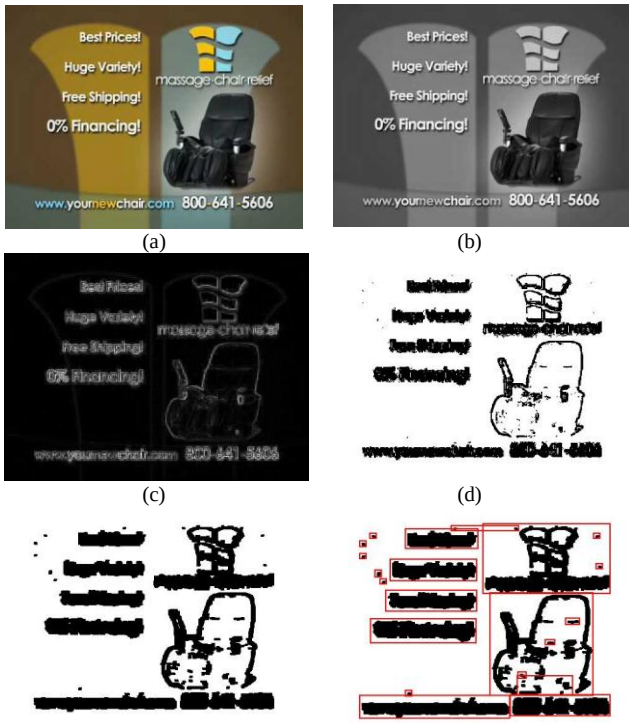


Figure 3. Process in pre-processing, Original (a), Grayscale (b) Homogeneity Edge detection (c), Binarization (d), Morphology Erosion (e), Blob Extraction (f).

C. Blob Extraction

After pre-processing, we detect all connected pixels with same color index as a separate blob object. Blob extraction is done using same blob extraction function implemented in our Comic text extraction method [11]. This process produces text blobs and also non text blobs. To classify a blob as a text blob or non text blob, we extract some features from text blob candidate for classification. In text blob extraction using blob extraction function, we select only blob with minimal size that selected as candidate of text blob. The minimal size of the text candidate blob width is [Image.Width]/40 and the text blob height is [Image.Height]/40. Parameter of [Image] corresponds to input image size. Figure 3.f show sample of detected blob in a commercial image after pre-processing that contain text blob, non-text blob and noise blob. Then we implement text blob selection to select only text blob and to remove noise or non-text blob, and text blob classification to classify blob into text word or sentence based on text blob position.

$$TB[i].YCenter = TB[i].Top + (TB[i].Height/2) \quad (1)$$

Where;

TB[i] corresponds to i^{th} blob of detected text blob.

TB[i].Left corresponds to left point position parameter of detected text blob

TB[i].Width corresponds to width parameter of detected text blob

For selecting and classifying text blob into horizontal text region, we use some rules for classification as follows:

1. Blob size is smaller than half of image size.
2. Classify all blobs with same vertical center point and relatively same height size into the same cluster. Assuming difference distance between centers is BD_{min} and difference size between 2 blobs is BS_{min}.

   If (TB[i].XCenter +- BDmin) ∈ Column[j] Then TB[i] member of Column[j] (2)

3. Position of blob center Y (*Blob.Cy*) is in range *(Blob[i].Cy – Blob[i].Height/2 < Blob[j].Cy)* and *(Blob[i].Cy + Blob[i].Height/2 > Blob[j].Cy)* (3)
4. Ignore blob cluster with width size smaller than its height size.

Minimal distance BD_{min} is approximately less than half of average text blob height and BS_{min} is around 40% of difference width between two blobs. Figure 4 Show the sample of text detection process with text blob detection (4.b) and text blob classification (4.c)

D. Text Extraction and Enhancement

After all text blob candidate localized, then extracted into separate text blob candidate. Extraction blob is getting from original input image without any other processing. We extract every blob from original image using position information (top, left, width, height) of the blob.

Figure 4. Sample of Commercial Screenshot (a), Blob detection (b), Text blob classification (c),

After all blob extracted, then we should enhance the text blob before using OCR for recognizing the text. We implement simple text enhancement with binarization (threshold) using Otsu [13] method. Otsu method tries to find minima between two peaks in histograms. Otsu's method we exhaustively search for the threshold that minimizes the intra-class variance, defined as a weighted sum of variances of the two classes. Figure 5 shows the sample of original extracted text blob and after pre-processing for enhancement.

Figure 5. Sample of text blob originally extracted (left), and with pre-processing (right)

IV. Knowledge Based Method for Information Recognition

For recognizing and classifying the information from extracted text, we design a TV commercial information extraction knowledge based system with rules and pattern matching. We design a knowledge based system with specific pattern for each type of information that we want to extract. The important information to extract from TVCM is: phone number, URL information, price information and product name. After detected and extracted text from TVCM screenshot, we should extract the ASCII text representation of image text using OCR application. The accuracy of OCR results is depending on OCR application that not covers in this paper. Then by using knowledge based system we try to classify the information based from rules in knowledge based.

We should extract some features from text blob before put in selection process by using knowledge based system. The features should be extracted for each blob word as follows: text ASCII representation, text blob position (top, left), text blob size (width, height), and relative position (top | middle | bottom and left | center | right), character type (text, number, currency, or URL keyword). Then, by using data of extracted features we can classify text based on our knowledge based rules and pattern.

Figure 6 shows some example of extracted text blob from screenshot of TV commercial video.

Figure 6. Sample of extracted text blob from text extraction process

The knowledge based rule system and specific pattern for each type of information is described in the following sections:

A. Rules for Phone Number:

Definition rules for selection a text combination as a phone number is:

- Surround with text related to phone information: Call | Now | Phone
- Should only contain: numbers [0-9] and optional for: hyphen [-]

For Phone Type 1:

- Match the string pattern: xxxx-xxx-xxx where x is numbers.
- Match the first 4 number combinations: 0120 (Japan).
- Total numeric string number is 10 without hyphen.
- Minimal Two hyphen (-) detected.

For Phone Type 2:

- Match the string pattern: x-xxx-xxx-xxxx where x is numbers
- Match the first 4 number combinations: 1-800 (Japan)
- Total numeric string number is 11 without hyphen.
- Three hyphen detected

B. Rules for Price Information:

Definition rules for selection a text combination as price information is:

- Match the string pattern: x,xxx or xx,xxx or xxx,xxx
- Should only contain numbers [0-9], and optional for [.][,]
- String size is relatively large
- Surround with text related to price information: [Only | Price | Now]
- Surround with character related to currency sign [$, \]

C. Rules for Web Site Information:

Definition rules for selection a text combination as web site information is:

- Match the string pattern (only one, combination or all): [http:// | www. | .com | .net | .org | .jp]
- Web site URL name check function

D. Rules for Product Name:

Definition rules for selection a text combination as web site information is:

- String size is relatively larger than others
- Position is relatively in center top
- Found more than x times in video frame screenshot, while x > 50% of processing frame number in one commercial.

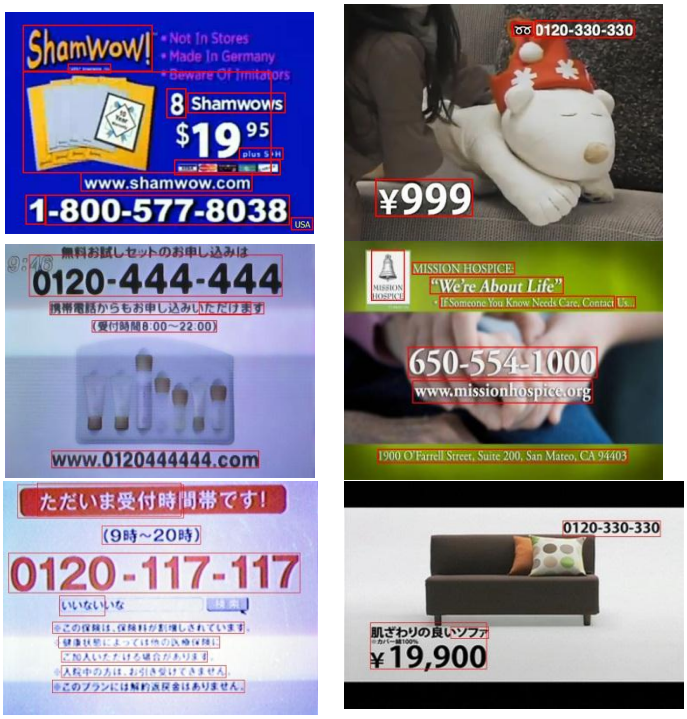


Figure 7. Sample of TV Commercial Screenshot Image with detected text in red rectangle

V. Experimental Results

We implement our method for text detection and extraction based on AForge open source image processing tools running on Visual Studio 2008. The proposed approach has been evaluated through data set containing different type of commercial screenshot images taken from Japan TV and some TV commercial screenshot randomly from internet. The based blob extraction method is the same blob extraction function we use in our text extraction method from digital Manga comic [11]. The performance of the system is evaluated on the text box level in terms of precision and recall. We also develop system for information recognition and evaluating it in real results from OCR and in ideal condition assuming OCR has no error on recognition the text. Figure 7 show the samples of TV commercial screenshot images with detected text located in red rectangle.

A. Evaluation of Text Extraction

The results for the experiments on text extraction are summarized in Table 2 where the number of existing text lines, the number of detected text lines, the number of false alarms and the corresponding values for recall and precision are listed. We use about 50 screenshot images from different TV commercial scene from image with simple text on background to text with complex background.

TABLE II. Experimental result for Text Extraction.

Number of images	50
Number of text lines	250
#correct detected	243
#false positives	19
Recall (%)	92.75%
Precision (%)	97.20%

Recall is defined as:

$$Precision = \frac{Correct\ Detected}{(Correct\ Detected + Missed\ Text\ Lines)}$$

Whereas precision is defined as:

$$Recall = \frac{Correct\ Detected}{(Correct\ Detected + False\ Positives)}$$

A text line is considered as detected correctly, if a text line consists in an extracted blob text, while a detected text blob is considered as a false alarm, if no text appears in that extracted text. The text extraction algorithm achieved a recall of 92.75% and a precision of 97.20%. The precision of our method is relatively better than Leinhart's [9] method and Gllavata's [3] method for detection text on commercial video as shown in Table 3.

TABLE III. COMPARISON OF PRECISION AND RECALL OF OUR METHOD WITH OTHER METHODS.

Method	Recall (%)	Precision (%)
Lienhart and al.Method [9]	-	66.00%
Gllvata and al.Method [3]	87.00%	90.00%
Our Method	92.75%	97.20%

B. Evaluation of OCR & Text Enhancement

Implementation of the character recognition process currently still uses commercial OCR application. We use Softi free OCR application. We evaluate the accuracy of character recognition without enhancement and with enhancement. We recognize only roman character within commercial text candidate blobs, and ignored Japanese character recognition at this time. There are around 750 characters in 50 samples of images, excluding space character. The results in Table 4 shown that by implementing the enhancement process, we can improve about 10% accuracy of the recognition.

TABLE IV. EVALUATION RESULTS OF CHARACTER RECOGNITION WITH/WITHOUT ENHANCEMENT

	Without Enhancement	With Enhancement
Total Character	748	748
Correct Detection	565	636
Miss Detection	183	112
Accuracy (%)	**75.53**	**85.03**

C. Evaluation of Information Extraction Knowledge Based Systems

Evaluation of the accuracy of knowledge based rule system is conduct through 2 conditions. First is real condition with real results from OCR, and second condition is ideal condition while assuming that we have 100% accuracy results from OCR. While first condition evaluate for the whole systems, second condition is only for evaluate the accuracy of knowledge based rule system assuming that there is no error from OCR. Table 5 show samples of text blob, its extracted information and type of information by the knowledge based rules. Table 6 shows the results of the accuracy evaluation of knowledge based. The term Tot_1 means the total of product information data from real OCR; the term Tot_2 means the total of product information data from perfect OCR; the term C means the correct detection; the term M means the miss detection; and the term FP means the False Positive. The accuracy of knowledge based system (only) for: *phone number* recognition: 86.36%; *URL address* recognition: 78.57% and *price information* recognition: 73.33%. And the accuracy of knowledge based system in whole system with original condition of OCR is: 75% accuracy of *phone number* recognition; 70% accuracy of *URL address* recognition and 60% accuracy of *price information*.

TABLE V. SAMPLE OF TEXT BLOB INFORMATION CLASSIFICATION.

Text Blob	Extracted Information	Type of Inf.
0120-330-330	0120-330-330	Phone
www.drylandcorp.com	www.drylandcorp.com	URL
¥1,900	¥1.900	Price
www.0120444444.com	www.0120444444.com	URL
Self Medication	Self Medication	-
954-534-4573	954-543-4573	Phone
nissen.jp	nissen.jp	URL
mightywallets.com	mightywallets.com	URL
9:48	9:48	-
1-8 September	1-8 September	-
$15.00 with Free Shipping*	$15.00 with Free Shipping	Price

TABLE VI. EVALUATION RESULTS OF KNOWLEDGE BASED RULE SYSTEM

Information	Tot_1	C	M	FP	Tot_2	C	M	FP
Phone Number	16	12	4	2	22	19	3	0
URL Address	10	7	3	1	14	11	3	0
Price Information	10	6	4	2	15	11	4	1

D. Discussion

Based on our experimental results, accuracy of recall on text extraction is not so high because the occurrence of many false positive, that is some non text objects are detected as text in our approach. Since the text blob candidate will then send to OCR process for recognition, non-text object is not a significant problem, because non-text object has no text output from OCR process. Also, if OCR generated results for non-text object, usually only a little text with no meaning that can be ignored. Although our experimental results only calculate the text from commercial video screenshot image, it is still possible to implementing in real time analysis for TV commercial video.

The accuracy results of information recognition are also depending on the accuracy of OCR application. From the evaluation process while assuming using the perfect OCR, knowledge based system accuracy for product phone number is about 86.36%, it is mean that our method for information recognition knowledge based rules and pattern matching method should be improved for better results.

VI. CONCLUSION & FUTURE WORK

In this paper, we have proposed an approach to automatically extracted text appearing in commercial screenshot images and recognize the product information based on the pattern matching method. We see that our method on

text extraction have good performance for localization of text in commercial image, also our information recognition has high accuracy on classifying and recognizing product information from TV commercials. This is notable that in the future, we can improve the method for implementing a real time product information extraction from TV commercial application, for using in a set top box TV system, as an application for helping people automatically retrieve important information from live streaming TV video contents.

REFERENCES

[1] K. Jung, K.I. Kim, and A.K. Jain, "Text Information Extraction in Images and Video: A Survey," Pattern Recognition, Vol. 37, pp. 977-997, 2004.

[2] G. Aghajari, J. Shanbehzadeh, and A. Sarrafzadeh, "A Text Localization Algorithm in Color Image via New Projection Profile", Proceedings of International MultiConference of Engineers and Computer Scientists 2010, Vol II, pp. 1486-1489, 2010

[3] J. Gllavata, R. Ewerth, and B. Freisleben, "A Robust Algorithm for Text Detection in Images", Proc. of 3rd Int'l Symposium on Image and Signal Processing and Analysis, Rome, pp. 611-616, 2003.

[4] J. Gllavata, R. Ewerth and B. Freisleben, "Finding Text in Images via Local Thresholding", International Symposium on Signal Processing and Information Technology, Darmstadt, Germany, pp. 539-542, 2003.

[5] M. Cai, J. Song, and M. R. Lyu, "A New Approach for Video Text Detection", Proc. of IEEE Int'l Conference on Image Processing, Rochester, New York, USA, pp. 117-120 , 2002.

[6] A. K. Jain, and B. Yu, "Automatic Text Location in Images and Video Frames", Pattern Recognition Vol 31, No. 12, pp. 2055-2076, 1998.

[7] Q. Ye, Q. Huang, W. Gao and D. Zhao, "Fast and robust text detection in images and video frames", Image Vision Comput. 23, pp. 565–576, 2005.

[8] J. Gllavata, R.Ewerth and B. freisleben, "A text detection, localization and segmentation system for OCR in images," in Proc. IEEE Sixth International Symposium on Multimedia Software Engineering, pp. 425-428, 2004.

[9] Rainer Lienhart and Wolfgang Effelsberg, "Automatic text segmentation and text recognition for video indexing". Multimedia Syst. Vol 8, No. 1, pp 69-81, 2000.

[10] Xiaojun Li, Weiqiang Wang, Shuqiang Jiang, Qingming Huang, Wen Gao, "Fast and Effective Text Detection", IEEE International Conference on Image Processing, San Diego, USA, pp. 969-972, 2008.

[11] Kohei, A., Tolle, H., "Method for Real Time Text Extraction from Digital Manga Comic", International Journal of Image Processing Vol 4, No. 6, pp. 669-676, 2011.

[12] Kohei Arai, Tolle Herman, "Text Extraction from TV Commercial using Blob Extraction Method", International Journal of Research Review on Computer Science (IJRRCS), Vol 2, No.3, pp. 895-899, 2011.

[13] N. Otsu. A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man, and Cybernetics. 1979, 9(1): 62-66

[14] Yijun Li, Danqing Zhang, Xiangmin Zhou, and Jesse S. Jin. A confidence based recognition system for TV commercial extraction. In *Proceedings of the nineteenth conference on Australasian database - Volume 75* (ADC '08), Vol. 75. Australian Computer Society, Inc., Darlinghurst, Australia, Australia, 57-64. 2008.

[15] Pritam Singh Negi, M M S Rauthan and H S Dhami. Article: "Text Summarization for Information Retrieval using Pattern Recognition Techniques". International Journal of Computer Applications 21(10):20-24, 2011.

[16] Ralph Grishman, "Information Extraction: Techniques and Challenges", . In Information Extraction: A Multidisciplinary Approach to an Emerging Information Technology, in M.T. Pazienza, (ed.), Springer, 97.

AUTHORS PROFILE

Kohei Arai, He received BS, MS and PhD degrees in 1972, 1974 and 1982, respectively. He was with The Institute for Industrial Science and Technology of the University of Tokyo from April 1974 to December 1978 and also was with National Space Development Agency of Japan from January, 1979 to March, 1990. During from 1985 to 1987, he was with Canada Centre for Remote Sensing as a Post Doctoral Fellow of National Science and Engineering Research Council of Canada. He moved to Saga University as a Professor in Department of Information Science on April 1990. He was a councilor for the Aeronautics and Space related to the Technology Committee of the Ministry of Science and Technology during from 1998 to 2000. He was a councilor of Saga University for 2002 and 2003. He also was an executive councilor for the Remote Sensing Society of Japan for 2003 to 2005. He is an Adjunct Professor of University of Arizona, USA since 1998. He also is Vice Chairman of the Commission A of ICSU/COSPAR since 2008. He wrote 26 books and published 227 journal papers.

Herman Tolle, He graduated Bachelor degree in Electrical Engineering from Brawijaya University, Malang in 1998, and also graduated Master degree in Telecommunication Information System from Bandung Institute of Technology (ITB), Bandung in 2002. He is with Engineering Faculty of Brawijaya University from 2002 to present. He is now a Doctoral student in Department of Information Science, Faculty of Science and Engineering, Saga University Japan. He has major concern of research in image analysis, multimedia, content adaptation and web engineering.

Efficient Cancer Classification using Fast Adaptive Neuro-Fuzzy Inference System (FANFIS) based on Statistical Techniques

K.AnandaKumar
Assistant Professor,
Department of Computer Applications
Dr.SNS Rajalakshmi College of Arts and Science
Coimbatore,TamilNadu

Dr.M.Punithavalli
Director
Department of Computer Studies
Dr.SNS Rajalakshmi College of Arts and Science
Coimbatore,TamilNadu

***Abstract*- The increase in number of cancer is detected throughout the world. This leads to the requirement of developing a new technique which can detect the occurrence the cancer. This will help in better diagnosis in order to reduce the cancer patients. This paper aim at finding the smallest set of genes that can ensure highly accurate classification of cancer from micro array data by using supervised machine learning algorithms. The significance of finding the minimum subset is three fold: a) The computational burden and noise arising from irrelevant genes are much reduced; b) the cost for cancer testing is reduced significantly as it simplifies the gene expression tests to include only a very small number of genes rather than thousands of genes; c) it calls for more investigation into the probable biological relationship between these small numbers of genes and cancer development and treatment. The proposed method involves two steps. In the first step, some important genes are chosen with the help of Analysis of Variance (ANOVA) ranking scheme. In the second step, the classification capability is tested for all simple combinations of those important genes using a better classifier. The proposed method uses Fast Adaptive Neuro-Fuzzy Inference System (FANFIS) as a classification model. This classification model uses Modified Levenberg-Marquardt algorithm for learning phase. The experimental results suggest that the proposed method results in better accuracy and also it takes lesser time for classification when compared to the conventional techniques.**

Keyword- Gene Expressions, Cancer Classification, Neural Networks, Neuro-Fuzzy Inference System, Analysis of Variance, Modified Levenberg-Marquardt Algorithm

I. INTRODUCTION

MICRO array data analysis has been successfully applied in a number of studies over a broad range of biological disciplines including cancer classification [3, 10] by class discovery and prediction , identification of the unknown effects of a specific therapy , identification of genes relevant to a certain diagnosis or therapy , and cancer prognosis.

The multivariate supervised classification techniques such as support vector machines (SVMs) [13] and multivariate statistical analysis method such as principal component analysis (PCA), singular value decomposition (SVD) [9] and generalized singular value decomposition (GSVD) cannot be applied to data with missing values. The finding of missing value is an essential preprocessing step. Because of various reasons, there may be some loss of data in gene expression [8, 11, 12] e.g. inadequate resolution, image corruption, dirt or scratches on the slides or experimental error during the laboratory process. Several algorithms have been developed for recovering data because it is costlier and time consuming to repeat the experiment. Moreover, estimating unknown elements in the given data has many potential applications in the other fields. There are several approaches for the estimating the missing values. Recently, for missing value estimation, the singular value decomposition based method (SVDimpute) and weighted k-nearest neighbors imputation (KNNimpute) has been introduced. It has been shown that KNNimpute shows better performance on non-time series data or noisy time series data, whereas, SVDimpute works well on time series data with low noise levels. Considering as a whole, the weighted k-nearest neighbor based imputation offers a more robust method for missing value estimation than the SVD based method.

In this paper, Fast Adaptive Neuro-Fuzzy Inference System (FANFIS) is used along with gene ranking technique called Analysis of Variance (ANOVA). The learning technique used in this paper is Modified Levenberg-Marquardt algorithm.

II. RELATED WORKS

Isabelle *et al.*,[1] proposed the Gene Selection for Cancer Classification using Support Vector Machines. In this paper, the author address the problem of selection of a small subset of genes from broad patterns of gene expression data [4, 5], recorded on DNA micro-arrays.

Using available training examples from cancer and normal patients, the approach build a classifier suitable for genetic diagnosis, as well as drug discovery. Previous attempts to address this problem select genes with correlation techniques. The author proposes a new method of gene selection utilizing Support Vector Machine methods based on Recursive Feature Elimination (RFE). It is experimentally demonstrated that the genes selected by our techniques yield better classification [14] performance and are biologically relevant to cancer. Jose *et al.*,

[2] presents a Genetic Embedded Approach for Gene Selection [15, 16] and Classification of Microarray Data [7, 17].

Murat *et al*., [6] gives the early prostate cancer diagnosis by using artificial neural networks. The aim of this study is to design a classifier based expert system for early diagnosis of the organ in constraint phase to reach informed decision making without biopsy by using some selected features. The other purpose is to investigate a relationship between BMI (body mass index), smoking factor, and prostate cancer. The data used in this study were collected from 300 men (100: prostate adenocarcinoma, 200: chronic prostatism or benign prostatic hyperplasia). Weight, height, BMI, PSA (prostate specific antigen), Free PSA, age, prostate volume, density, smoking, systolic, diastolic, pulse, and Gleason score features were used and independent sample t-test was applied for feature selection. In order to classify related data, the author have used following classifiers; scaled conjugate gradient (SCG), Broyden-Fletcher-Goldfarb-Shanno (BFGS) and Levenberg-Marquardt (LM) training algorithms of artificial neural networks (ANN).

III. METHODOLOGY

Cancer classification proposed in this paper comprises of two steps. In the first step, all genes in the training data set are ranked using a scoring scheme. Then genes with high scores are retained. This paper uses Analysis of Variance (ANOVA) method for ranking. In the second step, the classification capability of all simple two gene combinations among the genes selected are tested in this step using Fast Adaptive Neuro-Fuzzy Inference System (FANFIS) in which the training is performed using Modified Levenberg-Marquardt algorithm.

Step 1: Gene Importance Ranking

This step performs the computation of important ranking of each gene by means of Analysis of Variance (ANOVA) method.

Step 2: Finding the minimum gene subset

This step attempts to classify the data set with single gene after selecting several top genes in the important ranking list. Each selected gene is given as an input to the classifier. When good accuracy is not obtained, it is required to classify the data set with all possible 2 gene combination within the selected genes.

Even if the good accuracy is not obtained, this procedure is repeated with all of the 3 gene combinations and so on until the good accuracy is obtained.

Adaptive Neuro-Fuzzy Inference System (ANFIS)

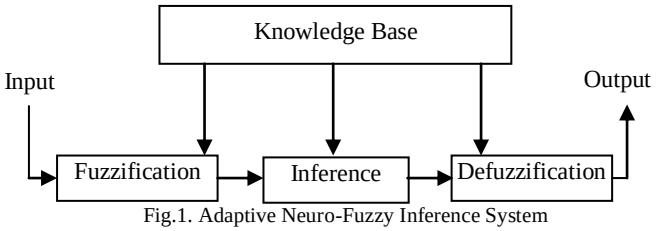


Fig.1. Adaptive Neuro-Fuzzy Inference System

The fuzzy inference system that we have considered is a model that maps

- Input characteristics to input membership functions,
- Input membership function to rules,
- Rules to a set of output characteristics,
- Output characteristics to output membership functions, and
- The output membership function to a single-valued output, or
- A decision associated with the output.

Architecture of ANFIS

The ANFIS is a framework of adaptive technique to assist learning and adaptation. This kind of framework formulates the ANFIS modeling highly organized and not as much of dependent on specialist involvement. To illustrate the ANFIS architecture, two fuzzy if-then rules according to first order Sugeno model are considered:

$$Rule\ 1: If(x\ is\ A_1) and\ (y\ is\ B_1) then\ (f_1 = p_1x + q_1y + r_1)$$
$$Rule\ 2: If(x\ is\ A_2) and\ (y\ is\ B_2) then\ (f_2 = p_2x + q_2y + r_2)$$

where x and y are nothing but the inputs, A_i and B_i represents the fuzzy sets, f_i represents the outputs inside the fuzzy region represented by the fuzzy rule, p_i, q_i and r_i indicates the design parameters that are identified while performing training process.

The ANFIS architecture to execute these two rules is represented in figure 2, in which a circle represents a fixed node and a square represents an adaptive node.

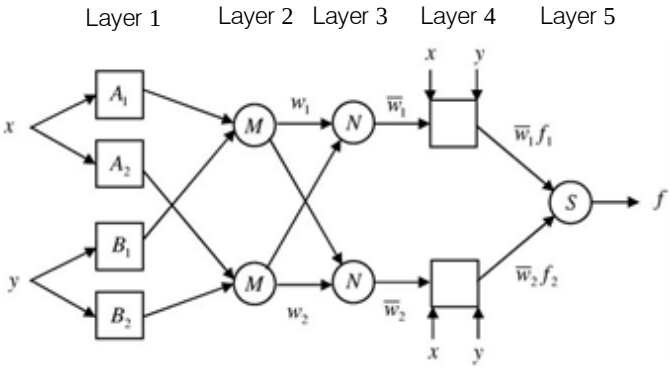


Fig.2. ANFIS Architecture

In the first layer, every node is adaptive node. The outputs of first layer are the fuzzy membership grade of the inputs that are represented by:

$$O_i^1 = \mu_{A_i}(x)\ \ i = 1,2 \quad (1)$$

$$O_i^1 = \mu_{B_{i-2}}(y)\ \ i = 3,4 \quad (2)$$

where $\mu_{A_i}(x), \beta_{B_{i-2}}(y)$, can accept any fuzzy membership function. For example, if the bell shaped membership function is employed, $\mu_{A_i}(x)$ is represented by:

$$\mu_{A_i}(x) = \frac{1}{1 + \left\{\left(\frac{x - c_i}{a_i}\right)\right\}^{b_i}} \quad (3)$$

where a_i, b_i and c_i represents the parameters of the membership function, controlling the bell shaped functions consequently.

In layer 2, the nodes are fixed nodes. These nodes are labeled with M, representing that they carry out as a simple multiplier. The outputs of this layer can be indicated by:

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \;\; i = 1,2 \quad (4)$$

which are the called as firing strengths of the rules.

The nodes are fixed in layer 3 as well. They are labeled with N, representing that they are engaged in a normalization function to the firing strengths from the earlier layer.

The outputs of this layer can be indicated as:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} \quad i = 1,2 \quad (5)$$

which are the called as normalized firing strengths.

In layer 4, all the nodes are adaptive nodes. The output of the every node in this layer is merely the product of the normalized firing strength and a first order polynomial. Therefore, the outputs of this layer are provided by:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i(p_i x + q_i y + r_i) \quad i = 1,2 \quad (6)$$

In layer 5, there exists only one single fixed node labeled with S. This node carries out the operation like summation of every incoming signal. Therefore, the overall output of the model is provided by:

$$O_i^5 = \sum_{i=1}^{2} \bar{w}_i f_i = \frac{\sum_{i=1}^{2} w_i f_i}{w_1 + w_2} \quad (7)$$

It can be noted that layer 1 and the layer 4 are adaptive layers. Layer 1contains three modifiable parameters such as a_i, b_i, c_i that is associated with the input membership functions.

These parameters are called as premise parameters. In layer 4, there exists three modifiable parameters as well such as {p_i, q_i, r_i}, related to the first order polynomial. These parameters are called consequent parameters.

Learning algorithm of ANFIS

The intention of the learning algorithm is to adjust all the modifiable parameters such as{a_i, b_i, c_i} and {p_i, q_i, r_i}, for the purpose of matching the ANFIS output with the training data.

If the parameters such as a_i, b_i and c_i of the membership function are unchanging, the outcome of the ANFIS model can be given by:

$$f = \frac{w_1}{w_1 + w_2} f_1 + \frac{w_2}{w_1 + w_2} f_2 \quad (8)$$

Substituting Eq. (5) into Eq. (8) yields:

$$f = \bar{w}_1 f_1 + \bar{w}_2 f_2 \quad (9)$$

Substituting the fuzzy if-then rules into Eq. (15), it becomes:

$$f = \bar{w}_1(p_1 x + q_1 y + r_1) + \bar{w}_2(p_2 x + q_2 y + r_2) \quad (10)$$

After rearrangement, the output can be expressed as:

$$f = (\bar{w}_1 x)p_1 + (\bar{w}_1 y)q_1 + (\bar{w}_1)r_1 + (\bar{w}_2 x)p_2 + (\bar{w}_2 y)q_2 + (\bar{w}_2)r_2 \quad (11)$$

which is a linear arrangement of the adjustable resulting parameters such as p_1, q_1, r_1, p_2, q_2 and r_2. The least squares technique can be utilized to detect the optimal values of these parameters without difficulty. If the basis parameters are not adjustable, the search space becomes larger and leads to considering more time for convergence. A hybrid algorithm merging the least squares technique and the gradient descent technique is utilized in order to solve this difficulty. The hybrid algorithm consists of a forward pass and a backward pass. The least squares technique which acts as a forward pass is utilized in order to determine the resulting parameters with the premise parameters not changed. Once the optimal consequent parameters are determined, the backward pass begins straight away. The gradient descent technique which acts as a backward pass is utilized to fine-tune the premise parameters equivalent to the fuzzy sets in the input domain. The outcome of the ANFIS is determined by using the resulting parameters identified in the forward pass.

The output error is utilized to alter the premise parameters with the help of standard backpropagation method. It has been confirmed that this hybrid technique is very proficient in training the ANFIS.

Modified Levenberg-Marquardt algorithm

A Modified Levenberg-Marquardt algorithm is used for training the neural network.

Considering performance index is $F(w) = e^T e$ using the Newton method we have as:

$$W_{K+1} = W_K - A_K^{-1}.g_K \quad (12)$$

$$A_k = \nabla^2 F(w)\Big|_{w=w_k} \quad (13)$$

$$g_k = \nabla F(w)|_{w=w_k} \quad (14)$$

$$[\nabla F(w)]_j = \frac{\partial F(w)}{\partial w_j} = 2\sum_{i=1}^{N} e_i(w).\frac{\partial e_i(w)}{\partial w_j} \quad (15)$$

The gradient can write as:

$$\nabla F(x) = 2J^T e(w) \tag{16}$$

Where

$$J(w) = \begin{bmatrix} \frac{\partial e_{11}}{\partial w_1} & \frac{\partial e_{11}}{\partial w_2} \dots & \frac{\partial e_{11}}{\partial w_N} \\ \frac{\partial e_{21}}{\partial w_1} & \frac{\partial e_{21}}{\partial w_2} \dots & \frac{\partial e_{21}}{\partial w_N} \\ & \vdots & \\ \frac{\partial e_{KP}}{\partial w_1} & \frac{\partial e_{KP}}{\partial w_2} \dots & \frac{\partial e_{KP}}{\partial w_N} \end{bmatrix} \tag{17}$$

$J(w)$ is called the Jacobian matrix.

Next we want to find the Hessian matrix. The k, j elements of the Hessian matrix yields as:

$$\begin{aligned} \left[\nabla^2 F(w)\right]_{k,j} &= \frac{\partial^2 F(w)}{\partial w_k \partial w_j} \\ &= 2\sum_{i=1}^{N} \left\{ \frac{\partial e_i(w)}{\partial w_k} \frac{\partial e_i(w)}{\partial w_j} \right. \\ &\quad \left. + e_i(w).\frac{\partial^2 e_i(w)}{\partial w_k \partial w_j} \right\} \end{aligned} \tag{18}$$

The Hessian matrix can then be expressed as follows:

$$\nabla^2 F(w) = 2J^T(W).J(W) + S(W) \tag{19}$$

$$S(w) = \sum_{i=1}^{N} e_i(w).\nabla^2 e_i(w) \tag{20}$$

If S(w) is small assumed, the Hessian matrix can be approximated as:

$$\nabla^2 F(w) \cong 2J^T(w)J(w) \tag{21}$$

Using (13) and (21) we obtain the Gauss-Newton method as:

$$\begin{aligned} &W_{k+1} = \\ &W_k - \left[2J^T(w_k)\cdot J(w_k)\right]^{-1} 2J^T(w_k)e(w_k) \\ &\cong W_k - \left[J^T(w_k)\cdot J(w_k)\right]^{-1} J^T(w_k)e(w_k) \end{aligned} \tag{22}$$

The advantage of Gauss-Newton is that it does not require calculation of second derivatives.

There is a problem the Gauss-Newton method is the matrix $H = J^T J$ may not be invertible. This can be overcome by using the following modification.

Hessian matrix can be written as:

$$G = H + \mu I \tag{23}$$

Suppose that the eigenvalues and eigenvectors of H are $\{\lambda_1, \lambda_2, \dots\dots, \lambda_n\}$ and $\{z_1, z_2, \dots\dots, z_n\}$.Then:

$$\begin{aligned} Gz_i &= [H + \mu I]z_i \\ &= Hz_i + \mu z_i \\ &= \lambda_i z_i + \mu z_i \\ &= (\lambda_i + \mu)z_i \end{aligned} \tag{24}$$

Therefore the eigenvectors of G are the same as the eigenvectors of H, and the eigen values of G are $(\lambda_i + \mu)$.

The matrix G is positive definite by increasing μ until $(\lambda_i + \mu) > 0$ for all i therefore the matrix will be invertible.

This leads to Levenberg-Marquardt algorithm:

$$w_{k+1} = w_k - \left[J^T(w_k)J(w_k) + \mu I\right]^{-1} J^T(w_k)e(w_k) \tag{25}$$

$$\Delta w_k = \left[J^T(w_k)J(w_k) + \mu I\right]^{-1} J^T(w_k)e(w_k) \tag{26}$$

As known, learning parameter, μ is illustrator of steps of actual output movement to desired output. In the standard LM method, μ is a constant number. This paper modifies LM method using μ as:

$$\mu = 0.01e^T e \tag{27}$$

Where e is a $k \times 1$ matrix therefore $e^T e$ is a 1×1 therefore $[J^T J + \mu I]$ is invertible.

Therefore, if actual output is far than desired output or similarly, errors are large so, it converges to desired output with large steps. Likewise, when measurement of error is small then, actual output approaches to desired output with soft steps. Therefore error oscillation reduces greatly.

IV. Experimental Result

Lymphoma Data Set

Lymphoma data set [18] contains 42 samples obtained from diffuse large B-cell lymphoma (DLBCL). Among these, 9 samples are from follicular lymphoma (FL), 11 samples are from chronic lymphocytic leukaemia (CLL). The whole data set contains the expression data of 4026 genes. In this data set, a small portion of data is missing. A k-nearest neighbor technique was utilized to fill those missing data. In the initial step, the 62 samples are randomly seperated into 2 groups in such a way those 31 samples for training, and 31 samples for testing. Next the complete 4026 genes are ranked with the help of ANOVA technique. Then 100 genes are taken from them with the highest rank. Then the proposed ANFIS technique is applied for classification.

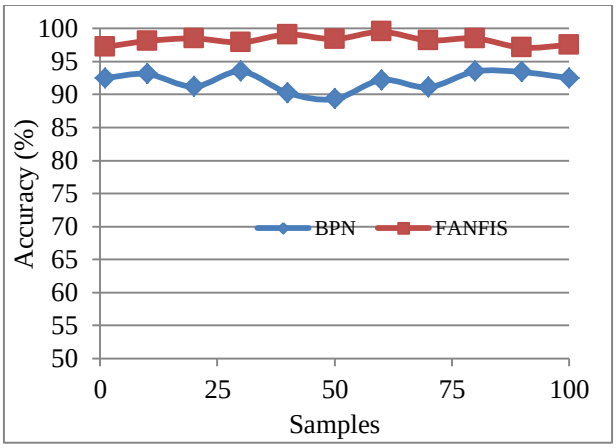


Fig.3.: Classification Accuracy for Lymphoma Data Set with ANOVA Ranking

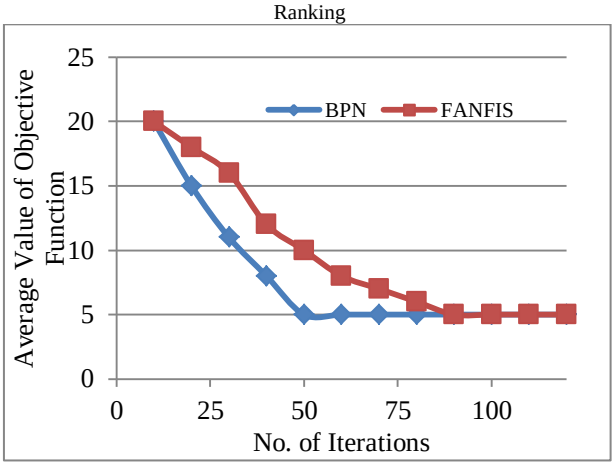


Fig.4. Convergence Behavior for Lymphoma Data Set with ANOVA Ranking

Figure 3 represents the resulted for classifying the lymphoma data set and figure 4 represents the convergence behavior of lymphoma data set.

Liver Cancer Data Set

The liver cancer data set [19] contains two classes, i.e. the nontumor liver and HCC. The data set consista 156 samples and the expression data of 1648 important genes. In that, 82 are HCCs and the remaining 74 are nontumor livers.

The data is randomly separated into 78 training samples and 78 testing samples. In this data set, there are some missing values. K-nearest neighbor technique is utilized to fill those missing values. Initially, 100 important genes are chosen in the training data set. Next all possible 1-gene and 2-gene combinations are tested within the 100 important genes.

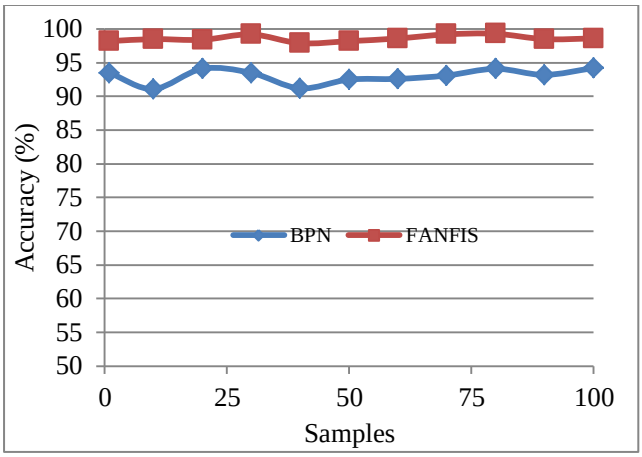


Fig.5.: Classification Accuracy for Liver Cancer Data Set with ANOVA Ranking

Figure 5 represents the resulted for classifying the liver cancer data set and figure 6 represents the convergence behavior of liver cancer data set. From these results, it can be observed that the proposed technique results in better accuracy of classification and it takes lesser time to converge.

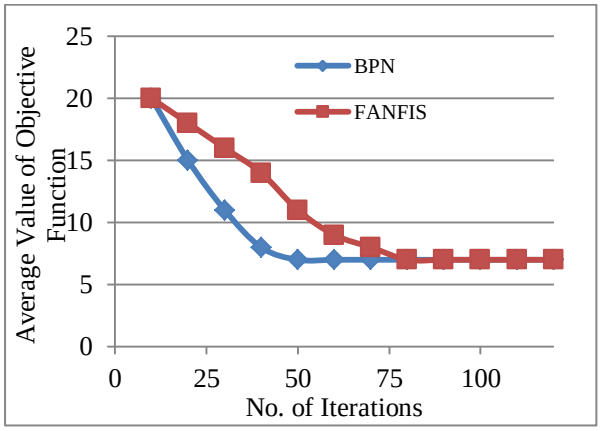


Fig.6. Convergence Behavior for Liver Cancer Data Set with ANOVA Ranking

V. CONCLUSION

This paper suggests a better technique for classification of cancer. In the proposed technique, the ANOVA ranking technique is initially applied to the dataset in order to find the higher ranked genes.

After ranking the genes, Adaptive Neuro-Fuzzy Inference System is used in used for classification which has both the advantages of neural network and fuzzy logic. But, it takes more time for classification.

To overcome this paper uses Fast Adaptive Neuro-Fuzzy Inference System (FANFIS). The learning is performed using the Modified Levenberg-Marquardt algorithm. The proposed technique is tested using two dataset namely, Lymphoma dataset and Liver cancer dataset. The experimental result shows that the proposed technique results in better accuracy of classification and also takes lesser time for convergence.

REFERENCES

[1] Isabelle Guyon, Jason Weston, Stephen Barnhill and Vladimir Vapnik," Gene Selection for Cancer Classification using Support Vector Machines", 2002.

[2] Jose Crispin Hernandez Hernandez, Béatrice Duval and Jin-Kao HaoGuyon, Jason Weston, Stephen Barnhill and Vladimir Vapnik," A Genetic Embedded Approach for Gene Selection and Classification of Microarray Data", 2007.

[3] Cun-gui Cheng, Lu-yao Cheng, Run-sheng Xu ,"Classification of FTIR Gastric Cancer Data Using Wavelets and SVM",ICNC '07: Proceedings of the Third International Conference on Natural Computation - Volume 01, 2007.

[4] Mingjun Song and Sanguthevar Rajasekaran, "A greedy algorithm for gene selection based on SVM and correlation", International Journal of Bioinformatics Research and Applications, 2010.

[5] Chen Liao and Shutao Li, "A support vector machine ensemble for cancer classification using gene expression data", ISBRA'07: Proceedings of the 3rd international conference on Bioinformatics research and applications, 2007.

[6] Murat Cınar, Mehmet Engin, Erkan Zeki Engin and Y. Ziya Atesci ,"Early prostate cancer diagnosis by using artificial neural networks and support vector machines", Expert Systems with Applications: An International Journal, April 2009.

[7] Kim H, and Park H, Multi class gene selection for classification of cancer subtypes based on generalized LDA

[8] Shipp M. A et al,.Diffuse large B-cell lymphoma outcome prediction by gene expression profiling and supervised machine learning, Nat. Med., 8,68-74, 2002.

[9] Alter O., Brown P.O. , and Botstein D., "Generalized Singular value decomposition for comparative analysis of genome-scale expression datasets of two different organisms", Proceedings of Natural academy of Science, USA, 100(6), 3351-3356, 2003.

[10] Y. Lee and C. K. Lee, "Clasication of multiple cancer types by Multicategory Support Vector Machines using gene expression data", Bioinformatics, 19, 1132-1139, 2003.

[11] S. Chen, S. R. Gunn and C. J. Harris, "The relevance vector machine technique for channel equalization application," IEEE Trans on Neural Networks, Vol. 12, No. 6, pp. 1529-1532, 2001.

[12] Tipping M. E. "Sparse Bayesian Learning and the Relevance Vector Machine", Journal of Machine Learning Research, pp. 211-244, 2001.

[13] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," IEEE Trans. on Medical Imaging, vol. 21, 1552-1563, 2002.

[14] L. Carin and G. J. Dobeck, "Relevance vector machine feature selection and classification for underwater targets," Proceedings of OCEANS 2003, Vol. 2, pp. 22-26, 2003.

[15] Shutao Li, Xixian Wu and Xiaoyan Hu, " Gene selection using genetic algorithm and support vectors machines", Springer-Verlsg, Soft Computing - A Fusion of Foundations, Methodologies and Applications, Feb 2008.

[16] Chen Liao, Shutao Li and Zhiyuan Luo," Gene Selection Using Wilcoxon Rank Sum Test and Support Vector Machine for Cancer Classification", Springer-Verlag, Computational Intelligence and Security, april 2007

[17] Chaoyang Zhang, Peng Li, Arun Rajendran and Youping Deng," Parallel Multicategory Support Vector Machines (PMC-SVM) for Classifying Microcarray Data", IMSCCS '06: Proceedings of the First International Multi-Symposiums on Computer and Computational Sciences, 2006.

[18] http://llmpp.nih.gov/lymphoma

[19] http://genome-www.stanford.edu/hcc

AUTHORS PROFILE

Mr.K.AnandaKumar was born in TamilNadu, India on March 1975. He received the B.Sc Degree in Physics from Bharathiar University in 1995. He received his MCA Degree in Computer Applications from Bharathiar University in 1998. He received his M.Phil from Periyar University in 2006 and he is doing Doctor of Philosophy in Computer Science and Engineering from Bharathiar University, Coimbatore. He had 13 years of teaching experience. Currently he is working as HOD in Computer Applications Department, Dr.SNS Rajalakshmi College of Arts and Science College, Coimbatore. His Professional activities include…Guided Twenty PG projects and Ten M.Phil and guiding Ten PG and Six M.Phil projects. Published and presented 8 papers in International and National Conferences and 6 national and international journals.

Dr. M. Punithavalli received the Ph.D degree in Computer Science from Alagappa University, Karaikudi in May 2007. She is currently serving as the Director of the Computer Science Department, Sri Ramakrishna college of Arts and Science for Women, Coimbatore. Her research interest lies in the area of Data mining, Genetic Algorithms and Image Processing. She has published more than 10 Technical papers in International, National Journals and conferences. She is Board of studies member various universities and colleges. She is also reviewer in International Journals. She has given many guest lecturers and acted as chairperson in conference. Currently 10 students are doing Ph.D under her supervision

Clustering Student Data to Characterize Performance Patterns

Bindiya M Varghese
Dept. of Computer Science,
Rajagiri College of Social Sciences, Kalamassery
Kerala, India

Jose Tomy J
Rajagiri College of Social Sciences, Kalamassery
Kerala, India

Unnikrishnan A
Sciencist G,
NPOL Kochi

Poulose Jacob K
Dean, Dept. of Computer Science,
CUSAT Kochi

***Abstract*— Over the years the academic records of thousands of students have accumulated in educational institutions and most of these data are available in digital format. Mining these huge volumes of data may gain a deeper insight and can throw some light on planning pedagogical approaches and strategies in the future. We propose to formulate this problem as a data mining task and use k-means clustering and fuzzy c-means clustering algorithms to evolve hidden patterns.**

Keywords- Data mining; k-means Clustering; Fuzzy C-means; Student performance analysis.

I. INTRODUCTION

Data mining techniques are used to extract useful and valid patterns from huge databases. Large amount of data is accumulated in universities and colleges concerning the students. The proactive knowledge gained by these techniques will help the stakeholders for decision making that likely to effect on student's learning outcomes. The model developed helps achieve measurable student progress monitoring process and identifies the features that profoundly influence the performance, thus benefiting stakeholders in the educational system and the wider community.

II. CLUSTERING

Clustering is a method to group data into classes with identical characteristics in which the similarity of intra-class is maximized or minimized. Clustering is a descriptive task that seeks to identify homogeneous groups of objects based on the values of their attributes [1][2]. Current clustering techniques can be broadly classified into three categories; partitional, hierarchical and locality-based algorithms. Given a set of objects and a clustering criterion, the partitional clustering obtains a partition of objects into clusters such that the objects in a cluster are more similar to the objects inside the cluster than to objects in different clusters. Partitional clustering algorithms attempt to decompose the dataset directly into a set of k disjoint clusters, provided k is the number of initial clusters. An iterative optimization is done to emphasize the local structure of data, which involves minimizing some measure of dissimilarity in the objects within the cluster, while maximizing the dissimilarity of different clusters. Partitional algorithms are generally iterative in nature and converge to some local optima. Given a set of data points $xi \in \Re d$, i = 1,..,N , partitional clustering algorithms aim to organize them into K clusters {C1, …, CK} while maximizing or minimizing a pre-specified criterion function J.

A. K-Means Clustering Algorithm

K-means is one of the simplest unsupervised learning algorithms used for clustering. K-means partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean [3]. This algorithm aims at minimizing an objective function, in this case a squared error function. The algorithm aims to minimize the objective function. K-means is one of the simplest unsupervised learning algorithms used for clustering. K-means partitions n observations into k clusters in which each observation belongs to the cluster with the nearest mean. This algorithm aims at minimizing an objective function, in this case a squared error function. The algorithm aims to minimize the objective function $J = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^j - c_j \right\|^2$ where $\left\| x_i^j - c_j \right\|^2$ is a chosen distance measure between a data point x_i^j and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centres.

B. Fuzzy C- Means Algorithm

Fuzzy c-means clustering allows one data element to belong to two or more clusters. Given a finite set of data, X, the problem of clustering in X is to find several cluster centres that can properly characterize relevant classes of X. In classical cluster analysis, these classes are required to form a partition of X such that the degree of association is strong for data within blocks of the partition and weak for data in different blocks. However, this requirement is too strong in many practical applications, and it is thus desirable to replace it with a weaker requirement.

When the requirement of a crisp partition of X is replaced with a weaker requirement of fuzzy partition we refer to the emerging problem area as fuzzy clustering. Fuzzy pseudo partitions are often called fuzzy c-partitions, where c designates the number of fuzzy classes in the partition. This method was developed by Dunn in 1973 and improved by Bezdek in 1981. It is based on minimization of the following objective function:

$$J_m = \sum_i^N \sum_j^C u_{ij}^m \left\| x_i - c_j \right\|^2, \; 1 \leq m < \infty,$$

where m is any real number greater than 1, u_{ij} is the degree of membership of x_i in the cluster j, x_i is the ith of d-

dimensional measured data, c_j is the d-dimension center of the cluster.

Fuzzy partitioning is carried out through an iterative optimization of the objective function shown above, with the update of membership u_{ij} and the cluster centers c_j by:

$$u = \frac{1}{\sum_{k-1}^{c}\left\{\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right\}^{\frac{2}{m-1}}},$$

$$\text{given } c_j = \frac{\sum_{i-1}^{N} u_{ij}^{m}.x_i}{\sum_{i-1}^{N} u_{ij}^{m}}.$$

This iteration will stop when $max_{ij}\{|u_{ij}^{k+1} - u_{ij}^{k}|\}? < \varepsilon$ where ε is a termination criterion between 0 and 1, whereas k are the iteration steps. This procedure converges to a local minimum or a saddle point of J_m.

Let $X=\{x_1,x_2,x_3,\cdots,x_n\}$ be a set of given data. A fuzzy pseudo partition or fuzzy c-partition of X is a family of fuzzy subsets of X, denoted by $P=\{A_1,A_2,A_3,\cdots,A_c\}$ which satisfies $\sum_{i=1}^{c} A_i(x_k) = 1$ for all $k \in \mathbb{N}_n$ and $0 < \sum_{k=1}^{n} A_i(x_k) < n$ for all $i \in N_c$, where c is a positive integer.

III. Dataset

The dataset consisted of details of students of five consecutive years. The main features are the following attributes for each course attended by the student

i. Attendance
ii. Internal mark assessment
iii. Seminar assessment
iv. Class assignment assessment
v. University marks scored

The dataset consisted of approximate 8000 records. The attributes internal assessment, seminar assessment and the class assignment were transformed and consolidated into proper normal forms appropriate for mining. Normalization was done on these attributes so that data should fall within a small specified range and hence does not outweigh the measurement of other attributes.

IV. Results and Discussion

Both, k-means and Fuzzy C-means were applied on the dataset. The prominent results from both the experiments are shown below.

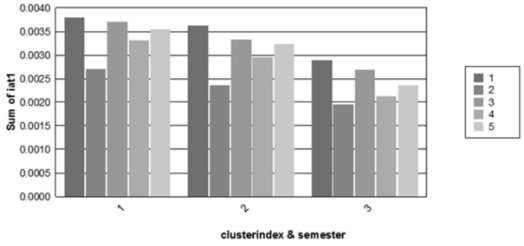


FIGURE 1.A. FIRST INTERNAL ASSESMENT TEST (K-MEANS)

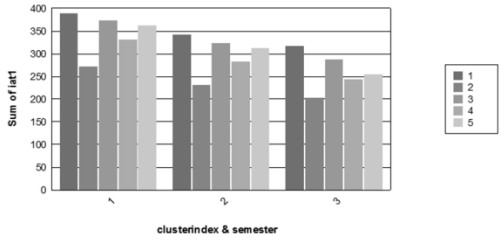


FIGURE 1.B. FIRST INTERNAL ASSESMENT TEST (FUZZY C-MEANS)

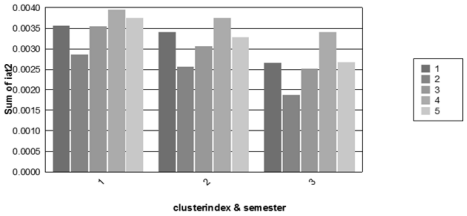


FIGURE 2.A. SECOND INTERNAL ASSESMENT TEST (K MEANS)

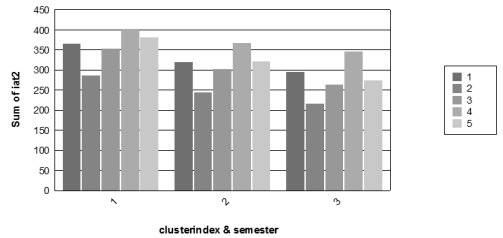


FIGURE 2.B. SECOND INTERNAL ASSESMENT TEST (FUZZY C-MEANS)

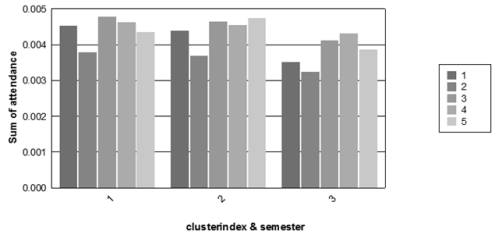


FIGURE 3. A. ATTENDANCE IN EACH SEMESTER (K-MEANS)

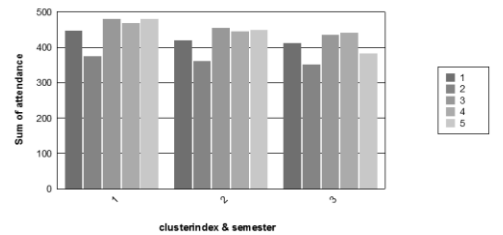


FIGURE 3. B. ATTENDANCE IN EACH SEMESTER (FUZZY C-MEANS)

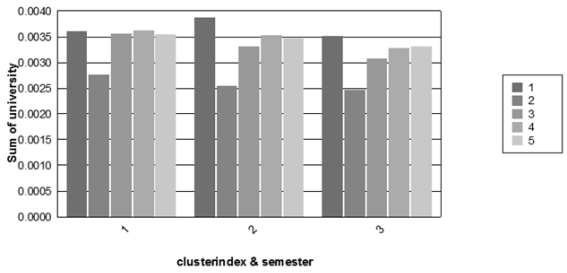


FIGURE 4.A. UNIVERSITY RESULTS (K MEANS)

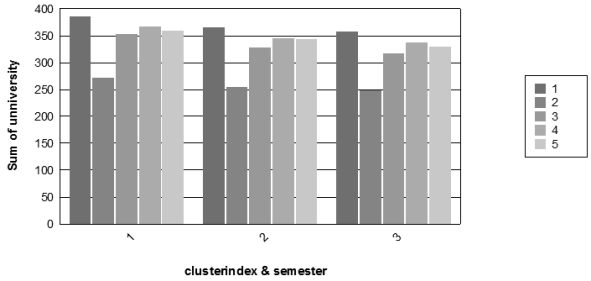


FIGURE 4.A. UNIVERSITY RESULTS (FUZZY C-MEANS)

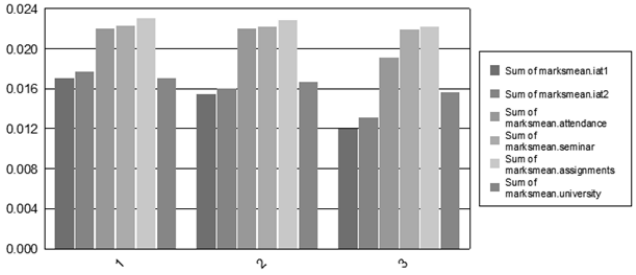


FIGURE 5.A. AGGREGATE PERFORMANCE (K-MEANS)

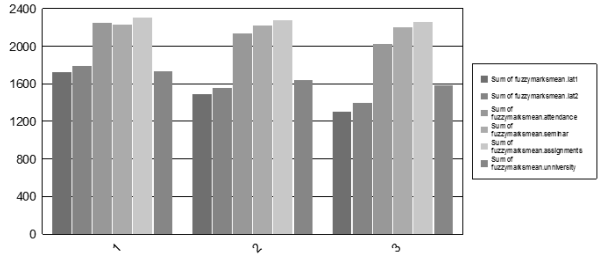


FIGURE 5.A. AGGREGATE PERFORMANCE (FUZZY C-MEANS)

Following interpretations are made out of the results. The fluctuations in internal assessment marks decreases with each passing semester. The graphs indicate a directly proportional link between attendance and student aggregate performance and that the performance decreases with decreasing attendance. Except for second semester all university marks are equal to or just below first semester university marks. The first semester mark can be considered as an indicator of what can be expected of a student in further semesters.

The graphs of both the algorithms support the same facts that students score more in second internal assessment which is conducted after 30 sessions of the semester than in the first internal assessment conducted after 15 sessions and university mark is almost the same or just above internal assessment marks

This pilot study provides fundamental inferences to develop basic heuristics for the course. The clustering process provides us with different perspectives which can made use while preparing the schedule for internal assessments and the curriculum. The Internal Assessment Scores are clearly an indicator for the student's academic performance and at the end of First Internal Assessment remedial classes can be designed and implemented. As a future research, clustering can be directly applied to more expounded data, so that more relations between the different attributes are emerged.

REFERENCES

[1] Ester, M., Frommelt, A., Kriegel, H.-P., and Sander, J. 1998. Algorithms for characterization and trend detection in spatial databases. In Proc. 4th Int. Conf. on Knowledge Discovery and Data Mining, New York City, NY, pp. 44-50.

[2] Kaufman, Leonard and Peter J. Rousseeuw. Finding Groups in Data: An Introduction to Cluster Analysis (Wiley Series in Probability and Statistics). Wiley-Interscience, March 2005.

[3] MacQueen, J. B. (1967). "Some Methods for classification and Analysis of Multivariate Observations". 1. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. pp. 281–297. MR0214227. Zbl 0214.46201. Retrieved 2009-04-07.

[4] http://en.wikipedia.org/wiki/Data_clustering#Fuzzy_c-means_clustering

[5] Richard Nock, Frank Nielsen, "On Weighting Clustering," IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 1223-1235, August, 2006

AUTHORS PROFILE

Bindiya M Varghese is a researcher in the field of data ming, specilizing on fuzzy quadtree based algorithms. Currently works as the Assistant Professor in the Department of Computer Science, Kalamassery Kerala, India.

Jose Tomy is currently with Tata Consultancy Services, Kochi as a software Engineer.

Unnikrishnan A Ph.D is with Naval Physical Oceanographic Laboratory , an organization under defence ministry of India, as Senior Scienctist.

Poulose Jacob K Ph.D is the Dean and HOD of the department of Computer Sciences, Cochin University of Science and Technology Kerala.

Comparative Analysis of Various Approaches Used in Frequent Pattern Mining

Deepak Garg, Hemant Sharma
Thapar University, Patiala

***Abstract*—Frequent pattern mining has become an important data mining task and has been a focused theme in data mining research. Frequent patterns are patterns that appear in a data set frequently. Frequent pattern mining searches for recurring relationship in a given data set. Various techniques have been proposed to improve the performance of frequent pattern mining algorithms. This paper presents review of different frequent mining techniques including apriori based algorithms, partition based algorithms, DFS and hybrid algorithms, pattern based algorithms, SQL based algorithms and Incremental apriori based algorithms. A brief description of each technique has been provided. In the last, different frequent pattern mining techniques are compared based on various parameters of importance. Experimental results show that FP- Tree based approach achieves better performance.**

Keywords- Data mining; Frequent patterns; Frequent pattern mining; association rules; support; confidence; Dynamic item set counting.

I. INTRODUCTION

Frequent patterns are item sets, subsequences, or substructures that appear in a data set with frequency no less than a user-specified threshold. Frequent pattern mining is a first step in association rule mining. One of the major uses with association rules is to analyze large amount of supermarket basket transactions [1-3]. Recently, association rules have been applied to other areas like outlier's detection, classification, clustering etc [4, 6, 8].

Association rules mining can formally be defined as follows. Let I = $\{i_1, i_2, i_3, .., i_m\}$ be a set of attributes called items. Let D be a set of transactions. Each transaction t ϵ D consists of a set of items such that t $\subseteq$ I. A transaction t is said to contain an item set X if and only if all items within X are also contained in t. Each transaction also contains a unique identifier called transaction identification (TID). Support of an item set is normalized number of occurrences of the item set within the dataset. An item set is considered as frequent or large, if the item set has a support that is greater or equal to the user specified minimum support [25-26].

The most common form of association rules is implication rule which is in the form of X $\Rightarrow$ Y, where X $\subset$ I, Y $\subset$ I and X $\cap$ Y = Φ. The support of the rule X Y is equal to the percentage of transactions in D containing X $\Rightarrow$ Y. The confidence of the rule X Y is equal to the percentage of transactions in D containing X also containing Y. Once the required minimum support and confidence are specified, association rule mining becomes finding all association rules that satisfy the minimum support requirements. The problem can be further broken down into two steps: mining of frequent item sets and generating association rules.

The number of possible combinations of item sets increases exponentially with I and the average transaction length. Therefore it is infeasible to determine the support of all possible item sets. When counting the supports of item sets, there are two strategies. The first strategy is to count the occurrences directly, whenever an item set is contained in a transaction, the occurrence of the item set is increased. The second strategy is to count the occurrences indirectly by intersecting TID set of each component of the item set. The TID set of a component X, where X can be either item or item set, is denoted as X.TID. The support of an item set S = X $\cup$ Y is obtained by intersecting X.TID $\cap$ Y.TID = S.TID and the support of S equals S.TID [28, 29].

II. VARIOUS FREQUENT PATTERN MINING TECHNIQUES

A. *Apriori-based Algorithms*

The first published frequent item set mining algorithm is Apriori [1]. Apriori uses breadth first search (BFS). At each level, Apriori reduces the search space by using downward closure property of item set. If an item set of length k is not frequent, none of its superset patterns can be frequent. Candidate frequent item sets, C_k where k is the length of the item set, are generated before each data scan. The supports of candidate frequent item sets are counted. Candidate k item sets, C_k are generated with frequent (k – 1) item sets. Apriori algorithm achieves good performance by reducing the candidate item sets iteratively. The problem, however, associated with Apriori is it requires k data scans to find all frequent k-item sets. It is very much expensive to scan the large data base. Dynamic Item set Counting, (DIC) relaxes the strict separation between generating and counting of item sets [4]. DIC starts counting the support of candidate frequent item sets as soon as they are being generated. By overlapping counting and candidate item set generation, DIC reduces the overall data scans required. Orlando et al. [13] proposed an algorithm that combines transaction reduction and direct data access. At the end of each scan, transactions that are potentially useful are used for the next iteration. A technique called scan reduction uses candidate 2 item sets to generate subsequent candidate item sets [12]. If all intermediate data can be held in the main memory, only one scan is required to generate all candidate frequent item sets. Another data scan is required to verify whether the candidate frequent item sets are frequent or not.

With all of those improvements, the number of data scans required by Apriori based algorithms has been reduced significantly. However, the cost of generating candidate frequent item sets has not been fully addressed by Apriori based algorithms. This problem becomes visible when there are huge numbers of frequent 1 or 2 item sets.

B. Partition-based Algorithms

Partition-based Algorithms [15] solves the problem of high number of database scans, associated with Apriori-based algorithm. It requires two complete data scan to mine frequent item sets. The Partition algorithm divides the dataset into many subsets so that each subset can be fitted into the main memory. The basic idea of the Partition-based algorithm is that a frequent item set must be frequent in at least one subset. Partition-based algorithm generates local frequent item sets for each partition during the first data scan. Since the whole partition can be fitted into the main memory, the complete local frequent item sets can be mined without any disk I/O operations. The local frequent item sets are added to the global candidate frequent item sets. In the second data scan, false candidates are removed from the global candidate frequent item sets. In a special case where each subset contains identical local frequent item sets, Partition algorithm can mine all frequent item sets with a single data scan. However, when the data is distributed unevenly across different partitions, this algorithm may generate a lot of false candidates from a small number of partitions. By employing the knowledge collected during the mining process, false global candidate frequent item sets are pruned when they are found that they cannot be frequent. In addition, those algorithms reduce the number of scans in the worse case to (2b- 1)/b where b is the number of partitions.

C. DFS and Hybrid Algorithms

Eclat and Clique [16] combine both depth first search (DFS) and intersection counting. Since intersection counting is used, no complicated data structure is required. These hybrid algorithms reduces the memory requirement, since only the TID sets of the path item sets from the root to the leaves have to be kept in the memory simultaneously. Intersection of TID sets can be stopped as soon as the remaining length of the shortest TID set is shorter than the required support minus the counted support value. The intersection of TID sets of 1-item set to generate frequent 2 item sets is expensive.. The maximal hyper graph clique clustering is applied to 2-frequent item sets to generate a refined set of maximal item sets. Hipp et al. [10] pointed out that DFS cannot prune candidate k item sets by checking frequent (k – 1) item sets, because DFS searches from the root to the leaves of the tree without using any subsets relationship. A hybrid approach of BFS and DFS is proposed in [11]. It is cheaper to use item set counting with BFS to determine the supports, when the number of candidate frequent item sets is small. When the number of candidate frequent item sets is relatively large, the hybrid algorithm switches to TID set intersection with DFS, since simple TID set intersection is more efficient than occurrence counting when the number of candidate frequent item sets is relatively large. This results in additional costs to generate TID sets. The authors proposed [11] to use hash-tree-like structure to minimize the cost of transition. However, the authors do not provide an algorithm to determine the best condition to switch the strategy. In the evaluation, the authors provide parameters to change in strategy. However, those parameters may not be generalized enough for all kinds of datasets. Incorrect timing of changing strategy may decrease the performance of hybrid algorithm.

D. Pattern-Growth Algorithms

Two major costs of Apriori based algorithms are the cost to generate candidate frequent item sets and the cost associated with I/O operations. The issues related to I/O have been addressed, but the issues related to candidate frequent item sets generation remain open. If there are n frequent 1 item sets, Apriori based algorithms would require to generate approximately $n^2/2$ candidate frequent item sets. Secondly, the memory required to hold the candidate frequent item sets and their supports could be substantial. For example, when n equals 10,000, there would be more than 10^8 length 2 candidate frequent item sets. If we assume that it requires 4 bytes to hold the support and 4 bytes to hold the item sets, approximately 0.5 gigabytes of main memory would be needed to store the information [18]. Furthermore, the memory required does not include the overhead associated with the data structure. Also the cost required to count the support of candidate item sets may be large. As far as Apriori-based algorithms are concerned, the run time increases as the support value decreases. Therefore, the cost of candidate frequent item sets generation of Apriori based algorithms will exceeds than the cost of I/O [24]. Han et al. [9] proposed a data structure called frequent pattern tree or FP Tree. FP-growth mines frequent item sets from FP-Tree without generating candidate frequent item sets. FP-Tree is an extension of prefix tree structure. Only frequent items get stored in the tree. Each node contains the item's label along with its frequency. The paths from the root to the leaves are arranged according to the support value of the items with the frequency of each parent is greater than or equal to the sum of its children's frequency. The construction of FP-Tree requires two data scans. In the first scan, the support value of each item is found. This calculated support values are used in the second scan to sort the items within transactions in descending order. If two transactions share a common prefix, the shared portion is merged and the frequencies of the nodes are incremented accordingly. Nodes with the same label are connected with an item link. The item link is used to facilitate frequent pattern mining. In addition, each FP-Tree has a header that contains all frequent items and pointers to the beginning of their respective item links. FP-growth partitions the FP-Tree based on the prefixes. FP-growth traverses the paths of FP-Tree recursively to generate frequent item sets. Pattern fragments are concatenated to ensure all frequent item sets are generated properly. Thus FP-growth avoids the costly operations for generation and testing operations of candidate item sets. When the data is sparse, the compression achieved by the FP-Tree is small and the FP Tree is bushy. As a result, FP-growth would spend a lot of effort to concatenate fragmented patterns with no frequent item sets being found. A new data structure called H-struct is introduced in [14]. In this, transactions are sorted with an arbitrary ordering scheme. Only frequent items are projected in the H-struct. H-struct consists of projected transactions and each node in the projected transactions contains item label and a hyper link pointing to the next occurrence of the item. A header table is created for H-struct. The header contains frequencies of all items, their supports and hyper link to the

first transaction containing given item. H-mine mines the H-struct recursively by building a new header table for each item in the original header with subsequent headers omitting items that have been mined previously. For each sub- header, H-mine traverses the H-struct according to the hyper links and finds frequent item sets for the local header. At the same time, H-mine builds links for items that have not been mined in the local header. Those links are used to find conditional frequent patterns within the local header. The process is repeated until all frequent item sets have been mined. In case of a dense dataset, H-struct is not as efficient as FP-Tree because FP-Tree allows compression.

E. Incremental Update with Apriori-based Algorithms

Complete dataset is normally huge and the incremental portion is relatively small compared to the complete dataset. In many cases, it is not feasible to perform a complete data mining process while transactions are being added continuously. Therefore, incremental data mining algorithms have to reuse the existing information as much as possible, so that either computational cost and/or I/O cost can be reduced. A general incremental mining algorithm called Fast Update 2 (FUP2), that allows both addition and deletion of transactions was proposed in [7]. The major idea of FUP2 is to reduce the cost of candidate frequent item sets generation. Incremental portion of the dataset is scanned; frequent patterns in the incremental data are compared with the existing frequent item sets in the original dataset. Previous frequent item sets are removed if they are no longer frequent after the incremental portion of the data is added or removed. The supports of previous frequent item sets that are still frequent are updated to reflect the changes. In those ways, previous frequent item sets that are still frequent are not required to be checked for their supports again. New (k + 1) candidate frequent item sets are generated from frequent k item sets. The entire updated dataset is scanned to verify those newly added candidate item sets if they are indeed frequent. The process is repeated until the set of candidate frequent item set becomes empty. FUP2 offers some benefits over the original Apriori algorithm. However, it still requires multiple scans of the dataset. Another incremental Apriori based algorithm is called Sliding Window Filtering (SWF) [12]. SWF incorporates the main idea of Partition algorithm with Apriori to allow incremental mining. SWF divides the dataset into several partitions. During the scan of partitions, a filtering threshold is employed in each partition to generate candidate frequent 2 item sets. When a candidate 2 item set is found to be frequent in the newly scanned partition, the partition number and the frequency of the item set are stored. Cumulative information about candidate frequent 2 item sets is selectively carried over toward subsequence partition scans. Cumulative frequencies of previous generated candidate frequent 2 item sets are maintained as new partitions are being scanned. False candidate frequent item sets are pruned when the cumulative support of the candidate frequent item sets fall below required proportional support since they have become frequent. Once incremental portion of the dataset is scanned, scan reduction techniques are used to generate all subsequence candidate frequent items sets [5]. Another data scan over the whole dataset is required to confirm the frequent item sets. In the case of data removal, the partition to be removed are scanned, the cumulative count and the start partition number of candidate length 2 item sets are modified accordingly. Although SWF achieves better performance than pervious algorithms, the performance of SWF still depends on the selection of partition size and removal of data can only be done at partition level.

F. SQL-based algorithms

DBMS can facilitate data mining to become an online, robust, scalable and concurrent process by complementing the existing querying and analytical functions. The first attempt to the particular problem of integrated frequent item set mining was the SETM algorithm [10, 17], expressed as SQL queries working on relational tables. The Apriori algorithm [1] opened up new prospects for FIM. The database- coupled variations of the Apriori algorithm were carefully examined in [19]. The SQL-92 based implementations were too slow, but the SQL implementations enhanced with object-relational extensions (SQL-OR) performed acceptable. The so- called Cache-Mine implementation had the best overall performance, where the database-independent mining algorithm cached the relevant data in a local disk cache [21-23]. SQL based frequent mining using FP-tree provide best performance than other SQL based techniques [20]. Although an FP-tree is rather compact, it is unrealistic to construct a main memory- based FP-tree when the database is large. However using RDBMSs provides us the benefits of using their buffer management systems specially developed for freeing the user applications from the size considerations of the data. And moreover, there are several potential advantages of building mining algorithms to work on RDBMSs. An interesting alternative is to store a FP-tree in a table. There are two approaches in this category - FP, EFP (Expand Frequent Pattern). They are different in the construction of frequent pattern tree table, named FP. FP approach checks each frequent item whether it should be inserted into a table FP or not one by one to construct FP. EFP approach introduces a temporary table EFP, thus table FP can generate from EFP. According to the properties of FP-tree, FP-tree can be presented by a table FP with three column attributes: item identifier (item), the number of transactions that contain this item in a sub- path (count), and item prefix sub-tree (path). The field path is beneficial not only to construct the table FP but also to find all frequent patterns from FP. In the construction of table FP, the field path is an important condition to judge if an item in frequent item table F should be insert into the table FP or update the table FP by incrementing the item's count by 1. If an item does not exist in the table FP or there exist the same items as this item in the table FP but their corresponding path are different, insert the item into table FP. In the process of constructing conditional pattern base for each frequent item, only need to derive its entire path in the table FP as a set of conditional paths, which co-occurs with it.

III. COMPARISON OF VARIOUS FREQUENT PATTERN MINING TECHNIQUES

Comparison of different FPM techniques is given in Table 1, where A is length of maximal frequent item set and B is number of partitions. As Shown In the table, various algorithms are compared against four parameters, number of database scans required for the generation of frequent item set, the candidate generation technique used, whether the frequent item generation approach is incremental or not, and how the

algorithm is sensitive to the change in user parameters. Apriori-based methods use efficient technique for pruning the candidate item sets, but they require lots of computational time as well as multiple database scans to generate candidate item sets. Partition-based methods limit the size of candidate item sets. Partition algorithm may generate a lot of false candidates from a small number of partitions. FP-Tree based methods require only two database scans in order to generate frequent patterns. These methods use a compact tree- structure to represent the entire database. They do not require candidate generation, reducing the computational cost.

IV. Comparison of Apriori and Primitive Association Rule Mining

Comparison of the algorithms, Apriori and Primitive Association Rule Mining is given in this section. There are many advantages of Primitive Association Rule Mining over Apriori.

Apriori uses candidate Generate function for generating every candidate k-item sets and it takes enormous amount of time to generate candidate k+1-item sets from large k item sets. However, Primitive Association Rule Mining does not use this function; instead it uses graph based approach after generating of large 2–item sets.

In primitive association, a graph is constructed with large two item sets. Using graph, large three item sets can be generated easily without scanning the database.

At each pass in primitive association, it is enough to use graph with k large item sets for generating k+1 candidate item sets. Traversal of one link list (adjacency list) takes less time as compared to Apriori generation function.

Secondly in Apriori approach we are accessing transaction as a whole or we can divide into parts but it takes lot of memory whereas in Primitive Association Rule Mining, transactions are converted into bit vector which is based on items. Bit vector representation takes very less times as well as memory, theoretically 32 times less. Primitive Association Rule Mining takes less time since transactions are represented in bit vector form, and we are using logical AND, OR operation which is very fast. Further the bit representation consumes less memory also.

A. Comparison of AprioriTid and Apriori Hybrid

AprioriTid and AprioriHybrid are just variations of Apriori. In Apriori, at every step, we have to find candidate k-item sets, and we have to scan whole database at each k, which is time consuming. So, AprioriTid algorithm has given a solution for finding candidate k-item sets without scanning whole transaction. This algorithm works on the basis of transaction Id that is associated with every transaction. Apriori Hybrid is combined approach of Apriori and Apriori Tid, in which if some part of transactions (which is stored in other place) do not fit into the memory then use Apriori algorithm, otherwise swap Apriori algorithm to Apriori Tid.

In general, using Apriori AprioriTid and AprioriHybrid algorithms we can find frequent item sets, whereas we assume that items and transactions have equal weights. Sometimes, it is important to know that whether every items have equal weights or not, if all items have equal weights then Apriori and their variation can do good job for finding frequent item sets, and if weights of items are not equal, then Apriori and their variations do not work. So, to solve this problem, we have two solutions, whether we can assign weights to items and transactions, or to use some algorithms, so that it can give weights of items and weights of transactions. If we have weights of items in the beginning in the database then we can find frequent item sets using weighted association rule of mining, otherwise we can use association rule of mining without pre-assign weights, which gives weights of items and weights of transactions using HITS algorithm.

TABLE I. Comparison Of Various Frequent Pattern Mining

	Apriori-Based	Partition Based	Incremental Apriori	FP Tree	SQL Based
Number of Database Scan For Best Case Scenario	2	1	2	2	1
Number of Database Scan For Worst Case	A+1	(2B-1)/B	A+1	2	1
Candidate Generation Needed or Not	Yes	Yes	Yes	No	No
Incremental Mining Possible	No	No	Yes	No	No
Sensitive to Change in User Parameter	Yes	Yes	Yes	Yes	Yes

B. Experimental Results

Following are real life datasets which were taken, these are:

Kosarak- The kosarak dataset comes from the click-stream data of a Hungarian online news portal, Number of Instances =990,002, Number of Attributes= 41,270.

TABLE II. KOSARAK DATABASE

Large Item Sets	Time taken by Apriori	Time taken by Primitive
3	0.46	0.29
4	2.166	1.86
5	12.04	11.11

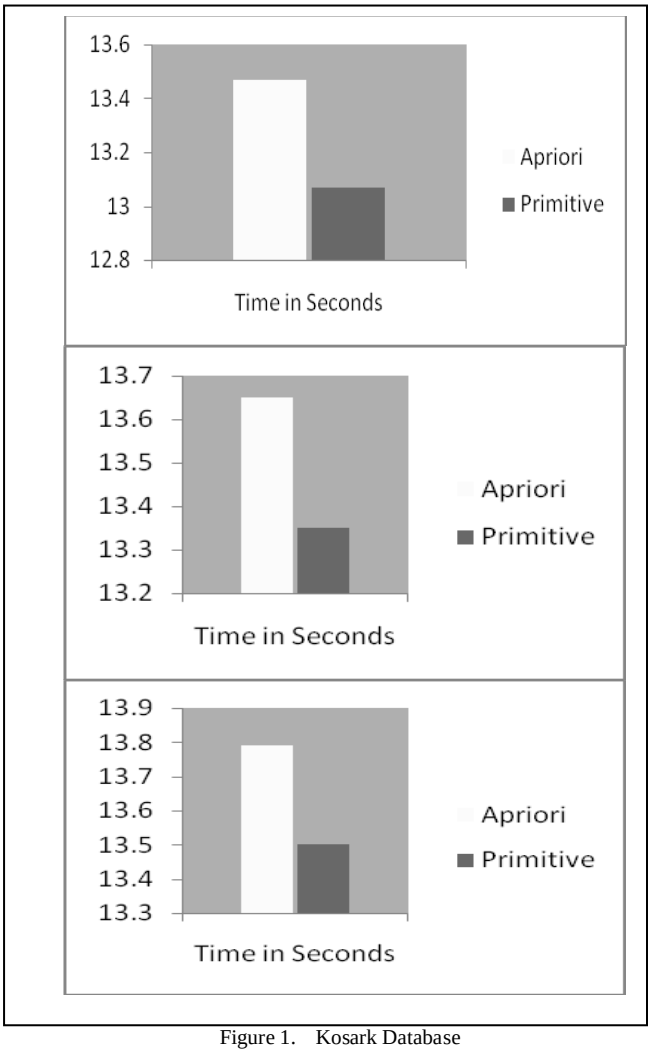


Figure 1. Kosark Database

The results clearly show that Primitive algorithm is taking less time as compared to the apriori algorithm.

Mushroom- This data set includes descriptions of hypothetical samples corresponding to 23 species of gilled mushrooms. Each species is identified as definitely edible, definitely poisonous, or of unknown edibility and not recommended. This latter class was combined with the poisonous one. The Guide clearly states that there is no simple rule for determining the edibility of a mushroom. Number of Instances = 8124, Number of Attributes = 22.

TABLE III. FOR MUSHROOM DATABASE

Large Item Sets	Time taken by Apriori	Time taken by Primitive
3	13.47	13.07
4	13.65	13.35
5	13.79	13.5

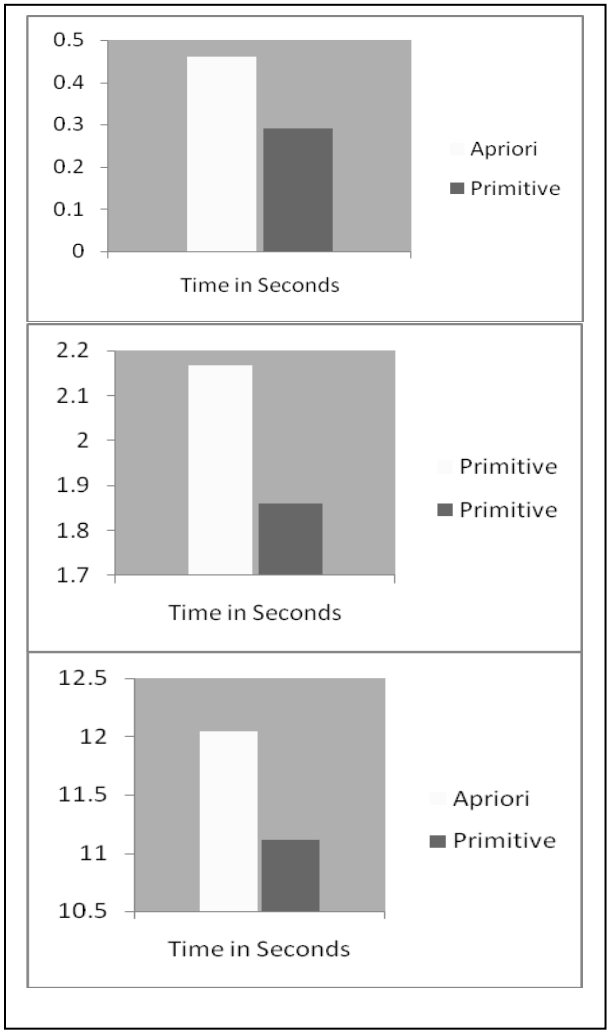


Figure 2. Mushroom Database

Experimental result clearly shows that Apriori is taking more time.

Chess - A game datasets.

Attribute Information: Classes (2): -- White-can-win ("won") and White-cannot-win ("nowin"). It believes that White is deemed to be unable to win if the Black pawn can safely advance. Number of Instances= 3196, Number of Attributes=36.

TABLE IV. FOR CHESS DATABASE

Large Item Sets	Time taken by Apriori	Time taken by Primitive
3	0.41	0.37
4	4.64	4.04
5	50.43	43.43

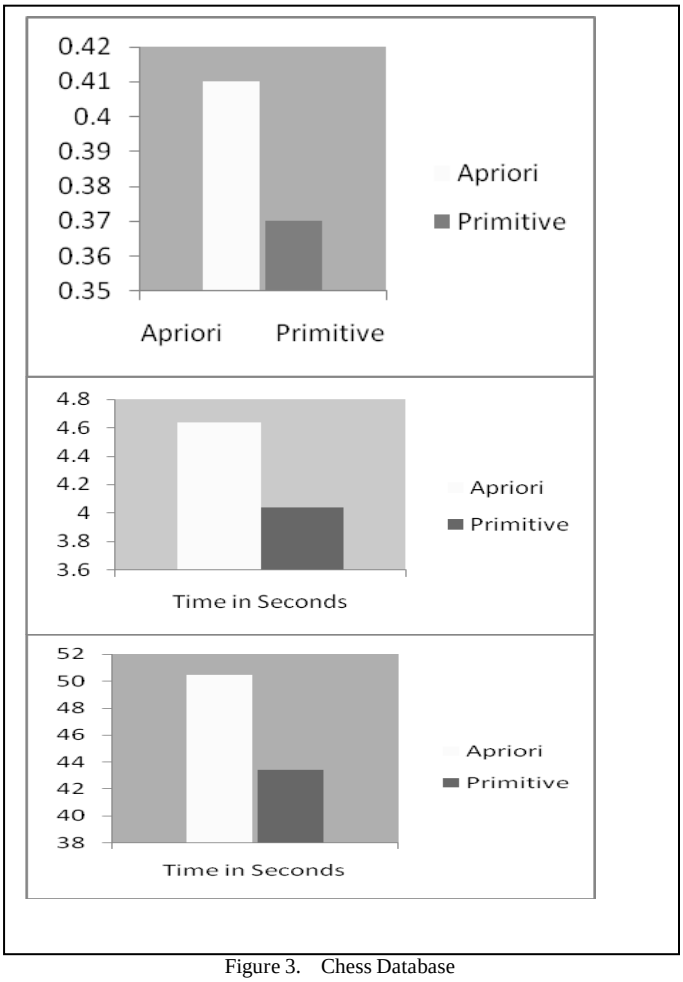


Figure 3. Chess Database

By looking at the above results it is clear that FP- Tree based approach are showing a clear edge because the number of database scans required are less which in turn reduces the computational time. Because the database is represented in tree structures which are taking less space so the overall memory requirement reduces.

CONCLUSION

Frequent pattern mining is the first step for association rule mining. Association rule mining has found many applications other than market basket analysis, including applications in marketing, customer segmentation, medicine, e-commerce, classification, clustering, web mining, bioinformatics and finance. Various techniques have been found to mine frequent patterns.

Each technique has its own pros and cons. Performance of particular technique depends on input data and available resources. Among all of the techniques discussed above, FP-Tree based approach achieves better performance by requiring only two database scans hence reducing the computational time. It takes less memory by representing large database in compact tree-structure. But a word of caution here that association rules should not be used directly for prediction without further analysis or domain knowledge. They are, however, a helpful starting point for further exploration & understanding of data. Experimental results have shown advantages of Primitive Association Rule Mining over Apriori.

REFERENCES

[1] Agrawal Rakesh, Imilienski T., and Swami Arun. Mining association rules between sets of items in large datasets. SIGMOD, 207-216, 1993

[2] Bayardo Roberto J. Efficiently Mining Long Patterns from Databases. SIGMOD, 83-93, Seattle, Washington, June 1998

[3] Brin Sergey, Motwani Rajeev, and Silverstein Craig. Beyond market baskets: Generalizing association rules to correlations. SIGMOD, 265-276, Tucson, AZ, USA, May 1997

[4] Brin Sergey, Motwani Rajeev, Ullman Jeffrey D., and Tsur Shalom. Dynamic itemset counting and implication rules for market basket data. SIGMOD, Tucson, AZ, USA, May 1997

[5] Chen Ming Syan, Park J. S., and Yu P. S. Efficient Data Mining for Path Traversal Patterns. IEEE Transactions on Knowledge and Data Engineering 10(2), 209-221, 1998

[6] Chen Xiaodong and Petrounias Ilias. Discovering temporal association rules: Algorithms, language and system. 2000 IEEE 16th International Conference on Data Engineering,San Diego, CA, USA, February 2000

[7] Cheung David W., Lee S. D., and Kao Benjamin. A General Incremental Technique for Maintaining Discovered Association Rules. Proc.International Conference On Database Systems For Advanced Applications, April 1997

[8] Han Jiawei, Pei Jian, Mortazavi-Asl Behzad, Chen Qiming, Dayal Umeshwar, and Hsu Mei-Chun. FreeSpan: Frequent pattern-projected sequential pattern mining. Boston, Ma, August 2000

[9] Han Jiawei, Pei Jian, and Yin Yiwen. Mining Frequent Patterns without Candidate Generation. SIGMOD, 1-12, Dallas, TX, May 2000

[10] Hipp Jochen, Güntzer Ulrich, and Nakhaeizadeh Gholamreza. Algorithms of Association Rule Mining - A General Survey and Comparison. SIGKDD Explorations 2(1), 58-64, 2000

[11] Hipp Jochen, Güntzer Ulrich, and Nakhaeizadeh Gholamreza. Mining Association Rules: Deriving a Superior Algorithm by Analyzing Today's Approaches. 159-168, Lyon, France, September 2000

[12] Lee Chang Hung, Lin Cheng Ru, and Chen Ming Syan. Sliding Window Filtering: An Efficient Method for incremental Mining on a Time-Variant Database. Proceedings of 10th International Conference on Information and Knowledge Management, 263-270, November 2001

[13] Orlando Salvatore, Palmerini P., and Perego Raffaele. Enhancing the Apriori Algorithm for Frequent Set Counting. 3rd International Conference on Data Warehousing and Knowledge Discovery, Germany, September 2001

[14] Pei Jian, Han Jiawei, Nishio Shojiro, Tang Shiwei, and Yang Dongqing. H-Mine: Hyper- Structure Mining of Frequent Patterns in Large Databases. Proc.2001 Int.Conf.on Data Mining, San Jose, CA, November 2001

[15] Savasere Ashok, Omiecinski Edward, and Navathe Shamkant. An Efficient Algorithm for Mining Association Rules in Large Databases. Proceedings of the Very Large Data Base Conference, September 1995

[16] Zaïane Osmar R. and Oliveira Stanley R. M. Privacy preserving frequent itemset mining. Workshop on Privacy, Security, and Data Mining, in conjunction with the IEEE International Conference on Data Mining, Japan, December 2002

[17] M. Houtsma and A. Swami. Set-oriented data mining in relational databases. Data Knowl. Eng.,245–262, 1995

[18] Christian Borgelt, An Implementation of the FP-growth Algorithm, OSDM'05, 2005

[19] S. Sarawagi, S. Thomas, and R. Agrawal. Integrating association rule mining with relational database systems: alternatives and implications. In SIGMOD, International conference on Management of data, pages 343–354, 1998

[20] X. Shang, K.-U. Sattler, and I. Geist. SQL based frequent pattern mining with fp-growth. In INAP/WLP, pages 32–46, 2004

[21] R. Agrawal and K. Shim. Developing tightly-coupled data mining application on a relational database system. In Proc.of the 2nd Int. Conf. on Knowledge Discovery in Database and Data Mining, Portland,Oregon, 1996

[22] J. Han, Y. Fu, W. Wang, K. Koperski, and O. Zaiane. DMQL: A data mining query language for relational database. In Proc. Of the 1996 SIGMOD workshop on research issues on data mining and knowledge discovery, Montreal, Canada, 1996

[23] R. Meo, G. Psaila, and S. Ceri. A new SQL like operator for mining association rules. In Proc. Of the 22nd Int. Conf. on Very Large Databases, Bombay, India, 1996

[24] Wang Ke, Tang Liu, Han Jiawei, and Liu Junqiang. Top down FP-Growth for Association Rule Mining. Proc.Pacific- Asia Conference, PAKDD 2002, 334-340, Taipei, Taiwan, May 2002

[25] Ozden Banu, Ramaswamy Sridhar, and Silberschatz Avi. Cyclic association rules. The 1998 14th International Conference on Data Engineering, 412-421, Orlando, FL, USA, February 1998

[26] Wang Jiinlong, Xu Conglfu, Cben Weidong, Pan Yunhe, Survey of the Study on Frequent Pattern Mining in Data Streams, 5917-5920, IEEE International Conference on Systems, Man and Cybernetics, 2004

[27] Jiawei Han, Hong Cheng, Dong Xin, Xifeng Yan, Frequent pattern mining: current status and future directions, 57-60, Data Mining Knowledge Discovery, 2007

[28] R. srikant, R. Agarwal, Mining sequential patterns: generalization and performance improvements, 1-15, IBM Research report, 1996

[29] Balazs Racz,Ferenc Bodon,Lars Schmidt-Thieme., On Benchmarking Frequent Itemset Mining Algorithms, from Measurement to Analysis, 37-42, Chicago, Illinois, USA, august 2005.