



International Journal of Advanced Computer Science and Applications

Volume 6 Issue 5

May 2015



ISSN 2156-5570(Online)

ISSN 2158-107X(Print)



www.ijacsa.thesai.org



W H E R E W I S D O M S H A R E S

INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS



THE SCIENCE AND INFORMATION ORGANIZATION

www.thesai.org | info@thesai.org

OAlster

getCITED



arXiv.org

DOAJ DIRECTORY OF
OPEN ACCESS
JOURNALS

IET InspecDirect

INDEX COPERNICUS
INTERNATIONAL



EBSCO
HOST
Research
Databases

Editorial Preface

From the Desk of Managing Editor...

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

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

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

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

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

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

Thank you for Sharing Wisdom!

Managing Editor
IJACSA
Volume 6 Issue 5 May 2015
ISSN 2156-5570 (Online)
ISSN 2158-107X (Print)
©2013 The Science and Information (SAI) Organization

Editorial Board

Editor-in-Chief

Dr. Kohei Arai - Saga University

Domains of Research: Technology Trends, Computer Vision, Decision Making, Information Retrieval, Networking, Simulation

Associate Editors

Chao-Tung Yang

Department of Computer Science, Tunghai University, Taiwan

Domain of Research: Software Engineering and Quality, High Performance Computing, Parallel and Distributed Computing, Parallel Computing

Elena SCUTELNICU

"Dunarea de Jos" University of Galati, Romania

Domain of Research: e-Learning, e-Learning Tools, Simulation

Krassen Stefanov

Professor at Sofia University St. Kliment Ohridski, Bulgaria

Domains of Research: e-Learning, Agents and Multi-agent Systems, Artificial Intelligence, Big Data, Cloud Computing, Data Retrieval and Data Mining, Distributed Systems, e-Learning Organisational Issues, e-Learning Tools, Educational Systems Design, Human Computer Interaction, Internet Security, Knowledge Engineering and Mining, Knowledge Representation, Ontology Engineering, Social Computing, Web-based Learning Communities, Wireless/ Mobile Applications

Maria-Angeles Grado-Caffaro

Scientific Consultant, Italy

Domain of Research: Electronics, Sensing and Sensor Networks

Mohd Helmy Abd Wahab

Universiti Tun Hussein Onn Malaysia

Domain of Research: Intelligent Systems, Data Mining, Databases

T. V. Prasad

Lingaya's University, India

Domain of Research: Intelligent Systems, Bioinformatics, Image Processing, Knowledge Representation, Natural Language Processing, Robotics

Reviewer Board Members

- **Abassi Ryma**
Higher Institute of Communications Studies of Tunis
, Iset'com
- **Abbas Karimi**
Islamic Azad University Arak Branch
- **Abdelghni Lakehal**
Université Abdelmalek Essaadi Faculté
Polydisciplinaire de Larache Route de Rabat, Km 2 -
Larache BP. 745 - Larache 92004. Maroc.
- **Abdel-Hameed A. Badawy**
Arkansas Tech University
- **Abdur Rashid Khan**
Gomal University
- **Abeer Mohamed ELkorany**
Faculty of computers and information, Cairo
Univesity
- **ADEMOLA ADESINA**
University of the Western Cape
- **Aderemi A. Atayero**
Covenant University
- **Ahmed S.A AL-Jumaily**
Ahlia University
- **Ahmed Boutejdar**
- **Ahmed Nabih Zaki Rashed**
Menoufia University
- **Akbar Hossain**
- **Akram Belghith**
University Of California, San Diego
- **Albert Alexander S**
Kongu Engineering College
- **Alci-nia Zita Sampaio**
Technical University of Lisbon
- **Alexandre Bouënard**
Sensopia
- **Ali Ismail Awad**
Luleå University of Technology
- **Amitava Biswas**
Cisco Systems
- **Anand Nayyar**
KCL Institute of Management and Technology,
Jalandhar
- **Andi Wahju Rahardjo Emanuel**
Maranatha Christian University
- **Andrews Samraj**
Mahendra Engineering College
- **Anirban Sarkar**
National Institute of Technology, Durgapur
- **Antonio Formisano**
- **Anuranjan misra**
Bhagwant Institute of Technology, Ghaziabad, India
- **Appasami Govindasamy**
- **Arash Habibi Lashkari**
University Technology Malaysia(UTM)
- **Aree Ali Mohammed**
Directorate of IT/ University of Sulaimani
- **Aris Skander Skander**
Constantine 1 University
- **Ashok Matani**
Government College of Engg, Amravati
- **Ashraf Mohammed Iqbal**
Dalhousie University and Capital Health
- **Ashraf Hamdy Owis**
Cairo University
- **Asoke Nath**
St. Xaviers College(Autonomous), 30 Park Street,
Kolkata-700 016
- **Ayad Ghany Ismaeel**
Department of Information Systems Engineering-
Technical Engineering College-Erbil Polytechnic
University, Erbil-Kurdistan Region- IRAQ
- **Ayman EL-SAYED**
Computer Science and Eng. Dept., Faculty of
Electronic Engineering, Menofia University
- **Babatunde Opeoluwa Akinkunmi**
University of Ibadan
- **Badre Bossoufi**
University of Liege
- **BASANT KUMAR VERMA**
JNTU
- **Basil Hamed**
Islamic University of Gaza
- **Basil M Hamed**
Islamic University of Gaza
- **Bhanu Prasad Pinnamaneni**
Rajalakshmi Engineering College; Matrix Vision
GmbH
- **Bharti Waman Gawali**
Department of Computer Science & information T

- **Bilian Song**
LinkedIn
- **Brahim Raouyane**
FSAC
- **Bright Keswani**
Associate Professor and Head, Department of
Computer Applications, Suresh Gyan Vihar
University, Jaipur (Rajasthan) INDIA
- **Brij Gupta**
University of New Brunswick
- **C Venkateswarlu Venkateswarlu Sonagiri**
JNTU
- **Chandrashekhar Meshram**
Chhattisgarh Swami Vivekananda Technical
University
- **Chao Wang**
- **Chao-Tung Yang**
Department of Computer Science, Tunghai
University
- **Charlie Obimbo**
University of Guelph
- **Chien-Peng Ho**
Information and Communications Research
Laboratories, Industrial Technology Research
Institute of Taiwan
- **Chun-Kit (Ben) Ngan**
The Pennsylvania State University
- **Ciprian Dobre**
University Politehnica of Bucharest
- **Constantin Filote**
Stefan cel Mare University of Suceava
- **Constantin POPESCU**
Department of Mathematics and Computer
Science, University of Oradea
- **CORNELIA AURORA Gyorödi**
University of Oradea
- **Dana - PETCU**
West University of Timisoara
- **Deepak Garg**
Thapar University
- **Dheyaa Kadhim**
University of Baghdad
- **Dong-Han Ham**
Chonnam National University
- **Dr K Ramani**
K.S.Rangasamy College of Technology,
Tiruchengode
- **Dr. Harish Garg**
Thapar University Patiala
- **Dr. Sanskruti V Patel**
Charotar Univeristy of Science & Technology,
Changa, Gujarat, India
- **Dr. Santosh Kumar**
Graphic Era University, Dehradun (UK)
- **Dr. JOHN S MANOHAR**
VTU, Belgaum
- **Dragana Becejski-Vujaklija**
University of Belgrade, Faculty of organizational
sciences
- **Driss EL OUADGHIRI**
- **Duck Hee Lee**
Medical Engineering R&D Center/Asan Institute for
Life Sciences/Asan Medical Center
- **Elena Camossi**
Joint Research Centre
- **Elena SCUTELNICU**
Dunarea de Jos University of Galati
- **Eui Chul Lee**
Sangmyung University
- **Evgeny Nikulchev**
Moscow Technological Institute
- **Ezekiel Uzor OKIKE**
UNIVERSITY OF BOTSWANA, GABORONE
- **FANGYONG HOU**
School of IT, Deakin University
- **Faris Al-Salem**
GCET
- **Firkhan Ali Hamid Ali**
UTHM
- **Fokrul Alom Mazarbhuiya**
King Khalid University
- **Frank AYO Ibikunle**
Botswana Int'l University of Science & Technology
(BIUST), Botswana.
- **Fu-Chien Kao**
Da-Y eh University
- **Gamil Abdel Azim**
Suez Canal University
- **Ganesh Chandra Sahoo**
RMRIMS
- **Gaurav Kumar**
Manav Bharti University, Solan Himachal Pradesh,
- **George Mastorakis**
Technological Educational Institute of Crete
- **George D. Pecherle**

- University of Oradea
- **Georgios Galatas**
The University of Texas at Arlington
 - **Gerard Dumancas**
Oklahoma Baptist University
 - **Ghalem Belalem Belalem**
University of Oran 1, Ahmed Ben Bella
 - **Giacomo Veneri**
University of Siena
 - **Giri Babu**
Indian Space Research Organisation
 - **Govindarajulu Salendra**
 - **Grebenisan Gavril**
University of Oradea
 - **Gufran Ahmad Ansari**
Qassim University
 - **Gunaseelan Devaraj**
Jazan University, Kingdom of Saudi Arabia
 - **GYÖRÖDI ROBERT STEFAN**
University of Oradea
 - **Hadj Hamma Tadjine**
IAV GmbH
 - **Hamid Mukhtar**
National University of Sciences and Technology
 - **Hamid Alinejad-Rokny**
The University of New South Wales
 - **Hamid Ali Abed AL-Asadi**
Department of Computer Science, Faculty of Education for Pure Science, Basra University
 - **Hany Kamal Hassan**
EPF
 - **Harco Leslie Hendric SPITS WARNARS**
Surya university
 - **Hazem I. El Shekh Ahmed**
Pure mathematics
 - **Hesham G. Ibrahim**
Faculty of Marine Resources, Al-Mergheb University
 - **Himanshu Aggarwal**
Department of Computer Engineering
 - **Hossam Faris**
 - **Huda K. AL-Jobori**
Ahlia University
 - **Iwan Setyawan**
Satya Wacana Christian University
 - **JAMAIAH HAJI YAHAYA**
NORTHERN UNIVERSITY OF MALAYSIA (UUM)
 - **James Patrick Henry Coleman**
Edge Hill University
 - **Jatinderkumar Ramdass Saini**
Narmada College of Computer Application, Bharuch
 - **Jayaram A M**
 - **Ji Zhu**
University of Illinois at Urbana Champaign
 - **Jia Uddin Jia**
Assistant Professor
 - **Jim Jing-Yan Wang**
The State University of New York at Buffalo, Buffalo, NY
 - **John P Sahlin**
George Washington University
 - **JOSE LUIS PASTRANA**
University of Malaga
 - **Jyoti Chaudhary**
high performance computing research lab
 - **K V.L.N.Acharyulu**
Bapatla Engineering college
 - **Ka-Chun Wong**
 - **Kashif Nisar**
Universiti Utara Malaysia
 - **Kayhan Zrar Ghafoor**
University Technology Malaysia
 - **Khin Wee Lai**
Biomedical Engineering Department, University Malaya
 - **KITIMAPORN CHOOCHOTE**
Prince of Songkla University, Phuket Campus
 - **Kohei Arai**
Saga University
 - **Krasimir Yankov Yordzhev**
South-West University, Faculty of Mathematics and Natural Sciences, Blagoevgrad, Bulgaria
 - **Krassen Stefanov Stefanov**
Professor at Sofia University St. Kliment Ohridski
 - **Labib Francis Gergis**
Misr Academy for Engineering and Technology
 - **Lazar Stošic**
Collegefor professional studies educators Aleksinac, Serbia
 - **Leandros A Maglaras**
University of Surrey
 - **Leon Andretti Abdillah**
Bina Darma University
 - **Lijian Sun**

- Chinese Academy of Surveying and
- **Ljubomir Jerinic**
University of Novi Sad, Faculty of Sciences,
Department of Mathematics and Computer Science
- **Lokesh Kumar Sharma**
Indian Council of Medical Research
- **Long Chen**
Qualcomm Incorporated
- **M. Reza Mashinchi**
Research Fellow
- **M. Tariq Bandy**
University of Kashmir
- **Manas deep**
Masters in Cyber Law & Information Security
- **Manju Kaushik**
- **Manoharan P.S.**
Associate Professor
- **Manoj Wadhwa**
Echelon Institute of Technology Faridabad
- **Manpreet Singh Manna**
Associate Professor, SLIET University, Govt. of India
- **Manuj Darbari**
BBD University
- **Marcellin Julius Antonio Nkenlifack**
University of Dschang
- **Maria-Angeles Grado-Caffaro**
Scientific Consultant
- **Marwan Alseid**
Applied Science Private University
- **Mazin S. Al-Hakeem**
LFU (Lebanese French University) - Erbil, IRAQ
- **MD RANA**
University of Sydney
- **Md. Zia Ur Rahman**
Narasaraopeta Engg. College, Narasaraopeta
- **Mehdi Bahrami**
University of California, Merced
- **Messaouda AZZOUZI**
Ziane AChour University of Djelfa
- **Milena Bogdanovic**
University of Nis, Teacher Training Faculty in Vranje
- **Miriampally Venkata Raghavendra**
Adama Science & Technology University, Ethiopia
- **Mirjana Popovic**
School of Electrical Engineering, Belgrade University
- **Miroslav Baca**
- University of Zagreb, Faculty of organization and
informatics / Center for biometrics
- **Mohamed Ali Mahjoub**
Preparatory Institute of Engineer of Monastir
- **Mohamed A. El-Sayed**
Faculty of Science, Fayoum University, Egypt.
- **Mohamed Najeh LAKHOUA**
ESTI, University of Carthage
- **Mohammad Ali Badamchizadeh**
University of Tabriz
- **Mohammad Hani Alomari**
Applied Science University
- **Mohammad Azzeh**
Applied Science university
- **Mohammad Jannati**
- **Mohammad Haghighat**
University of Miami
- **Mohammed Shamim Kaiser**
Institute of Information Technology
- **Mohammed Sadgal**
Cadi Ayyad University
- **Mohammed Abdulhameed Al-shabi**
Associate Professor
- **Mohammed Ali Hussain**
Sri Sai Madhavi Institute of Science & Technology
- **Mohd Helmy Abd Wahab**
Universiti Tun Hussein Onn Malaysia
- **Mona Elshinawy**
Howard University
- **Mostafa Mostafa Ezziyani**
FSTT
- **Mourad Amad**
Laboratory LAMOS, Bejaia University
- **Mueen Uddin**
University Malaysia Pahang
- **Murthy Sree Rama Chandra Dasika**
Geethanjali College of Engineering & Technology
- **Mustapha OUJAOURA**
Faculty of Science and Technology Béni-Mellal
- **MUTHUKUMAR S SUBRAMANYAM**
DGCT, ANNA UNIVERSITY
- **N.Ch. Sriman Narayana Iyengar**
VIT University,
- **Nagy Ramadan Darwish**
Department of Computer and Information Sciences,
Institute of Statistical Studies and Researches, Cairo
University.

- **Najib A. Kofahi**
Yarmouk University
- **Natarajan Subramanyam**
PES Institute of Technology
- **Nazeeruddin - Mohammad**
Prince Mohammad Bin Fahd University
- **NEERAJ SHUKLA**
ITM UNiversity, Gurgaon, (Haryana) Inida
- **Nestor Velasco-Bermeo**
UPFIM, Mexican Society of Artificial Intelligence
- **Nidhi Arora**
M.C.A. Institute, Ganpat University
- **Ning Cai**
Northwest University for Nationalities
- **Noura Aknin**
University Abdelamlek Essaadi
- **Oliviu Matei**
Technical University of Cluj-Napoca
- **Om Prakash Sangwan**
- **Omaima Nazar Al-Allaf**
Asesstant Professor
- **Osama Omer**
Aswan University
- **Ousmane THIARE**
Associate Professor University Gaston Berger of
Saint-Louis SENEGAL
- **Paresh V Virparia**
Sardar Patel University
- **Poonam Garg**
Institute of Management Technology, Ghaziabad
- **Prabhat K Mahanti**
UNIVERSITY OF NEW BRUNSWICK
- **PROF DURGA PRASAD SHARMA (PHD)**
AMUIT, MOEFDRE & External Consultant (IT) &
Technology Tansfer Research under ILO & UNDP,
Academic Ambassador for Cloud Offering IBM-USA
- **Professor Ajantha Herath**
- **Qifeng Qiao**
University of Virginia
- **Rachid Saadane**
EE departement EHTP
- **Raed Kanaan**
Amman Arab University
- **Raghuraj Singh**
Harcourt Butler Technological Institute
- **Rahul Malik**
- **Raja Sarath Kumar Boddu**

- LENORA COLLEGE OF ENGINEERNG
- **Rajesh Kumar**
National University of Singapore
- **Rakesh Chandra Balabantaray**
IIIT Bhubaneswar
- **Rakesh Kumar Dr.**
Madan Mohan Malviya University of Technology
- **Rashad Abdullah Al-Jawfi**
Ibb university
- **Rashid Sheikh**
Shri Aurobindo Institute of Technology, Indore
- **Ravi Prakash**
University of Mumbai
- **Ravisankar Hari**
CENTRAL TOBACCO RESEARCH INSTITUE
- **Rawya Y. Rizk**
Port Said University
- **Reshmy Krishnan**
Muscat College affiliated to stirling University.U
- **Ricardo Ângelo Rosa Vardasca**
Faculty of Engineering of University of Porto
- **Ritaban Dutta**
ISSL, CSIRO, Tasmaniia, Australia
- **Ruchika Malhotra**
Delhi Technoogical University
- **SAADI Slami**
University of Djelfa
- **Sachin Kumar Agrawal**
University of Limerick
- **Sagarmay Deb**
Central Queensland Universiry, Australia
- **Said Ghoniemy**
Taif University
- **Sandeep Reddivari**
University of North Florida
- **Sasan Adibi**
Research In Motion (RIM)
- **Satyendra Prasad Singh**
Professor
- **Sebastian Marius Rosu**
Special Telecommunications Service
- **Seema Shah**
Vidyalankar Institute of Technology Mumbai,
- **Selem Charfi**
University of Pays and Pays de l'Adour
- **SENGOTTUVELAN P**
Anna University, Chennai

- **Senol Piskin**
Istanbul Technical University, Informatics Institute
- **Sérgio André Ferreira**
School of Education and Psychology, Portuguese Catholic University
- **Seyed Hamidreza Mohades Kasaei**
University of Isfahan,
- **Shafiqul Abidin**
Northern India Engineering College (Affiliated to GGS I P University), New Delhi
- **Shahanawaj Ahamad**
The University of Al-Kharj
- **Shaiful Bakri Ismail**
- **Shawki A. Al-Dubae**
Assistant Professor
- **Sherif E. Hussein**
Mansoura University
- **Shriram K Vasudevan**
Amrita University
- **Siddhartha Jonnalagadda**
Mayo Clinic
- **Sim-Hui Tee**
Multimedia University
- **Simon Uzezi Ewedafe**
Baze University
- **Siniša Opic**
University of Zagreb, Faculty of Teacher Education
- **Sivakumar Poruran**
SKP ENGINEERING COLLEGE
- **Slim BEN SAOUD**
National Institute of Applied Sciences and Technology
- **Sohail Jabbar**
Bahria University
- **Sri Devi Ravana**
University of Malaya
- **Sudarson Jena**
GITAM University, Hyderabad
- **Suhas J Manangi**
Microsoft
- **SUKUMAR SENTHILKUMAR**
Universiti Sains Malaysia
- **Sumazly Sulaiman**
Institute of Space Science (ANGKASA), Universiti Kebangsaan Malaysia
- **Sumit Goyal**
National Dairy Research Institute
- **Suresh Sankaranarayanan**
Institut Teknologi Brunei
- **Susarla Venkata Ananta Rama Sastry**
JNTUK, Kakinada
- **Suxing Liu**
Arkansas State University
- **Syed Asif Ali**
SMI University Karachi Pakistan
- **T C.Manjunath**
HKBK College of Engg
- **T V Narayana rao Rao**
SNIST
- **T. V. Prasad**
Lingaya's University
- **Taiwo Ayodele**
Infonetmedia/University of Portsmouth
- **Tarek Fouad Gharib**
Ain Shams University
- **Thabet Mohamed Slimani**
College of Computer Science and Information Technology
- **Totok R. Biyanto**
Engineering Physics, ITS Surabaya
- **Touati Youcef**
Computer sce Lab LIASD - University of Paris 8
- **Uchechukwu Awada**
Dalian University of Technology
- **Urmila N Shrawankar**
GHRCE, Nagpur, India
- **Vaka MOHAN**
TRR COLLEGE OF ENGINEERING
- **Vinayak K Bairagi**
AISSMS Institute of Information Technology, Pune
- **Vishnu Narayan Mishra**
SVNIT, Surat
- **Vitus S.W. Lam**
The University of Hong Kong
- **VUDA SREENIVASARAO**
PROFESSOR AND DEAN, St.Mary's Integrated Campus,Hyderabad.
- **Wei Wei**
Xi'an Univ. of Tech.
- **Xiaoqing Xiang**
AT&T Labs
- **Yi Fei Wang**
The University of British Columbia
- **Yihong Yuan**

University of California Santa Barbara

- **Yilun Shang**
Tongji University
- **Yu Qi**
Mesh Capital LLC
- **Zacchaeus Oni Omogbadegun**
Covenant University
- **Zairi Ismael Rizman**
Universiti Teknologi MARA
- **Zenzo Polite Ncube**
North West University

- **Zhao Zhang**
Department of EE, City University of Hong Kong
- **Zhixin Chen**
ILX Lightwave Corporation
- **Ziyue Xu**
National Institutes of Health, Bethesda, MD
- **Zlatko Stapic**
University of Zagreb, Faculty of Organization and Informatics Varazdin
- **Zuraini Ismail**
Universiti Teknologi Malaysia

CONTENTS

Paper 1: A New Bio-Informatics Framework: Research on 3D Sensor Data of Human Activities

Authors: Sajid Ali, *Wu Zhongke, Muhammad Saad Khan, Ahmad Tisman Pasha, Muhammad Adnan Khalid, Zhou Mingquan

PAGE 1 – 11

Paper 2: A Generic Adaptive Multi-Gene-Set Genetic Algorithm (AMGA)

Authors: Adi A. Maaita, Jamal Zraqou, Fadi Hamad, Hamza A. Al-Sewadi

PAGE 12 – 18

Paper 3: Real-Time Digital Image Exposure Status Detection and Circuit Implementation

Authors: Li Hongqin, Wu Jianzhen, Zhang Liping, Ning Jun

PAGE 19 – 24

Paper 4: Quantifying the Relationship between Hit Count Estimates and Wikipedia Article Traffic

Authors: Tina Tian, Ankur Agrawal

PAGE 25 – 28

Paper 5: Applying Linked Data Technologies for Online Newspapers

Authors: Tsvetanka Georgieva-Trifonova, Tihomir Stefanov

PAGE 29 – 33

Paper 6: Fall Monitoring Device for Old People based on Tri-Axial Accelerometer

Authors: Jing Luo, Bocheng Zhong, Dinghao Lv

PAGE 34 – 39

Paper 7: The Fir Digital Filter Design based on Iwpso

Authors: Xinnan Hu, Yujia Wang, Kun Su

PAGE 40 – 44

Paper 8: Analysis and Research of Communication Interrupt Fault for Shanghai Metro Data Transmission System

Authors: Jianru Liang, Xinyuan Lu, Cong Shi, Minglai Yang

PAGE 45 – 48

Paper 9: A Fuzzy PI Speed Controller based on Feedback Compensation Strategy for PMSM

Authors: Ou Sheng, Liu Haishan, Liu Guoying, Zeng Guohui, Zhan Xing, Wang Qingzhen, Liu Haishan

PAGE 49 – 54

Paper 10: CPLD-Based Circuit Design of IGBT Dead-Time Compensation

Authors: Qing-zhen WANG, Guo-hui ZENG, Jin LIU, Xing ZHAN

PAGE 55 – 59

Paper 11: Detection and Removal of Gray, Black and Cooperative Black Hole Attacks in AODV Technique

Authors: Hosny M. Ibrahim, Nagwa M. Omar, Ebram K. William

PAGE 60 – 70

Paper 12: Adoption of e-Government in Pakistan: Demand Perspective

Authors: Zulfiqar Haider, Chen Shuwen, Zareen Abbassi

PAGE 71 – 80

Paper 13: Applying Topology-Shape-Metric and FUZZY Genetic Algorithm for Automatic Planar Hierarchical and Orthogonal Graphs

Authors: Nahla F.Omran, Sara F. Abd-el ghany

PAGE 81 – 87

Paper 14: Performance Enhancement of Scheduling Algorithm in Heterogeneous Distributed Computing Systems

Authors: Aida A. NASR, Nirmeen A. EL-BAHNASAWY, Ayman EL-SAYED

PAGE 88 – 96

Paper 15: Numerical Evaluation of the Effect of Gradient on Reflection Coefficient of Continuously Graded Layer

Authors: Ahmed Markou, Hassan Nounah

PAGE 97 – 102

Paper 16: Automatic Ferrite Content Measurement based on Image Analysis and Pattern Classification

Authors: Hafiz Muhammad Tanveer, Hafiz Muhammad Tahir Mustafa, Waleed Asif, Munir Ahmad, Muhammad Anjum Javed, Maqsood Ahmad

PAGE 103 – 108

Paper 17: Development of Eye-Blink and Face Corpora for Research in Human Computer Interaction

Authors: Emmanuel Jadesola Adejoke., Ibiyemi Tunji Samuel

PAGE 109 – 111

Paper 18: Comparison Fractal Color Image Compression using YIQ and YUV Color Model

Authors: Eman A. Al-Hilo, Rusul Zehwar

PAGE 112 – 116

Paper 19: A 'Cognitive Driving Framework' for Collision Avoidance in Autonomous Vehicles

Authors: Alan J. Hamlet, Carl D. Crane

PAGE 117 – 124

Paper 20: Denoising CT Images using wavelet transform

Authors: Lubna Gabralla, Hela Mahersia, Marwan Zaroug

PAGE 125 – 129

A New Bio-Informatics Framework: Research on 3D Sensor Data of Human Activities

Sajid Ali

College of Information Science and Technology, Beijing
Normal University, Engineering Research Center of Virtual
Reality and Application, Ministry of Education, Beijing,
China

*Wu Zhongke

College of Information Science and Technology, Beijing
Normal University, Engineering Research Center of Virtual
Reality and Application, Ministry of Education, Beijing,
China

Muhammad Saad Khan

School of Electronic Engineering
Beijing University of Post Telecommunication, Beijing,
China

Ahmad Tisman Pasha

Department of Information Technology,
Institute of Computing,
Bahauddin Zakariya University
Multan, Pakistan

Muhammad Adnan Khalid

Institute of Automation
Chinese Academy of Sciences,
Beijing, China

Zhou Mingquan

College of Information Science and Technology, Beijing
Normal University, Engineering Research Center of Virtual
Reality and Application, Ministry of Education, Beijing,
China

Abstract—Due to increasing attraction of motion capture systems technology and the usage of captured data in wide range of research-oriented applications, a framework has developed as an improved version of MOCAP TOOLBOX in Matlab platform. Firstly, we have introduced a faithful script to deal with public motion capture data, which will be friendly for us. Various functions through dynamic programming, by using the Body Segment Parameters (BSP) are edited and they configured the position of markers according to data. It is used to visualize and refine without the MLS view and the C3D editor software. It has opened a valuable way of sensor data in many research aspects as gait movements, marker analysis, compression and motion pattern, bioinformatics, and animation. As a result, performed on CMU and ACCAD public mocap data, and achieved higher corrected configuration scheme of 3D markers when compared with the prior art, especially for C3D file. Another distinction of this work is that it handles the extra markers distortion, and provides the meaningful way to use captured data.

Keywords—Sensor data; Bioinformatics behavior; Optical actives; BSP domain; Marker Configuration

I. INTRODUCTION

Due to improvement in motion, capture technologies such as optical, mechanical, or magnetic sensors are attached to human joints and their movements are record. Such systems depend on an active source, which emits pulses of infrared light with a high frequency, which is reflected by small spherical markers, or LEDs attached to track the subject (e.g. contribution of subject walking, or running). In motion capture system, every camera captures the position of reflective markers in two-dimensional cameras. Network of system computes position data in 3-dimensions. Now more and more researchers are interested to use 3D-mocap data with different

types of application; such as retarget [1-3], analysis [4], animation [5] and surveillance systems [6]. In addition, they have demanded for different types of toolbox to utilize the variety of data ASF/AMC, BVH [7], and C3D [8] file format for multiple research purposes. Some of them have interested to develop mocap toolbox; which are helpful to other scientists. They are focused to use the captured data in various research directions. Recently, Jeroen and Boxel [9] have developed a toolbox (Biomotion toolbox) in the Matlab environment that can read and display the different types of mocap data by using Psychtoolbox-3[10]. This third party toolbox (Psychtoolbox-3) is not exactly suitable for biomotion. However, some features of the biomotion are limited, and have specific designs to display and manipulate point-light displays (PLD). Charles Vernon [11], developed a toolbox with a limited number of functions. It provides a graphical user interface (GUI). The major mocap toolbox in MathWorks platform is mainly dealing with recorded data by infrared marker based optical motion capture system.

The MathWork provides precompiled functions[12], and they are used on different types of data. Some of the functions are used as part of mocap toolbox, such as PCA and ICA packages, Signal Processing and Statistical Toolbox. So user can design functions and scripts in the Matlab environment according to their requirements. Recently, Burger and Petri [13] developed the MoCap Toolbox having 64 functions, excluding other toolbox packages. These functions have been used to visualize and analyze captured data and have the capability to read the different types of data format. They notified three parts: 1)-Motion Data Structure (MDS), 2)-Segment Data Structure (SDS) and 3)-Normal Data Structure. These structures have interconnection and processed. Their

computation implied statistical and mathematical methods in order to propose a homogeneous framework and their analysis and simulation (animation). They also claimed that their work can read C3D file, but still it has some indispensable issues to trade with such type of data. It cannot read the public mocap data. We have inspired of work [14], our focus is to refine functions and use them on data (C3D format) which will be used for different purposes such as in clinical field, retarget motion and animation. In [13], we have big issues and challenges for mocap data researchers. These are: 1)-reading the C3D file not displaying in human skeleton shape (see Figure 1 a), 2)-Markers configuration which shows the motion data structure of the human skeleton, 3)-Normalized the data of the human skeleton (see Figure 1 b, c); and 4) -required Qualisys software to manage the .dat and.mat data format. We have tried hard to resolve these individual issues see.

```
??? Error using ==> fopen % if "DEC" selected in
Invalid machine format. export c3d options in IQ
you
Error in ==> readc3d at 40
fid=fopen(fname,'r',machinetype);
will get an error, change to PC
Error in ==> mcreadc3d at 13
data = readc3d(fn);
Error in ==> mcread at 57
d = mcreadc3d(fn);
Error in ==> try_mc at 28
walk1= mcread(files{i,:});
```

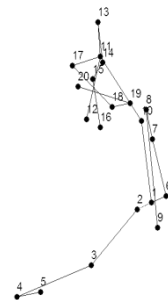
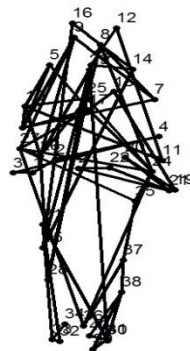
```
Output= mcread ('E: \walk.c3d'); (1)
```

We use some of the existing functions after embedding script (C3D_VaxD2PC) into the toolbox and read the data successfully by (1). The following functions have been used to display calibrated 3D markers position and form them into the human skeleton shape.

However, it is not like human skeleton. It looks like network connection between marker nodes because some existing toolbox functions are not operated according to marker positions (see Figure-1 (b) & (c)). The functions used on mocap data before modification are:

```
mcplotframe(walk1, 160, mapar);
mcplotframe(walk2j, 160, japar);
```

```
type: 'MoCap data'
filename: 'E:\MoCapToolbox\mocaptoolbox\walk_8\08_01.c3d'
nFrames: 772
nCameras: 0
nMarkers: 41
freq: 120
nAnalog: []
anaFreq: []
timerOrder: 0
markerName: {41x1 cell}
data: [277x123 double]
analogdata: []
other: [1x1 struct]
```



(a)

(b)

(c)

Fig. 1. (a) Indicates the errors to in reading the C3D file, before introducing a script. After removing, the errors and we still have a tremendous data problem such as frame issue, (b) existing setting markers position in function.

Keeping the above issues, we have refined some functions of Mocap Toolbox to stem these issues and compiled them successfully according to data [15]& [16]. It can be associated with other toolboxes in the Matlab environment such as Mocap136 [17] and Robotics [18]. The pictorial structure of the improved version of Toolbox has been given in Figure 2.

Lot of research works have been done on mocap data. Improved version can be used in different research fields such as joint analysis, design the locomotion pattern, retarget motion, human skeleton animation, motion classification, 3D

pose estimation and human identification by using 3D mocap databases plausible with the human body segment parameters (BSP).

The rest of this paper is set up as follows. Section II gives a concise story about motion capture systems, and data preprocessing. Section III describes the configuration scheme of sensors and data visualization functions. Brief descriptions of the human Body Segment Parameters (BSP) and data normalization are discussed in section IV. The experimental results, conclusion and future work are explained in section V.

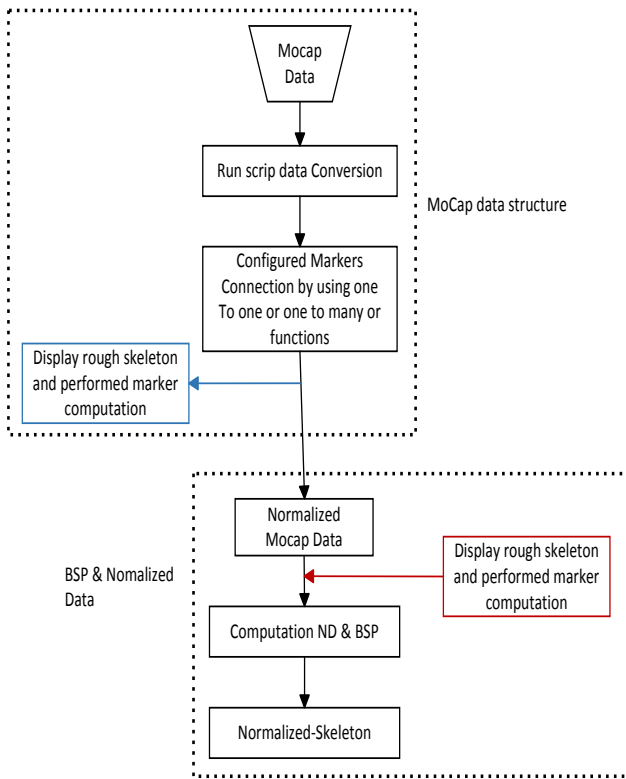


Fig. 2. Flowchart of the improved version mocap toolbox

II. MOCAP DATA SYSTEMS

Motion Capture Sensor Systems (MCSS) record the MOCAP data. Such systems are highly sophisticated, and require a certain number of motion capture session during subject activities. They are classified into two categories (1) Optical and (2) Non-optical Capture Sensor Systems. They provide different type data format. They are *. C3D, *. BVH, *.txt,*.tvt, and *. ASF/AMC.

A. Mocap Data Preprocessing

Preprocessing of data is a primary step to get accurate results in any scientific research [19]. Data captured by ubiquitous sensors based on subject movement is stored as the C3D file format with necessary attribute such as start and end times, sensor id and sensor values. It depends upon several types of hardware platform issues such as DEC (Digital Equipment Corporation), SGI/MIPS (Microprocessor without Interlocked Pipeline Stages)) and Intel. They represent the different floating-point numbers and are stored accordingly in hardware (VAX-D, IEEE-LE and IEEE-BE Vicon call them a "C3D_VaxD2PC". It establishes the strong connection between MOCAP TOOLBOX function and data. It is used before the mcread () function, its syntax as

```
Output =C3D_VaxD2PC ('Convert',' data location'); (2)
```

It converts the C3D files into PC format, because these files depend on several types of hardware and floating-point. After compiling data through (2). MocapToolbox mcread () function, read C3D files, and arrange them into a reliable database. Here its name is called 'Processed Data for MocapToolbox'.

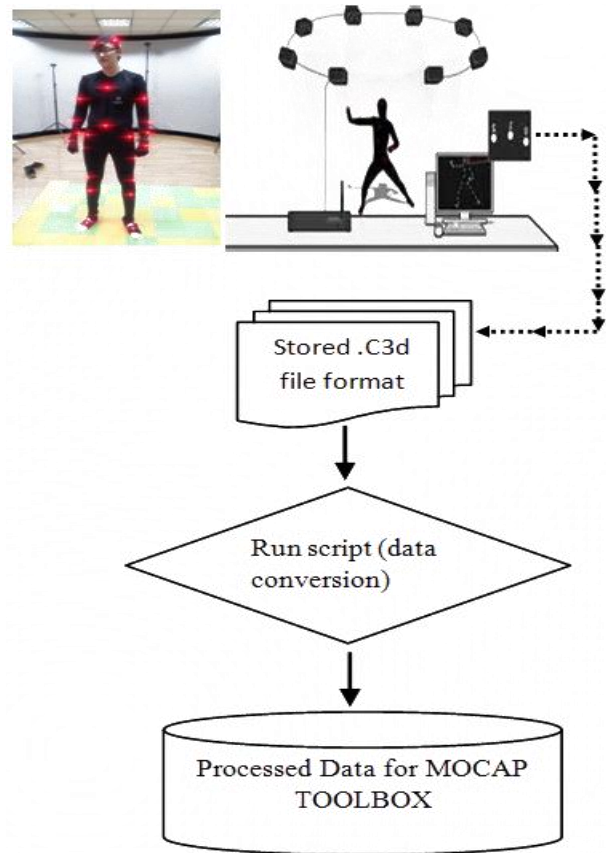


Fig. 3. Procedure of processing public domain mocap data for enhanced MOCAP TOOLBOX

As shown in Figure 1 (a), an error has resolved by emitting this function. The graphical representation of the database can be seen in Figure 3.

After reading the C3D file successfully, some remaining typical issues have shown in Figure 1 (c & d). They have affected the standard human skeleton; attached markers on human body and joint positions. Both of these issues will be addressed in the section III through mapar.conn and japar.conn functions.

III. MOTION SENSORS CONFIGURATION AND VISUALIZATION

A. Motion Sensors Configuration

The markers or motion sensor configuration on the human body has many possible ways for motion recording one of them is shown in Figure-4.

In Figure-4, 41 sensors have been configured on the human body to follow the configuration scheme of CMU motion capture system and are described in the Vicon 512 manual [20]. We analyzed the sensor labels of the C3D file by using 3MAX software. Keeping to these labels, we assessed the sensor values after editing the mcreadC3D function according to the markers order of template (see Figure-A.1). Data keeps maintaining some dumping markers during the motion capture session of human activities such as walking, running, dancing etc.

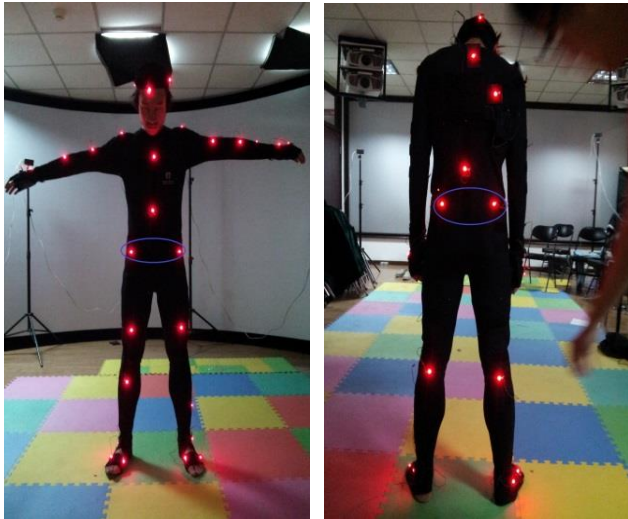


Fig. 4. An example, markers, or sensor configuration scheme on the human body according of CMU data

We modify the `mcread C3D` function by using the following steps:

1) First access the index of motion sensors which are placed on the human body by using the code (Appendix A.1. 1):

For example, accessing the `L_finger` index of motion sensor from C3D file of CMU database (see Appendix A.1).

Similarly have accessed the other remaining 40 or 41 labels from file.

2) Some extra markers have stored during the recording we handled and assigned zero values; and adjusted them with marker labels.

3) Many mocap data, scientists use C3D format. They feel hard to export and construct raw motion data to the desired model. They have unmatched position between the calibration motion sensors. These positions are unable to make human model. For this matter, we used a `mapar.conn()`; and configure according to public mocap data. It is configured by using the mathematical methods. They are as one to one and one to many (see Figure 5). The following function parameters can be driven (see Appendix A.1.1 for code of accessing sensor labels),

```
mapar.conn=[12 16;16 9;9 8;12 8;14 7;7 4;4 11;11 19;19 21;11 24;21 24;...
5 10;10 6;6 27;27 26;27 3;3 1;1 26; 22 23;22 20; 20 2; 2 23;31 15;15 28;...
28 34;34 36;36 32;34 33;33 32;35 37;37 38;38 30;30 39;30 41;41 40;39 40;...
23 31;31 2;23 15;2 15;20 35;22 35;20 37;22 37;5 23;2 5;22 14;20 14;14 13;...
13 5;23 25;2 25;20 17;22 17;13 29;29 25;29 17;17 18;25 18;5 25;25 14;...
17 5; 17 14; 20 25; 5 18; 18 14];
```

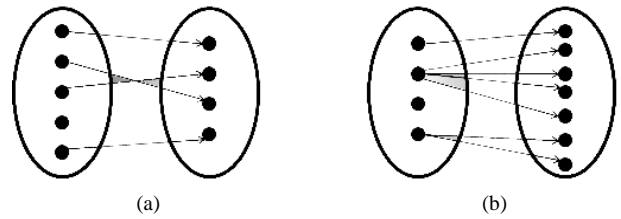


Fig. 5. Configuration scheme between markers, (a) One to one; and (b) One to many mapping

After successful compilation steps 1, 2, and 3 on data, step to construct the human skeleton model by using the following function, seen in Figure 6.

```
mcplotframe (walkm,180, mapar); (3)
```

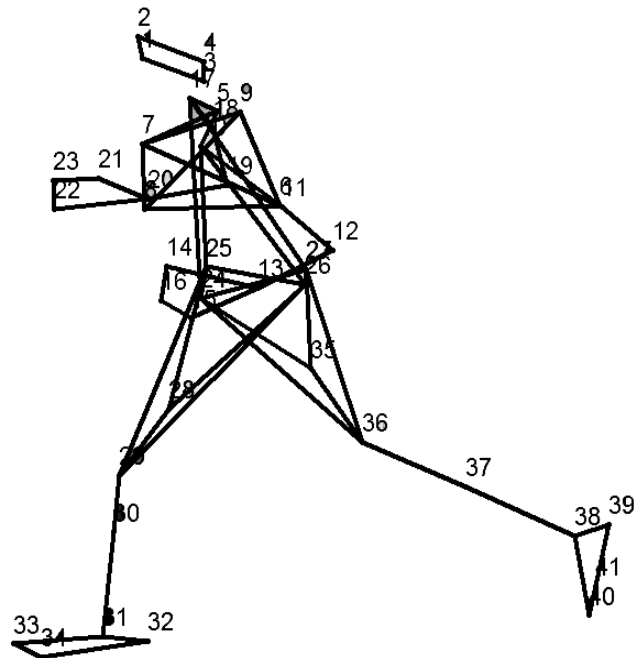


Fig. 6. Specific pose of skeleton after steps 1, 2, and 3 by using function (3)

IV. HUMAN BODY SEGMENT PARAMETERS AND DATA NORMALIZATION

There are many methods [21-31] to compute the human body segment parameters. They have trendy source of BSP knowledge of human body. The BSP and mocap data should comply on each other.

A. Human Body Segment Parameters

Body segment parameters perform an important role to generate motion of human activities by motion capture systems. The `mcgetsegmpar()` function parameters of the Mocap Toolbox use the Dempster computed BSP data [32]. We replaced BSP parameters [32] of `mcgetsegmpar()` function to Zatsiorsky and Seluyanov adjusted by the de Leva [33], and their computation is [34]. For instance, an example of the BSP computation of specific human body part can be seen Eqs A.1 to A.6 in Appendix A.1.3. These parameters and 3D motion

capture data used to visualize the standard human skeleton with 20 joints (see Figure 7).

B. Mocap Data Normalization and Visualization of Skeleton Refinement

In order to investigate the marker's positions of C3D file and their indexing, we configured the `mapar.conn()` and `japar.conn()` function according to indices. The source code of markers indexing configuration (Appendix A.1.2).

The `m2jpar()` depends on `mcm2j()` function. It has the knowledge to compute the translation from markers to joint representation. The joint position has computed by applying the center method around placed marker on human body joints. For instance, the root position of joint is found-out between the 22,23,2 and 20 marker (markers marked with blue ellipses see Figure 4). Similarly, other joint positions are computed. We label each joint with specific name. They can be seen in Table 1. The following function uses to initialize required joint parameters.

```
japar = mcinitanimpar;
```

It contains information that will be helpful to initialize joint parameters and assigns the attributes of `japar()` structure. One of the fields of this structure is to edit the parameter by putting joint index of `m2jpar` and make the connection between these indices by applying the faithful methods (see Figure-5). These functions are used by applying `japar.conn()` Parameters can be accessed (see source code in Appendix A.1.3) and attain human skeleton (see Figure-7 with a specific pose and generated animation motion frame see Figure-8).

```
japar.conn=[1 2; 2 3; 3 4; 4 5...
1 6;6 7;7 8;8 9;...
1 10; 10 11; 11 12;...
11 13; 13 14; 14
15; 15 16...
11 17; 17 18; 18 19; 19 20];
```

Expect of two joints 11 and 1, used the one to one function definition. The following function visualize the skeleton having 20 joints (see Figure 7),

```
mcplotframe (walk2j,180, japar);
```

The following function performs the animation pose of extracted skeleton of the C3D file from CMU database. It also has ability to create animation of the mocap data after editing some mocap Toolbox functions. The Figure 8 is created by using the `mcanimate()` function as,

```
mcanimate (walk2j, 15, japar); (4)
```

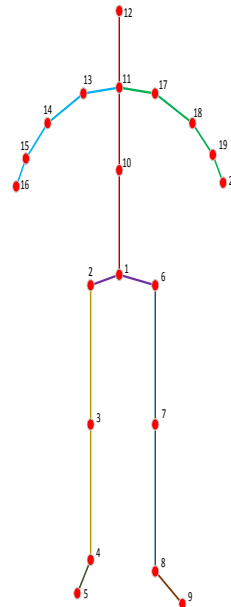


Fig. 7. Normalized Skeleton

TABLE I. SKELETON JOINT NAMES

1	root
2	lhip
3	lknee
4	lankle
5	ltoe
6	rhip
7	rknee
8	rankle
9	rtoe
10	midtors
11	neck
12	head
13	lshoulder
14	lelbow
15	lwrist
16	lfinger
17	rshoulder
18	relbow
19	rwrist
20	rfinger

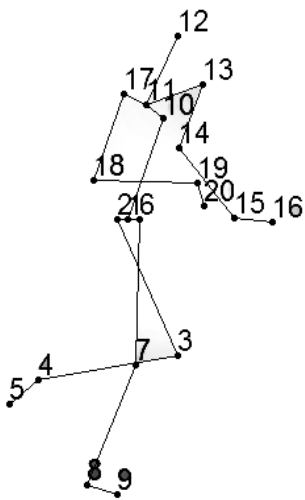


Fig. 8. An example of specific animated pose by using the function (4)

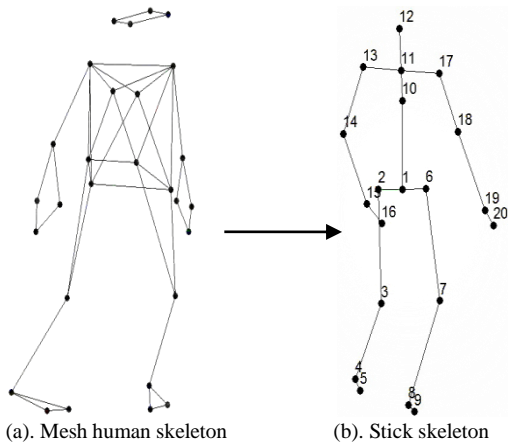


Fig. 9. (a) Triangular mesh skeleton contain 28 sensors, and (b) human skeleton with 20 joints [34]

After modification of toolbox functions, it can be enabled to read the C3D file successfully from public 3D mocap databases. Some of the existing function performances will be discussed in the section V.

V. RESULTS & DISCUSSION

In this paper, we will improve the Mocap Toolbox by introducing new script and editing some functions, it can be used as input mocap data such as CMU, and ACCAD (Advance Computing Centre for the Arts and Design) databases. Earlier, Toolbox demonstrates 28 sensors and formed a triangular mesh human model and extract a skeleton (see Figure-9 a & b) by applying the structure of connection matrix. It had been collected by using the Qualisys motion capture system. After that, we used public mocap data (C3D files) and found the fundamental errors (see Figure 1). This shows that data could not support because of more than 28 sensors.

Meanwhile, the public data has at least 41, or 42 markers, it is as standard form but more than that such as 80, 90, and 356 create the ambiguity and difficult trouble to understand in many research fields. This issue can be handled by adding the

code into mcreadC3Dc3d (). This code would be seen in (Appendix A.1.2.). We modified a list of functions of the Mocap Toolbox according to CMU and ACCAD mocap database. The list of functions is as mcread C3Dc3d (), mapar.conn (), japar.conn , mcm2j() (), mcgetsegmpar (), and m2jpar (). The visualization of placed markers on the human body and is transferred into skeleton by using some editing functions. It is illustrated in Figure 10. It shows the effects of the edited functions. The rest of functions also can also be used for public mocap data. Some examples are as follows:

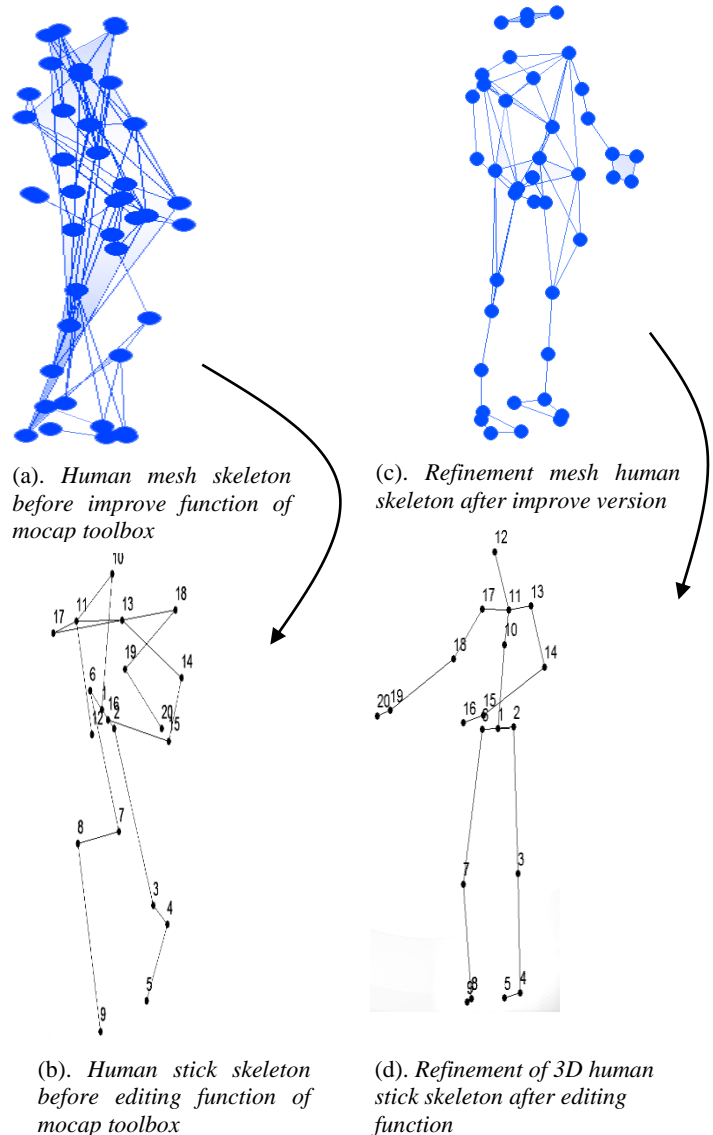


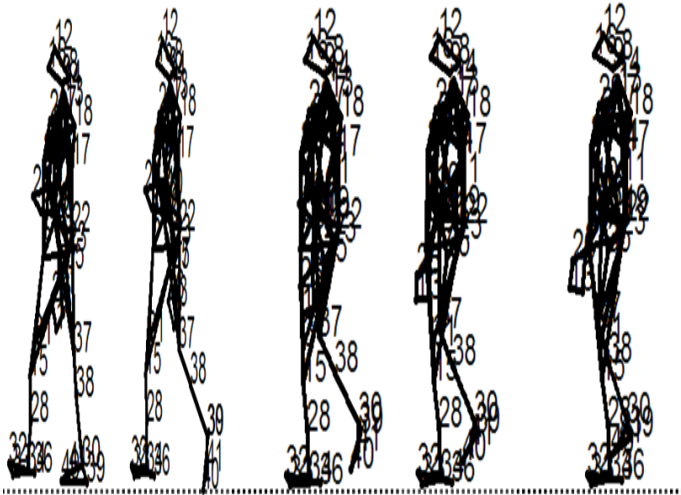
Fig. 10. (a) & (b) describe the sensor connection through mapar.conn structure field to construct a mesh human skeleton and human stick skeleton extracted by using mcm2j function is used to transform these sensors information to the joint positions;(c) and (d) proved the edited functions that explained in section III, IV

```
newpar1 = mcanimate(walk1, mapar); (5)
newpar2 = mcanimate(walk2j, japar); (6)
```

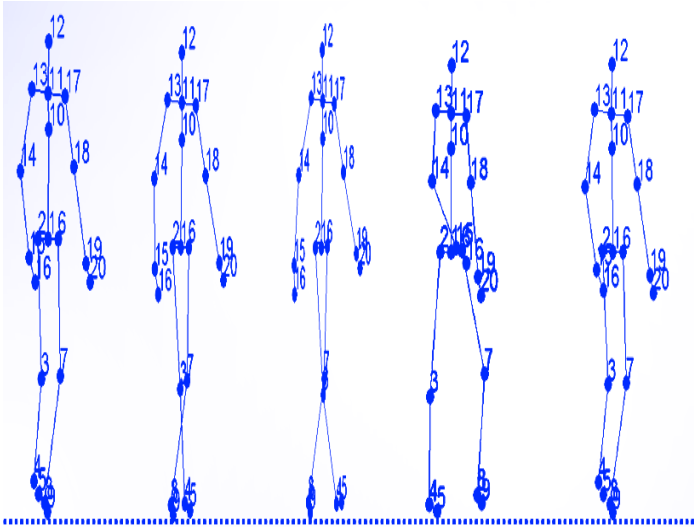
These two functions (5 & 6) executed successfully on mentioned data, and demonstrate the following animation

poses. It will be used to analyze the performance of placement marker pose by pose of human movements.

As shown in Figure 11. (a), the successful result of animation poses of the human mesh skeleton, between markers interconnection, which complied with all required parameters in function (8). Figure 11. (b) shows the successful result of animation poses of normalized human skeleton, which extracted from 41 markers, and compiled all required parameters in function (9). The parameters (walk1 and mapar) deal with placing markers on the human body; and they (walk2j, japar) deal with human joint positions.



(b)An example walking animation poses of triangular mesh human model between motion sensors connection



(a)An example of animation poses of normalized human stick model

Fig. 11. After modification functions results (a) marker positions of animation poses of the human waking results, and (b) normalized human skeleton of walking animation poses of CMU mocap data

The following functions are helpful to assess the accuracy of placing sensors on the human body during the motion recording sessions giving the information of missing markers, which are very useful for researchers. (See Figure 12). In

Figure-12, black shaded colors indicate the missing markers and number of frames from a C3D file in CMU database.

```
[mf, mm, mgrid] = mcmissing (walk1);
subplot(3,1,1),
bar(mf), xlabel('Marker'), ylabel('Num.
of Missing frames')
subplot(3,1,2),
bar(mm), xlabel('Frame'), ylabel('Num. of
Missing markers')subplot(3,1,3),
imagesc(-mgrid), colormap gray,
xlabel('Frame'), ylabel('Marker')
```

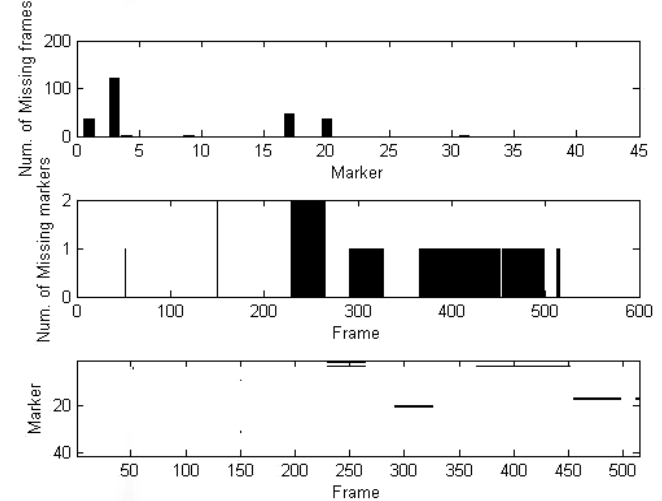


Fig. 12. Analysis of sensors and frames

The rest of other existing functions in Mocap toolbox (framework) can be applied on mocap data, and some of them can plot with respect to time as a `mplottimeseries` function. For example, hip joint sensor with xyz coordinates missing frames, the missing area marked with a circle (see Figure 13) information is related to female subject no B20 in ACCAD mocap database.

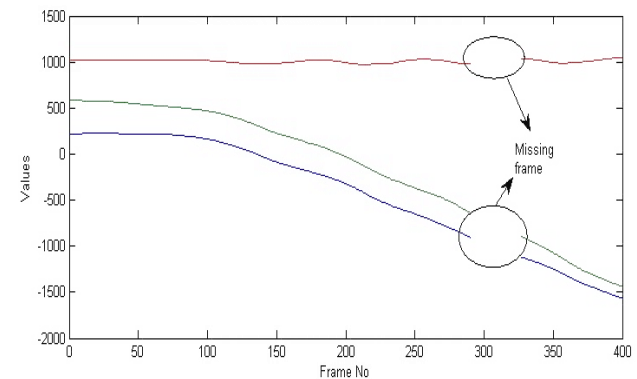


Fig. 13. An example of hip joint movement missing frame

We compared some of the above functions (see Table 2) to early version of the toolbox with respected to response time under a platform. We performed this evaluation on two public mocap databases and existing data of the toolbox. The details of these experiments are presented above. We have performed all above experiments under some specifications such as 2 GB

RAM, Intel(R) Core™ i5 CPU M520 @ 2.4 GHz Dell i5 Intel core 2 CPU 2.4 and window 7 ultimate 64-bit and MATLAB R2012a. In Table 2, fourth column contains two colors red and blue ellipse. The red denotes the edited functions and blue indicate early function response time. Finally, we concluded that red has taken more time to blue because these function use public data but opened for everyone who want to test mocap data for multiple applications. The conclusion is formed in Figure 14 with functions time response.

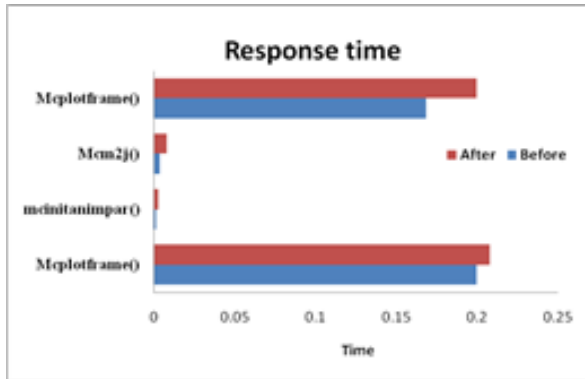


Fig. 14. Response timing comparison between different function. The red bar indicates our editing function according to public mocap data, and blue indicates the early functions with 28 markers configuration

VI. CONCLUSION AND FURTUR WORK

In this study, improved version is used partly due to the inherent structure information from Mocap toolbox. The key assumption of this improved version is that there is a high probability of using public mocap data. It is in native C3D file format. It has rich and plentiful information of human, which offers various types of numerical treatment. In future, we hope that it will be allowed for a prosperous variety of functions used for motion capture data for numerous research fields like animation, human joint analysis, gender and human identification (bioinformatics), retarget motion in real time environment.

Our improved version is proved to be effective in finding human mesh and stick skeleton model from public data as compared to capture with previous development platform [34] because it support only 28 3D markers. These skeleton models will be useful for joint movements' analysis and multiple purposes in indoor and outdoor environment. In addition, it gives the information about the quality of mocap-captured data (see Figure 12, 13).

Based on the investigation, we conclude that which type of data is more reliable. The Xense¹ has introduced a state-of-the-art of inertial high sensory quality systems to capture the motion of subjects with wide range of area. It is capable of supporting and generating different motion formats, such as BVH, C3D etc. These types of data is very different from the image and video data. It has real and truthful relics of the human and object activities.

¹ www.xsens.com

TABLE II. TIME EVALUATION OF BEFORE AND AFTER EDITING TOOLBOX FUNCTIONS

S. No	Function Name	3D MoCap Data	CPU Elapsed time With respect to functions & data	Comments
1	mcreadc3d()	CMU & ACCA D database s	0.204703s	Before improved of it, it could not be operated on public Mocap data. It prepare data by Qualisys Mocap system then author used. After improved Toolbox, we used public Mocap data. This is response time to read C3D file from Mocap database.
2	Mcplotframe()	Qualisys Mocap System with 28 makers	0.199483s	This is response time of Mcplotframe(), and display See (Figure 9.a). It contains 28 sensor moreover, shows Triangular human mesh model.
3	Mcplotframe()	CMU & ACCA D database s	0.207661s	This is response time of Mcplotframe(), and display See (Figure 10.c). It contain 41 or 42 sensor moreover, shows Triangular human mesh model.
4	mcinitanimpar	Qualisys Mocap System with 28 makers	0.000762s	It is response time to initialize the parameter to 28 joints
5	mcinitanimpar	CMU & ACCA D database s	0.002590s	It is response time to initialize the parameter to 41 or 42 joints
6	Mcm2j()	Qualisys Mocap System with 28 makers	0.003352s	It is response time transfer 28 sensors to 20 joints position
7	Mcm2j()	CMU & ACCA D database s	0.007916s	It is response time our configured 41 or 42. They transfer into 20 joints position.
8	Mcplotframe()	Qualisys Mocap System with 28 makers	0.168140s	This is response time of mcplotframe(walk2j,100.jar). It is extracted stick skeleton human model From 28 markers display See (Figure 9.b). It contains 20 joints and, Shows human stick model.
9	Mcplotframe()	CMU & ACCA D database s	0.199701s	This is response time of mcplotframe(walk2j,50.jar). It is extracted stick skeleton human model from 41 markers display See (Figure 10.d). It contains 20 joints and, shows human stick model.

ACKNOWLEDGMENT

We are thank [15-16]; which provide the data for public usage. We are also thankful to the National Natural Science Foundation of China under Grant No.61170203, 61170170, The National Key Technology Research and Development Program of China under Grant No.2012BAH33F04, Beijing Key Laboratory Program of China under Grant No.Z111101055281056, supports sharing with us.

REFERENCES

- [1] M. K. Hsieh, B. Y. Chen, & M. Ouhyoung, "Motion retargeting and transition in different articulated figures", Proceedings of the 9th International Conference on Computer Aided Design and Computer Graphics, 2005, Hong Kong, China, pp. 6-pp.
- [2] Yang, G. H., Won, J., & Ryu, S., "Online retargeting for multi-lateral teleportation", 10th International Conference on In Ubiquitous Robots and Ambient Intelligence (URAI), 2013.
- [3] W.Gao, Y. Chen, G. Fang, C.Yang, D. Jiang, C. Ge, & C., Wang, "A Motion Retargeting Method for Topologically Different Characters", Proceedings of the sixth International Conference in Computer Graphics, Imaging and Visualization, 2009, Tianjin, China, pp. 96-100.
- [4] J. Tilmanne, R. Sebbe, & T. Dutoit, "A database for stylistic human gait modeling and synthesis", Proceedings of the Workshop on Multimodal Interfaces, 2008, Paris, France.
- [5] X. Xu, C. Leng, & Z. Wu, "Rapid 3D Human Modeling and Animation Based on Sketch and Motion Database", Proceedings Workshop on Digital Media and Digital Content Management, 2011, Hangzhou, Zhejiang, China, pp. 121-124.
- [6] S. Ali, Z. Mingquan, W. Zhongke, A. Razzaq, M. Hamada, & H. Ahmed, "Comprehensive use of Hip Joint in Gender Identification Using 3-Dimension Data", TELKOMNIKA Indonesian Journal of Electrical Engineering, 2013, 11.6, 2933-2941.
- [7] Meredith M., and Maddock S., "Motion capture file formats explained," in Proceedings of Technical Report of Department of Computer Science, University of Sheffield, 2001
- [8] The explain the information of c3d file format www.c3d.org
- [9] Jeroen J.A and Van Boxtel "A biological motion toolbox for reading displaying, and manipulating motion capture data in research settings," Journal of Vision, 2013, 13(12), pp 1-16.
- [10] Kleiner, Mario, et al. "What's new in Psychtoolbox-3." Perception 36.14 (2007): 1-1.
- [11] C. Verron, Traitement et Visualisation de Vonnées Gestuelles Captées par Optotrak. IDMIL Report, 2005.
- [12] Matlab platform for precompiled Matlab toolbox: www.mathworks.com, accessed 2014
- [13] Burger, Birgitta, and Petri Toiviainen. "MoCap Toolbox-A Matlab toolbox for computational analysis of movement data." Proceedings of the Sound and Music Computing Conference 2013, SMC 2013, Logos Verlag Berlin.
- [14] Mocap systems <http://www.qualisys.com/>, accessed 2014.
- [15] Motion capture lab (Advance Computing Centre for the Arts and Design) 2014. <http://accad.osu.edu/research/mocap/mocapdata.htm>
- [16] Carnegie Mellon University Graphics Lab. CMU Motion Capture Database, 2014. mocap.cs.cmu.edu.
- [17] MocapToolbox136, <http://staffwww.dcs.shef.ac.uk/people/N.Lawrence/mocap/>
- [18] RoboticsToolbox www.petercorke.com/Robotics_Toolbox.html.
- [19] Mitchell, T. Machine Learning; McGraw Hill: Columbus, OH, USA, 1997.
- [20] Marker placement guide on human body <http://users.aber.ac.uk/hoh/CS390/512ViconSWManual.pdf>.
- [21] M. Yeadon and M. Morlock, "The appropriate use of regression equations for the estimation of segmental inertia parameters." Journal of Biomechanics, vol. 22, no. 6-7, pp. 683-689, 1989.
- [22] P. de Leva, "Adjustments to zatsiorsky-seluyanov's segment inertia parameters," Journal of Biomechanics, vol. 29, no. 9, pp. 1223-1230, 1996.
- [23] D. Pearsall and J. Reid, "Inertial properties of the human trunk of males determined from magnetic resonance imaging," Annals of Biomedical Engineering, vol. 22, pp. 692-706, 1994.
- [24] J. Durkin and J. Dowling, "Analysis of body segment parameter differences between four human populations and the estimation errors of four popular mathematical models," Biomechanical Engineering, vol. 125, no. 4, pp. 515-522, 2003.
- [25] C. Cheng, H. Chen, C. Chen, C. Chen, and C. Chen, "Segment inertial properties of Chinese adults determined from magnetic resonance imaging," Clinical Biomechanics, vol. 15, no. 8, pp. 559-566, 2000.
- [26] D. Pearsall, J. Reid, and L. Livingston, "Segment inertial parameters of the human trunk as determined from computed tomography," Annals of Biomedical Engineering, vol. 24, pp. 198-210, 1996.
- [27] V. M. Zatsiorsky and V. M. Seluyanov, "The mass and inertia characteristics of the main segments of the human body," Biomechanics VIII B (Edited by Matsui, H. and Kobayashi, K.), pp. 1152-1159, 1983.
- [28] J. Durkin, J. Dowling, and D. Andrews, "The measurement of body segment inertial parameters using dual energy x-ray absorptiometry," Journal of Biomechanics, vol. 35, no. 12, pp. 1575-1580, 2002.
- [29] P. L. Davidson, S. J. Wilson, B. D. Wilson, and D. J. Chalmers, "Estimating subject-specific body segment parameters using a 3-dimensional modeller program," Journal of Biomechanics, vol. 41, no. 16, pp. 3506-3510, 2008.
- [30] J. Wicke, G. Dumas, and P. Costigan, "A comparison between a new model and current models for estimating trunk segment inertial parameters," Journal of Biomechanics, vol. 41, no. 4, pp. 861-867, 2008.
- [31] J. A. Reinbolt, R. T. Haftka, T. L. Chmielewski, and B. J. Fregly, "Are patient-specific joint and inertial parameters necessary for accurate inverse dynamics analyses of gait?" IEEE Transactions on Biomedical Engineering, vol. 54, no. 5, pp. 782-793, May 2007.
- [32] www.health.uottawa.ca/biomech/csb/Archives/dempster.pdf.
- [33] De leva, "Adjustments to zatsiorsky-seluyanov's segment inertia parameters", Biomechanics, vol.29, Page.1223-1230, 1996.
- [34] The detail of human body (BSP) computation "http://regonstate.edu/instruct/exss323/CM_Lab/Center%20of%20Mass.htm"

APPENDIX:

A. Markers Configuration and BSP Computation

1) Source Code for retrieving the maker index

We edited the `mcreadc3d()` function as follow

```
d= mcreadc3d (fn);
% Here fn is file name of the data

[poss jointj] = size(d.markerName)

% This is for L_finger-----1
for labelInd=1: poss %- This is for L_finger
if(strcmpi(d.markerName{labelInd,1}, 'liu:LFIN')
|| strcmpi(d.markerName{labelInd,1}, 'steve:LFIN')
|| strcmpi(d.markerName{labelInd,1}, 'TakeoMonday
:LFIN') || strcmpi(d.markerName{labelInd,1}, 'Just
in:LFIN') || strcmpi(d.markerName{labelInd,1}, 'Ta
iChi:LFIN') || strcmpi(d.markerName{labelInd,1}, '
Style1:LFIN') || strcmpi(d.markerName{labelInd,1}
, 'Style2:LFIN') || strcmpi(d.markerName{labelInd,
1}, 'Style4:LFIN') || strcmpi(d.markerName{labelInd
```

```
d,1},'Style7:LFIN') || strcmpi(d.markerName{label  
Ind,1},'rory6:LFIN') || strcmpi(d.markerName{labe  
lInd,1},'Female1:LFIN') || strcmpi(d.markerName{l  
abelInd,1},'male2:LFIN')  
dataindex_1=labelInd end
```

end

2) Source Code for handling extra markers index

We add the following code to **mcreadc3d()** function and manage the extras marker index during reading c3d files, those were stored during motion capture session

```
[Pose numjoints]=size(d.data);  
numjoints_maker =numjoints;  
total_joints= numjoints_maker/3  
if total_joints > 41_use_able_joints  
Extrajoints=total_joints - 41_use_able_joint  
set_extract_joints_zerosuse=3*Extrajoints;  
tempmatrix=zeros(1,  
set_extract_joints_zerosuse)  
save('tempmatrix')  
end  
if numjoints/3==41 % This is for 41 markers or  
more  
for pose = 1: size(d.data(:,:),1)  
tem_data = reshape (d.data (pose,:),3,41)';  
d.data(pose,:)= [tem_data(dataindex_1,:),tem_dat  
a(dataindex_2,:),tem_data(dataindex_3,:),tem_da  
ta(dataindex_4,:),tem_data(dataindex_5,:),tem_d  
ata(dataindex_6,:),tem_data(dataindex_7,:),tem_  
data(dataindex_8,:),tem_data(dataindex_9,:),tem_  
_data(dataindex_10,:),tem_data(dataindex_11,:),  
tem_data(dataindex_12,:),tem_data(dataindex_13,  
:),tem_data(dataindex_14,:),tem_data(dataindex_  
15,:),tem_data(dataindex_16,:),tem_data(dataind  
ex_17,:),tem_data(dataindex_18,:),tem_data(data  
index_19,:),tem_data(dataindex_20,:),tem_data(d  
ataindex_21,:),tem_data(dataindex_22,:),tem_dat  
a(dataindex_23,:),tem_data(dataindex_24,:),tem_  
data(dataindex_25,:),tem_data(dataindex_26,:),t  
em_data(dataindex_27,:),tem_data(dataindex_28,:  
,tem_data(dataindex_29,:),tem_data(dataindex_3  
0,:),tem_data(dataindex_31,:),tem_data(datainde  
x_32,:),tem_data(dataindex_33,:),tem_data(datai  
ndex_34,:),tem_data(dataindex_35,:),tem_data(da  
taindex_36,:),tem_data(dataindex_37,:),tem_data  
(dataindex_38,:),tem_data(dataindex_39,:),tem_d  
ata(dataindex_40,:),tem_data(dataindex_41,:)];  
end  
elseif jsize>41 % set zero values to extra makers  
for pose = 1: size(d.data(:,:),1)
```

```
tem_data = reshape (d.data (pose,:),3,jsize)';  
d.data(pose,:)= [tem_data(dataindex_1,:),tem_dat  
a(dataindex_2,:),tem_data(dataindex_3,:),tem_da  
ta(dataindex_4,:),tem_data(dataindex_5,:),tem_d  
ata(dataindex_6,:),tem_data(dataindex_7,:),tem_  
data(dataindex_8,:),tem_data(dataindex_9,:),tem_  
_data(dataindex_10,:),tem_data(dataindex_11,:),  
tem_data(dataindex_12,:),tem_data(dataindex_13,  
:),tem_data(dataindex_14,:),tem_data(dataindex_  
15,:),tem_data(dataindex_16,:),tem_data(datain  
dex_17,:),tem_data(dataindex_18,:),tem_data(d  
ataindex_19,:),tem_data(dataindex_20,:),tem_  
data(dataindex_21,:),tem_data(dataindex_22,:),  
tem_data(dataindex_23,:),tem_data(dataindex_24,  
:),tem_data(dataindex_25,:),tem_data(dataindex_  
26,:),tem_data(dataindex_27,:),tem_data(datain  
dex_28,:),tem_data(dataindex_29,:),tem_data(d  
ataindex_30,:),tem_data(dataindex_31,:),tem_d  
ata(dataindex_32,:),tem_data(dataindex_33,:),te  
m_data(dataindex_34,:),tem_data(dataindex_35,:)  
,tem_data(dataindex_36,:),tem_data(dataindex_37  
,:),tem_data(dataindex_38,:),tem_data(dataindex  
_39,:),tem_data(dataindex_40,:),tem_data(datain  
dex_41,:),tempmatrix(1,:)] ;
```

end

3) Source code for normalized skeleton refinement between 41 markers (see Figure 4 & A.1 (a)) of mocap data (C3D).

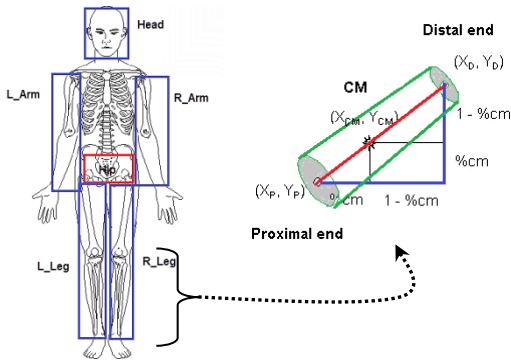
```
m2jpar = mcinitm2jpar;  
m2jpar.markerName{1} = 'root';  
m2jpar.markerNum{1} = [22:23 2 20];  
m2jpar.markerName{2} = 'lhip';  
m2jpar.markerNum{2} = [23 2];  
m2jpar.markerName{3} = 'lknee';  
m2jpar.markerNum{3} = [31 15 28];  
m2jpar.markerName{4} = 'lankle';  
m2jpar.markerNum{4} = [34];  
m2jpar.markerName{5} = 'ltoe';  
m2jpar.markerNum{5} = [32:34 36];  
m2jpar.markerName{6} = 'rhip';  
m2jpar.markerNum{6} = [20 22];  
m2jpar.markerName{7} = 'rknee';  
m2jpar.markerNum{7} = [35 37:38];  
m2jpar.markerName{8} = 'rankle';  
m2jpar.markerNum{8} = [30];  
m2jpar.markerName{9} = 'rtoe';  
m2jpar.markerNum{9} = [39:41 30];  
m2jpar.markerName{10} = 'midtorso';  
m2jpar.markerNum{10} = [18 29 25 17];  
m2jpar.markerName{11} = 'neck';  
m2jpar.markerNum{11} = [14 13 5 29];  
m2jpar.markerName{12} = 'head';  
m2jpar.markerNum{12} = [8:9 12 16];  
m2jpar.markerName{13} = 'lshoulder';  
m2jpar.markerNum{13} = [5];  
m2jpar.markerName{14} = 'lelbow';
```

```

m2.jpar.markerNum{14} = [10 6];
m2.jpar.markerName{15} = 'lwrist';
m2.jpar.markerNum{15} = [26 1 27 3];
m2.jpar.markerName{16} = 'lfinger';
m2.jpar.markerNum{16} = [1];
m2.jpar.markerName{17} = 'rshoulder';
m2.jpar.markerNum{17} = [14];
    
```

Marker Number	Position on body	Main Body part
1	L_scent head	Head
2	R_scent head	
3	L_back head	
4	R_back head	
5	C7	Left_Arm
6	T10	
7	CLAV	
8	STRN	
9	RBAC	
10	L_SHO	
11	L_UPA	
12	L_ELB	
13	L_FRM	
14	L_WRA	
15	L_WRB	
16	L_FEN	
17	R_SHO	Right_ARM
18	R_UPA	
19	R_ELB	
20	R_FRM	
21	R_WRA	
22	R_WRB	
23	R_FEN	
24	L_FWT	Hip
25	R_FWT	
26	L_BWT	
27	R_BWT	
28	L_THI	Left_Leg
29	L_KNE	
30	L_SHN	
31	L_ANK	
32	L_HEE	
33	L_TOE	
34	L_MTS	
35	R_THI	Right_Leg
36	R_KNE	
37	R_SHN	
38	R_ANK	
39	R_HEE	
40	R_TOE	
41	R_MTS	

(a).Marker labels



(b). Human body segment (c). Segmental CM

Fig. 15. A. Human Anatomical landmarks Template

4) Computation of BSP

The information of human body segments is used to determine the location of each segment's center of mass and each segment's mass. Each of these body segment parameters (BSP) usually expressed as a percentage value. Computing the center of each segment and each segment mass. It can be computed by the following equations:

$$x_{CM} = x_D (\%cm) + x_p (\%cm) \quad A.1$$

$$Y_{CM} = Y_D (\%cm) + Y_p (\%cm) \quad A.2$$

Where (XCM, YCM) = X & Y coordinates of the segmental CM, (XD, YD) = coordinates of the **Distal end** of the segment, (XP, YP) = coordinates of the **proximal end**, and %cm = CM.

In short, the general formulas for computing the position of each segment center of mass and each segment's mass:

$$x_{cm} = \frac{\sum_{s=1}^n (B_{seg_s} \times x_{seg_{cm_s}})}{B_{body}} \quad A.3$$

$$Y_{cm} = \frac{\sum_{s=1}^n (B_{seg_s} \times Y_{seg_{cm_s}})}{B_{body}} \quad A.4$$

Where B is the mass of the body, X, Y is a position coordinate and n is the number segments and each segment parameter (BSP) is routinely indicated as a percentage values.

The body center mass can be estimated from the CMs and the masses of the segments:

$$x = \frac{\sum_i (B_i * x_i)}{\sum_i B_i} \quad A.5$$

$$Y = \frac{\sum_i (B_i * Y_i)}{\sum_i B_i} \quad A.6$$

From the equations A.5 & A.6 describe as (X, Y) = coordinates of the body CM, i = segment number, (Xi, Yi) = the X & Y coordinates of the CM of segment i, and Bi = mass of segment i. We can say in other words, the body CM coordinates are equal to the sum of the segmental mass products and segmental CM coordinates divided by the body mass ($\sum B_i$).

A Generic Adaptive Multi-Gene-Set Genetic Algorithm (AMGA)

Adi A. Maaita, Jamal Zraqou, Fadi Hamad and Hamza A. Al-Sewadi

Faculty of Information Technology
Isra University
Amman, Jordan

Abstract—Genetic algorithms have been used extensively in solving complex solution-space search problems. However, certain problems can include multiple sub-problems in which multiple searches through distinct solution-spaces are required before the final solution combining all the sub-solutions is found. This paper presents a generic design of genetic algorithms which can be used for solving complex solution-space search problems that involve multiple sub-solutions. Such problems are very difficult to solve using basic genetic algorithm designs that utilize a single gene-set per chromosome. The suggested algorithm presents a generic solution which utilizes both multi-gene-set chromosomes, and an adaptive gene mutation rate scheme. The results presented from experiments done using an automatic graphical user interface generation case study, show that the suggested algorithm is capable of producing successful solutions where the common single-gene-set design fails.

Keywords—Genetic algorithm; Multi-gene-set; Single-gene-set; Artificial Intelligence; Generic algorithm; Generic architecture

I. INTRODUCTION

Nature inspired algorithms have been presenting astonishing results in solving problems that are not structured in nature. Genetic algorithms (GAs) are a category of nature inspired algorithms that have been used extensively in solving problems requiring an advanced form of heuristic search throughout a solution space, besides numerical and combinatorial optimization problems [1].

GAs are inspired from the natural processes of sexual reproduction and natural selection [1, 2]. The complete characteristics of a living being is miraculously encoded within its chromosomes, which have those life codes stored as Deoxyribonucleic Acid (DNA) molecules [3]. These huge molecules contain the life codes as characteristic encoding genes comprised of combinations of the primary nucleobases which are cytosine, guanine, adenine, and thymine [4].

During the process of sexual reproduction, each of the parents contributes genetic characteristics through providing half of their child's chromosomes. These chromosomes then undergo a process called crossover which causes parent chromosomes to break and then recombine into chromosomes with a gene set contributed by both parents. Hence, this child combines characteristics from both parents [3, 4]. The chromosomes and the genes they hold are referred to as the Genotype [5]. As for the characteristics that result from those genes, they are referred to as the Phenotype [6]. Moreover, sometimes genetic mutations occur. Such genetic mutations are

changes in the original sequence of genes and can lead to evolution, health problems, or may have no effect [7].

After the introduction to GAs in section 1, a brief account of previous work is listed in section 2. The general structure of the GA is described in section 3. The proposed Adaptive Multi-gene-set Genetic Algorithm (AMGA) is outlined in section 4 then section 5 details the architecture of this algorithm. Implementation and results are included in section 6 and finally section 7 concludes the paper.

II. PREVIOUS WORK

GAs found a wide area of applications for which to generate useful solutions. These applications did not spare any direction such as physics, mathematics, chemistry, medical, economics, computer, bioinformatics, pharmacology, etc. In the remainder of this section, a brief literature review of applications utilizing GAs for obtaining solutions is presented.

Shimamoto et al. [8] utilized a GA for flexible real-time dynamic routing control for traffic changes in broadband networks. It generates the exact solution for finding a routing arrangement that keeps the traffic loss-rate below a target value.

Lienig [9] proposed a novel parallel GA approach for performance-driven VLSI routing running on a distributed network and optimizes physical constraints such as nets size, crosstalk and delay.

Chun et al. [10], examined heuristic algorithms as the search tools for diverse optimization problems. The examined algorithms included the Immune algorithm (IA), the GA and the evolution strategy (ES).

Chang and Ramakrishna [11] examined GAs for the shortest path routing solution of the traveling salesman problem TSP, proving that the GA is one of the best heuristic algorithms to solve this problem. Variable-length chromosomes and their genes were utilized for encoding the problem, and partial chromosomes crossover with curing of all the infeasible chromosomes by a simple repair function were used for creating the diversity within the population.

Juang [12] proposed a recurrent fuzzy network for dynamic systems processing by using a neural network and a GA for optimizing the neural network. The fuzzy network was called TSK. It implements a series of recurrent fuzzy if-then rules with TSK type consequent parts with supervised learning. It proved superiority when applied to dynamic system.

Ozpineci et al. [13] proposed Harmonic optimization of multilevel converters using a GA. Optimum switching angles for cascaded multilevel inverters was achieved for eliminating some higher order harmonics while maintaining the required fundamental voltage.

Chowdhury et al. [14] designed an Encryption and Decryption algorithm for communication networks using a GA to robustly speed up and secure the total cryptography process. One point crossover and block cipher techniques were implemented for the simplification of the GA cryptosystem technique.

Mahdad et al. [15] presented a combined GA and fuzzy logic rules to enhance the optimal power flow with consideration of multi shunt flexible AC transmission systems. The presented method was effective in giving a near optimal solution and remarkably reduced the computation time.

Malhotra et al. [16] proposed a GA as an optimization tool for heuristic search applied to optimize process controllers for using natural operators. Their work explores the well-established methodologies of the literature to realize the workability and applicability of GAs for process control applications. GAs are applied to direct torque control of induction motor drive, speed control of gas turbines, and speed control of DC servo motors.

Kabeer et al. [17] used Boosted Feature Subset Selection (BFSS) as a preprocessing step to provide a gene subset that is fed to a GA, thus reducing the feature subset to smaller numbers and helping to generate a better optimal subset of genes. They claim that their hybrid approach shows better results compared to other well-known approaches when applied to leukemia, colon and lung cancer benchmarked datasets.

Ahmed [18] developed a simple GA using sequential constructive crossover to obtain heuristic solutions to the Bottleneck traveling salesman problem. He proposed a hybrid GA that incorporates 2-optimization search, another proposed local search and immigration to the simple GA for obtaining better solutions.

Umbarkar et al. [19] proposed Dual Population GA for solving Constrained Optimization Problems. It is based on maximum constraints satisfaction applied as a constraints handling technique and a Dual Population GA used as a meta-heuristic. It achieved close to optimum rather than exact optimum solution as compared with the Ant Colony algorithm, the Bee Colony algorithm, the Differential Evolution algorithm and the GA that have been used for solving the same problem set.

Moin et al. [20] proposed a hybrid GA with multi parents crossover for job shop scheduling problem (JSSP). The search space is reduced by generating a full-active schedule that satisfies precedence constraints, a neighborhood search is applied to exploit the search space for better solutions and to enhance the GA. Simulation suggests sustainability of this hybrid GA in solving JSSP.

Sankaran et al. [21] proposed a GA based parallel optimization technique aiming to improve the performance of

batch schedule of two massively parallel application codes; a turbulent combustion flow solver (S3D) and a molecular dynamics code (LAMMPS). Experiments have shown a significant deviation from ideal weak scaling and variability in performance. This technique showed significant improvement in solving speed, besides improvement in variability and scalability.

The work in this paper suggests the use of a multi-gene-set chromosome GA with an adaptive gene mutation rate scheme, aiming to deal with multiple sub-problems in order to find a viable solution for complex solution-space problem. Such problems are usually very difficult to solve using the traditional single-gene-set chromosomes GA designs.

III. THE GENERAL STRUCTURE OF GENETIC ALGORITHMS

GAs were inspired from nature to solve problems to which we have no structured solutions that can be coded into algorithms. Examples of such problems are building and refining a set of production rules [22], then creating and adapting computer programs [23].

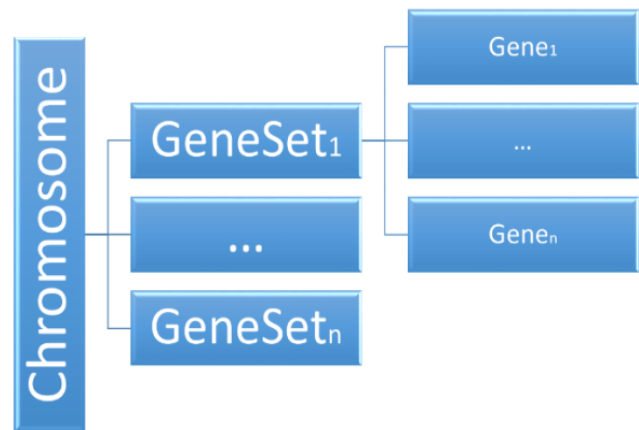


Fig. 1. A chromosome with multi-gene-sets

These kinds of problems require gradually evolved solutions rather than simply calculated ones. This is where GAs show their true power. They are evolutionary algorithms that use a number of initial solutions and attempt to evolve them in a way that eventually leads to the sought solution.

In order to perform an evolutionary search into the solution space, a GA requires a number of data structures, some specific rules and certain procedures. The main data structure of a GA is the chromosome. The chromosome represents the genotype of a solution. This genotype is usually encoded as a string of 0's and 1's, however, it can be encoded using other types of data as well [1].

Another important part of a GA's structure is the algorithm for calculating a chromosome's fitness value. This algorithm is called the fitness function and it is responsible for providing a measure of a chromosome's quality as a solution to the problem at hand. It is important to note that a fitness function is problem specific. That means we need to create a new fitness calculating algorithm for each different problem. A fitness value is usually a measure of how close a certain solution is to the required solution [1, 2].

IV. THE PROPOSED ADAPTIVE MULTI-GENE-SET GENETIC ALGORITHM

The problem under consideration in this research considers a complex system with multiple objectives. Hence, an algorithm is proposed that is capable of looking into a complex solution space. It suggests an Adaptive Multi-gene-set Genetic Algorithm that has multiple heterogeneous solution fitness aspects. In other words, the problem requires searching for a complex solution that requires a genotype that contains multiple types of genes. The main challenge with such a genotype would lie in the complex crossover process. We simply cannot exchange gene data between genes containing different data semantics even if the data types match.

To visualize the multiple gene sets for a certain chromosome, a class diagram is shown in fig 4. Furthermore, neighboring multi-gene sets chromosomes would have some influence on each other through genetic operations such as crossover and mutations.

The crossover process is meant to create a new set of solutions that mix aspects of existing solutions. For that reason, we need a proper encoding of the genotype that preserves the purpose of crossover. The proposed algorithm suggests the use of multiple gene sets within a single chromosome. This would enable a complex problem to be divided into multiple smaller sub-problems which will be handled separately, and then those sub-solutions are gathered in order to create a complete final solution. Fig 2 illustrates a sketch for the expected crossover process amongst multi-gene set chromosomes.

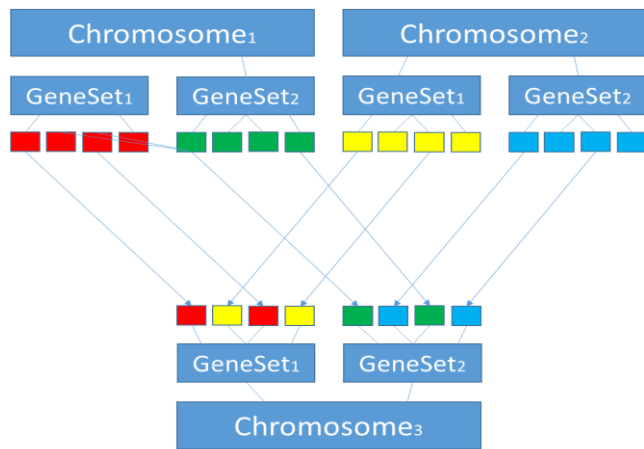


Fig. 2. Crossover amongst multi-gene-set chromosomes

Each sub-problem would require its own crossover process which is capable of generating offspring sub-solutions with data that are semantically correct. The fitness of each sub-solution needs to be calculated in order to determine whether that part of the complete solution needs further search for the current generation of the algorithm execution or not. This means that a sub-solution that may seem optimal for the sub-problem might not be suitable for the complete solution. Moreover, this would require searching for a new sub-solution during the coming generations until a suitable complete solution is found.

Such an architecture of the GA also allows for multi-threading to be efficiently used. This is made possible by separating the sub-solution search operations into distinct threads that run simultaneously and independently in parallel. This enables very efficient implementations of the algorithm to be created when compared to the traditional GA architecture.

The fitness of the complete solution needs to be calculated in each generation to determine how each of the sub-solutions search algorithms should operate. Fig 3 clarifies the contribution of Gene-set fitness to the calculation of Chromosomes fitness.

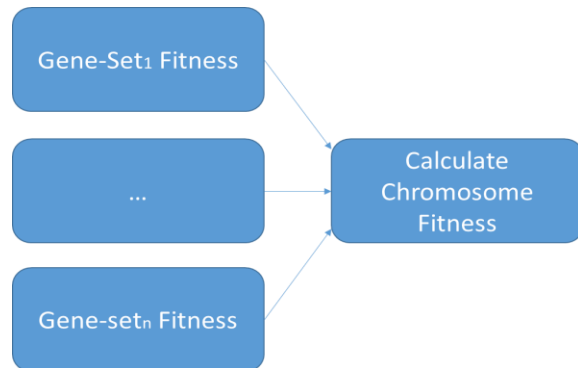


Fig. 3. Gene-set fitness values contribute towards the total chromosome fitness

A. Adaptive mutation rates

When the search results seem to converge towards unsuccessful solutions then it is time to add new genetic material to the mating pool. This is where mutation comes into play. Mutation is a process that changes the data of the genotype by creating new values that are usually randomly generated within predefined constraints. Mutation rate is usually predefined prior to the execution of the GA. This limits the behavior of the GA to a semi-static form when it comes to steering the search process towards a solution. Hence, the GA would not be able to increase the rate of introduction of new genetic material when the search converges towards unsuccessful solutions in order to escape that convergence. The GA will also be unable to reduce that rate of new genetic material introduction when the search seems to converge towards successful solutions. This reduction is very helpful when there is a need to concentrate the search within the existing genetic material that represents successfully evolving solutions.

In order to succeed, a GA needs to be fine-tuned by setting a proper mutation rate through a process of trial and error. If an unsuitable mutation rate is used, the search may never be able to converge successfully towards a solution. A higher than needed mutation rate would cause the algorithm to search the solution space blindly. A lower than needed mutation rate may lead the algorithm to converge towards unsuccessful solutions. This would be the result of existing genes representing the current mating pool of solutions being overly dominant and if those solutions have low fitness values then the new generated solutions are most likely to have a low fitness values as well.

On the other hand, an adaptive mutation rate would steer the search conducted by the GA towards successful solutions more efficiently than using a static rate. In this adaptive scheme, the mutation rate is adapted for each sub-solution according to statistics showing convergence towards successful solutions, or convergence towards unsuccessful solutions. Each mutation rate is re-evaluated for each generation. According to that evaluation, the mutation rate is increased or decreased according to equation (1), where mr_{new} , mr_{old} , and r_c are the new mutation rate, old mutation rate and convergence rate, respectively.

$$mr_{new} = mr_{old} + (mr_{old} \times r_c) \quad (1)$$

It is important to note that the rate of convergence is represented as negative values when the convergence is towards less successful solutions and it is positive when the convergence is towards more successful solutions.

V. THE ARCHITECTURE OF THE ADAPTIVE MULTI-GENE-SET GENETIC ALGORITHM

The class diagram representing the architecture of a traditional single-gene-set chromosome GA is shown in fig 4. The GA consists of a population of chromosomes that accommodate genes and data structures representing gene-data.

The proposed adaptive multi-gene-set chromosome GA is completely different from the traditional generic model as clarified in the class diagram shown in fig 5. The algorithm is designed to be generic in nature as it requires minimal modifications in order to be used for solving any evolutionary search problem. For example, an interested developer using this architecture would only need to create new classes implementing the interfaces IGene and IGeneData in order to solve any type of problem requiring chromosomes with multiple gene sets.

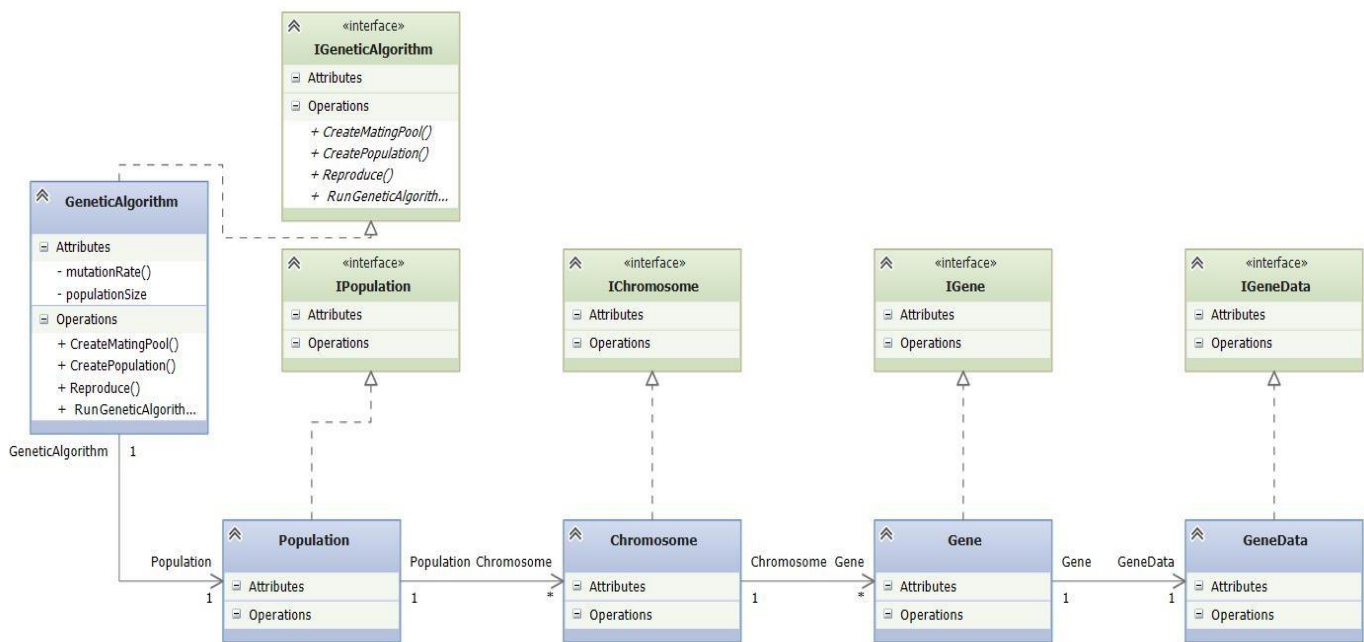


Fig. 4. Class diagram for the single-gene-set GA architecture

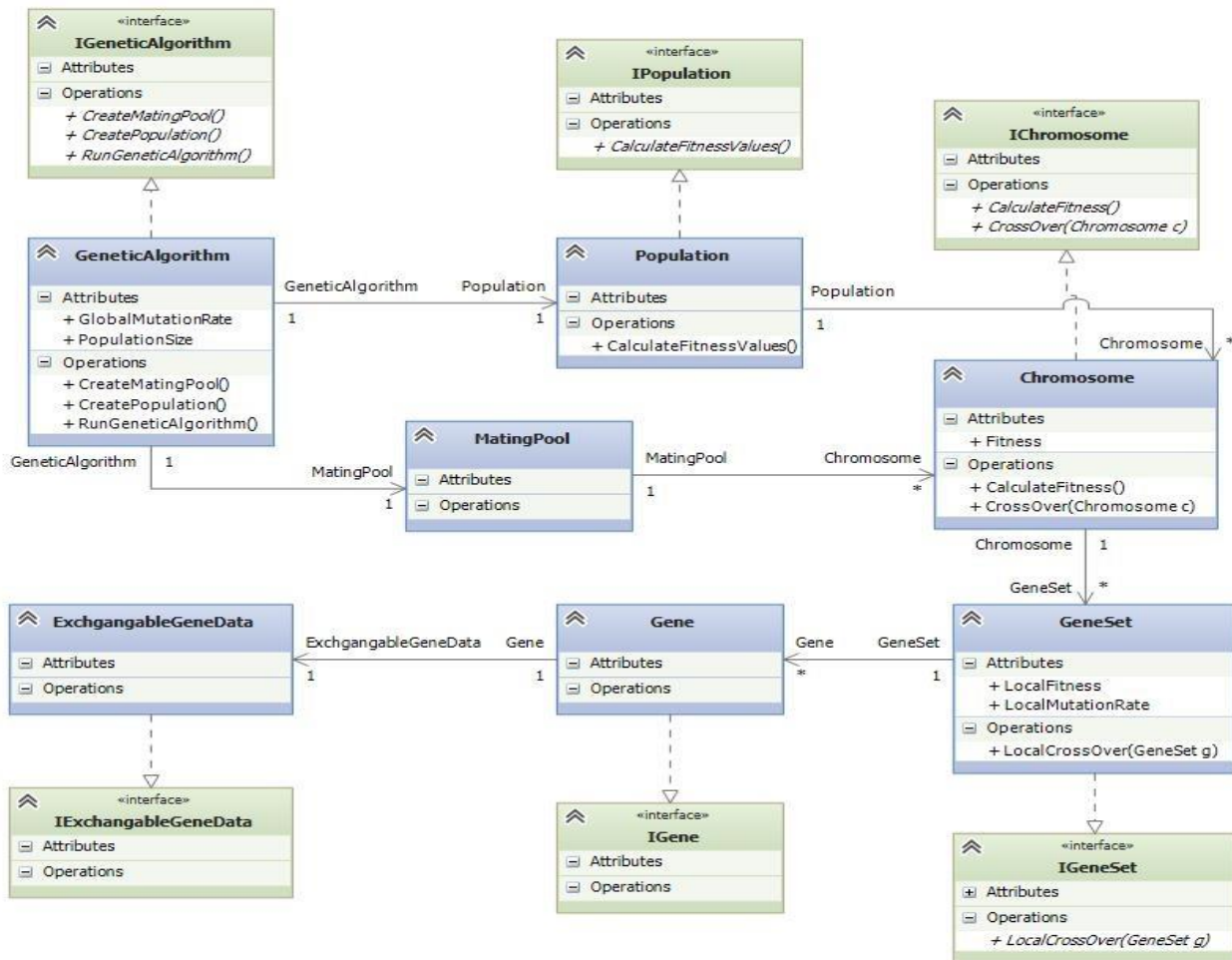


Fig. 5. Class diagram of the multi-gene-set GA architecture

VI. IMPLEMENTATION AND RESULTS

To demonstrate the effectiveness of the proposed AMGA, an experiment is conducted to implement it in the process of graphical user interface generation. The results are then compared to those obtained when conducting the same experiment using a static mutation rate single-gene-set GA.

The problem of automated graphical user interface generation using GAs was discussed in [24-26].

The genetic material of the graphical user interface generation problem consists of two types of gene-sets per chromosome. The first gene-set contains information on the containers used to host controls on a form. The second gene-set contains information on the controls that are used to create the graphical user interface.

The single-gene-set GA relies on a single gene-set to represent both the controls and the containers. In order to do that, the containers were treated as controls, and each regular control contained an attribute that stores a serial number representing its parent control. Each control was represented as a gene containing the following data:

- 1) X-axis location in pixels.
- 2) Y-axis location in pixels.

- 3) Width in pixels
- 4) Height in pixels.
- 5) Margin in pixels.
- 6) Padding in pixels.
- 7) Parent container serial number.
- 8) Dock location on the form (top, bottom, left, right, full, none)

Using that gene data, the algorithm needs to search for candidate formations of the supplied controls in order to accommodate the following criteria:

- 1) Controls of the same type should be located within the same container.
- 2) Controls should be horizontally stacked if the parent container is vertically oriented and should be vertically stacked if the parent container is horizontally oriented.
- 3) Controls should be left aligned if they exist in a vertically oriented container and should be center aligned if they exist in a horizontally aligned container.
- 4) Controls of the same type within the same container should have the same height value in pixels.

The multi-gene-set GA relies on two separate gene sets for representing the containers and the controls respectively. This requires creating two gene types. The first type represents the containers and it contains the following data:

- 1) Container serial number.
- 2) Container type.

While the exchangeable data include:

1) Container dock location on the form (top, bottom, left, right, full, none).

- 2) Container width.
- 3) Container height.

As for the second type, it represents the controls and it contains the following data:

- 1) Control name.
- 2) Control type.

While the exchangeable data include:

- 1) X-axis location in pixels.
- 2) Y-axis location in pixels.
- 3) Width in pixels.
- 4) Height in pixels.
- 5) Margin in pixels.
- 6) Padding in pixels.
- 7) Parent container serial number.

The obtained best fitness values for the case study when implementing the single-gene-set algorithm running for a maximum number of 1000 generations and having 5 containers with different numbers of controls (e.g. 10, 20 & 30), are summarized in Tables I. While the obtained best fitness values for implementing the AMGA running for the same parameters, are summarized in table II.

TABLE I. RUNNING THE SINGLE-GENE-SET GA

Single-gene-set Algorithm			
Max No. of generations = 1000			
No. of Containers	No. of Controls	Generation of termination	Best fitness value
5	10	1000	71.70%
5	20	1000	72.60%
5	30	1000	61.90%

TABLE II. RUNNING THE MULTI-GENE-SET GA

Multi-gene-set Algorithm			
Max No. of generations = 1000			
No. of Containers	No. of Controls	Generation of termination	Best fitness value
5	10	81	100%
5	20	556	100%
5	30	1000	93.70%

Comparison of the two tables indicates that the AMGA was able to achieve successful solutions where the single-gene-set GA failed for the same number of generations.

The fitness values were calculated for both, AMGA algorithm and single-gene-set GA as a function of the number of generations for different numbers of controls, (namely 10, 20 and 30). The results are plotted in fig 6 - 8.

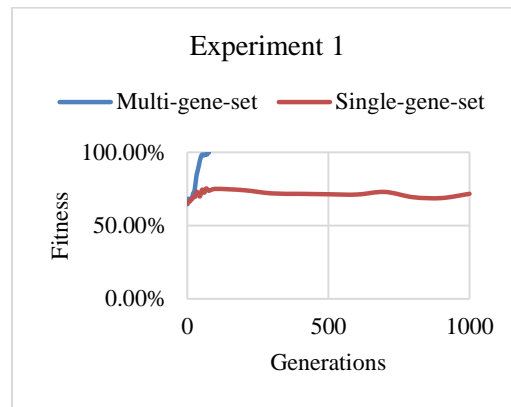


Fig. 6. Performance of the multi-gene-set algorithm when compared to its single-gene-set counterpart for the automated GUI generation problem with 5 containers and 10 controls

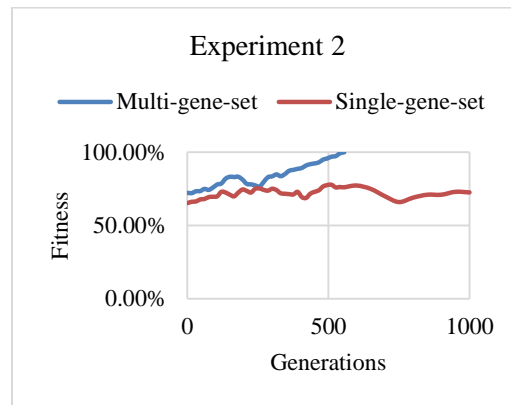


Fig. 7. Performance of the multi-gene-set algorithm when compared to its single-gene-set counterpart for the automated GUI generation problem with 5 containers and 20 controls

It is noticed that in the case of the multi-gene-set GA, convergence towards the best fitness (100% fitness) is achieved after 81 generations for 10 controls. But as the number of controls increases, the convergence gets slower as the available space on the canvas or form is too little to host all the required controls according to the required conditions, or it would even be impossible to place all the required controls on the canvas as the totality of the controls' areas

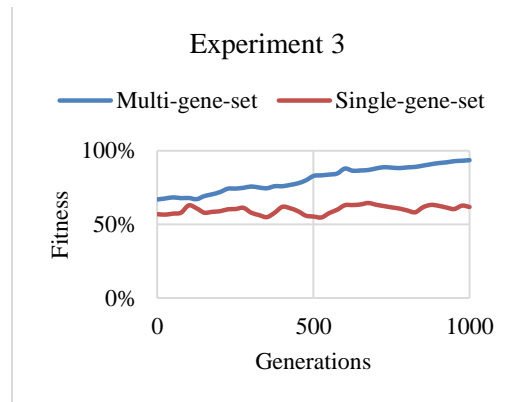


Fig. 8. Performance of the multi-gene-set algorithm when compared to its single-gene-set counterpart for the automated GUI generation problem with 5 containers and 30 controls

would be larger than the actual available space on the canvas. On the other hand, it can be noticed that no convergence was achieved at all in the case of the single-gene-set GA.

VII. CONCLUSIONS

The results presented in this paper show that the AMGA is capable of converging towards solutions more efficiently than its single-gene-set counterpart. The results presented also show that the algorithm is also capable of recovering from divergence from solutions adaptively by increasing the rate of mutation when that is required. It is also capable of reducing the rate of mutation when the algorithm begins to converge towards successful solutions using the existing genetic material.

Future work relating to that presented in this paper would involve using meta-heuristic optimization techniques as part of the generic architecture of the AMGA. This would allow for more complex types of problems to be solved by avoiding the problem of the algorithm getting stuck on a local maxima. Such optimizations would also allow for better performance for the AMGA if implemented correctly.

REFERENCES

- [1] Mitchell M., "An Introduction to Genetic Algorithms", 1998: MIT Press. 209.
- [2] Luger G.F., "Artificial Intelligence: Structures and Strategies for Complex Problem Solving", 2008: Addison-Wesley Publishing Company. 784.
- [3] Brown T., "Introduction to Genetics: A Molecular Approach", 1st Ed. 2011: Garland Science.
- [4] Brooker, R.J., "Genetics: Analysis and Principles", 3rd Ed. 2008: McGraw-Hill Higher Education.
- [5] Sahni N., Song Yi, Q. Zhong, N. Jaikhani, B. Charleaux, M. E. Cusick and M. Vidal, "Edgotype: a fundamental link between genotype and phenotype", Elsevier, Current Opinion in Genetics & Development 2013, Vol. 23, pp. 649–657
- [6] Cooper D. N., Michael Krawczak, Constantin Polychronakos, Chris Tyler-Smith, and Hildegard Kehrer-Sawatzki, "Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease", Hum Genet 2013, Vol. 132, PP1077–1130
- [7] "The New Genetics, National Institute of General Medical Sciences National Institutes of Health U.S. Department of Health and Human Services", NIH Publication No.10-662 Revised April 2010, <http://www.nigms.nih.gov>.
- [8] Shimamoto, N., Hiramatsu, A. and Yamasaki, K. (2000). A dynamic routing control based on a genetic algorithm, IEEE International Conference on Neural Networks, 1993, Vol.2, pp. 1123-1128.
- [9] Lienig, J. (1997). A parallel genetic algorithm for performance-driven VLSI routing, IEEE Transactions on Evolutionary Computation, Vol. 1, No. 1, pp. 29-39.
- [10] Chun J. S., Hyun-Kyo Jung and Song-Yop Hahn, "A study on comparison of optimization performances between immune algorithm and other heuristic algorithms", IEEE Transactions on Magnetics, Vol. 34, No 5, 1998, pp. 2972-2975.
- [11] Chang W. A. and Ramakrishna, R.S., "A genetic algorithm for shortest path routing problem and the sizing of populations", IEEE Transactions on Evolutionary Computation, Vol. 6, No. 6, 2002, pp. 566-579.
- [12] Juang C. F., "A TSK-type recurrent fuzzy network for dynamic systems processing by neural network and genetic algorithms", IEEE Transactions on Fuzzy Systems, Vol. 10, No. 2, 2002, pp. 155-170.
- [13] Ozpineci B., L. M. Tolbert, and J. N. Chiasson, "Harmonic optimization of multilevel converters using genetic algorithms", IEEE 35th Annual Power Electronics Specialists Conference, vol.5, 2004, pp. 3911-3916.
- [14] Chowdhury S., S. K. Das, and A. Das, "Application of Genetic Algorithm in Communication Network Security", International Journal of Innovative Research in Computer and Communication Engineering (An ISO 3297: 2007 Certified Organization) Vol. 3, No. 1, January 2015.
- [15] Mahdad B., T. Bouktir and K. Srairi, "Optimal power Flow of the Algerian Network using Genetic Algorithms/Fuzzy Rules", Power and Energy Society General Meeting - Conversion and Delivery of Electrical Energy in the 21st Century, 2008, pp. 1-8.
- [16] Malhotra R., N. Singh & Y. Singh, " Genetic Algorithms: Concepts, Design for Optimization of Process Controllers", Computer and Information Science Vol. 4, No. 2; March 2011, www.ccsenet.org/cis
- [17] Kabeer S. J., M. T.Moin, M. A. Rahman, A. Mottalib, M. H. Kabir, "BFSSGA: Enhancing the Performance of Genetic Algorithm using Boosted Filtering Approach", IJCA Journal, Volume 51 - Number 19. 2012.
- [18] Ahmad, Z. H., "A Hybrid Genetic Algorithm for the Bottleneck Traveling Salesman Problem", ACM Transactions on Embedded Computing Systems, Vol. 12, No. 1, Article 9, Publication date: January 2013.
- [19] A. J. Umbarkar A. J., M. S. Joshi, and P. D. Sheth, "Dual Population Genetic Algorithm for Solving Constrained Optimization Problems", International Journal of Intelligent Systems and Applications (IJISA, IJISA Vol. 7, No. 2, January 2015, PP.34-40, DOI: 10.5815/ijisa.2015.02.05
- [20] Moin N. H., O. C. Sin, and M. Omar, "Hybrid Genetic Algorithm with Multiparents Crossover for Job Shop Scheduling Problems, Mathematical Problems in Engineering Volume 2015 (2015), Article ID 210680, 12 pages.<http://dx.doi.org/10.1155/2015/210680>
- [21] Sankaran R., J. Angel and W. M. Brown, "Genetic algorithm based task reordering to improve the performance of batch scheduled massively parallel scientific applications, published on line on 8 APR 2015, DOI: 10.1002/cpe.345, to appear in Concurrency and Computation, Practice and Experience.
- [22] Holland, J. H., "Escaping brittleness: the possibilities of general-purpose learning algorithms applied to parallel rule-based systems, in Computation & intelligence", F.L. George, Editor. 1995, American Association for Artificial Intelligence. pp. 275-304.
- [23] Koza, J. R., "Genetic programming: on the programming of computers by means of natural selection", 1992: MIT Press. 680.
- [24] G., O.A.M.N.V., "Interactive Design Of Web Sites With A Genetic Algorithm", 2002: Lisbon, Portugal. pp. 355 - 362.
- [25] Quiroz, J.C.R.L., S.J. ; Shankar, A. ; Dascalu, S.M., "Interactive Genetic Algorithms for User Interface Design", 2007, IEEE Singapore. pp. 1366 - 1373.
- [26] Plessis M. Cd. and L. Barnard, "Incorporating layout managers into an evolutionary programming algorithm to design graphical user interfaces", Proceedings of the 2008 annual research conference of the South African Institute of Computer Scientists and Information Technologists on IT research in developing countries: riding the wave of technology. 2008, ACM: Wilderness, South Africa. pp. 41-47

Real-Time Digital Image Exposure Status Detection and Circuit Implementation

Li Hongqin

School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, CHINA

Zhang Liping

School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, CHINA

Wu Jianzhen

School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, CHINA

Ning Jun

Verisilicon Microelectronics (Shanghai) co., Ltd.
Shanghai, CHINA

Abstract—Auto exposure is an important part of digital image signal processing. We studied the detection of the exposure status in this paper, and fast and parallel detection method was presented. The method comprises the following steps: first obtaining the current image, counting the numbers of pixels in bright and dark regions of the image and obtaining these pixels brightness; then determining exposure parameters based on the proportions of the counted numbers of pixels in bright and dark regions with preset value respectively; if the actual proportion is lower than preset value, then continue to adjust exposure parameters until pixel brightness value reaches preset brightness threshold. Experiments show that the computational complexity and operation demand is low, which can quickly determine the exposure status of the image, improving the real-time capability in image exposure control. The proposed method will make the whole digital image signal processing system works smoothly and be reliable. The circuit implementation of this method is simple with high real-time controllability. This method has been applied for China patent successfully.

Keywords—real-time; auto exposure; fast and parallel method

I. INTRODUCTION

With the development of digital photography technology, digital video cameras have been used in many different environments. Among them, one function of a digital video camera is automatic exposure. Automatic exposure is the default setting of a digital video camera. The camera will automatically control the image's exposure in the automatic exposure mode and users do not need to do anything. Camera or video camera in the sensor can set exposure parameters such as exposure time, gain, aperture value automatically according to the intensity of light reflected from the object. However, if images from the imaging equipment are under-exposed or over-exposed, it will have impact on the subsequent image processing. How to determine the status of current exposure and real-time control imaging equipment by feedback to get an image of higher quality is an important part of image processing tasks [1]. As a necessary part, automatic exposure control determines the display quality of the whole

image system [2], which is a technology we are very concerned about and is a hot research field.

In general, conventional image exposure status detection methods are calculating luminance histogram of the entire image through software [3, 4], determining exposure status based on the distribution of pixels luminance value in the histogram, and then controlling the exposure. However, detection of image exposure state by this method is of high computational complexity with higher hardware performance requirements for running the image processing software, a longer time lag in the exposure state detection and poor real-time performance. It also requires additional storage space and poor real-time performance is harm to the subsequent image processing. As to the hardware implementation method, selecting an area in the image to process not only has high computational complexity, but also involves how to select a region in the image. It will cause the problem of taking a part for the whole and drawback of inaccurate detection as well as long calculation reciprocating cycle [5,6,7].

In view of the above shortcomings, an automatic exposure adjustment system is established in this paper. A fast and parallel image exposure detection algorithm is proposed to adjust brightness of the whole image in real-time with the advantages of simple calculation, accurate detection and high real-time response. The problems of complicated exposure status detection, detection lag and poor real-time response in existing technology [8,9] can be solved.

II. RESEARCH ON FAST PARALLEL IMAGE EXPOSURE STATUS DETECTION ALGORITHM

A. Fast and Parallel Exposure Status Detection Process

The fast parallel image exposure status detection process is as follows:

- (1) Get current image, and calculate the pixel numbers and their brightness in dark and bright regions of the image, where the bright and dark regions have a preset brightness threshold range;

(2) Calculate the proportions of pixels in bright and dark regions respectively based on the above statistical results. Proportions of pixels in bright and dark regions are calculated respectively according to the proportions of pixels numbers in the bright and dark regions in the total number of pixels;

(3) If either the proportions of pixels in bright & dark regions is larger than or equal to the preset proportion, adjust exposure parameters of the imaging device according to the brightness of the pixels counted in the last step and the preset brightness threshold;

(4) If the proportions of pixels in bright & dark regions are lower than the preset proportion, continue to carry out statistics until pixel brightness reach the preset brightness threshold.

B. Design of Fast and Parallel Exposure Status Detection Algorithm

The flow chart of fast parallel image exposure status detection is shown in Fig.1.

Bright and dark regions in the image should be preset in step 1. Bright regions consist of pixels in values from preset boundary value to 255 while dark regions consist of pixels in values from a minimum 0 to the preset boundary value. Bright and dark regions are consistent with image Poisson distribution theory and can be supposed to account for 10% of the entire image each. So the boundary values of bright regions can be supposed to be 255 to $256 \times 90\%$ while dark regions to be 0 to $256 \times 10\%$. Of course, these two boundary

values can also be changed. According to the image inputted in real-time, the numbers of pixels whose values are within the ranges of bright and dark regions are counted respectively.

It is indicated in steps 2, 3 and 4 that the algorithm will count the numbers of pixels in bright regions and dark regions of current image after the image is inputted. Different pixel numbers indicates different results:

1) *If the number of pixels in bright regions is greater than the preset number of bright regions, and the number of pixels in dark regions is less than or equal to the preset number of dark regions, the current image is supposed to be over-exposed, and the pixel values equal to the preset number of bright regions is recorded;*

2) *If the number of pixels in bright regions is less than or equal to the preset number of bright regions, and the number of pixels in dark regions is greater than the preset number of dark regions, the current image is supposed to be in under-exposed, and the pixel values equal to the preset number of dark regions is recorded;*

3) *If the number of pixels in bright regions is greater than the preset number of bright regions, and the number of pixels in dark regions is greater than the preset number of dark regions, the current image is supposed to be extreme exposed, and the pixel values equal to the preset numbers of dark & bright regions is recorded respectively;*

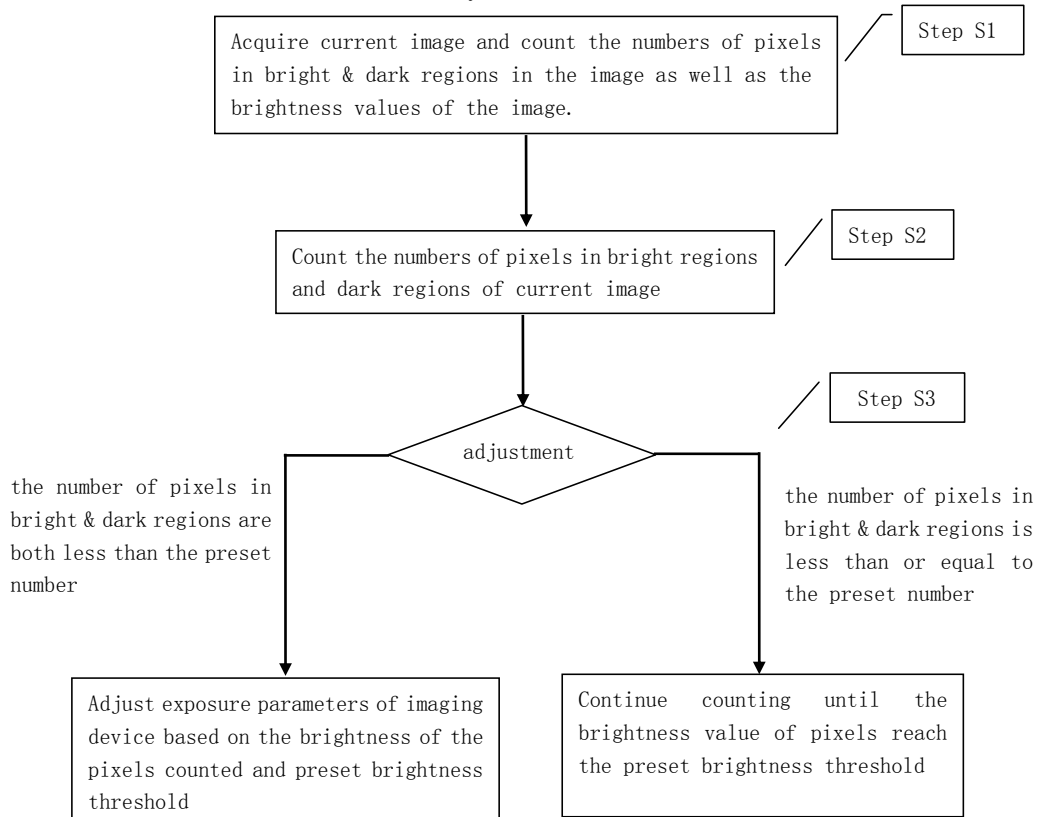


Fig. 1. The flow chart of fast parallel image exposure status detection

4) If the number of pixels in bright regions is less than or equal to the preset number of bright regions, and the number of pixels in dark regions is less than or equal to the preset number of dark regions, the current image is supposed to be in normal exposure condition.

C. Exposure Feedback Controlling Method after Fast Parallel Exposure Status Detection

After the above detection is finished, different exposure feedback control methods will be taken based on exposure state.

If it is over-exposed, use the ratio of recorded pixel values equaling to the preset number and the preset boundary value of bright regions as the gain parameter of light in time for adjusting the CMOS image sensor; if it is supposed to be under-exposed, use the ratio of the preset boundary value and recorded pixel values equaling to the preset number and of dark regions as the gain parameter of light in time for adjusting the CMOS image sensor; if it is supposed to be over-exposed and under-exposed, use the new gain factor calculated through weight processing of the above two ratios as the gain parameter of light in time for adjusting the CMOS image sensor.

After detection factor are set in the above three conditions, continue to calculate the real-time input image adjusted until the image is in normal exposure.

III. HARDWARE ARCHITECTURE DESIGN FOR FAST PARALLEL EXPOSURE STATUS DETECTION & CONTROL

Hardware architecture for fast parallel image exposure status detection & control is shown in Fig.2, including system bus interface, system control status register banks, exposure state machine, pixels counter banks, pixel interface as well as control logic and etc.

A. System Bus Interface And System Control Status Register Banks

An AHB interface of common SoC is utilized as the system bus interface for easy system integration and debugging. System control status register banks communicate with host computer primarily through system bus interface to make the configuration and status query of the module by host computer. It includes control registers of automatic exposure control module, such as image resolution, module enable and etc. It also includes some status information of automatic exposure control module such as the exposure status of the current frame.

B. Pixel Interface And Control Logic Module

Pixel interface and control logic module is mainly the luminance component interface for receiving pixels and transfer these pixels to the following bright and dark counter banks of pixel numbers divided by 32 to count the information on bright and dark regions in real-time. Meanwhile, the interface is also responsible for receiving the feedback signals from the system over-exposed, under-exposed and extreme exposure state machine. When status of current frame is over-exposed, under-exposed or extreme exposed, the luminance

component interface of pixels in subsequent two frame is masked to achieve power saving. Because under normal circumstances, once the current frame is detected in non-normal exposure state, the state should be reported to the host system and host system will re-adjust the exposure parameters of the image sensor. The processing of subsequent two frames can be temporarily ignored, and this kind of treatment is also reasonable.

C. Division by 32 Counter Banks

Pixel data will be transmitted simultaneously to the bright and dark counter banks of pixel numbers divided by 32. These two modules will independently calculate statistical information of bright and dark regions in parallel, and the results will be sent to the corresponding state machines of over-exposed bright and dark regions. The circuits of these two counter banks are shown in Fig.3 and Fig.4 respectively.

As can be seen from Fig.3 and Fig.4, two counter groups consist of 33 (since divided by 32) counters of 28 bits, increment in 1. Each counter is responsible for count the number of pixels in corresponding pixel values.

Division by 32 of bright region is corresponded to the maximum brightness of 255 in physical world and the preset lower threshold in bright regions. Division by 32 of dark region is corresponded to the minimum brightness of 0 in physical world and the preset higher threshold in dark regions.

Two counter groups simultaneously count pixels in a frame, and give specific statistical information to the two state machines during frame interval. The statistic results will be transferred simultaneously to over-exposed state machine of bright regions and under-exposed state machine of dark regions to independently determine in parallel whether the current frame is over-exposed or under-exposed. The results will be sent to the next system over-exposed, under-exposed and extreme exposure state machines.

D. System Over-, Under-, and Extreme Exposure State Machines

If pre-stage judgment indicates that neither over-exposed nor underexposed is exist, then conclusion will be that system is in normal exposure condition and no action will be taken.

If pre-stage judgment indicates over-exposure and without under-exposure, then conclusion will be that system is in over-exposure condition, the corresponding status and interrupt bits will be reset and actions in the following two frames will be masked.

If pre-stage judgment indicates under-exposure and without over-exposure, then conclusion will be that system is in under-exposure condition, the corresponding status and interrupt bits will be reset and actions in the following two frames will be masked.

If pre-stage judgment indicates under-exposure and over-exposure, then conclusion will be that system is in extreme exposure condition, the corresponding status and interrupt bits will be reset and actions in the following two frames will be masked.

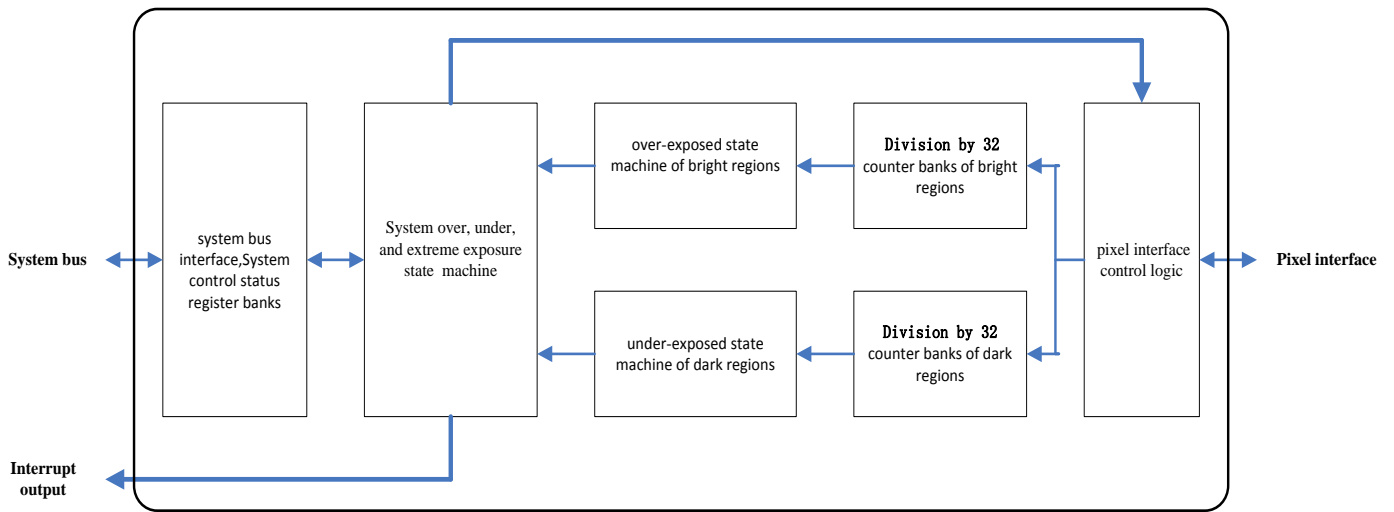


Fig. 2. Hardware architecture of fast parallel image exposure status & control

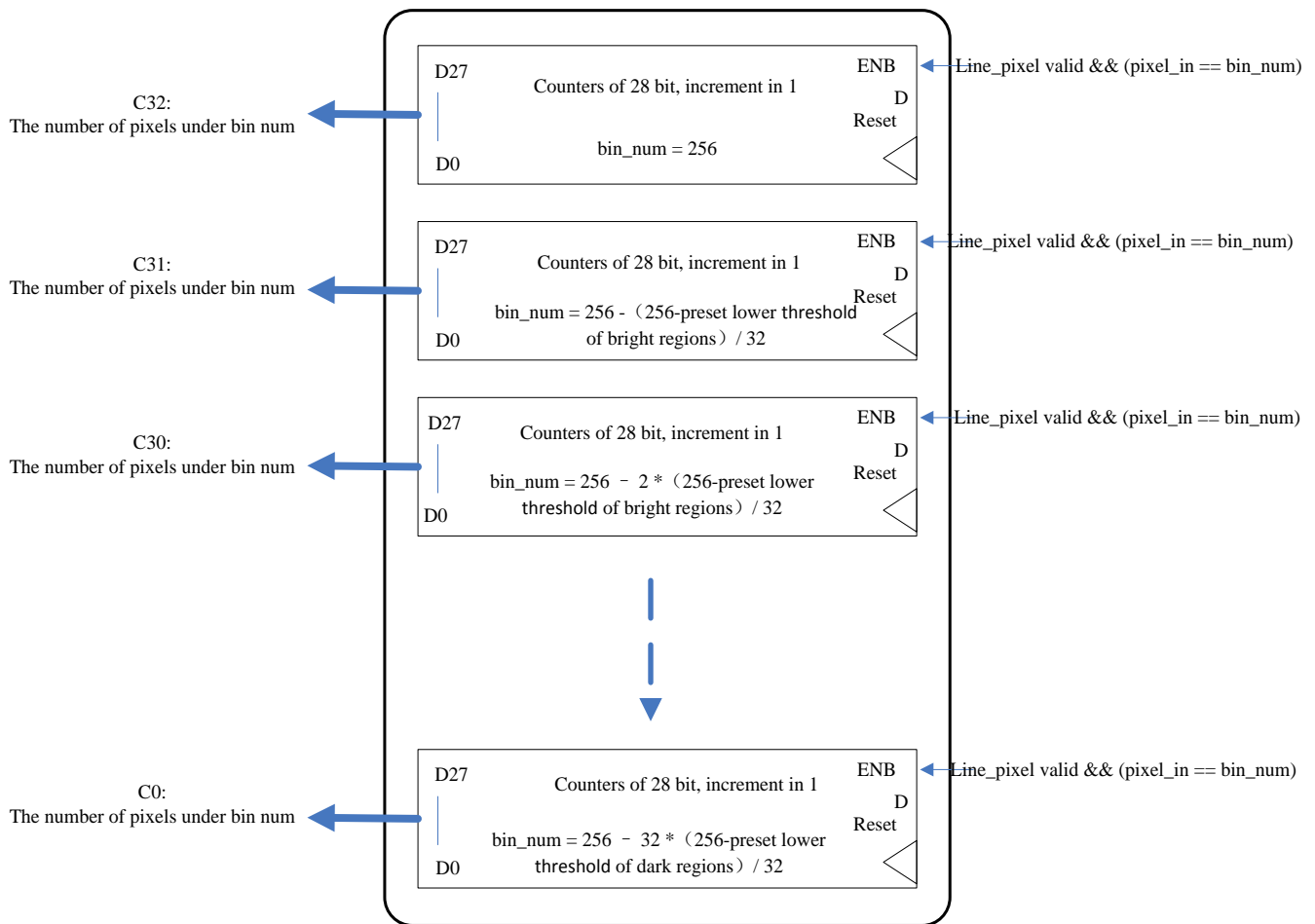


Fig. 3. Pixel counters in each bright region divided by 32

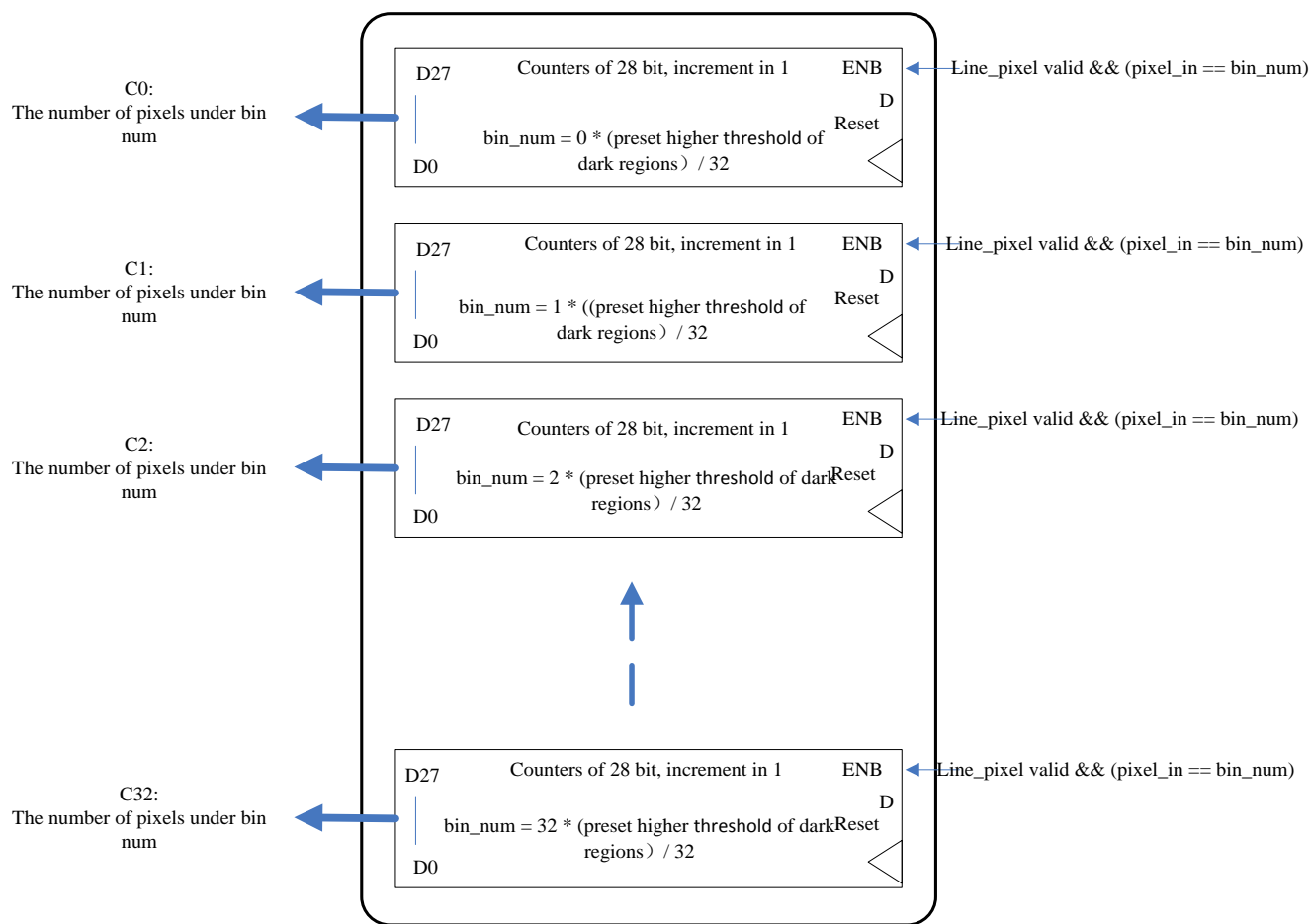


Fig. 4. Pixel counters in each dark region divided by 32

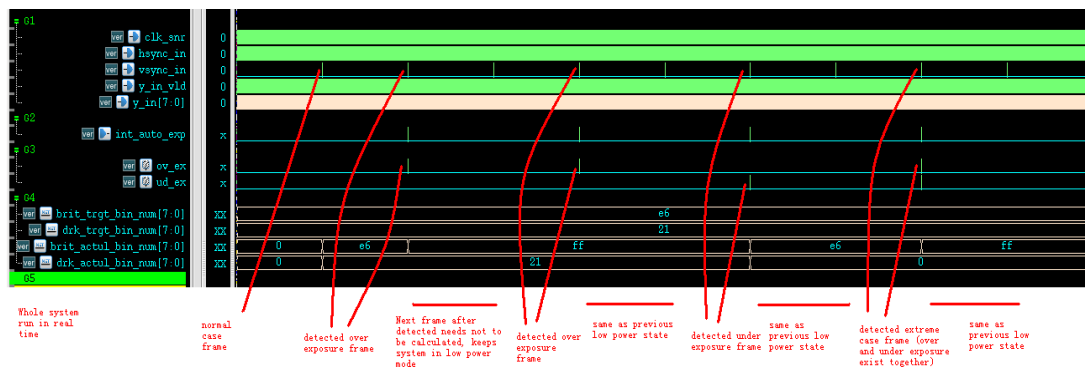


Fig. 5. Simulation results of fast parallel exposure status detection

E. Simulation Results of Fast Parallel Exposure Status Detection

Simulation results using hardware emulation software VCS are given in Fig.5. It can be derived from simulation results that the entire hardware system can running in real-time, without frequent software intervention, thus reducing software overhead. In addition, it can accurately indicate exposure status while dynamically adjust power consumption.

Compared with the traditional image exposure status detection methods, the proposed method count the number of

pixels in bright and dark regions and obtain these pixels brightness value ,then calculate the proportions of pixels numbers in bright regions with pixels numbers in dark regions. If the calculated proportion is different with preset value, compensation measures should be taken. The proposed method has lower computational complexity and easy detection process, which can not only quickly determine image exposure state, but also greatly improve the real-time response of image exposure control.

When the proportions of pixels numbers in bright regions or dark regions is larger than or equal to preset value, or the

current brightness value of pixels counted reaches the preset brightness threshold, the counting process will be automatically stopped, avoiding unnecessary computational steps ,thus reducing the total computational works of the system.

IV. CONCLUSIONS

The paper proposed a fast and real-time parallel exposure status detection and control algorithm with the implementation of detailed hardware circuits. The proposed method explore the maximum luminance value of the bright regions and the minimum luminance value of the dark regions, count the numbers of pixels in bright and dark regions and these pixels bright values. According to the proportions of pixels numbers in bright regions or dark regions, image exposure status can be determined to be as under-exposure, over-exposure or both. Through the compensation of brightness and darkness, the problems of the existing technology are solved such as complicated detection on image exposure status, detection lag, and poor real-time performance in this algorithm. The traditional method limitations on only over-exposure state detection are overcome. With advantages simple calculation, simple structure, strong real-time response, the proposed algorithm is very suitable for applications in intelligent mobile devices.

REFERENCES

- [1] Bu Chin Wang, "Digital signal processing techniques and applications in radar image processing," John Wiley, 2008.
- [2] Tao Jiang, Kuhnert K.D. , Duong Nguyen, Kuhnert. L, "Multiple templates auto exposure control based on luminance histogram for onboard camera," Computer Science and Automation Engineering (CSAE), 2011 IEEE International Conference, vol.3, pp.237-241, 2011.
- [3] R Gonzalez,R Woods, Digital Image Processing(Second Edition), New Jersey:Prentice Hall, 2002.
- [4] Xin Po Wang, Ming Yang, Tian Tian Meng, " The exposure method for digital microscopic image," Conference Anthology, 2013 IEEE International Conference, pp.1-4, 2013.
- [5] Wang Min, Huang Zhan-Hua , "New concepts of network camera for measuring images," Image Analysis and Signal Processing 2009, IEEE International Conference, pp.389-391, 2009.
- [6] E.Reinhard, M.Stark, P.Shirley, and J.Ferwerda, "Photographic tone reproduction for digital images.ACM Transactions on Graphics, " vol.21(3), pp.267-276, 2002
- [7] H. T. Yang, Y. L. Chang, J. Wang, and J. Y. Huo, "A new automatic exposure algorithm for video cameras using luminance histogram," Frontiers of Optoelectron, China, vol. 1(3), pp.285-291, 2008.
- [8] Wang Min, Huang Zhan-Hua , "New concepts of network camera for measuring images," Image Analysis and Signal Processing 2009, IEEE International Conference, pp.389-391, 2009.
- [9] Vassilios Vonikakis and Ioannis Andreadis, "Fast automatic compensation of under/over- exposed image regions," Lecture Notes in Computer Science, vol.4872, pp.285-291, 2013.

Quantifying the Relationship between Hit Count Estimates and Wikipedia Article Traffic

Tina Tian, Ankur Agrawal
Department of Computer Science
Manhattan College
New York, USA

Abstract—This paper analyzes the relationship between search engine hit counts and Wikipedia article views by evaluating the cross correlation between them. We observe the hit count estimates of three popular search engines over a month and compare them with the Wikipedia page views. The strongest cross correlations are recorded with their delays in days. We present the results in both graphs and quantitative data among different search engines. We also investigate the predicting trends between the hit counts and Wikipedia article traffic.

Keywords—hit count estimations; search engines; Wikipedia article traffic; cross correlation; positive delay, negative delay; prediction of Web hosting trend

I. INTRODUCTION AND RELATED WORK

When a user searches for a term, the search engines return the estimated number of Web pages related to the keyword searched, named the search engine hit count [1]. Search engine results are now widely used for measurement purposes in Webometrics. Researchers have used hit counts as input for many studies of Web information, e.g., to determine how many pages in one country link to another [2]. Cilibrasi and Vitanyi used search engine hit counts to measure semantic similarities of words [3].

Wikipedia is the largest encyclopedia in existence and it is among the fastest growing sites on the web [4]. Wikipedia has appeared in many research papers as a valuable data source of study. For example, Ponzetto and Strube used Wikipedia for computing semantic relatedness [5]. Cucerzan presents a system for the recognition and semantic disambiguation of entities based on information extracted from Wikipedia [6]. In the field of health information, Laurent and Vickers investigated whether Wikipedia article traffic correlated with epidemiological factors and compared page views statistics to a major online health encyclopedia [7].

This paper studies the correlation between search engines' hit counts and the Wikipedia article traffic. It also analyzes the predicting trends between the two resources, which can be a useful tool for business and Web publishers to promote their websites.

The rest of the paper is organized as follows. Section II describes the process of mining hit counts and Wikipedia article views and describes the approach to evaluating their correlation. In Section III, we present the results by graphing the correlations and we investigate if traffic to Wikipedia

articles can predict the Web hosting trend, or vice versa. Section IV concludes the paper and proposes future work.

II. METHODS

The experiment was based on 400 popular search terms from four different categories, including medicine, people, science and technology. We collected 100 terms for each category. Each term was queried against the official site of Wikipedia article traffic statistics [8], which measures the page views of a given article in a given month. Besides the total number of views in a month, the site also provides daily views in JSON format [8]. A program was developed to extract the Wikipedia page views over a period of one month and to store them in a database.

In the meantime, each search term was sent to popular search engines to collect their daily hit count. Three search engines were selected in this research, including Google, Yahoo! and Bing. Programs were built to retrieve search engines' hit counts through their APIs. Google's JSON/Atom Custom Search API returns search results of a term through RESTful requests from Google Custom Search [9]. The returned results are presented in JSON or Atom format, which can be parsed with a program.

Bing's hit counts were retrieved in a similar way using the Bing Search API, which returns the search results in XML or JSON format [10]. We used Bing's results to represent both Bing and Yahoo!, since the latter is powered by Bing's search engine [11].

We queried the search terms against the search engines' APIs and mined their Wikipedia page views. The results were observed over a month and were stored in a database. Outliers with 0 hits from a search engine were removed.

In order to analyze the relationship between the search hit counts and the Wikipedia article traffic, the cross correlation of the two was calculated. Cross correlation is a standard method of estimating the degree to which two series are correlated [12]. It is a measure of similarity of two series as a function of the lag of one relative to the other. Given two N-element series $x[i]$ and $y[i]$ where $i=0,1,2,\dots,N-1$, the cross correlation r at delay d is defined as in equation (1),

$$r(d) = \frac{\sum_i ((x[i] - mx)(y[i - d] - my))}{\sqrt{\sum_i (x[i] - mx)^2} \sqrt{\sum_i (y[i - d] - my)^2}} \quad (1)$$

where m_x is the mean of series x and m_y stands for the average of series y . We used the x series to represent the search engine's monthly hit counts of a search term and we used the y series for the number of views of the according Wikipedia article. Cross correlations have been observed with different days of delays assigned, ranging from $-N$ to N . The maximum cross correlation was then selected and stored together with its delay for further analysis.

The pseudocode below illustrates the procedure of calculating the strongest cross correlation, knowing the search engine's monthly hit counts of a term and its Wikipedia monthly traffic.

```
MAX_CROSS_CORRELATION(x[], y[]){
    max_r = 0;
    FOR (delay = -N; delay <= N; delay++){
        Calculate the cross correlation r at delay
        IF (r > max_r){
            max_r = r;
            max_delay = delay;
        }
    }
    return [max_r, max_delay];
}
```

A cross correlation with 0 delay has the same value as the Pearson correlation coefficient. A negative delay represents a delay in the search engine hit count, while a positive delay stands for a delay in the Wikipedia article traffic. For example, Figure 1 shows Google's monthly hit counts for search term "Jared Cohen." Figure 2 plots the series of the monthly views of article "Jared Cohen."

The cross correlation series with a maximum delay of 30 is shown in Figure 3. One can observe that the strongest correlation occurs at delay of about 1. In other words, the series of Google hit counts from day 1 correlated with the Wikipedia article traffic from day 2. Thus, for search term "Jared Cohen," the decline of Google's hit count predicted the traffic of the same titled article on Wikipedia.

Java was used to extract hit counts from search engines and page views from Wikipedia. We wrote programs in C++ to calculate and store the cross correlations between the two. At the time of writing, Google's Custom Search API provides 100 search queries per day for free with additional queries at a cost [9]. The Bing Search API charges if there are more than 5,000 transactions per month [10].

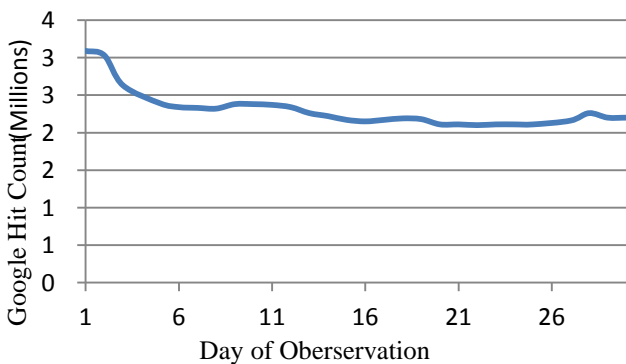


Fig. 1. Monthly Google hit counts for term "Jared Cohen"

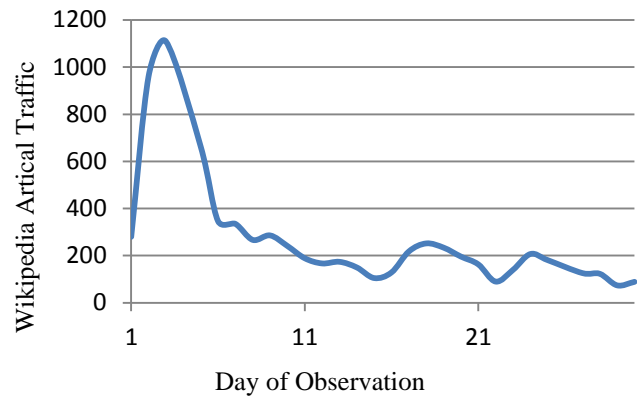


Fig. 2. Monthly Wikipedia page views of "Jared Cohen"

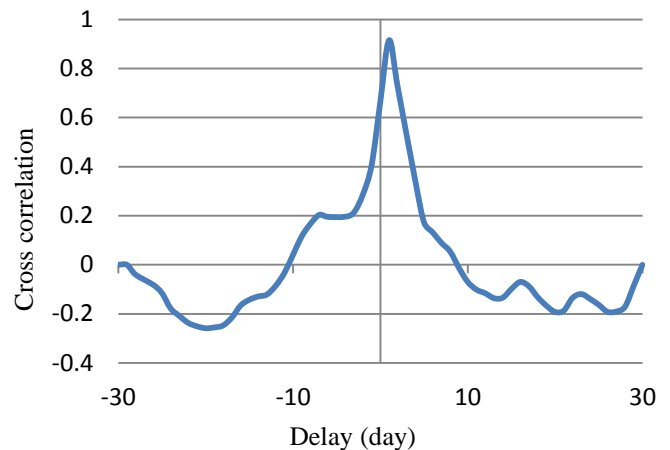


Fig. 3. Cross correlation between Google hit counts and Wikipedia page views of Jared Cohen

III. RESULTS

We calculated the strongest cross correlations for terms in the four categories and recorded them together with their delays. The results are shown in Figure 4 and Figure 5. The horizontal axis represents the delay in days and the vertical axis represents the strongest correlation caused by the according delay. We can see from the figures, that most of the cross correlations lie between 0.2 and 0.8 with delays ranging from -20 days to 20 days.

Table I displays the average and the standard deviation of the strongest cross correlations between search engine hits and Wikipedia article views. We compare the results among terms from different categories. Yahoo! and Bing show a stronger correlation than Google for terms in medicine, science and technology, while Google has a stronger correlation for queries about people. Google, in general, has a larger standard deviation than Bing and Yahoo!.

Terms in the medicine field result in the lowest average cross correlation between the search engine hits and Wikipedia page views. One of the reasons is that, unlike terms in other categories, medical terms often have synonyms. For example, bovine spongiform encephalopathy has a synonym of mad cow disease, which is more commonly known. At the time of writing, search term "bovine spongiform encephalopathy"

results in 444,000 Google hits, while query “mad cow disease” returns 1,640,000 Web pages in Google. The former result fails to represent the total number of pages on the Web regarding the disease. One possible solution to this problem is to include the hits of synonyms, which we will discuss in Section IV.

Table II shows the average and the standard deviation of delays that result in the maximum cross correlation. It is interesting to see that Bing/Yahoo! has an average positive delay except for the science category, while Google has negative delays for all categories. Comparing with Bing and Yahoo!, Google has a larger standard deviation.

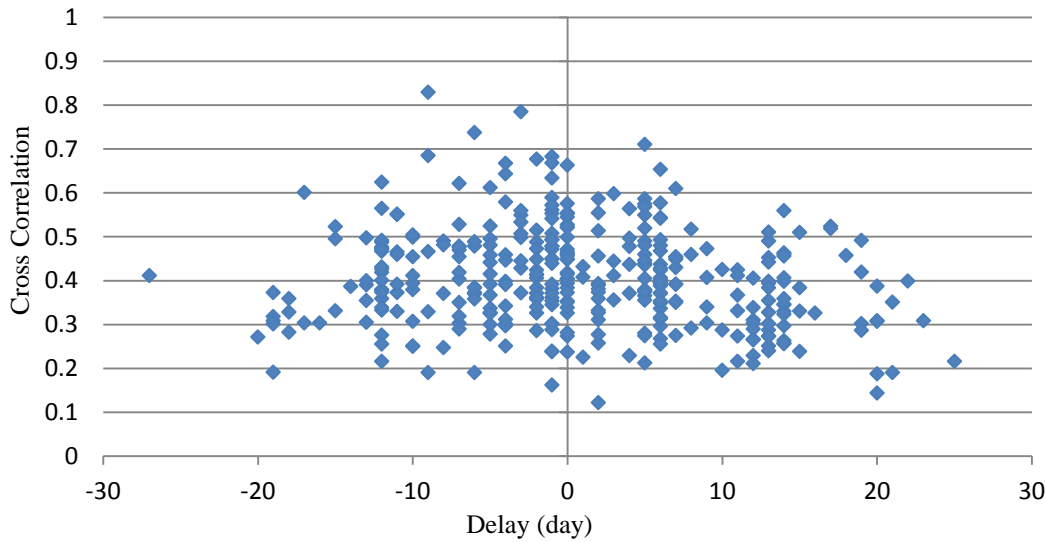


Fig. 4. Maximum cross correlations with delays between Bing/Yahoo!'s hit counts and Wikipedia article traffic of all terms

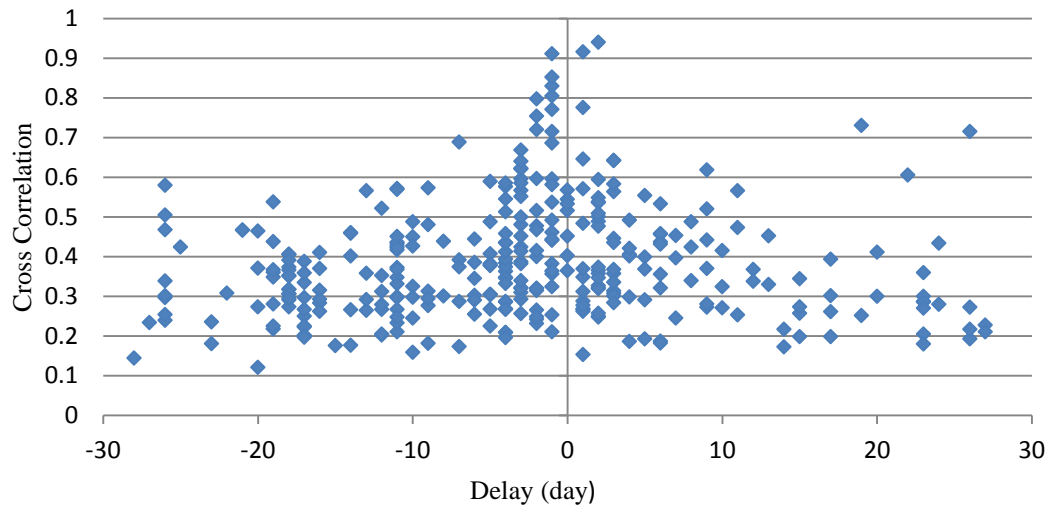


Fig. 5. Maximum cross correlations with delays between Google's hit counts and Wikipedia article traffic of all terms

TABLE I. AVERAGE AND STANDARD DEVIATION OF THE STRONGEST CROSS CORRELATIONS

	Average Cross Correlation (Bing/Yahoo!)	Standard Deviation (Bing/Yahoo!)	Average Cross Correlation (Google)	Standard Deviation (Google)
Medicine	0.37	0.10	0.34	0.11
People	0.41	0.11	0.46	0.17
Science	0.41	0.11	0.35	0.11
Technology	0.44	0.12	0.42	0.17
Overall	0.41	0.11	0.39	0.15

TABLE II. AVERAGE AND STANDARD DEVIATION OF DELAYS

	Average Delay (Bing/Yahoo!)	Standard Deviation (Bing/Yahoo!)	Average Delay (Google)	Standard Deviation (Google)
Medicine	0.04	9.61	-4.19	10.75
People	2.92	9.48	-1.60	10.32
Science	-3.15	9.48	-3.31	13.82
Technology	2.36	7.94	-2.90	11.51
Overall	0.78	9.37	-3.06	11.56

As mentioned in Section II, a positive delay in the correlation represents a delay in the Wikipedia page views, while a negative delay indicates a delay in the number of websites. In other words, Wikipedia article traffic can be predicted by the number of related Web pages, if there is a positive delay. Similarly, the Web hosting trend can be predicted by the growth/decline of the Wikipedia article views, if there is a negative delay. Among all the search terms, Bing and Yahoo! produce 53.3% positive delays and 46.7% negative delays while calculating the cross correlation. On the other hand, 62.7% of the cross correlations with Google were generated by negative delays, as shown in Table III.

TABLE III. DISTRIBUTION OF POSITIVE AND NEGATIVE DELAYS

	Positive Delays	Negative Delays
Bing/Yahoo!	53.3%	46.7%
Google	37.3%	62.7%

IV. CONCLUSIONS AND FUTURE WORK

In this paper, we presented a method to analyze the correlation between search engines' hit counts and Wikipedia's article traffic. We collected 400 popular search terms from fields of medicine, people, science and technology. Each term was sent to major search engines, including Google, Bing and Yahoo!, to retrieve the number of related pages on the web. We also extracted the number of views on the term's Wikipedia article. A month of data was collected and used to calculate the cross correlations between the search engine hits and Wikipedia page views. Most cross correlations lie between 0.2 and 0.8, which represent a moderate to strong positive relationship.

We also analyzed the predicting trends between the hits on the Web and the views in Wikipedia. 62% of the cross correlations between Google and Wikipedia were caused by negative delays, which indicates a leading trend in the Wikipedia article traffic in predicting the inclination of Web hosting. As mentioned in Section III, a term with synonyms may cause inaccuracy in calculating the correlation. In the future, we plan to use a synonym search API, such as [13], to include the search engine hit counts returned by synonyms.

ACKNOWLEDGMENT

We thank Rocco Pascale for his contribution to the collection of the search engine hit counts and the Wikipedia article views.

REFERENCES

- [1] T. Tian, S.A. Chun, and J. Geller, "A prediction model for Web search hit counts using word frequencies," *Journal of Information Science*, Sage Publishing Co., vol. 37, issue 5, pp. 462-475, 2011.
- [2] L. Yuen, M. Chang, Y.K. Lai, C.K. Poon, "Excalibur: a personalized meta search engine," the 28th Annual International Computer Science Software and Applications Conference, vol. 2, pp. 49-50, 2004.
- [3] R.L. Cilibrasi, and P. Vitanyi, "Normalized Web distance and word similarity," in: *Handbook of Natural Language Processing*, 2nd ed., N. Indurkha and F.J. Damerau, Eds. Boca Raton, FL: CRC Press, Taylor and Francis Group, 2010.
- [4] D. Milne, and I. H. Witten, "Learning to link with Wikipedia," the 17th ACM Conference on Information and Knowledge Management, pp. 509 - 518, 2008.
- [5] S.P. Ponzetto, and M. Strube, "Knowledge derived from Wikipedia for computing semantic relatedness," *Journal of Artificial Intelligence Research*, vol. 30, pp. 181-212, 2007.
- [6] S. Cucerzan, "Large-scale named entity disambiguation based on Wikipedia data," the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Prague, pp. 708-716, June 2007.
- [7] M.R. Laurent, T.J. Vickers, "Seeking health information online: does Wikipedia matter?" *Journal of the American Medical Informatics Association*, vol. 16, issue 4, pp. 471-479, July 2009.
- [8] Wikipedia article traffic statistics, <http://stats.grok.se/>, retrieved on 04/15/2015.
- [9] Google's JSON/Atom Custom Search API, <https://developers.google.com/custom-search/json-api/v1/overview>, retrieved on 04/15/2015.
- [10] Bing Search API, <http://datamarket.azure.com/dataset/bing/search>, retrieved on 04/15/2015.
- [11] D. Milne, and I. H. Witten, "Learning to link with Wikipedia," the 17th ACM Conference on Information and Knowledge Management, pp. 509 - 518, 2008.
- [12] M. Andrews, "Searching the Internet," *IEEE Software*, vol. 29, issue 2, pp. 13- 16, 2012.
- [13] R. Bracewell, "Pentagram Notation for Cross Correlation," *The Fourier Transform and Its Applications*, New York, McGraw-Hill, pp. 46 and 243, 1965.
- [14] Stands4 API, <http://www.abbreviations.com/api.asp>, retrieved 05/17/2015.

Applying Linked Data Technologies for Online Newspapers

Tsvetanka Georgieva-Trifonova
Department of Mathematics and Informatics
University of Veliko Tarnovo
Veliko Tarnovo, Bulgaria

Tihomir Stefanov
Department of Mathematics and Informatics
University of Veliko Tarnovo
Veliko Tarnovo, Bulgaria

Abstract—The constantly growing data volume at the companies along with the necessity for finding information for the shortest possible time span involves methods of information search different from the ones conventionally used. The semantic technologies, developed in the late 90s and the beginning of the new century, are viewed as a new generation of databases and as text analyzing technologies. The present paper deals with researching the opportunities and examining the advantages of applying the semantic web technologies and the linked data for online newspapers. Besides, a RDF-based ontology for the purposes of a system for study and evaluation of online editions of regional daily newspapers is proposed. SPARQL endpoint is implemented to access the RDF data.

Keywords—*semantic web; ontology; linked data; SPARQL endpoint; RDF dataset; online newspaper*

I. INTRODUCTION

The constantly growing data volume in companies along with the necessity for finding information within the minimum required time entails methods of information search different from conventional ones. The contemporary information society structure becomes more and more complex and the requirements to the effectiveness of the information processing algorithms are growing accordingly. The most popular technologies in this aspect recently are data mining, knowledge discovery in databases, machine learning. They provide theoretical and methodological basis for the study, analysis and rationalization of huge databases, but on account of the specifics of the Web data structure they themselves are not effective enough.

As a result, the semantic web popularity increases with its possibility to help in the dissemination of knowledge embedded in documents provided that the process of semantic interpretation becomes at least partially automated. Therefore, it is necessary to describe and present in sufficiently formalized form the data relevant to a specific area.

The present paper deals with researching the opportunities and introducing the advantages of applying the semantic web and linked data technologies for regional online newspapers in Bulgaria. The created RDF dataset *LinkedNewsData* which is linked to DBpedia and Europeana has been described. The data access is ensured by providing a SPARQL endpoint.

The rest of the paper is organized as follows. In Section 2, a

review of the semantic web and linked data technologies is made. In Section 3, the related works on the application of semantic web technologies for digital newspaper archive maintenance are surveyed. In addition, the available linked data received from online newspapers along with SPARQL endpoint for data access are studied. In Section 4, the construction of the dataset *LinkedNewsData*, extracted from a system of study and evaluation of online regional newspaper editions, is motivated and described.

II. SEMANTIC WEB TECHNOLOGIES

The semantic technologies developed in the late 90s and the beginning of the new century, are viewed as a new generation databases and text analyzing technologies. The semantic web concept was proposed for the first time in 2001 by Tim Berners-Lee [1] – the founder of the World Wide Web. It consists in the automated process of conversion into semantic meaning of the data received from different network resources. The processing and exchange of information should be carried out not by people but by special agents (programs, distributed in the network). In order to be able to interact, these agents shall present the incoming data from each source in common form. Semantic web languages have two basic aspects. First, they shall have formal syntax and semantics to allow automated content processing. And second, a standard dictionary shall be provided, connected with the real world semantics, accessible for information sharing to both people and automated agents.

The Resource Description Framework (RDF) [2] standard, developed by the World Wide Web Consortium (W3C), allows for the semantic description of web resources and their relations in a way understandable to both people and machines, with the option XML presentation format (eXtensible Markup Language). RDF Schema is RDF extension, defining resource classes, their properties and their relations. OWL (Web Ontology Language) [3] is an extension of the RDF and RDFS, aimed specifically at description of reusable definitions of specific problem areas, called *ontologies*.

One of the main advantages of using RDF and OWL for the presentation of information is their reusability and their development through integration and upgrade of ontologies already built by other developers for specific areas, accessible on the Web. The development of semantic web languages is illustrated in Figure 1.

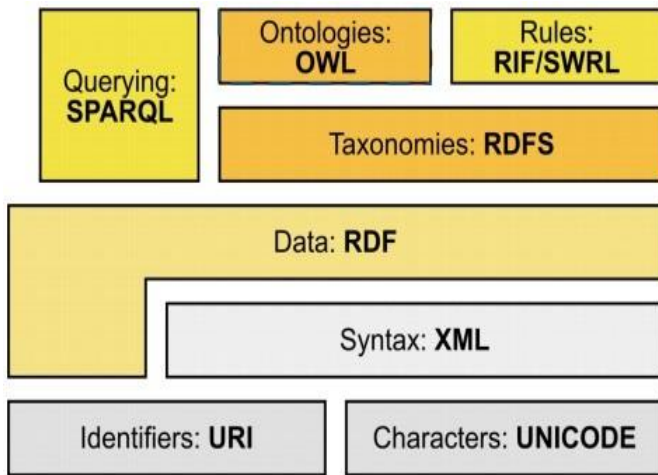


Fig. 1. Development of languages for semantic web

With the growing popularity of RDF and RDFS in the business companies and scientific circles, greater amount of RDF web content has been created which arises a necessity for standard access to the data stored. The language designed to allow the execution of queries to RDF data is SPARQL (SPARQL Protocol and RDF Query Language).

One of the most important concepts related to the semantic web, is the so-called *linked data*. Its goal is the publication of structured data, so that they can be easily linked to each other and thus, more useful. It is built on the basis of standard web technologies such as HTTP, URI, RDF, but develops them further so as to enable the sharing of information in a way comprehensible to computers. Some of the basic principles of linked data are:

- Using URI to designate the resources;
- Using HTTP URI for reference and search of resources by people and machines;
- The inclusion of links to related resources by using their URI when publishing a resource in the Web.

III. APPLYING THE SEMANTIC WEB TECHNOLOGIES FOR MAINTENANCE OF THE DIGITAL ARCHIVE OF A NEWSPAPER

Usually, the technologies traditionally used provide opportunities for keyword search, for viewing the text of the published newspaper article, navigation through static hand-made hyperlinks between news materials (for example, links to previous topic related articles). Aspects that can be improved are:

- Search by keywords has limited expressive capability;
- Weak relationship between archive items: users may need to manually perform a few indirect requests before they can get answers to complex queries;
- The lack of common standard for the presentation by news archive sharing between the newspapers;
- Lack of consent about the terminology used in content description among journalists and between the journalists and archivists;

- Lack of commitment on the part of reporters in the archive creation process.

Neptuno project [4] is directed to the usage of the semantic web technologies to improve the exploitation and maintenance of the digital archive of a newspaper. The goal is to develop high-quality semantic archive for the Diari SEGRE newspaper, where:

- reporters and archivists have more expressive means to describe and annotate news items;
- reporters and readers are provided with better search and browsing options than those available;
- the archive maintenance system is open to integration in electronic marketplaces of news products.

As great investments have been made in the present news management systems, it is advisable to make this transition gradually. On account of this, in [5] the construction of an ontological framework based on existing journalistic and multimedia standards is proposed. These standards are based on XML technologies. Attached is the approach for the conversion of XML Schema into OWL, combined and supplemented with XML into RDF conversion. The main advantage of this approach is that it allows the reuse of existing metadata, which facilitates data integration, management and retrieval of previously stored news. The ontological framework is applied in the Diari Segre Media Group, which produces press, radio and television content.

A typical example illustrating the benefits of semantic technologies, is the BBC media group, which in its technological architecture replaces the MySQL with the semantic database OWLIM Enterprise of Ontotext, applied to the website, dedicated to the Football World Championship in 2010 and the Olympic Games in London 2012 [6]. Ontotext [7] is a Bulgarian company for semantic technologies, part of Sirma Group Holding.

Besides, BBC performs a successful integration and links data from different areas through applying the linked data technologies [8]. The linked data are accessible from the BBC SPARQL endpoint: <http://api.talis.com/stores/bbc-backstage/services/sparql>.

Some of the SPARQL endpoints providing access to linked data are:

- DBpedia endpoint: <http://dbpedia.org/sparql>;

DBpedia [9] is a dataset derived from Wikipedia, freely available through the use of the semantic web and the linked data technologies.

- Europeana endpoint: <http://europeana.ontotext.com/sparql>;

Europeana [10] is an electronic library in which scanned books, pictures, audio, video objects from museums and archives reflecting different aspects of European culture are being stored.

- British National Bibliography endpoint: <http://bnb.data.bl.uk/sparql>;

The British National Bibliography dataset [11] is designed to store information about the publishing of United Kingdom and the Republic of Ireland since 1950. The dataset includes metadata for published books, periodicals, magazines, newspapers, etc.

- Endpoints for science fiction datasets.

Dataset RKBExplorer [12] describes the publications, authors, institutions, conferences, etc.

- IEEE Papers (RKBExplorer) SPARQL Endpoint: <http://ieee.rkbexplorer.com/sparql>;
- DBLP Computer Science Bibliography (RKBExplorer) SPARQL Endpoint: <http://dblp.rkbexplorer.com/sparql>;
- ACM Bibliography (RKBExplorer) SPARQL Endpoint <http://acm.rkbexplorer.com/sparql>.

For the present paper, published linked data with a given SPARQL access endpoint have been studied. According to the study, there is a deficiency of datasets containing news from newspapers published in Bulgarian.

IV. LINKEDNEWSDATA – LINKED DATA RECEIVED FROM ONLINE NEWSPAPERS

Except for the maintenance of the digital archive of the newspaper, the application of semantic web technologies is interesting in terms of constructing a system designed for the comparison and evaluation of online newspaper editions. The results generated from such a system, can be used from the owners of online editions in order that they be analyzed and thus become helpful in taking proper decisions for improvement of consumer access to information and satisfaction.

Below is shown a summary of the advantages of the semantic web data model to a database created for the purpose of such type of system.

- Without availability of semantic data, the owners of specific websites in which the search is based on SQL (Structured Query Language), are supposed to focus on the widely used data formats in order that the information shared is comprehensible to others. Deficiency is that in this way the sharing can not occur automatically and entails the human intervention.
- In terms of the data pattern in the semantic web, one and the same ontology, intended for the expression data meaning can be used by various websites that provide search through SPARQL.

The advantages mentioned above, are the basic reason to convert the existing relational database *SiteDB* into RDF data and their publishing as linked data.

The relational database *SiteDB* is designed and created for the purpose of a web-based system for examination and evaluation of online editions of regional daily papers, described in [13]. The database is implemented using the system for the management of relational databases MySQL.

In this paper, the *LinkedNewsData* ontology has been proposed, obtained after the conversion of the relational database *SiteDB* into RDF. *LinkedNewsData* is a set of RDF data associated with DBpedia and Europeana. Sample data for the study of the defined structure are accessible through the use of:

- SPARQL user applications – SPARQL endpoint: <http://newspaper.byethost18.com/arc2-starter-pack/endpoint.php>;
- Semantic web browsers – initial point: <http://newspaper.byethost18.com/arc2-starter-pack/news.rdf>.

The implementation of SPARQL endpoint is based on ARC RDF Store [14].

TABLE I. TYPES AND PROPERTIES IN *LINKEDNEWSDATA*

RDF type	Property	Data types of the property values	Description
Site	hasName	Text	Newspaper name
	owl:sameAs	URL	URL of the newspaper in DBpedia
	owl:sameAs	URL	URL of the newspaper in Europeana
	hasURL	URL	Website address of the newspaper
	whenUpdated	Datetime	Date and time of website latest update
	hasOutlinks	Integer	Number of external links referring to the URL of the Website
Page	ofSite	URL of sample type <i>Site</i>	Website of the first newspaper that published the news
	hasTitle	Text	Page title
	hasItemsCount	Integer	Number of news articles
	hasContent	Text	News content
	hasPageURL	URL	URL of the news article
	whenPageCreated	Datetime	Date and time of news article publishing
	hasExternalLinks	Integer	Number of external links
	hasWordsCount	Integer	Word count
Category	hasNewsImage	URL	URL of the news article picture
	inCategory	URL of sample type <i>Category</i>	Page category
	hasCategoryName	Text	Name of the category
Word	owl:sameAs	URL	URL of the category in DBpedia
	owl:sameAs	URL	URL of the category in Europeana
	hasWord	Text	A word
SiteWord	owl:sameAs	URL	URL of the word in DBpedia
	owl:sameAs	URL	URL of the word in Europeana
	inSite	URL of sample type <i>Site</i>	Website of the word containing newspaper
SiteWord	containWord	URL of sample type <i>Word</i>	The word contained on the Website
	hasRepeatCount	Integer	Number of repetitions

Table 1 contains resume of the RDF types, defined in *LinkedNewsData*; the properties; the data type of the property values and their brief description. Graphic illustration of each type of samples (*Site*, *Page*, *Word*, *Category*, *SiteWord*) and some of their properties is shown in Figure 2.

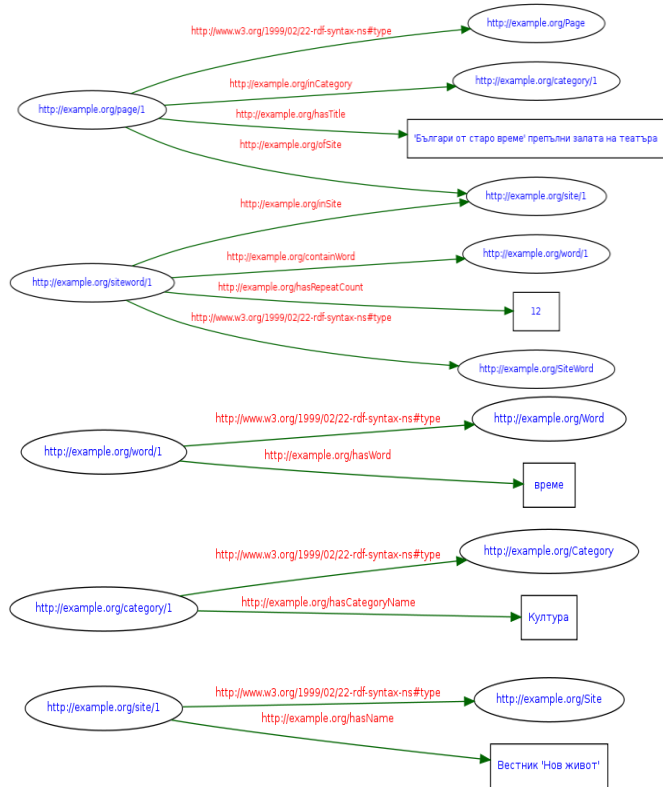


Fig. 2. RDF types *Site*, *Page*, *Word*, *Category*, *SiteWord* and some of their properties

Figure 3 shows Turtle language presentation of the fragment from Figure 2.

```

@prefix e: <http://example.org/> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .

<http://example.org/site/1> a e:Site ;
  e:hasName "Вестник \"Нов живот\"" .

<http://example.org/page/1> a e:Page ;
  e:ofSite <http://example.org/site/1> ;
  e:inCategory <http://example.org/category/1> ;
  e:hasTitle "\"Българи от старо време\" препълни залата на театъра" .

<http://example.org/category/1> a e:Category ;
  e:hasCategoryName "Култура" .

<http://example.org/siteword/1> a e:SiteWord ;
  e:inSite <http://example.org/site/1> ;
  e:containWord <http://example.org/word/7> ;
  e:hasRepeatCount "12"^^xsd:int .

<http://example.org/word/1> a e:Word ;
  e:hasWord "време" .
    
```

Fig. 3. Turtle language presentation

Figure 4 shows a view of the DataSet *LinkedNewsData* in OpenLink Data Explorer, which is a browser extension for semantic data browsing.

Some general data about RDF dataset *LinkedNewsData* can be collected through the execution of SPARQL queries, as shown in Table 2.

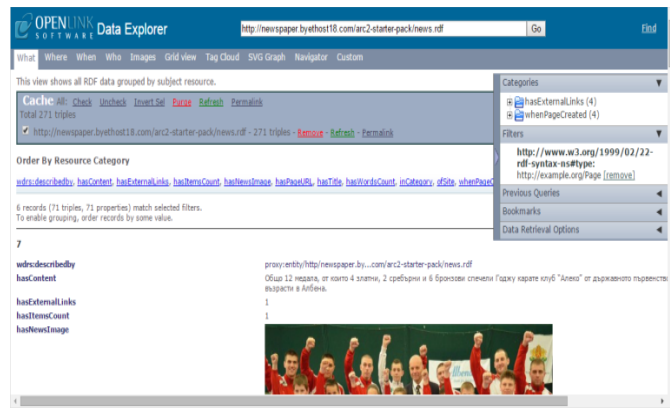


Fig. 4. View of *LinkedNewsData* in OpenLink Data Explorer

TABLE II. SPARQL QUERIES FOR RETRIEVING GENERAL DATA ABOUT *LINKEDNEWSDATA*

Number of triplets	SELECT (COUNT(*) AS ?c) WHERE { ?s ?p ?o. }
Number of newspapers	PREFIX rdf:<http://www.w3.org/1999/02/22-rdf-syntax-ns#> PREFIX np:<http://example.org/> SELECT (COUNT(?s) AS ?c) WHERE { ?s rdf:type np:Site. }
Number of pages per newspaper	PREFIX np:<http://example.org/> SELECT ?sn (COUNT(?p) AS ?c) WHERE { ?p np:ofSite ?s. ?s np:hasName ?sn. } GROUP BY ?sn
Number of pages per category	PREFIX np:<http://example.org/> SELECT ?cn (COUNT(?p) AS ?cp) WHERE { ?p np:inCategory ?c. ?c np:hasCategoryName ?cn. } GROUP BY ?cn
Number of words per newspaper	PREFIX np:<http://example.org/> SELECT ?sn (COUNT(?sw) AS ?c) WHERE { ?sw np:inSite ?s. ?s np:hasName ?sn. } GROUP BY ?sn
Total number of the repetitions of a word in all newspapers	PREFIX np:<http://example.org/> SELECT ?wn (SUM(?rc) AS ?sumrc) WHERE { ?sw np:inSite ?s; np:containWord ?w; np:hasRepeatCount ?rc. ?w np:hasWord ?wn. } GROUP BY ?wn ORDER BY DESC(?sumrc)
Number of links to Dbpedia, of links to Europeana	SELECT (COUNT(?l) AS ?c) WHERE { ?s owl:sameAs ?l. }

The linking of dataset *LinkedNewsData* with DBpedia and Europeana is carried out by means of Google refine. The connection is made through the names of the newspapers, the news categories and the words contained in online newspaper websites.

V. CONCLUSION

In this paper, the opportunities and the advantages of semantic web technologies usage are discussed in terms of maintenance of a digital archive of a newspaper.

The transformation of an existing relational database that stores data for the online editions of regional newspapers into RDF format and their publishing as linked data are dealt with. Our future work envisages creation of user friendly interface to access the data through a simple browser.

REFERENCES

- [1] T. Berners-Lee, J. Hendler, and O. Lassila, "The Semantic Web, *Scientific American*", 2001, pp. 35-43, available at: <http://www.w3.org/2001/sw> (accessed 5 May 2015).
- [2] D. Brickley and R.V. Guha, "RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation", 2004, available at: <http://www.w3.org/TR/rdf-schema> (accessed 5 May 2015).
- [3] D. L. McGuinness and F. Harmelen, "OWL Web Ontology Language Overview, W3C Recommendation", 2004, available at: <http://www.w3.org/TR/owl-features> (accessed 5 May 2015).
- [4] P. Castells, F. Perdrix, E. Pulido, M. Rico, R. Benjamins, J. Contreras, and J. Lorés, "Neptuno: semantic web technologies for a digital newspaper archive", *The Semantic Web: Research and Applications: First European Semantic Web Symposium, Berlin, Springer, 2004*, pp. 445-458.
- [5] R. García, F. Perdrix, R. Gil, and M. Oliva, "The Semantic Web as a Newspaper Media Convergence Facilitator", *Journal of Web Semantics*, Vol. 6, No. 2, 2008, pp. 151-161.
- [6] TechNews.bg, "BBC uses Semantic Technologies of the Bulgarian Company Ontotext", 2010, available at: http://technews.bg/article-18510.html#.U5RGPfl_utM (accessed 5 May 2015).
- [7] Ontotext, "Ontotext provides a complete set of semantic technologies including text mining and GraphDB™, an RDF triplestore that performs inferencing at scale", 2011, available at: <http://www.ontotext.com/company> (accessed 5 May 2015).
- [8] G. Kobilarov, T. Scott, Y. Raimond, S. Oliver, C. Sizemore, M. Smethurst, C. Bizer, and R. Lee, "Media Meets Semantic Web – How the BBC Uses DBpedia and Linked Data to Make Connections", *The Semantic Web: Research and Applications, Lecture Notes in Computer Science, Springer, Vol. 5554, 2009*, pp. 723-737.
- [9] J. Lehmann, R. Isele, M. Jakob, A. Jentzsch, D. Kontokostas, P. N. Mendes, S. Hellmann, M. Morsey, P. Kleef, S. Auer, and C. Bizer, "DBpedia - A Large-scale, Multilingual Knowledge Base Extracted from Wikipedia", *Semantic Web – Interoperability, Usability, Applicability, IOS Press, 2012*, pp. 1-28.
- [10] N. Ikononov, B. Simeonov, J. Parvanova, and V. Alexiev, "Europeana Creative. EDM Endpoint. Custom Views.", *Digital Presentation and Preservation of Cultural and Scientific Heritage, Vol. 3, No 1, 2013*, pp. 35-43.
- [11] C. Deliot, "Publishing the British National Bibliography as Linked Open Data", *Catalogue & Index, Issue 174, 2014*, pp. 13-18.
- [12] H. Glaser, I. Millard, and A. Jaffri, "Rkbexplorer.com: A knowledge driven infrastructure for linked data providers", *In European Semantic Web Conference, 2008*, pp. 797-801.
- [13] T. Stefanov and D. Tsvetkov, "A Model for Evaluation of Regional Electronic Media in terms of Efficiency Criteria and User Satisfaction", *Collection of Writings 'Days of Science 2014', Veliko Turnovo, 2014*, in press.
- [14] A. McIntyre and E. Durham, "ARC RDF Store Easy RDF and SPARQL for LAMP systems", *CSC 8711 Project 4, 2011*.

Fall Monitoring Device for Old People based on Tri-Axial Accelerometer

Jing Luo¹

School of Electronic and Electrical
Engineering
Shanghai University of Engineering
Science Songjiang District
Shanghai 201620, China

Bocheng Zhong²

School of Electronic and Electrical
Engineering
Shanghai University of Engineering
Science Songjiang District
Shanghai 201620, China

Dinghao Lv³

School of Electronic and Electrical
Engineering
Shanghai University of Engineering
Science Songjiang District
Shanghai 201620, China

Abstract—To be able to timely and effective judgment of the elderly fall, a fall monitoring device based on tri-axis accelerometer for elderly is designed. The device collects acceleration and the angle between elderly and horizontal plane of elderly people by MPU6050 tri-axial accelerometer, comparing the acceleration and angle that people and horizontal plane with threshold value to determine whether the old people fell. Delay for a period of time compared the angle and threshold again to judge the elderly in the fall still, finally send text messages to a mobile phone of guardian by GPRS module so that elderly can be helped.

Keywords—Fall monitoring; tri-axis acceleration sensor; threshold; GPRS module

I. INTRODUCE

Fairly rapid development of modern society an aging population is more and more serious. There are many reasons caused the problem such as public health level advances in technology have made extended life expectancy, production and living rhythm is fast, the birth rates is falling caused by life pressure, and so on. The outlook of world population report released by the United Nations learned that the current global showed a trend of aging population. Aging also is not a disease of the wealthy in the developed countries, developing countries also began to appear aging, and one of the most important reasons is low fertility rate overall global. These countries like United States, Japan, Korea, Australia and other western countries have unprecedented pension burden of an aging population. More and more elderly need to live in the compound and there are a lot of family support function weakened. To this problem, this paper designed a fall monitoring device that can be carried easily. This device can collect acceleration signals and angle signals and then through these signals to judge whether the elderly is fall, then the GPRS that controlled by the Microcontroller can send a message to guardian phone. The device can help the old man in the empty nest for rescue when they are in the fall.

Detection of falls using accelerometers and mobile phone technology [1-2] proposed a device that has two components: an intelligent mobile phone and an accelerometer. The accelerometer is responsible for collecting the acceleration signal and smartphone is responsible for judge whether the elderly is fall by acceleration threshold and send a message to guardian, but In daily life has many more vigorous exercise

acceleration will influence the judgment of the threshold algorithm such as rapid squat, so we cannot judge whether the elderly is fall by a single acceleration threshold. Accelerometer Placement for Posture Recognition and Fall Detection [3] put accelerometers in the four parts of body to detect human body falls and distinguish the gesture. However, too many accelerometers in the body can lead to less mobile when the elderly in action. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring [4] is to perform the vast majority of signal processing onboard the wearable unit using embedded intelligence, the system distinguishes between periods of activity and rest, recognizes the postural orientation of the wearer, detects events such as walking and falls, and provides an estimation of metabolic energy expenditure.

Real-Time Fall Detecting System Using a Tri-axial Accelerometer for Home Care [5] proposes a real-time detection system based on home, this system can distinguish between up to four different stumble like forward, backward, turn right and left, and it easy to carry, low cost and high accuracy. But only through accelerometer can't accurate judgment too violent action as quick squat, quick sit, et al. In order to ensure the security of the old people under the condition of not to affect the normal life of elderly, many researchers are working on better fall monitoring devices. But there are lots of problems in the elderly fall monitoring device, we should consider improving all aspects of the problem in safety, quickly, conservation and charges et al. This system adopted a tri-axial acceleration MPU6050 to collect the acceleration signal and angle signal of elderly action, then used the improved algorithm to judge the elderly fall, finally send the message of fall to guardian by GPRS module, so that rescue the elderly in time.

II. FALL DETECTION SYSTEM

A. System Architecture

The elderly fall monitoring system based on tri-axial acceleration sensor uses tri-axis acceleration sensor gather the acceleration and angle signal of the activity of old people and transmission to Microcontroller using threshold algorithm to judge whether the elderly people fall, last using a serial port sent AT command [6] control fall in the GPRS module to send message of old people falling to the guardian for help. System architecture is shown in figure 1:

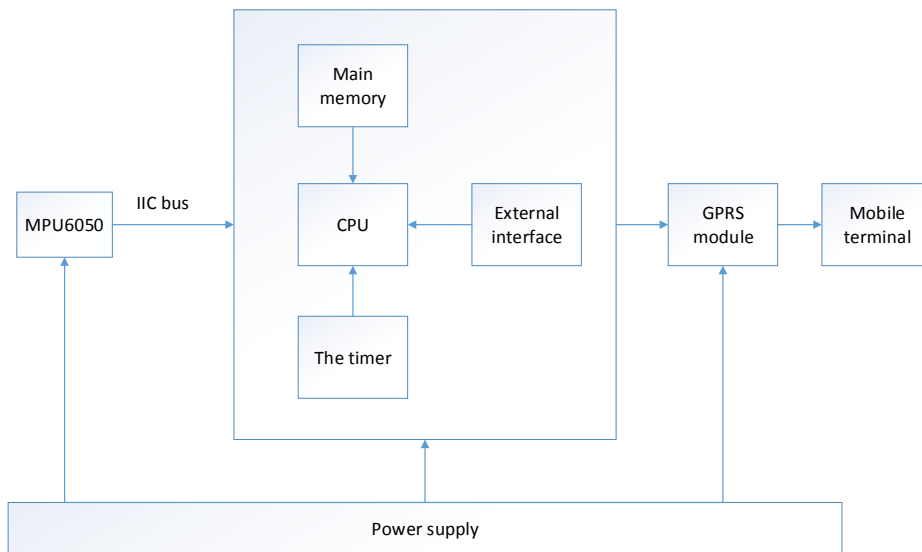


Fig. 1. System architecture



Fig. 2. signal acquisition and processing module

B. The Signal Acquisition And Processing Module

Signal acquisition and processing module of fall monitoring device is composed of three parts, respectively, tri-axis acceleration sensor MPU6050, filter circuit and microcontroller. Signal acquisition and processing module structure diagram as shown in figure 2.

The acceleration signal acquisition module using the tri-axis acceleration sensor MPU6050 collecting activities acceleration of the elderly. Tri-axis acceleration sensor MPU6050 the integration of the three axis accelerometer, gyroscope and extensible digital motion processor [7]. It can collect the acceleration and angular velocity signal in direction of X, Y and Z at the same time. It avoids the acceleration scheduler and the problem of the difference between the timeline of gyroscope, reduces the effect of the preset [8], sensor drift and reduce the complicated data fusion algorithms, motion processing operation of the operating system loads.

The MCU of data processing module is stm32f103rht6, it be used to receive the acceleration and angle signals that be collected by tri-axis acceleration sensor and carry on the judgment, in the final analysis to see if the elderly fall and in the danger. Stm32f103rht6 [9]adopt high-performance ARM Cortex - M3TM32 RISC core that can operate at 72 MHZ frequency, have high speed embedded memory and support three kinds of low power consumption mode. It can be achieve a best balance between low power consumption with short start time.

C. The Wireless Communication Module

The fall monitoring device of elderly people selects and uses the GPRS wireless communication module to send the message of old people falling, for the sake of this information

of elderly falling can be transmitted to guardian accurately and promptly. The GPRS wireless communication module that a mode of packet-switched data load and transmission is developed on the system of GSM [10]. By using the TCP/IP protocol, The GPRS wireless communication module can ensure the data's security and reliability in the transmission process of data. The way of GPRS wireless communication is channel multiplexing that can always keeping online and to avoid the serious consequences that brought by dropping. GPRS communication module is being paid by amount of information, that can be cut down the costs of users who using it.

III. SOFTWARE DESIGN OF FALL MONITORING DEVICE

This system with the aid of VC platform to write programs related that can control the MPU6050 to obtain the activities' acceleration and angle of elderly and estimate the outcome that dealing with the improved threshold algorithm. The fall monitoring device acquisition acceleration and angular velocity signal through the tri-axis acceleration sensor. The angle that between the elderly and horizontal plane can be get by integral the angular velocity. The angle that by transformation to obtain and acceleration are transmitted to MCU to processing and judgement by threshold algorithm. Once the result of judgement is the elderly fall, the MCU controlling the GPRS wireless communication module to send a text message to the guardian to rescue the elderly immediately.

IV. THE FALL MONITORING ALGORITHM

The acceleration can be considered as a three dimensional vector that acceleration signal of the elderly activities is gathered by tri-axis acceleration sensor. It will be said on the space rectangular coordinate system, and X, Y, Z axis

represents the direction of the acceleration. Owing to the body center of gravity direction is unpredictable, and the single direction of acceleration cannot judge whether the body falls, so threshold algorithm is based on the resultant acceleration (SVM) [11] as the basis of judgment. The SVM is expressed as

$$SVM = \sqrt{a_x^2 + a_y^2 + a_z^2}$$

The SVM is acceleration amplitude only associated with the intensity of exercise, has nothing to do with direction. Only according to the acceleration signal cannot accurately judge whether the elderly fall, in order to decrease the misjudgment.

In this paper, combining the acceleration and the angle between human body with horizontal coordinates with up to judge the elderly fall status, Moment of the experiment shows the body fell to the ground impact acceleration reach maximum

SVM, at the moment the angle of human body and horizontal coordinates is greater than 60°. This threshold algorithm will delay 2s after judgment fall again to judgment of Angle, to eliminate the circumstance that after elderly people fell can autonomous action and have no influence for normal activities. Algorithm process as shown in figure 3.

Combined the acceleration threshold judgment with angle threshold of the human body and horizontal coordinates can reduce misjudgment caused quickly squat action and so on. It can cut down the number of alarm for the elderly fall without any hurt and reduce the work burden of guardians that determine the size of angle and threshold again after a 2s delay. By tri-axial acceleration sensor acquisition the status SVM signal of human body such as: static, walking, running, fast squats, bend over and fall as shown in figure 4.

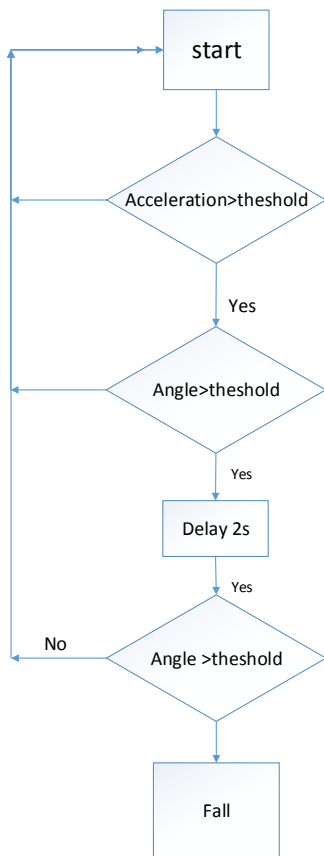


Fig. 3. Algorithm process

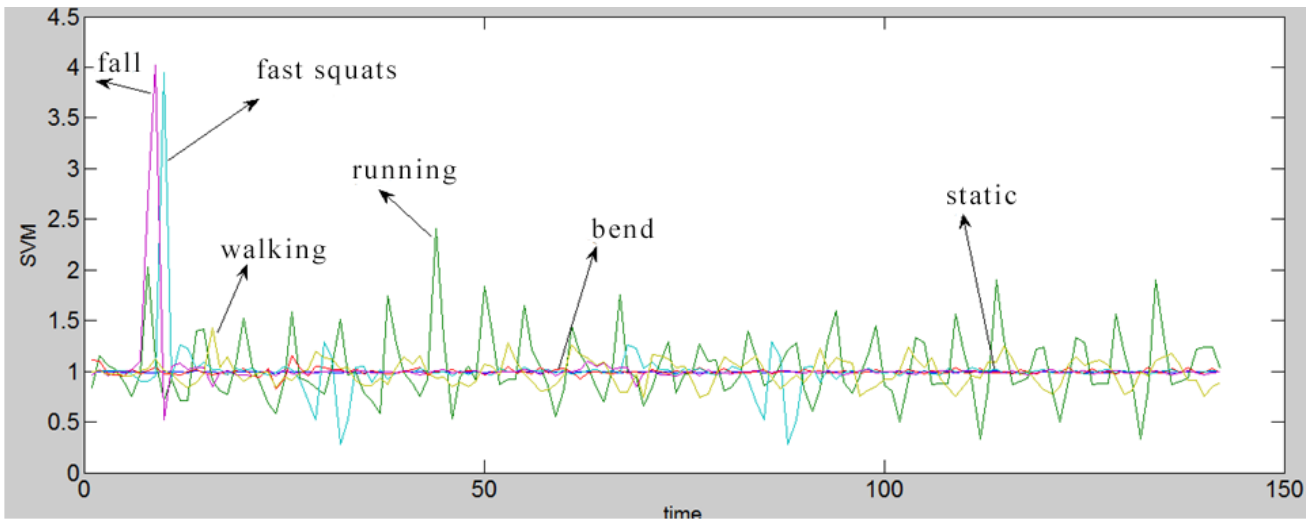


Fig. 4. 4 SVM

From the figure we can see the SVM of human in static is range from 0.98g to 1g. Resultant velocity only affected by the acceleration of gravity, it is consistent with the actual activity. The SVM of walk and bend down have a minor changes are range from 0.5g to 1.5g. Due to when elderly people is running have a certain degree of ups and downs that the action is more dramatic than walking and bend down, so its SVM to a maximum of 2.4g. And fast crouch down and fell the most severe serious in daily actions weightlessness, and the SVM can up to 4g. 3g can be selected as the SVM threshold to judge fall preliminary by decide in the experiment. From the figure 4 shows there are many actions like quickly squat and other will cause misjudgment in acceleration threshold value judgment, so only by judge the SVM cannot be completely accurate to judge whether the elderly fall, this paper puts forward a method that combine to the acceleration threshold algorithm and the angle threshold algorithm of the old people and the horizontal value of X, Y axis to judge whether the elderly fall. The angle that the human body and X, Y axis in horizontal direction of static, running, fast squat, fall, bend shown in figure 5-6.

Figure 5-6 shows the body at stationary state the angle that the human body and the horizontal X, Y direction is not more

than 10, it is a small fluctuations belong to deviation that human caused when he stationary, and it does not affect the judgment falls. In the process of running and fast squat the angle is rang from 15o to 25o and always in fluctuation, it indicating that the human body has a range of shaking in the activity, this is consistent with the normal behavior of the human body. Experiment enumerates the angle that can be get by simulation the human fall forward and fall to right, setting the angle between the body and X axis toward the side down, with the Y axis angle of toward the forward and backward. The figure shows that when the elderly fall its angle between human body and horizontal plane is more than 60o, according this angle to judge if the elderly is fall and he fall forward or toward the side. The algorithm which angle threshold combined with acceleration threshold can estimate whether the elderly fall accurately, ruled out the misjudgment caused by some actions such as rapid squat and other movement acceleration greater than the threshold. The algorithm will judge the size of angle and threshold that human under the current state and X, Y axis to confirm old people fall and lose the ability to act autonomously after a microcontroller judgement determine the elderly is fall.

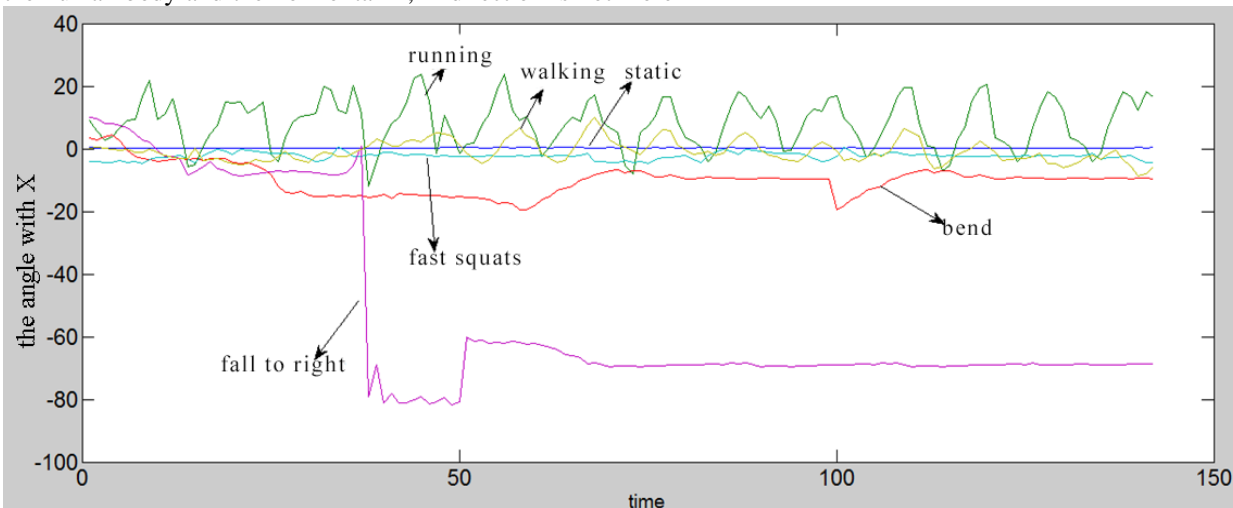


Fig. 5. The angle with X

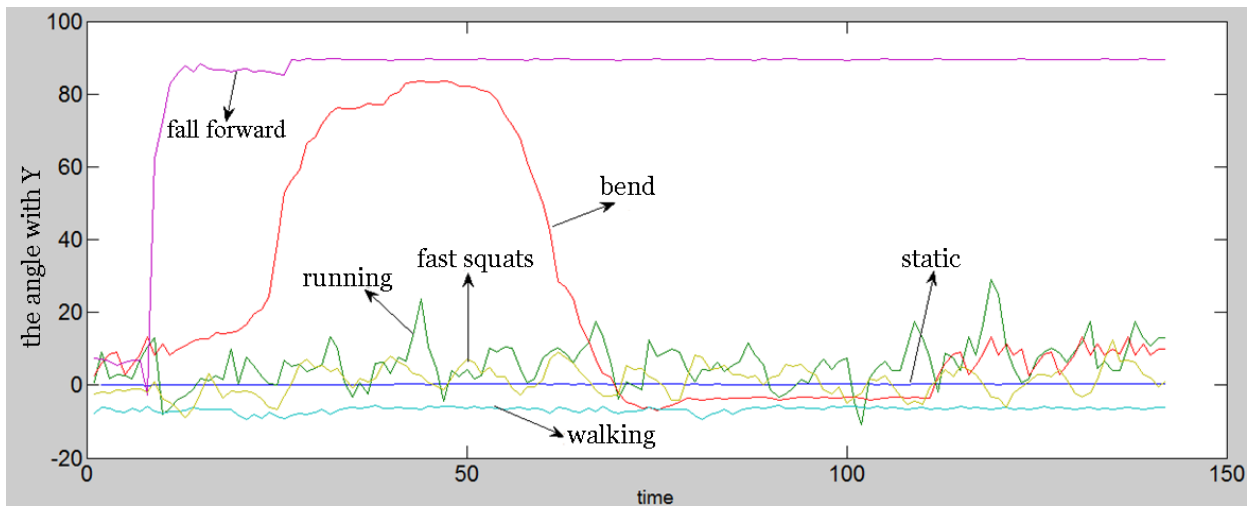


Fig. 6. The angle with Y

V. EXPERIMENTAL RESULTS AND ANALYSIS

Taking into account the physical condition of the elderly, the experiment used a ground of 20 to 25 years old young people to the imitate activities of old people. Experimental determines whether the old people fall and occurs some injuries though the acceleration and angle have been collected that in the actions like walking, running, fast squat, bend down and fall.

In the process of experiment, the young people simulated many activities of elderly, we can see there have some false positives if judged only by acceleration threshold, it cannot identify fast squat and other actions that the SVM greater than threshold whether is fall. The acceleration threshold comparison results are shown in figure 7.

The angle threshold comparison results are shown in figure 8 and 9.

Action	Static	Walk	Run	Fast squat	Bend	Fall forward	Fall to side
SVM>threshold	-	-	-	yes	-	yes	yes
Send message	-	-	-	yes	-	yes	yes
False positives	-	-	-	yes	-	no	no

Fig. 7. The acceleration threshold comparison results

Action	Static	Walk	Run	Fast squat	Bend	Fall forward	Fall to side
angle>threshold	-	-	-	-	-	-	yes
Send message	-	-	-	-	-	-	yes
False positivies	no	no	no	no	no	no	no

Fig. 8. The angle with X axis comparison results

Action	Static	Walk	Run	Fast squat	Bend	Fall forward	Fall to side
angle>threshold	-	-	-	-	yes	yes	-
Send message	-	-	-	-	-	yes	-
False positivies	no	no	no	no	yes	no	no

Fig. 9. The angle with Y axis comparison results

Action	Static	Walk	Run	Fast squat	Bend	Fall forward	Fall to side
angle>threshold	-	-	-	yes	-	yes	yes
Angle with X>threshold	-	-	-	-	-	-	yes
Angle with Y>threshold					yes	yes	-
Send message	-	-	-	-	-	yes	-
False positives	no	no	no	no	no	no	no

Fig. 10. Combine the acceleration and angle comparison results

By the experiment result, acceleration threshold and angle threshold comparison all can produce false positives. This paper combined the acceleration and angle threshold algorithm can distinguish the elderly fall accurately in the fast squat and other actions that acceleration greater than acceleration threshold, and the rate of false positives to 0. The device can comparative the size of angle and threshold again delay 2s after judgement in the state of fall by combined algorithm to analysis whether the old people fall and lose the ability to act independently, according to the specific situation of the elderly micro controller will control the GPRS module send text messages to a guardian mobile phones. After many experiments proved that the device can accurately diagnose fall and normal send text messages to mobile phones.

VI. CONCLUSION

This paper puts forward a fall monitoring devices, using MPU6050 tri-axial acceleration sensor to collect the acceleration and angle of elderly, through the threshold detection algorithm to judge whether the elderly fall and cause certain harm, in a timely manner through GPRS module send text messages to a guardian mobile phone timely treatment to help.

The device has some advantage: it has a small volume and convenient to carry; it has a low cost and can judge the situation of fall timely and effectively. The innovation of this device is use the MPU6050 tri-axial acceleration sensor, the monitoring method within can judge whether the elderly fall down and in danger, finally microcontroller controls the GPRS module send text messages to a guardian mobile phone. The device can judge in the timely and reduced the workload of the guardian and guardian has no necessary to worry about it.

REFERENCES

- [1] Lee, Raymond, Carlisle, Alison. Detection of falls using accelerometers and mobile phone technology[J]. Age and Ageing, 2011, 40 (6): 690—696.
- [2] Jiangpeng Dai, Xiaole Bai, Zhimin Yang; et al. PerFallD: A pervasive fall detection system using mobile phones[C]//8th IEEE International Conference on Pervasive Computing and Communications Workshops, 2010: 292—297.
- [3] Gjoreski, H, Lustrek, M, Gams, M. Accelerometer Placement for Posture Recognition and Fall Detection[C]//Intelligent Environments (IE), 2011 7th International Conference on. IEEE, 2011: 47-54.
- [4] Karantonis, D.M.; Narayanan, M.R; Mathie, M; et al. Implementation of a real-time human movement classifier using a tri-axial accelerometer for ambulatory monitoring [J]. Information Technology in Biomedicine, IEEE Transactions on, 2006, 10(1): 156-167.
- [5] Chien-Cheng Lan, Ya-Hsin Hsueh, Rong-Yuan Hu. Real-time fall detecting system using a tri-axial accelerometer for home care[C]//Biomedical Engineering and Biotechnology (iCBEB), 2012 International Conference on. IEEE, 2012: 1077-1080.
- [6] Zhan-bei Wang. Design of Soil Temperature and Humidity Remote Monitoring System of Tea Garden Based on GPRS [J]. Journal of Agricultural Mechanization Research, 2013, 356(6):973-976.
- [7] Zhijian Yin, Haojie Ning, Inoue Y, et al. A novel wireless motion sensor for analyzing golf swing[C]//SENSORS, 2013 IEEE. IEEE, 2013: 1-4.
- [8] Popelka V. A Self Stabilizing platform[C]//Control Conference (ICCC), 2014 15th International Carpathian. IEEE, 2014: 458-462.
- [9] Ting-zhen Dai, jun-hui Hu. A control system for the ultrasonic 3-DOF micro/nano manipulator[C]//Piezoelectricity, Acoustic Waves, and Device Applications (SPAWDA), 2014 Symposium on. IEEE, 2014: 25-28.
- [10] Rashdi Adnan, Malik Rafia, Rashid Sanam, et al. Remote energy monitoring, profiling and control through GSM network[J]. Arabian Journal for Science and Engineering, 2013, 38(11): 3249-3257.
- [11] Kwolek, Bogdan, Kepski, Michal. Human fall detection on embedded platform using depth maps and wireless accelerometer [J]. Computer Methods And Programs In Biomedicine, 2014, 117(3):489-501.

The Fir Digital Filter Design based on Iwpsso

Xinnan Hu¹

College of Electronic and Electrical Engineering
Shanghai University of Engineering Science Songjiang
District, Shanghai 201620, China

Yujia Wang²

College of Electronic and Electrical Engineering
Shanghai University of Engineering Science Songjiang
District, Shanghai 201620, China

Kun Su³

College of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Songjiang District, Shanghai 201620, China

Abstract—The essence of finite impulse response (FIR) digital filter design is the problem of the parameter optimization. Namely the optimal parameters of FIR digital filter are the core of the design. In due to the traditional design method of FIR digital filter is not only accuracy not high but also sideband frequency is difficult to determine. Improve Weight Particle Swarm Optimization (IWPSO) to design FIR digital filter has less calculation and fast convergence speed. The simulation results also demonstrate that the IWPSO has better approximation properties and band-pass characteristics. what's more, the convergence of IWPSO algorithm made good results in filter design efficiency.

Keywords—FIR filter design; parameter optimization; improve weight particle swarm optimization

I. INTRODUCTION

FIR digital filter can change its amplitude frequency randomly and can guarantee accurate linear phase at the same time, accordingly it has bright research prospect. FIR digital filter is a basic computing unit of digital signal processing[1]and plays an important role in communication field and in the processing of digital signal. The design core of the FIR digital filter[2]centers on the optimization of multidimensional variable[3]. The design method of FIR digital filters are mainly: window function method, Chebyshev and frequency sampling method etc. However, the window function method cannot properly handle transition band. The Frequency Sampling Method results in the fluctuation on the edge of passband and the sampling frequency is restricted to integral number of $2\pi/N$ which cannot be sure the value of the cut-off frequency. N should be taken into consideration if need choose any value of the cut-off frequency, but this increases the amount of calculation. The newly emerged methods such as Genetic Algorithm(GA)[4], neural network method[5] and Particle Swarm Optimization(PSO) algorithm[6], although those methods do have its inspiring effects but still have serious shortcomings such as high complexity and slow convergence rate etc.

For the defects of traditional FIR digital filter, this paper will introduce particle swarm algorithm into the design of FIR digital filter. Then, it will elaborate the process of the FIR digital filter design of using IWPSO. Finally, it draws the conclusion that IWPSO method has better approximation properties, algorithm convergence and optimal value by compare with traditional method and standard PSO algorithm.

The rest of the paper is arranged as follows. In section II, the FIR filter design problem is formulated. In section III briefly discusses on conventional PSO employed for the FIR filter design problems and the proposed IWPSO algorithm. Section IV describes the simulation results obtained by Hamming windows, Frequency sample, PSO and IPSO. Finally, Section V concludes the paper.

II. THE OPTIMAL DESIGN OF FIR DIGITAL FILTER

A FIR digital filter which can be described by difference equation,

$$y(n) = \sum_{k=0}^{N-1} h(k)x(n-k) \quad (1)$$

Where N is the order of the filter. $h(k)$ is the filter's impulse response. The values of $h(k)$ will determine the type of the digital filter i.e. LP, HP, BP, BS etc. FIR digital filter has serious linear sequence property[7], so the unit impulse response has to be odd-symmetry and even-symmetry, namely,

$$y(n) = \pm y(N-1-k) \quad (2)$$

Hypothesis $h(0), h(1), \dots, h(N-1)$ is unit sampling responses of N in FIR digital filter. By Z transform[8] of (1), it can draw the transmission function as below,

$$h(z) = \sum_{k=0}^{N-1} h(k)z^{-k} \quad (3)$$

If $z = e^{j\omega}$, it can draw the frequency response function as follow,

$$H(e^{j\omega}) = \sum_{n=0}^{N-1} h(n)e^{j\omega n} \quad (4)$$

Digital filter can be expressed as,

$$H(e^{j\omega}) = |H(e^{j\omega})| e^{j\theta(\omega)} = H(\omega)e^{j\theta(\omega)} \quad (5)$$

The designed digital filter has linear phase, requiring $\theta(\omega)$ is ω linear function, namely, $\theta(\omega) = -\tau\omega$, including τ is a constant.

According to the requirements of design a filter and the amplitude function $H_i(e^{j\omega})$, suppose the designed amplitude function as $H_d(e^{j\omega})$, so the weighted error[9-10] can be represented as follow,

$$E(\omega) = W(\omega) |H_i(e^{j\omega}) - H_d(e^{j\omega})| \quad (6)$$

Thus:

$$E(\omega) = W(\omega) \left| H_i(e^{j\omega}) - \sum_{n=0}^{N-1} h(n)e^{j\omega n} \right| \quad (7)$$

In the above, $W(\omega)$ is the weighting function used to provide different weights for the approximate errors in different frequency bands, $W(\omega)$ should be maximum in the high precision frequency band. $W(\omega)$ should be minimum in the low precision frequency band. The designing of FIR filter can be supposed as,

$$W(\omega) = \begin{cases} \frac{1}{k}, 0 \leq |\omega| \leq \omega_p, k = \delta_p / \delta_s \\ 0, \omega_s \leq |\omega| \leq \pi \end{cases} \quad (8)$$

In the above, δ_p is pass band ripple peak and δ_s is stop band ripple peak. The error fitness function given in (9) has been considered as fitness function in many literatures [11]. The error to be minimized is defined as:

$$F = \max_{\omega \leq \omega_p} (|E(\omega)| - \sigma_p) + \max_{\omega \geq \omega_s} (|E(\omega)| - \sigma_s) \quad (9)$$

Obviously, the coefficient of the filter gets better along with F becoming smaller. (9) represents the error fitness function to be minimize using IWPSO algorithms. The algorithms try to minimize this error. The core part of designing FIR digital filter is to find the optimal filter coefficient. In order to use IWPSO algorithm to solve $h(k)$, it

is better to encode $h(0), h(1), \dots, h(N-1)$ at first and transform them into particles of the IWPSO algorithm, namely $x_i = h = (h(0), h(1), \dots, h(N-1))$. The particle is distributed in a N dimensional search space.

III. THE OPTIMAL DESIGN OF FIR DIGITAL FILTER BASED ON IWPSO

A. The Basic Particle Swarm Optimization (PSO) Algorithm

Particle swarm based on the swarm intelligence theory is an optimal algorithm developed by Eberhart[12]. The particle traces the two extremes which local optimum and global optimum in each process of iterative search. Then it adjusts its position and velocity to achieve the goal of optimum.

The position and velocity formula of the particle swarm is given as,

$$\begin{cases} v_{id}^{k+1} = \omega v_{id}^k + c_1 r_1 (pbest_i^k - x_{id}^k) + c_2 r_2 (gbest_i^k - x_{id}^k) \\ x_{id}^{k+1} = x_{id}^k + v_{id}^{k+1} \end{cases} \quad (10)$$

In the formula (10), x_k^i is d dimensional velocity of i th in the k th iteration. x_{id}^k is d dimensional position of i th in the k th iteration. ω is the inertia weight coefficient which keeps the particle inertia and enable it can discover the new areas. c_1 and c_2 are acceleration constants which push every particle accelerates to $pBest$ and $gBest$. r_1 and r_2 are the random number between 0 and 1.

B. The Illustration of IWPSO

The paper [13] revised the above formula by introducing inertia weight factor. Then Van den Bergh and Engelbrech[14] pointed out,

$$\omega > \frac{1}{2}(c_1 + c_2) - 1 \quad (11)$$

Assure the particle's convergence. If the condition cannot be satisfied, it may lead to diverging or periodic actions. Contemporarily, researchers often adopt the linear decreasing weight proposed by Ismail A[15].

$$\omega = \omega_{max} - (\omega_{max} - \omega_{min}) * \frac{G}{G_{max}} \quad (12)$$

G_{max} is the maximum evolution algebra. ω_{max} is the maximum inertia weight. ω_{min} is the minimum inertia weight. Where $\omega_{max} = 0.9$, $\omega_{min} = 0.4$. With linear inertia weight decreasing small, A particle swarm optimization algorithm with the Strategy of Nonlinear decreasing inertia weight will converge significantly. The (12) can be improved as following,

$$\omega = \omega_{max} - (\omega_{max} - \omega_{min}) * \left(\frac{2 * G}{G_{max}} - \left(\frac{G}{G_{max}} \right)^2 \right) \quad (13)$$

The steps involved are given as follows.

Step1 Define the solution space, fitness function, inertia weight and population size.

Step2 The initial positions and velocities of all particles are generated randomly within the n-dimensional search space. Initialize $pBest$ and $gBest$ of population size.

Step3 Evaluate the fitness values of particles and store the position of particle having best fitness value as $gBest$.

Step4 Every particle follows the position update formula and velocity update formula (10) to adjust its velocity and position.

Step5 Evaluate the fitness values of updated particles, For newly generated particle $i(i = 1, 2, \dots, M)$, if $Fitness(i)$ better than $fitness(pBest)$, $pBest=i$. If $fitness(i)$ better than $Fitness(gBest)$, $gBest=i$.

Step6 If the iterations achieve to the Maximum number of iterations, the algorithm terminates, $Fitness(gBest)$ is good enough, The $gBest$ is Solution. Otherwise, go to setp2 for a new round of iteration.

IV. SIMULATION AND ANALYSIS

In order to compare the optimal algorithms in terms of the error fitness with traditional design method, it will draw the advantages of the improved one and prove the validity of the IWPSO algorithm. By comparing, it can count the ideal unit impulse response of the low pass filter as following,

$$H_d(e^{j\omega}) = \begin{cases} 1, & 0 \leq \omega \leq 0.3\pi \\ 0, & 0.4\pi \leq \omega \leq \pi \end{cases} \quad (14)$$

TABLE.1. shows the design parameter of all kinds digital filter design methods. In order to demonstrate the effectiveness of the proposed filter design method, several examples of FIR digital filter are constructed using Hamming widows, Frequency Sample, PSO and IWPSO. The MATLAB[16-17] simulation has been performed extensively to realize the FIR LP digital filter of the order 40.

When it comes to the improvement of weight's linear characteristic, it is easy to find that with the increase of iteration, the improved weight decreases rapidly. All these can prove that it can improve significantly the convergence of the algorithm.

Fig.2 shows the design characteristics of all kinds of filter design methods. By the following comparison from Fig.2 we can conclude the buffer zone of the window function decreases smoothly. But if the buffer zone is too wide, the decrease become slow. However, the Frequency Sampling Method leads to the dramatically fluctuating on the edge of pass-band which may influence the pass-band characteristic. When it compares with the traditional method, the improved method has the better pass-band and stop-band, so the inner side of the pass-band is smooth enough to help the signal pass without distortion. What's more, the improved method has a more narrow transition bandwidth, which should be solved quickly in the signal communication system.

IWPSO is an optimal method, by simulation, which can draw the following unit impulse response coefficient of IWPSO and PSO in TABLE 2.

TABLE I. DESIGN PARAMETER

Parameters	Hamming widows	Frequency Sample	PSO	IWPSO
Population size	—	—	40	40
Max Iteration	—	—	1000	1000
c_1, c_2	—	—	1.49445, 1.49445	1.49445, 1.49445
f_p, f_s	—	—	—	0.9, 0.4
$\omega_{max}, \omega_{min}$	0.3, 0.4	0.3, 0.4	—	—
ω	—	—	0.5	—
N order	40	40	—	—

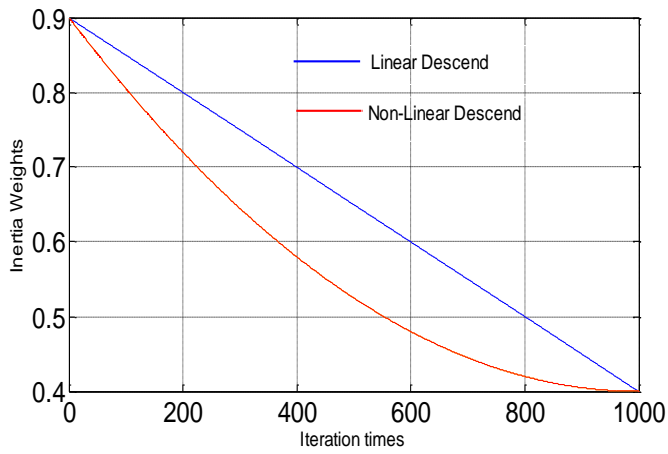


Fig. 1. The comparison of the two linear weight decrease curves

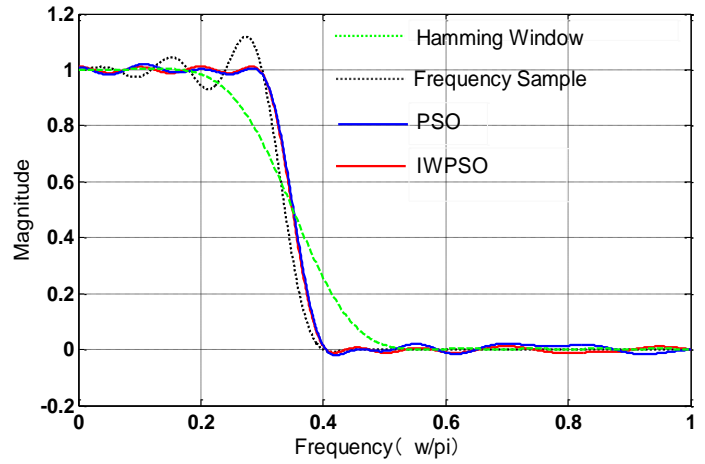


Fig. 2. The simulation result of different methods

TABLE II. OPTIMAL RESULT OF THE 40 ORDER UNIT IMPULSE RESPONSE

h(N)	PSO	IWPSO	h(N)	PSO	IWPSO
h(0)=h(20)	1.01202819879665	1.01993808983423	h(10)= h(30)	-0.0124072369246805	0.00978281196760191
h(1)=h(21)	0.998060873960858	0.984640179011981	h(11)= h(31)	-0.00979470870422273	-0.00196404276610325
h(2)= h(22)	0.991541707693223	1.00380591019032	h(12)= h(32)	2.59553011699883e-05	-0.0197163205167412
h(3)= h(23)	1.01092716101786	1.02127652877248	h(13)= h(33)	0.00598039898460515	0.00358347947328555
h(4)= h(24)	0.986452700263352	0.979822946574317	h(14)= h(34)	0.0136135775471002	0.00219139922982807
h(5)= h(25)	0.985670120151296	0.980068510678159	h(15)= h(35)	0.000925728610695822	-0.020888321837131
h(6)= h(26)	0.0149680758504600	0.0181362870809082	h(16)= h(36)	-0.00943301422868145	0.00151877405571810
h(7)= h(27)	0.00365264987363669	0.0211191618692504	h(17)= h(37)	-0.00960102613115888	0.00989747977171325
h(8)= h(28)	0.00586418920727231	-0.0117393209595076	h(18)= h(38)	0.0125120671876658	0.0113361221972772
h(9)= h(29)	-0.00963133762554668	0.00978281196760191	h(19)= h(39)	-0.00984607345164511	-0.0150980644781148

The Fig.3 and Fig.4 are coefficients optimal process of PSO and IWPSO respectively:

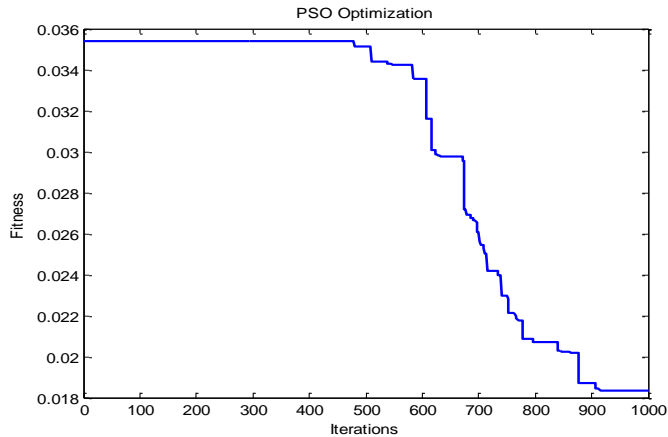


Fig. 3. The change of fitness value of PSO

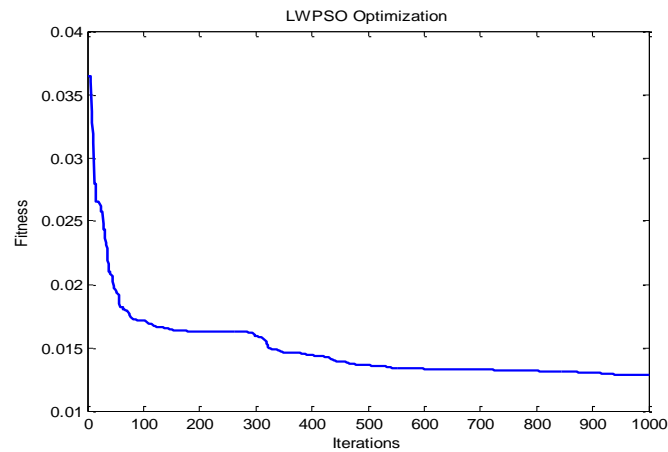


Fig. 4. The change of fitness value of IWPSO

After 1000 iterations, the TABLE.3 shows the simulation results of PSO and IWPSO. Both methods can achieve a good low-band effect. However, at the same coefficient design, IWPSO can get optimal result but use less time and smaller fitness function compared with PSO, which also can get a better optimal coefficient as TABLE. 3.

TABLE III. THE COMPARISON OF SIMULATION RESULTS BETWEEN PSO AND IWPSO

Optimization algorithm	PSO	IWPSO
Optimization of time	16.926600s	16.745021s
Minimum Fitness	2.372231e-02	2.076231e-02

V. CONCLUSION

To make up the deficiencies of traditional design of filter, the paper introduces IWPSO to improve the design method. To prove the validity of IWPSO method, the thesis compares to traditional method, PSO method, and the improved method. The major findings of the thesis are the IWPSO method has better convergence speed, better optimal coefficient, and approximation of the ideal filter. Therefore, the algorithm method has a brilliant application prospect in the field of signal process. Further analyze and improve the weights of PSO algorithm is primary aim in the future. what's more, the improved optimization algorithms applied to the moving average FIR filter, the dressing filter, the L-wave band filter and other modern filter are future research direction.

REFERENCE

- [1] Rao, P. Signals and Systems. McGraw-Hill,2008
- [2] Ababneh,J.I.,Bataineh,M.H., Linear phase FIR filter design using partical swarm optimization and genetic algorithms.Digital Signal Process.2008.
- [3] J. Branke, H. Schmeck, K. Deb, and R. S. Reddy, "Parallelizing multiobjective evolutionary algorithms: Cone separation," in Proc. Congr. Evol.Comput., Portland, OR, 2004, pp. 1952–1957.
- [4] H. J. C. Barbosa and A. M. S. Barreto, "An interactive genetic algorithm with co-evolution of weights for multiobjective problems," in Proc. 2001 Genetic Evol. Comput. Congr., 2001.
- [5] J. Kennedy, Eberhart R. Paticle Swarm Optimization[C]. Proceedings of IEEE International Conference on Neural Networks.1995.
- [6] Y. Shi and R. C. Eberhart,"Fuzzy adaptive particle swarm optimization,"in Proc. IEEE Congr. Evol. Comput.[A]. vol. 1. May 2001.
- [7] Oppenheim,A.,andSchafer,R.Discrete-Time Signal Processing,3rd edition.Prentice Hall,2011.
- [8] Yanfen Wang, Gang Wang, Xiaoguang Zhang. Digital signal processing and implementation[M]. Tsinghua University press.
- [9] Li K, Liu Y. "The FIR windows function design based on evolutionary algorithm" [C]. International conference on Mechatronic Science, Electric Engineering and computer (MEC),IEEE.2011.
- [10] KONAKAHARA T, SUYAMA K. Design of CSD Coefficients FIR Filters using PSO with Perturbation[J]. Technical Report of Ieice Cst, 2010, 110.
- [11] Ababneh,J.I., Bataineh, M.H.,Linear phase FIR filter design using particle swarm optimization and genetic algorithms.Digital Signal process.2008
- [12] Eberhart,R.,Shi,Y.,Comparison between genetic algorithms and particle swarm optimization.In: Proc.7th Ann.Conf.Evolutionary Computation,San Diefo.2000
- [13] Shi Y, Eberhart R.C., Empirical study of particle swarm optimization[C]//Proc.IEEE Congr.Comput.Intell.1998.
- [14] Van den Bergh F,Engelbrecht A P.A study of Particle swarm optimization particle trajectories[J].Information sciences,2006.
- [15] Ismail A, Engelbrecht A, The Self-adptive Comprehensive Learning Particle swarm Opyimizer[J]. Swarm Intelligence(Lecture Note in Computer Science),2012
- [16] FRED J. Taylor Digital Filters Principles and Applications with MATLAB[M]. Beijing: National Defense Industry Press.2013.
- [17] Ingle,V., and Proakis, J. Digital Signal Processing Using MALTLAB.CengageLearning.2007.

Analysis and Research of Communication Interrupt Fault for Shanghai Metro Data Transmission System

Jianru Liang

College of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

Xinyuan Lu

Communication and Signal Branch
Shanghai Rail Transit Maintenance Support Co Ltd.,
Shanghai, China

Cong Shi

Communication and Signal Branch
Shanghai Rail Transit Maintenance Support Co Ltd.,
Shanghai, China

Minglai Yang*

Shanghai Advanced Research Institute
Chinese Academy of Sciences
Shanghai, China

Abstract—A line of Shanghai metro has been put into use for nearly fifteen years. There are three times extended during this time. The existing line's data transmission system was modified over the last decades and has adopted many kinds of data transmission technology. By the analysis and research of communication interrupt of certain line's data transmission system, which usually occurred in some site, the maintainers can find various security hidden danger in time and take corresponding measures, which has improved the quality of metro operation.

Keywords—Data Transmission System (DTS); L3 Ethernet switch; Terminal Server; interlock

I. INTRODUCTION

One of Shanghai metro line has adopted a new data transmission technology, which provide a high speed, large capacity optical fiber transmission channel. Therefore, the data transmission system of shanghai metro has become more safe, reliable and fast. Recently, multi-station's signal system exists communication interrupt. Through the trouble analysis and research of this phenomena, the worker of metro maintenance can find various security hidden danger in time, and take corresponding corrective measures.

II. DTS SYSTEM NETWORK TOPOLOGY

Network topology refers to the transmission media interconnect devices of all kinds of physical layout, especially the position of the distribution of computer and the cross style of cable. There are many kinds of Network topology. Such as star, bus, ring, tree, distributed, network and cellular topology structure, etc. One line of Shanghai metro adopts ring topology, all nodes connect by end to end to form a closed ring communication lines. Information can be transmitted in one-way or two-way. Ring network structure has the following advantages: 1) when we increase or decrease the workstation, it only takes a simple connection operation; 2) it can use optical fiber to improve transmission distance; 3) once a node failure,

it can automatic bypass with high reliability. But the ring network structure also has some disadvantages: 1) any node failure in a ring network will cause the whole network paralysis; 2) it is difficult to detect trouble, because this connection does not centralized control, when trouble occurs ,it is essential to examine each node on the network, which is very difficult; 3) since the information is serial communication, too much nodes in the ring network will affect the transmission efficiency and make the corresponding time become longer. Due to the disadvantages of above, one line of Shanghai metro adopts double channel redundancy style, namely double loop network topology.

The double channel redundancy structure has the characteristics of continuity, real-time, limitless and quickness. Based on the double channel redundancy type, double loop network topology process switch automatically by the hardware, the fault detection and switch work are completed mainly by bottom layer, thus improving the process efficiency, reducing the fault recovery time of network communication, meet the real-time requirements.

One line of Shanghai metro's DTS (Data Transmission System) network topology structure as shown in Fig. 1. All loop is reverse two fiber optic fiber ring, which can provide redundancy. When one set of hardware occurs failure, the system can still continue to work. The loop transmit data through the redundant Ethernet switches and terminal server. Ethernet packets are processed by central control room and each signal equipment room through optical fiber industrial Ethernet switches, serial data through redundant serial and Ethernet device server for data processing.

In one line of Shanghai metro, the central server and each site combine together through the optical fiber to form a ring, the whole ring have S1, S2, two channels, they are redundant. each site with two three layer of Ethernet switches as relay nodes. In each site, man-machine dialogue workstation, terminal server and shielding door AP are unified management

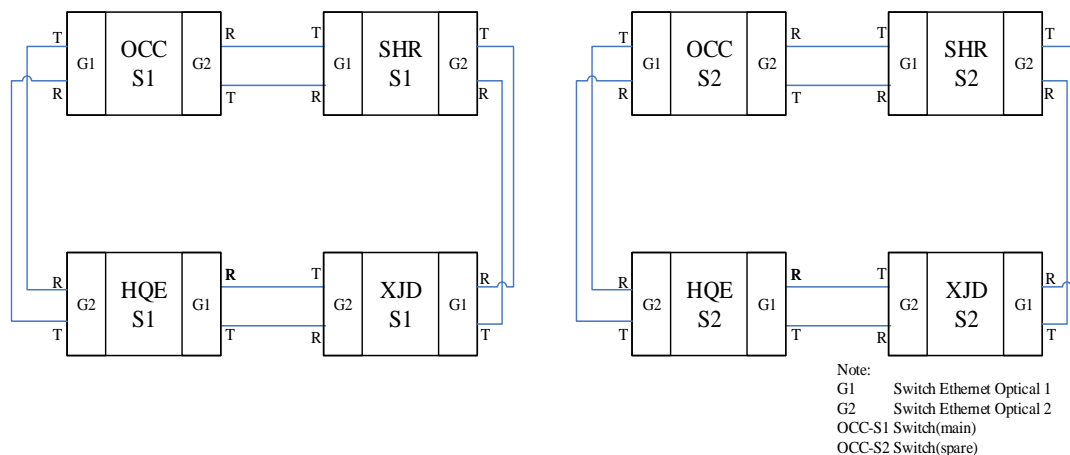


Fig. 1. The network topology of a Shanghai Metro line’s Data Transmission System

by the switch. RS910 terminal server is an industrial Ethernet switch, which has two serial port and two fast Ethernet port, meanwhile, has the function of limiting the port rate and inhibiting the broadcast storm. Due to the fault points usually occur in the terminal server, so the following content give a brief introduction about the work conditions and functions of the terminal server—RS910. The Fig. 2 is the network connection diagram of RS910 as a terminal server, the two terminal server connect A、B two switches separately, its main function is used for interlocking data transmission between the adjacent stations and make decision for redundancy (MI - terminal server - switches). The Third terminal server connect A、B two net segments at the same time, it is mainly used for site shielding door linkage control (MI - terminal server - switches - wayside AP - train MR), known as the PSD door terminal server.

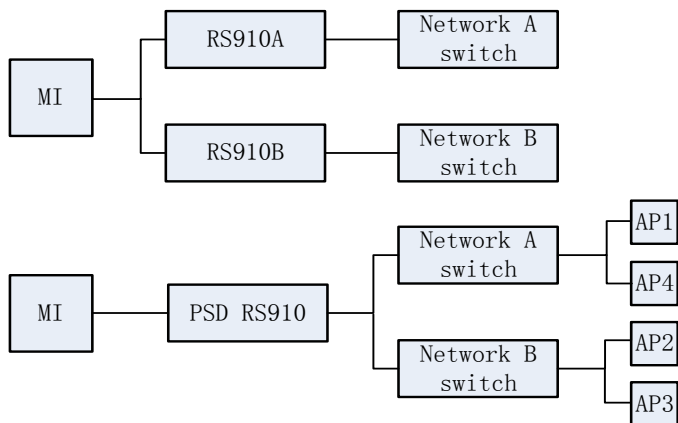


Fig. 2. Terminal Server’s Network connection diagram

III. FAULT PHENOMENON AND ITS INFLUENCE

Recently, DTS of Shanghai metro line appears multi-site communication interrupt trouble, which has led to the failure of interlock. Large interlock control area jumped to red tape, the station ATS and the center ATS show inconsistent, leading to approach cannot be arranged, the train stopped and affect the normal operation of the subway. As shown in the TABLE I.

IV. FAILURE ANALYSIS AND PROCESSING

Within a few minutes, the switches of A, B network segment downtime simultaneously at several stations of a subway line in December 9, 2013 and the 11th, the fundamental reason is the presence of large amounts of data network in a short time. The switching equipment utilization reached 100% by SNMP of the network monitoring station. In addition, the maintenance personnel found all the ports of switches including control ports of console equipment are not loading when they inspecting the switches in 11th December 9, 2013.

TABLE I. LIST OF COMMUNICATION INTERRUPT FAULTS IN DTS SYSTEM

Trouble phenomena	Trouble time	Trouble station	Reason analyze
Interlock invalidation	5:38AM 9th December,2014	H	Interlock communication channel interrupt
	10:25AM 11th December,2014	E	
Compel train stop	5:38AM 9th December,2014	I down line	Interlock loss communication and cut centralize control line
	10:25AM 11th December,2014	I up line	
Large interlock control area jumped to red tape and ATS shows off-line	10:25AM 11th December,2014	E	Interlock failure
	5:38AM 9th December,2014	E	
the station ATS and the center ATS show inconsistent	5:38AM 9th December,2014	D、I、J 和 K	Switch died and led to communication channel appear network storm
	8:12AM 9th December,2014	K	
approach cannot be arranged	10:25AM 11th December,2014	E	Interlock failure and make enter road cannot open
	10:31AM 11th December,2014	C down line	

The main reasons for switch downtime: 1) interlocking with RS910 terminal server firmware BUG resulting in serial does not work; 2) not enabled the broadcast suppression function of switch. Detailed analysis is as follows:

A. Interlocking with RS910 terminal server firmware bug resulting in serial does not work

The initial performance of these two failures are communication failure in the west of the K-station, after a few hours, several switches have downtime. In the 9th, for example, communication failure occurs in the K-station on the west side at 2:00 more, and E-station's switches occur downtime at 7:00. The reason for this is due to the presence of a terminal server bug of interlock system, and its availability decreased after a long working hours, which performance is the serial communication error. What's worse, it cause the release of large of data to the network in a short time, which result in utilization of switch very high and switch downtime at last. There is a surround bug of TCP sequence number in the RS910 terminal server. The sequence number account 4 bytes which range is $[0, 2^{32}-1]$, a total of 2^{32} (i.e 4 284 967 296). After increase to $2^{32}-1$, the next sequence number back to zero. That is to say, the sequence number use mod 2^3 algorithm. If the sequence number of connected TCP occur surround (final confirmation sequence number is actually a big unsigned figure, and the next sequence number is a small unsigned figure), the receive window will be permanently closed. This problem can be treated as a "Serial Lock". Even though the data cannot be transmitted to the serial port, the data received from the serial port can still be sent to the remote end.

RS910 Terminal Server the most basic role is the conversion of protocol and physical-line, shown in Fig. 3. The communication port of terminal server connect with computer interlocking MI is called serial port, and the communication port of terminal server connect with MOXA Layer 3 switch is called Ethernet port. It uses TCP transport protocol because of considering the protocol conversion between RS910 terminal server's serial port and Ethernet port. TCP transport protocol need to go through a three-step handshake confirmation before establish communication. The sequence number wraparound bug is the handshake signal generating TCP protocol sequence numbers wrap before establish TCP protocol data transmission, which will cause RS910 Terminal Server normally forward data received from local interlock to the remote end, and the remote end cannot transfer interlock communication data to interlock MI. Therefore, wraparound bug of RS910 terminal server firmware will cause the serial does not work.

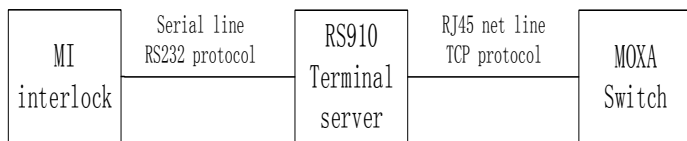


Fig. 3. Protocol conversion diagram of RS910 terminal server

B. Not enabled the broadcast suppression function of switch

During the 9th and 11th twice fails, broadcast storm protection option of the switches are disable, the switch does not use its own broadcast storm suppression.

In computer networks, data link layer and network layer use broadcasting technology, the former transmits the broadcast information to a plurality of physical devices, the latter transmits the broadcast information to a plurality of logical devices. There are two TCP / IP protocol broadcast:

- 1) full broadcast which send broadcast data to each host;
- 2) part broadcast which broadcast information to a particular broadcast network or subnet;

It may occur "broadcast storm" among multiple forwarding device during the broadcast process: a lot of broadcast data packet repeat transmission, degrade network performance and exhaust network bandwidth, which bring down the network.

The broadcast storm control principle is to allow the port to filter the broadcast storm appeared on the network. After broadcast storm control function is turned on, the port discard received broadcast frames automatically when broadcast frames port received add up to a predetermined threshold value. When this function is disabled or broadcast frame is not cumulative to the threshold, the broadcast frame is normally broadcast to other ports of the switch.

As show in Fig. 4, assume the host A send broadcast packet constantly to the outside until it paralysis. The broadcast packets send by host A need to be forwarded by switch. At this point, the switch will drop the broadcast packets (eg DHCP, ARP, RARP, NETBIOS, RIP, etc.) traffic of host A, so that other mainframe data traffic can be forwarded normally,so as to prevent broadcast packets sent by the host A effect other devices of broadcast domain.

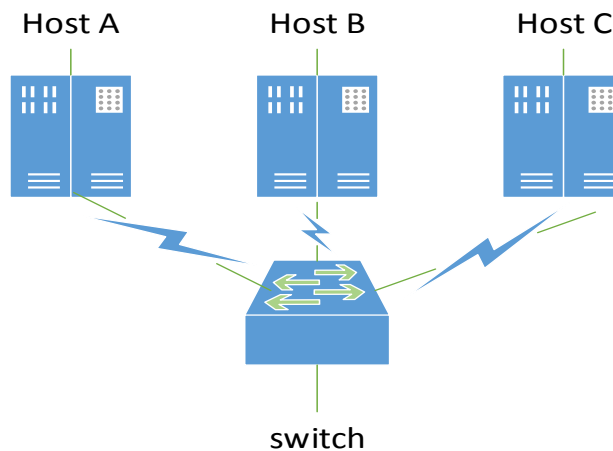


Fig. 4. The diagram of Broadcast Storm

If you do not suppress broadcast storms, the broadcast packets will be flooded to all tunnels. At the same time, a large number of broadcast reply will propagate in VLAN, causing broadcast storms.

Most switches support broadcast storm control at present, and broadcast packets of each port can be remain below a specific ratio which configure this feature in the future, so you can reserve bandwidth to be used, thereby inhibiting broadcast storms. pt-7828 switch have broadcast storm control function, as shown in Fig. 5.



Fig. 5. Broadcast storm protection

V. REFORM PROGRAM AND RESULTS

For RUGGEDCOM RS 910 TCP sequence number wraparound bug: it is necessary to upgrade software firmware of RS 910 terminal server.

Broadcast storm protection options of all switches is enable, so that the switch can automatically discard and protect operational security system when ports have a particularly large broadcast data.

For some stations, the adjacent station transmit interconnection information by try using a photoelectric converter iConverter to replace RS910 terminal servers and switches. The network connection of terminal Server contrast with iConverter photoelectric converter is shown in Fig. 6.

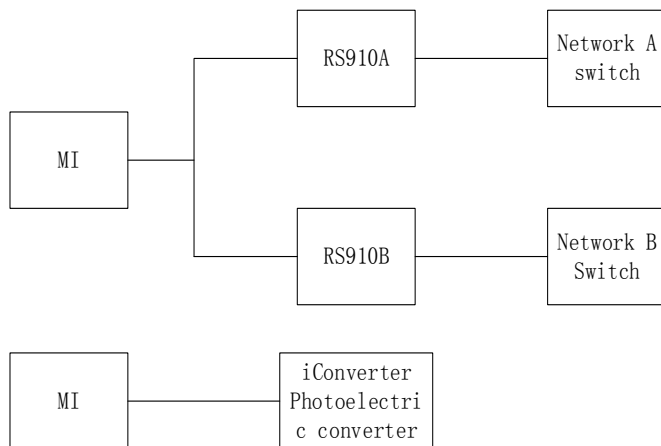


Fig. 6. Comparison diagram of network connection between terminal server and iConverter photoelectric converter

The main function of the communications device is to replace serial communication device, so that the adjacent station transmit interconnection information.

The iConverter photoelectric converter is mainly composed of three parts: 5-Module Chassis cage, RS485/422 cards and NMM2 photoelectric conversion modules. The cage built master / slave of two power modules, RS485/422 cards and NMM2 photovoltaic modules were installed in slot1 and slot2 ~ 5.

Take B station (middle) for example, slot 3 ~ 2 replace slot 5 ~ 4 of A station (west); slot 5 ~ 4 replace slot 3 ~ 2 of C station (east). Slot 1 is NMM2 module which connected to M1-P4 port of DTS switches.

Using iConverter photoelectric sensors replace RS910 terminal servers and switches, which avoid the software protocol conversion of a terminal server firmware effectively.

Since the end of December 2013, the metro took over the rectification program and DTS broadcast storm system has been effectively controlled.

VI. EPILOGUE

DTS system of shanghai metro a Line has experienced three extension and transformed DTS data transmission system of existing line and used a variety of new data transmission technology. The use of new equipment also raises some new questions: sequence number bug of terminal server firmware cause the serial does not work which result in broadcast storms, switches, and a series of chain reaction crash, leading to a neighboring station interlocking communication interrupted, or even the whole ring paralyzed and cause some impact on subway normal operation. To analyze the malfunction and develop appropriate corrective solution which can reduce the probability of occurrence of similar failures and ensure the normal operation of the subway. It can provide some maintenance experience for other urban rail transit maintenance unit. The on-site maintenance deal with the malfunction quickly when there is the same or similar failure and reduce the malfunction impact of the operating to a minimum.

ACKNOWLEDGMENT

JS-KY12R014-1: The development of portable train to wayside communications detector and monitoring system of Line 2 of Shanghai metro

REFERENCES

- [1] Wen Shao, Zhonghua Hao. Broadcast storm analysis and prevention of HFC bidirectional network [J]. Cable Television Technology, 2011(3) : 94-95.
- [2] Laiqiang Zhu, Xinzhi Chen, Zhiyong Gao. The broadcast storm prevention of digital substation [J]. Electric Power Automation Equipment, 2010, 30(11):98-101.
- [3] Yanhong Xue. Has the Broadcast Storm Reason and the Case Analysis [J]. Computer Knowledge and Technology, 2007, 1(3) : 724-725.
- [4] Wisitpongchan N, Tonguz O K, Parikh J S, etal. Broadcast storm mitigation techniques in vehicular ad hoc networks [J].IEEE Wireless Communications, 2007, 14(6) : 84-94.
- [5] Chunjun Wang. Design and Implementation of Redundant Dual-network Systems [D]. Dalian : Dalian University of Technology, 2008.
- [6] Hong Sun, Yang Chen. Analysis of computer network topology [J]. Economic and technical cooperation Information, 2013(29) : 88.
- [7] Yajun Wu. Computer network topology analysis [J]. Software Guide, 2011, 10(12) : 113-115.
- [8] Lianchun Zhuo. TCP / IP network broadcast storm control technology [J]. Applications of The Computer Systems, 2000(12):31-33.
- [9] Yongjin zou, Lian zhang. The use of broadcast stormto suppressIP MAN of switch [J]. Network administrators World, 2010(13) : 37-39.
- [10] Xiren Xie. Computer Network [M]. Beijing : Publishing House of Electronics Industry, 2007.
- [11] Yongcheng Lei. Broadcast storm suppression method and system : China, 103209092 A[P]. 2013-07-17.
- [12] Xu Li, Yuanquan Pang, Yunqi Wang, etal. A method and apparatus to suppress broadcast storm : China, 102196365 A[P]. 2011-09-21.
- [13] Qiang Wang, Jiang Zheng, Yazhou Liu. A method and apparatus to suppress broadcast storm of SDH services : China, 103368839 A[P]. 2013-10-23.

A Fuzzy PI Speed Controller based on Feedback Compensation Strategy for PMSM

Ou Sheng¹, Liu Haishan², Liu Guoying, Zeng Guohui, Zhan Xing, Wang Qingzhen
School of Electronic and Electrical Engineering
Shanghai university of engineering science
Shanghai, China

Liu Haishan²
School of Electronic and Electrical Engineering
Shanghai university of engineering science
Shanghai, China

Abstract—in order to solve the problem of robustness or anti-disturbance of the traditional PI speed controller in the permanent magnet synchronous motor. A fuzzy PI speed controller based on load torque feedback compensation is proposed for the permanent magnet synchronous motor. The combination of fuzzy PI control strategy and load feedback compensation method can enhance the robustness and disturbance rejection of the speed loop. According to the validated results of simulation and experiments, by using this PMSM speed controller, the robustness of the system speed control was enhanced markedly, and the capacity of anti-disturbance was also improved significantly.

Keywords—permanent magnet synchronous motor(PMSM); Fuzzy PI; torque feedback compensation; anti-disturbance

I. INTRODUCTION

In recent years, with the rapid development of high-performance permanent magnet materials technology, power electronics, microelectronics technology, vector control theory and automatic control theory, the permanent magnet synchronous motor control system has been developing rapidly. Due to its superior speed performance, it has overcome a series of restrictions DC servo motor mechanical commutator and some problems for the brushes. There are many advantages such as simple structure, reliable operation; and light weight, small size, high efficiency, small moment of inertia, overload strong, and power factor and so on, so it widely used in the field of high-precision, high-performance servo control, robotics.[1-2]

As the motor running in the actual process, there will be some changes in inertia or load changes cases, which likely to affect the system control performance. The servo control itself requires no output overshoot and quickly track the input command, which can hold the state steady and no static error at the same time. Therefore, the motor system needs to have a relatively strong robustness and disturbance rejection for parameter changes. To solve this problem, many researchers have proposed different control schemes, there are adaptive control [3-5], fuzzy control [6-7] and the sliding mode control scheme [8-9] to design a speed controller of the permanent magnet synchronous motors. Also, a variety of speed, torque observer and inertia identification method is also widely used in these schemes. The literature [6-7] analyzed the fuzzy PI speed controller in the permanent magnet synchronous motor

and demonstrates it can enhance the robustness of speed control.

In this paper, a fuzzy PI speed controller which is based on the load torque feedback compensation scheme is presented. The fuzzy PI control enables online self-tuning, so that the same conditions, the system more robust. Meanwhile fuzzy PI speed controller and PI speed controller have a same transfer function for speed changes caused by the disturbance, so they have the same anti-disturbance capability. Therefore the load torque feedback compensation strategy can adaptively compensate for the speed disturbance and increase the speed loop the disturbance rejection.

II. MATHEMATICAL MODELLING OF PMSM

Assuming PMSM stator winding is sinusoidal waveform of electromotive force, the gap is a positive sinusoidal magnetic field, ignoring the core saturation effects, excluding the eddy current and hysteresis losses, and while the permanent magnet and rotor without no damping effect, in the d-q coordinate system PMSM voltage equation is:

$$\begin{cases} u_d = L_d di_d/dt - \omega_e L_q i_q + R_s i_d \\ u_q = L_q di_q/dt + \omega_e L_d i_d + \omega_e \psi_f + R_s i_q \end{cases} \quad (1)$$

Flux equation:

$$\begin{cases} \psi_d = L_d i_d + \psi_f \\ \psi_q = L_q i_q \end{cases} \quad (2)$$

Mechanical equation:

$$J d\omega/dt = T_e - B\omega - T_L \quad (3)$$

Electromagnetic torque equation:

$$T_e = 3P\psi_f i_q / 2 = K_T i_q \quad (4)$$

where: u_d 、 u_q 、 ψ_d 、 ψ_q 、 i_d 、 i_q are voltage, flux and current at d-q axis. ω_e 、 R_s 、 ψ_f 、 ω 、 J 、 T_e 、 B 、 T_L 、 P are Electrical angular speed, the stator resistance,

magnetic flux, moment of inertia, the electromagnetic torque, friction coefficient, the load torque and the number of pole pairs.

III. FUZZY PI SPEED CONTROLLER BASED ON LOAD TORQUE FEEDBACK COMPENSATION STRATEGY

The literature [4] showed that it can improve system speed disturbance rejection control by the load torque compensation, reduce the oscillation amplitude perturbations for speed. The literature [7] showed that the fuzzy PI speed controller can effectively enhance the robustness of the speed control, such as the problem of motor just started speed overshoot in the case can be reduced. In this paper, it combine the fuzzy PI and

feedback compensation control for load torque, make it have a certain anti-disturbance capability, and enhance the robustness of the system speed control.

A. Feedback Compensation Control Strategy

In actual conditions, the disturbance reflected some load torque fluctuations. Therefore, the observed torque value use for feed forward compensation, in order to reduce volatility, thereby improving disturbance rejection. Block diagram of the control strategy as shown in Fig.1, using the load torque feedback compensation strategy to enhance the anti-disturbance capacity of the system speed loop. The control system block diagram as shown in Fig.2.

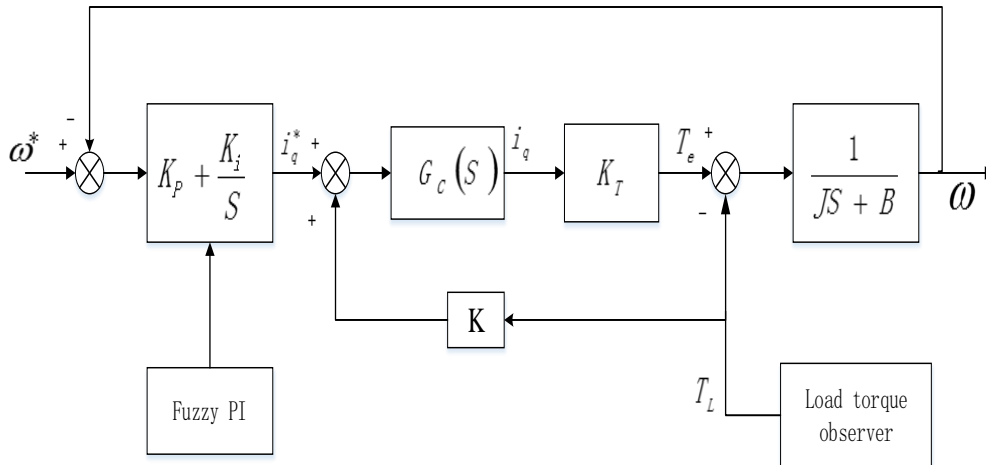


Fig. 1. Fuzzy PI speed controller with load torque feedback compensation

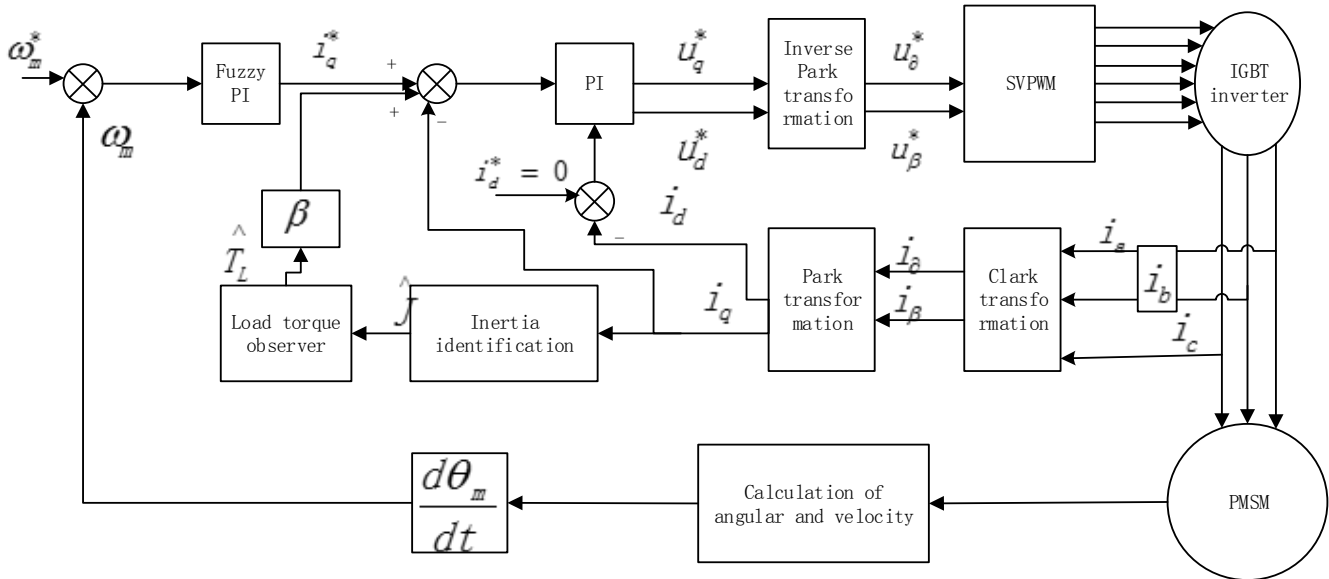


Fig. 2. Fuzzy PI control system based on feedback compensation

B. Load Reduced Order Observer

In this paper, a load reduced order observer is used [10], it can real time observe load torque changes to achieve feedback compensation strategies.

Since the change time of the load torque is much less than the sampling frequency of the system controller, it is generally believed the load torque does not substantially change in the control cycle, namely:

$$dT_L/dt = 0 \tag{5}$$

Then according to equation (3) and (6), the state equation can be obtained:

$$\begin{cases} dx/dt = Ax + Bu \\ y = Cx \end{cases} \tag{6}$$

There: $A = \begin{bmatrix} -B/J & -1/J \\ 0 & 0 \end{bmatrix}$; $B = \begin{bmatrix} 1/J \\ 0 \end{bmatrix}$; $X = \begin{bmatrix} \omega \\ T_L \end{bmatrix}$;
 $C = [1 \ 0]$; $u = T_e$; $y = \omega_e$.

That system input variables for the motor electromagnetic torque, the state variables for the load torque and mechanical acceleration, the output variable is the mechanical angular velocity. Reduced-order state observer model, the following equation (7):

$$\begin{cases} d\hat{x}/dt = A\hat{x} + Bu + K(y - \hat{y}) \\ \hat{y} = C\hat{x} \end{cases} \tag{7}$$

There: $\hat{x} = \begin{bmatrix} \hat{\omega} & \hat{T}_L \end{bmatrix}^T$ for the estimated state variables;

$K = [k_1 \ k_2]^T$ for State feedback gain matrix.

That can be obtained from the above formulas, equations of state motor:

$$\begin{cases} d\hat{\omega}/dt = (-B/J)\hat{\omega} + T_e/J - T_L/J + k_1(\omega - \hat{\omega}) \\ d\hat{T}_L/dt = k_2(\omega - \hat{\omega}) \end{cases} \tag{8}$$

That is according to formula (8), designing a load torque observer, the schematic diagram of the load torque observer as shown in Fig. 3.

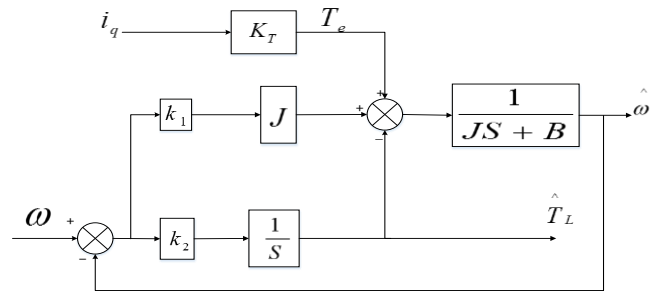


Fig. 3. The load torque observer

C. Fuzzy PI Speed Controller Analysis and Design

Speed control system block diagram as shown in Fig.4:

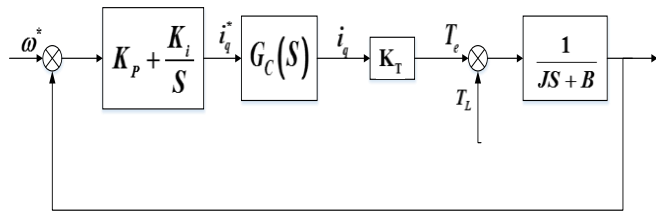


Fig. 4. Speed control system block diagram (PI controllers)

Transfer function of PI speed controller $G_S(S)$ are

$$G_S(S) = K_p + \frac{K_i}{S} \tag{9}$$

There K_p 、 K_i are proportional gain and integral gain coefficient.

Traditional PI control parameters of the speed controller is typically pre-set and remain unchanged during operation, and the fuzzy PI control is that two output of fuzzy controller are as the proportional, integral coefficient, adjust the controller parameters in real time according to the working conditions to achieve better control, control block diagram as shown in Fig.5.

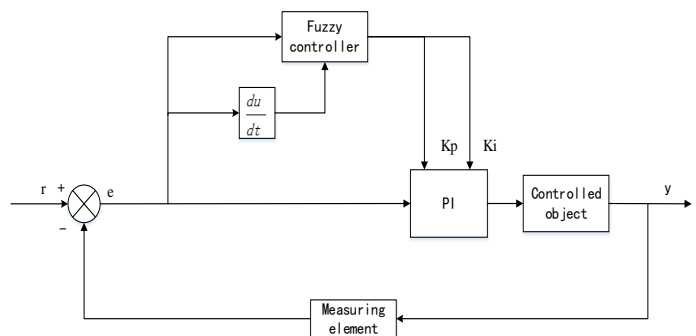


Fig. 5. Fuzzy PI controller

In this paper, fuzzy PI speed controller for PMSM speed deviation e and error change ec , adjusted in accordance with the fuzzy control rules, the amount of blur after processing the output K_p, K_i , in order to achieve real-time online PI parameter adjustment.

Fuzzy control core design is based on engineering and technical personnel technical knowledge and practical experience, establish appropriate fuzzy rule table. According to the fuzzy control table, K_p, K_i parameters were tuning. $e, ec = \{-3, -2, -1, 0, 1, 2, 3\}$, Its fuzzy subset, $e, ec = \{NB, NM, NS, Z0, PS, PM, PB\}$, where e, ec, K_p, K_i are normally distributed, Whereby each fuzzy subset membership is obtained, And in accordance with the membership table of each fuzzy subset and fuzzy control model of the parameters, using fuzzy synthesis to design fuzzy matrix of PI parameters, which make control system complete PI parameters tuning online by being analysis, looking up and operating the result of fuzzy logic rules.

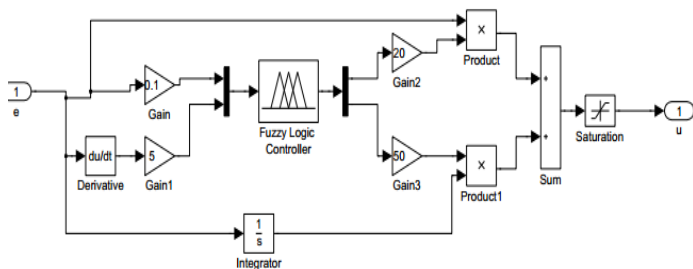


Fig. 6. Fuzzy PI speed controller emulation module

IV. SIMULATION AND EXPERIMENTAL RESULTS

A. Simulation Results

To verify the performance of fuzzy control PI speed controller for PMSM, using Matlab / Simulink control system to achieve a digital simulation, Fuzzy PI speed controller based on load torque feedback compensation for PMSM control system as shown in Fig.7.

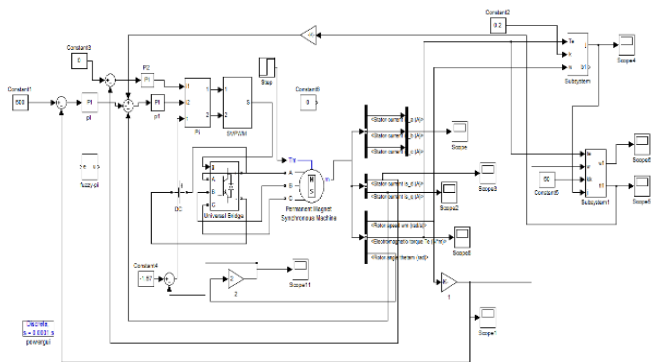


Fig. 7. PMSM simulation system diagram

PMSM parameters are as follows: 1.1kw, 220v, $L_d=L_q=8.5$ mH, $R=2.875$ Ω , $\psi=0.175$ Wb, $P=4$. The motor first run at 500r/min, at $t=0.1s$, load torque suddenly add $10N \cdot m$ to $20N \cdot m$, Speed estimation simulation waveforms as shown in Fig.8. As shown speed waveforms in Fig.8, Speed

estimates can be a good track real speed. The jitter of load torque feedback compensation waveform than uncompensated waveform's is smaller, which indicates that the system have better anti-disturbance performance.

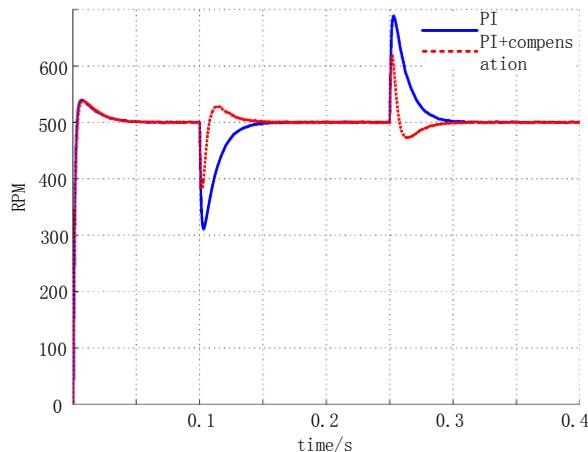


Fig. 8. Sudden increase in load speed estimation simulation waveforms

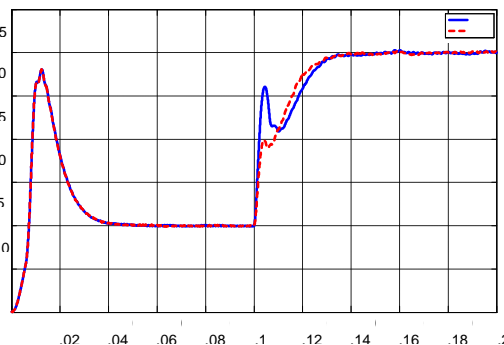


Fig. 9. Sudden increase in load torque estimation simulation waveforms

During the time $[0-0.1s]$ load torque is $10N \cdot m$, during the time $[01-0.2s]$ load torque is $30N \cdot m$. Speed fluctuation of fuzzy PI speed controller is small. When the disturbances occur, the oscillation amplitude is less than PI speed controller, which verify the fuzzy PI speed controller have more robustness than traditional PI controller.

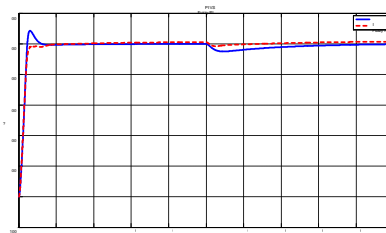


Fig. 10. Speed estimation simulation waveforms

Figures 11 is the torque estimation and speed estimation waveforms for given a rectangular wave command torque 0 and $20N \cdot m$, Period of 0.2s, running 800r / min speed.

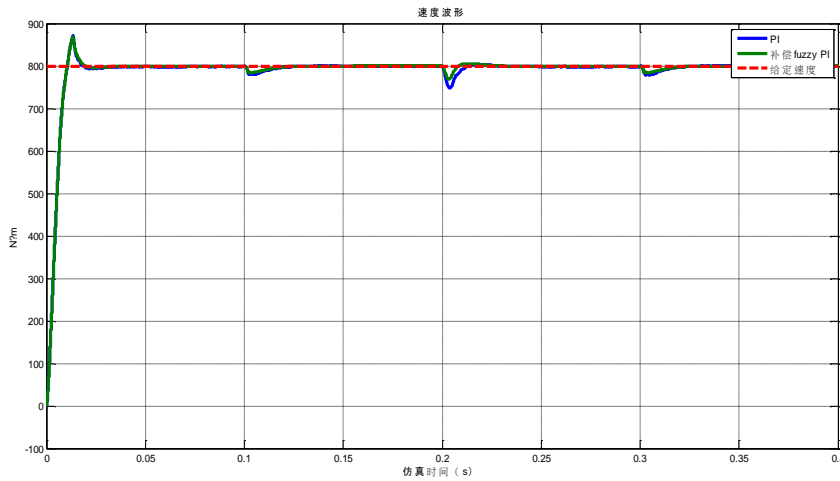


Fig. 11. Speed estimation simulation waveforms

Speed oscillation amplitude of PI speed controller is 50r / min, while Speed oscillation amplitude of the fuzzy PI speed controller speed is 25r / min. which indicates that the system have better anti-disturbance performance.

B. Experimental Results

In this paper, in order to verify the effectiveness of the control strategy, which was tested in adjustable load experimental platform, and by controlling the excitation power to achieve its load changes. At low speed 300r / min, it suddenly increase and decrease 5N • m load. There are some experimental waveforms, they are mechanical angular speed and the phase current waveforms for the traditional PI controller without feedback compensation and fuzzy PI controller with feedback compensation. Fuzzy PI controller combine with torque feedback compensation strategy, which play a role in the inhibition of the dynamic response of the speed overshoot, improving system anti-disturbance performance.

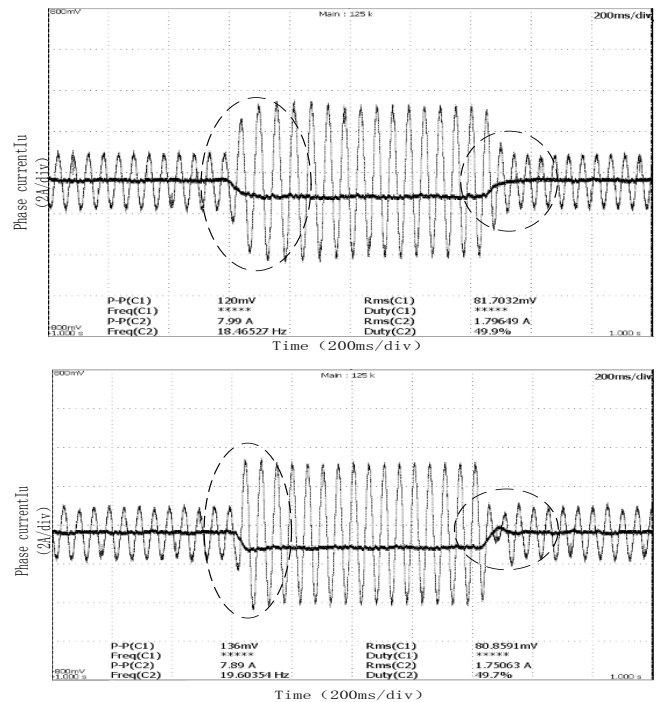
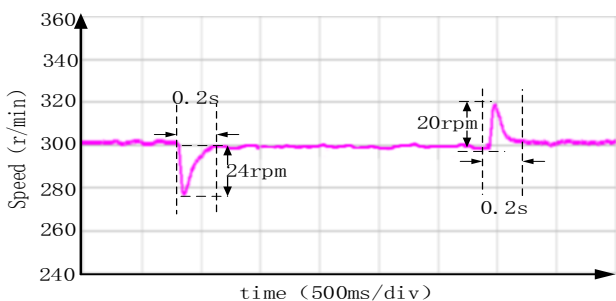
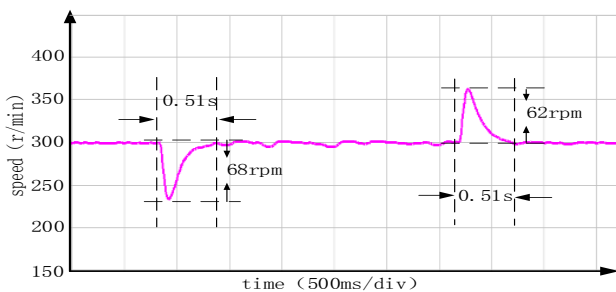


Fig. 12. at 300r / min, a sudden increase and a sudden reduction of the experimental results

V. CONCLUSION

In this paper, in order to solve the robustness, anti-disturbance ability and other issues of traditional PI speed controller for PMSM, a fuzzy PI control scheme based on load torque feedback compensation is proposed, which can improve system speed loop disturbance rejection performance and robustness of the system speed control.

Simulation results and experiment results show that using this control scheme that can significantly reduce the speed overshoot. When the load torque steps change, the speed controller significantly reduce the amplitude of the oscillation speed. Overall, the robustness and disturbance rejection capability has significantly improved.

REFERENCES

- [1] Kou Baoquan, Chen Shukang. AC servo motor and control [M]. Beijing, China Machine Press, 2008.
- [2] Yuandeng Ke, Taosheng Gui AC permanent magnet motor speed control system [M] Beijing: Mechanical Industry Press, 2011.
- [3] S. Bolognani, L. Sgarbossa, M. Zordan Self-tuning of MTPA current vector generation scheme in IPM Synchronous motor drives[J]. 2007 Conf. on Power Electronics and Applications : 1-10
- [4] Yasser Abdel, Rady Ibrahim Mohamed. Adaptive Self-Tuning Speed Control for Permanent-Magnet Synchronous Motor Drive With Dead Time[J]. 2006 Trans. On Energy Conversion Vol.21(4):855-862
- [5] Xu dong, Wang Tianmiao, Jing Liუმeng. Based on parameter identification of permanent magnet synchronous motors self-tuning speed control method [J] high-tech communications 2009 Vol.2:168-173
- [6] Li Wenshan. Based on permanent magnet synchronous motor torque ripple PI-type iterative learning control of suppression strategy [D], Tianjin University, 2011
- [7] Zhou Hui. study Based on implementation and performance fuzzy PI control of permanent magnet synchronous motor vector control system [D] Zhejiang University 2006
- [8] Jia Hongping, He Yikang. PMSM Sliding Mode Direct Torque Control [J]. China Electrotechnical Society, 2006,21 (1): 1-6.
- [9] Huang Lei, Cui Ying. Based on sliding mode observer of PMSM sensorless vector control [J]. Mechanical and Electrical Engineering, 2012,29 (11): 1303-1305.
- [10] Lu Wenqi, Hu Yuwen. Anti-disturbance adaptive control for permanent magnet synchronous motor servo system [J]. Proceedings of the CSEE . 2011, 31(3): 96-101.
- [11] Liu Zhigang, Li Shihua. Active disturbance rejection controller based on permanent magnetic synchronous motor model identification and compensation[J]. Proceedings of the CSEE,2008, 28(24): 118-123.
- [12] Sun Kai, Xu Zhenlin, Gai Kuo. A novel position controller of PMSM servo system based on active-disturbance rejection controller[J]. Proceedings of the CSEE, 2007, 27(15): 43-46.
- [13] Lai C K, Shyu K K. A novel motor drive design for incremental motion system via sliding mode control method[J]. IEEE Trans. on Industrial Electronics, 2005, 52(2) : 499-507.
- [14] Landau I D. Adaptive control : The model reference approach[M]. Wu Baifan Translation. Beijing : National Defence Industry Press, 1985 : 90-95.
- [15] Guo Qingding, Wang Chengyuan. AC servo system[M]. Beijing : China Machine Press, 1994 : 20-26.

CPLD-Based Circuit Design of IGBT Dead-Time Compensation

Qing-zhen WANG

School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

Jin LIU

School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

Guo-hui ZENG

School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

Xing ZHAN

School of Electronic and Electrical Engineering
Shanghai University of Engineering Science
Shanghai, China

Abstract—IGBT (insulated-gate bipolar transistors) dead-time compensation circuit has a very important significant for improving the output voltage waveform of the inverter, reducing the harmonic output current. Thus, many compensation strategies are reported in literatures and have been implemented in industrial drives recently. Overall, the method of dead-time compensation can be divided into hardware compensation and software compensation. Hardware compensation method needs additional hardware circuits, which means additional space and cost. Still more, the additional circuit is easy to interfere with others, which can reduce the compensation accuracy. While the software compensation method takes up a lot of memory space and additional input-output ports of processor, which often result to the added operation and heat dissipation of controller. In this paper, CPLD (complex programmable logic device)-based circuit design of dead-time compensation is presented to solve these existed drawbacks. It is verified that not only can the circuit simplify existed inverter dead-time compensation design, but also it has the advantages of small volume, strong anti-interference ability, and high compensation precision. The simulation results validate that this method is feasible and effective.

Keywords—Dead-time compensation; inverter; compensation method; circuit design; CPLD; IGBT

I. INTRODUCTION

Power electronic devices are not ideal, which its action has a finite turn-on and turn-off time. People join in the time delay, called dead-time control, of the gate drive in order to avoid the shoot-through of inverter arms. Dead time is consist of power switching dead time and dead-time control and effects caused by dead time called the dead-time effect of the inverter. The existence of dead-time effects has a great influence on the inverter. When the dead time is longer, the inverter output fundamental voltage loss will become greater and the voltage waveform distortion is more severe; when the magnitude of the load current is decreased, the current waveform distortion will more serious. The dead-time effects not only affect the amplitude of the output voltage, but also have an impact on its phase. However, the PWM waveform is no longer symmetrical to the center, so that the magnitude of the space voltage vector

will produce deviations and phase also change. In view of this, the impact of the dead time more and more caused people's attention; the dead-time compensation has become particularly important. Therefore, the domestic and foreign scholars carried out a lot of work to study and explore the problem of converter dead-time compensation.

Various compensation strategies have been proposed [1]–[16] to overcome the negative effect of dead time. The current research on the inverter dead-time compensation focused on the voltage feed-forward and current feedback compensation. Through the error analysis between the inverter reference voltage and the output voltage, paper [2] proposes a simple and feasible compensation method based on the dead time and the DC bus voltage; but due to lack of compensation for the power devices turn-on voltage drop, so there is the existence of compensation error. The inverter snubber and parasitic capacitance are considered when doing dead-time compensation in paper [3]. Based permanent magnet synchronous motor (PMSM) AC servo system, papers [6]–[7] use disturbance observer approach to achieve inverter dead-time compensation online for the dead time effect; the rotor d-q axis voltage caused by the non-ideal characteristic of the switching device and inverter dead time is regarded as the disturbance voltage to estimate online and then feedback to the inverter reference voltage to carry out feed-forward compensation. Because of the setting of disturbance observer, the method exist phase lag for compensation voltage, while the transmission gain selection of disturbance observer requires some experience. Paper [8] adopt a zero voltage vector splitting method to solve the zero current clamp problem caused by inverter dead time, however the dead time impact for voltage still need to compensate. Paper [13] takes into account the voltage compensation and elimination of zero current clamp effect, but because of correlation between the compensation voltage and the operating point, the zero current clamp phenomenon still exist compensation error. Summers T. et al [14] reduce the problem of zero current clamp used the predictive current control, but the systems is a local stability.

Currently, methods of achieving inverter IGBT dead-time compensation mainly exist two categories: the hardware

This work is supported by the Natural Science Foundation of Shanghai under Grant No. 14ZR1418400 and the Innovation Foundation of Shanghai Education Commission under Grant No. 13YZ111.

compensation and software compensation. Hardware compensation is needed to provide additional hardware circuit, and need to compare the actual voltage and voltage reference value, and then get the desired compensation voltage signal, so it is necessary to take up more hardware resources, each part of the circuit also generate interaction, causing the compensation precision is reduced. Software compensation need to add the dead-time compensation algorithm in the motor drive control program to achieve compensation for the dead zone, this method requires a large amount of memory resources occupied control chip, increasing the control processor's burden, affecting the running speed and the heat dissipation effect. These designs, CPLD dead-time compensation logic circuit, verify the feasibility and effectiveness through simulation tests in Modelsim.

Therefore, the power electronics technology need for a good inverter dead-time compensation circuit, it takes up a less hardware resources occupied various parts of the circuit and do not affect each other, the control chip run faster and excellent heat dissipation. This design uses the CPLD chip replace complex hardware circuit, using HDL (Hardware Description Language) to generate its internal logic circuit instead of software compensation algorithms, realizing inverter IGBT dead-time compensation. These designs, CPLD dead-time compensation logic circuit, verify the feasibility and effectiveness through simulation tests in Modelsim.

II. INTERNAL STRUCTURE AND DESCRIPTION OF THE CPLD

With the development of electronic technology and the continuous improvement of ASIC technology, the design of digital system is moving fast, large capacity, small volume, light weight in this direction. However, CPLD (Complex Programmable Logic Device) has been widely used in the design of electronic circuits and system because of its flexibility, modifiability, and short development cycle. Substitute for the traditional electronic circuit has incomparable advantages:

1)high integration, small discrete components, high anti-interference ability; 2)simple circuit design is and convenient, software design, easily modified; 3)Ease of product upgrading; 4) the use of high frequency oscillator, easy high speed processor (e.g. DSP) connection, high frequency control system. Based on this, the author uses ALTERA series MAX II CPLD to produce SPWM wave for IGBT bridge driver. The design operation is stable, strong anti-interference, safe and reliable.

The following describes the hardware configuration of the whole design.

The hardware structure in this design is mainly composed of the serial port module, dead-time compensation module, dead-time insertion module, protection module and output control module, as shown in Fig. 1.

The controller calculates the switching time of six IGBT switches and sent CPLD chip the messages via 6 data bus. The internal logic circuits of CPLD firstly compensate dynamic dead-time and then insert the dead-time. According to the control requirements, serial port module decides whether to compensate for the switching time.

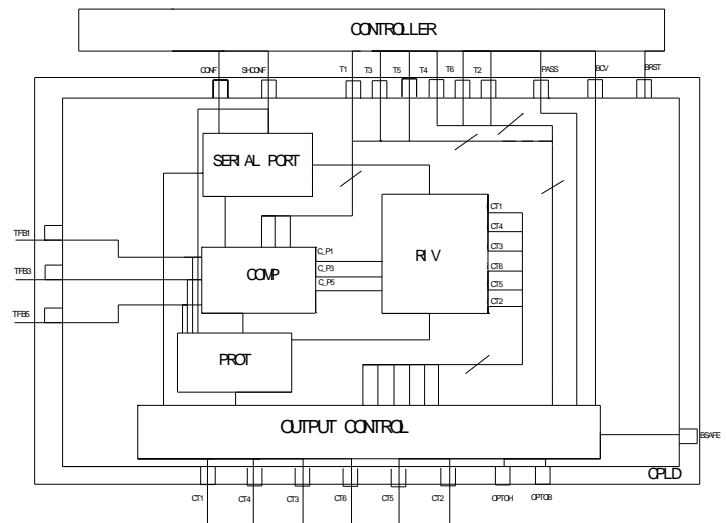


Fig. 1. CPLD internal structure

If it claims for compensation, the dead-time compensation module calculates the real-time dynamic compensation time to insert the dead-time. During the whole process, a real-time protection function can be done by protection module, such as the over-current, over-voltage and open circuit, etc.

III. PROGRAM REALIZATION

Verilog HDL and VHDL are currently most popular hardware description language (HDL: Hardware Description Language) in the world, are the IEEE standard and have been widely used in the project development based on programmable logic devices. Both are developed in the mid-1980s, the former by Gateway Design Automation, Inc. (acquired by Cadence in 1989) and the latter by the US military.

In text form to describe the hardware structure and behavior of digital systems, HDL is a language with a formal way to describe digital circuits and systems and describe their design ideas from the upper to the lower layer. In other words, it uses a series of sub-level modules to represent complex digital systems to verify the simulation layer by layer, and then the specific combination of modules converted into a gate-level netlist by the synthesis tool that put into a specific circuit structure by the layout tools. At present, this top-down approach has been widely used. Generally speaking, HDL contains the following main features:

- HDL structure contains some form of high-level programming language and also take into account the specific structure described the hardware circuit connection.
- HDL can describe the design at different abstract levels through the use of structural level behavioral description. It is a top-down digital circuit design method, including three areas and five abstract levels.
- HDL is processed in parallel and has the ability to perform multiple tasks at one time. This feature is

different from the general advanced design languages (such as C language, etc.) which use serial execution.

- HDL has a concept of timing. General, advanced programming language is no concept of timing, but the hardware circuit is always a time delay from input to output, it needs to introduce the concept of time delay in order to describe this feature. HDL can not only describe the function of the hardware circuit, but also describe the timing circuit.

In this design, using HDL writes programs to achieve CPLD internal function of each module.

A. Dead-time compensation

In order to make the compensation more accurate and effective, this bloc consider the time of both the transition from ON to OFF and effect of inverter snubber. The following introduced the proposed compensation method. These functions restore the time value pulses on the IGBT gates for matching the theoretical commands, after all distortions of them (due to dead time and distortion trough the command channel). Timing diagram is shown in Fig. 2.

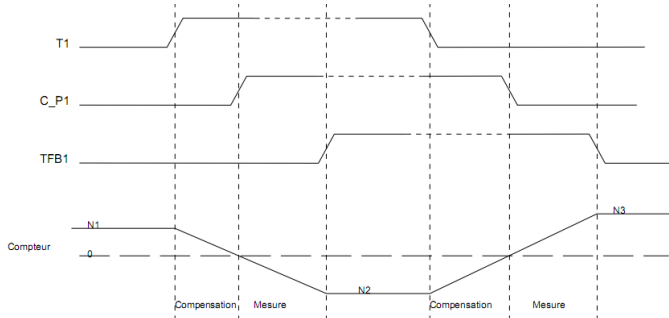


Fig. 2. Dead-time compensation timing diagram

At the rising edge of T1, the counter is at N1 value, this is the value to compensate, the counter is counting down till it reaches zero, at this time C_P1 is rising as the new theoretical command to apply, and the counter is still counting down till the rising edge of TFB1 which is the real application of the voltage on the phase. The time difference between C_P1 and TFB1 is due to the dead time and also the delay of all the components. At this time the counter is in stand-by, and it has N2 value which is the error to compensate (to add) at the next edge.

At the falling edge of T1, the counter is counting up from the N2 value till it reaches zero, at this time C_P1 is falling as the new theoretical command to apply, and the counter is still counting up till the falling edge of TFB1 which is the real application of the voltage on the phase. Now, the counter has N3 value which will be compensated at the next rising edge of T1, as explained with N1.

For example, the frequency of the clock signal is f and dead timer is n bits, then the compensation resolution is $1/f$ and the maximum compensation value is $2^n * (1/f)$. For this design, $f=40$ MHz, $n=10$, the maximum compensation value is 25us and the compensation resolution is 25ns. Flow chart is shown in Fig. 3.

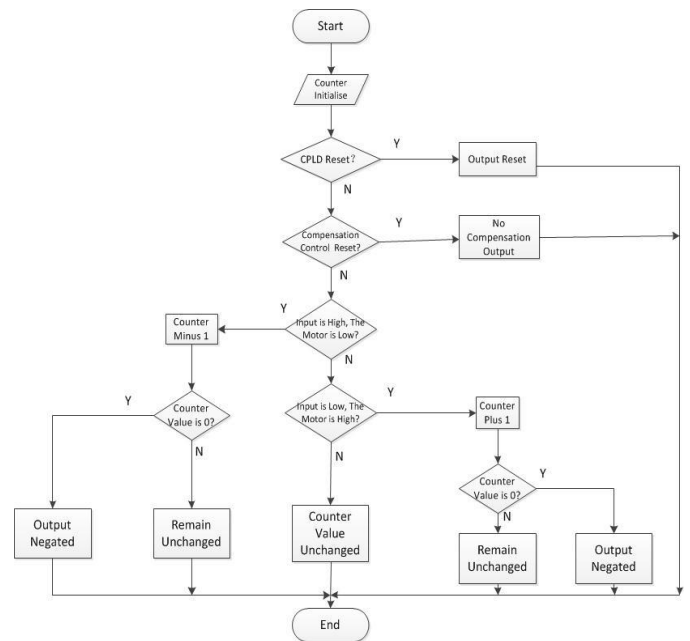


Fig. 3. Dead-time calculation flow chart

B. Dead-time insertion

This bloc adds the dead time at the compensated pulses, and drives directly the 6 outputs for transistors commands. Timing diagram is shown in Fig. 4.

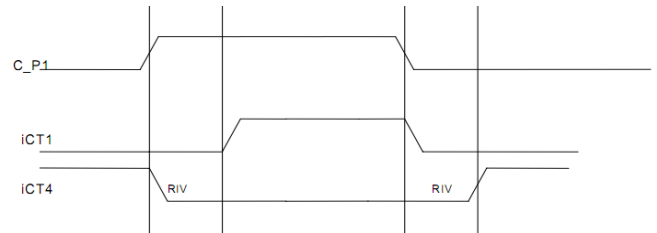


Fig. 4. Dead-time insertion timing diagram

For the same bridge arm of switch tube, at the rising edge of C_P1, iCT4 is falling, iCT1 is rising after the dead time INSE; at the falling edge of C_P1, iCT1 is falling, iCT4 is rising after the dead time INSE. It is the same rule for the others branches. Flow chart is shown in Fig. 5.

C. Serial port

A serial port offers the controller the capability to configure the CPLD by writing into the REGCONF register. SHCONF is used to shift signal for REGCONF register. CONF is data serial input and the configuration bits are applied to the MSB of the REGCONF register.

After RESET, the CPLD is in a configuration mode, the controller is able to write the configuration into the REGCONF register. In this mode the outputs CT1... CT6 are not driven. For getting out of this mode, the controller applies a falling edge on the BCV, and the CPLD goes in a normal mode, the REGCONF register is locked. But as soon as the circuit is in normal mode, the T1, T3, T5 inputs are interpreted and the outputs CTi can commute. Timing specifications are shown in Table 1.

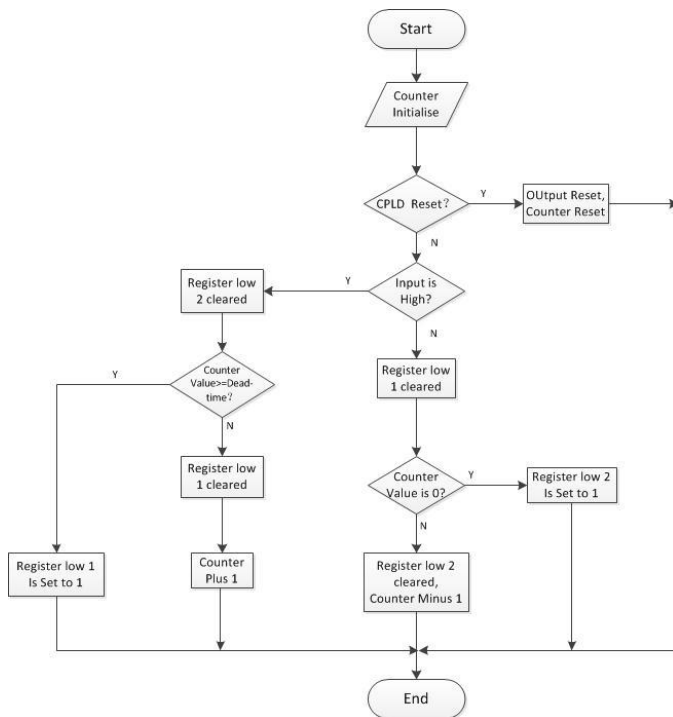


Fig. 5. Dead-time insertion flow chart

TABLE I. TIMING SPECIFICATIONS

Symbol	Description	Min	Unit
Ton	Time for SHCONF at level High	100	ns
Toff	Time for SHCONF at level Low	100	ns
Ts	Setup time for CONF on rising edge of SHCONF	0	ns
Th	Setup time for CONF on falling edge of SHCONF	0	ns

D. Output control

According to the states of the circuit, this Bloc applies the orders on Outputs CTi. In PASS state, all defaults detections are still activated. The Outputs CT1 ...CT6 can drive up to 24mA. As shown in Table 2.

TABLE II. STATES OF CIRCUIT

States	Conditions	Outputs CT1...CT6
Reset	BRST=0	0 0 0 0 0 0
PASS	PASS=1	0 0 0 0 0 0
BSAFE	BSAFE=0	0 0 0 0 0 0
Normal	BCV=0	iCT1...iCT6
CONFIGURATION	Until falling edge of BCV	0 0 0 0 0 0

IV. SIMULATION RESULTS

A. Theoretical waveforms

This paper is the inverter IGBT dead-time compensation circuit design based on CPLD. According to the description of each module and the port function, we can draw the waveform of this design theory. Theoretical waveforms are shown in Fig. 6.

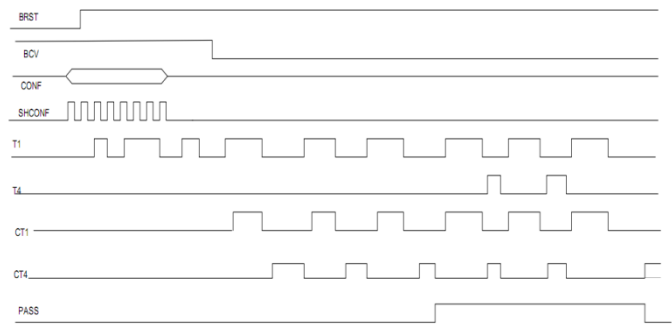


Fig. 6. Theoretical waveforms

B. Simulation waveforms

For the dead-time compensation method simulated test in Modelsim simulation platform. In the simulation of just one phase of bridge arm simulation, given the incentive program to simulate the actual circuit can be obtained by simulation results, the simulation waveform as shown in Fig. 7.

From the simulation waveform can be seen, at the falling edge of BCV, serial port module activates output control module, CT1 and CT4 are the output signal through an output port; the serial port module, dead-time compensation module, dead-time insertion module, protection module is not activated. Control output signal output control module for repeating the input signal. Output control module of control output signal ensures that the output signal is the input signal. At the rising edge of PASS, this function of dead-time compensation is inhibited.

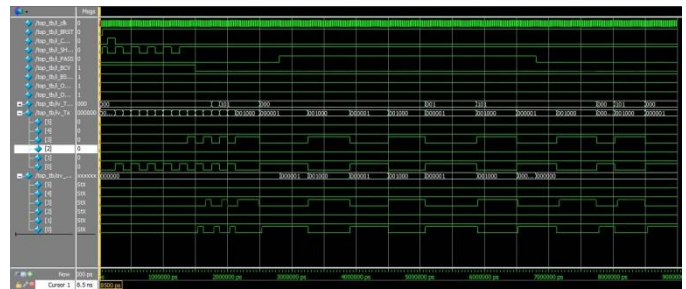


Fig. 7. Simulation waveforms

V. CONCLUSION

This paper mainly studies the problem of inverter IGBT dead-time compensation, proposed the dead-time compensation circuit design based on CPLD. First of all, starting from the hardware circuit describes the internal structure and function. Secondly, the software realization details programming ideas and flow chart.

Compared with the previous design, we can see that

- The design of the serial port module, dead-time compensation module, dead time insertion module, protection module and output control module is focused on a CPLD chip. Compared with the hardware compensation, each circuit section does not influence each other; the design takes less hardware resources and has the advantages of strong anti-interference ability, high compensation precision.

- Compared with the software compensation ,the motor control without adding driver dead-time compensation algorithm save control chip memory resources and the control chip run faster and excellent heat dissipation , and thus switch of the same bridge arm never occurs simultaneous conduction since switching speed , meanwhile have high compensation accuracy.

REFERENCES

- [1] Morohoshi T., Hoshi N., and Haruna J., "Dead-time compensation of adjustable dead-time controlled three-phase resonant snubber inverter for induction motor drive application," *Electrical Machines and Systems*, pp. 1766-1770,2013.
- [2] Jeong S., and Park M., "The analysis and compensation of dead-time effects in PWM inverters," *IEEE Transactions on Industrial Electronics*, pp. 108-114,1991.
- [3] Zhendong Zhang and Longya Xu., "Dead-Time Compensation of Inverters Considering Snubber and Parasitic Capacitance," *Power Electronics*, pp. 3179-3187,2014.
- [4] Choi J., and Sul S., "Inverter output voltage synthesis using novel dead time compensation," *IEEE Transactions on Power Electronics*, pp. 221-227, Nov. 2,1996.
- [5] Qingyi Wang, Xin Deng, Hui Luo, Quan Yin and Shuyun Wan, "A new dead-based compensation method SVPWM strategy," *Drives*,2008.
- [6] Urasaki N., Senjyu T. and Uezato K., "On-line dead-time compensation Method for permanent magnet synchronous motor drive," *IEEE International Conference on Industrial Technology*, pp. 268-273, Jan., 2002.
- [7] Hyun-Soo Kim, Hyung-Tae Moon and Myung-Joong Youn, "On-Line Dead-Time Compensation Method Using Disturbance Observer," *IEEE Transactions on power electronics*, Vol.18, NO.6, pp. 1336-1345, 2003.
- [8] Murai Y., Riyanto A., Nakamura H. and Matsui K., " PWM strategy for high frequency carrier inverters eliminating current clamps during switching dead-time," *Industry Applications Society Annual Meeting, Conference Record of the 1994 IEEE*, pp. 317-322, 1992.
- [9] Ilhami Colak, Ersan Kabalci and Ramazan Bayindir, "Review of multilevel voltage source inverter topologies and control schemes," *Energy Conversion and Management*, 2011.
- [10] Dong-Hee Lee and Jin-Woo Ahn, "A Simple and Direct Dead-Time Effect Compensation Scheme in PWM-VSI," *Industry Applications*, Vol. 50, pp. 3017-3025, 2014.
- [11] Bedetti N., Calligaro S. and Petrella R., "Self-commissioning of inverter dead-time compensation by multiple linear regression based on a physical model," *Energy Conversion Congress and Exposition (ECCE)*, pp. 242-249, 2014.
- [12] Choi J. and Sul S., "New dead time compensation eliminating zero current clamping in voltage-fed PWM inverter," *Industry Applications Society Annual Meeting, Conference Record of the 1994 IEEE*, pp. 977-984, 1994.
- [13] Summers T., Betz R., "Dead-time issues in predictive current control," *Industry Applications Society Annual Meeting, Conference Record of the 1994 IEEE*, pp. 2086-2093,2002.
- [14] Urasaki N., Senjyu T., Uezato K., and Funabashi T., "On-line dead-time compensation method for permanent magnet synchronous motor drive," *Industrial Technology, 2002 IEEE International Conference on Volume:1*, pp. 268-273, 2002.
- [15] Morohoshi T., Hoshi N. and Haruna J., "Dead-time compensation scheme for adjustable dead-time controlled three-phase resonant snubber inverter," *ECCE Asia Downunder (ECCE Asia)*, 2013 IEEE, pp. 836-841, 2013.
- [16] Jong-Lick Lin, "A new approach of dead-time compensation for PWM voltage inverters," *Circuits and Systems I: Fundamental Theory and Applications*, Volume: 49, pp. 476 - 483, 2002.

Detection and Removal of Gray, Black and Cooperative Black Hole Attacks in AODV Technique

Hosny M. Ibrahim, Nagwa M. Omar, Ebram K. William
Information Technology Department
Faculty of Computers and Information, Assiut University
Assiut, Egypt

Abstract—Mobile ad hoc network (MANET) is an autonomous self-configuring infrastructure-less wireless network. MANET is vulnerable to a lot of routing security threats due to unreliability of its nodes that are highly involved in the routing process. In this paper, a new technique is proposed to enhance the security of one of the most popular MANET routing protocols that is called Ad hoc on Demand Distance Vector (AODV) with minimum routing overhead and high packet delivery ratio. The proposed technique intends to detect and remove black, gray, and cooperative black hole AODV attacks depending on a mobile backbone network constructed from randomly moving regular MANET nodes based on their trust value, location, and power. The backbone network monitors regular nodes as well as each other to periodically estimate monitoring trust values which represent the reliability of each node in the network. The drop in the monitoring trust value of any node is used as a clue to its malicious behavior. The backbone network also tries to bait the malicious nodes to reply to a request for a route to fake destination address. The proposed technique uses the control packets of the AODV to exchange its control information which highly reduces the overhead. The simulation results show that the performance of the proposed technique is more secure than AODV and the other recently introduced techniques.

Keywords—MANET; AODV; Black Hole Attack; Gray Hole Attack; Cooperative Black Hole Attack

I. INTRODUCTION

Mobile ad hoc network (MANET) [1] is a set of mobile nodes communicate wirelessly to establish network without fixed infrastructure. MANET provides flexible communication when there are geographical or terrestrial constraints. Battlefields, military applications, emergencies and some disaster management situations need the existence of infrastructure-less network such as MANET [1].

MANET has a dynamic topology in which each node has unrestricted mobility, connectivity, and changes its links to other nodes frequently. In such networks the routing is not an easy task [1]. Routing in MANET is done cooperatively between nodes. Each node works as a router that forwards packets for other nodes. These infrastructure-less mobile nodes dynamically participate in an ad hoc route discovery process and create routes among themselves to form a wireless network on the fly. Due to the wireless communication nature and the collaboration of MANET nodes in finding routing paths, MANET is more vulnerable to security threats than ordinary wired networks [2]. Another characteristic of a MANET is its

resource constraints, i.e., limited bandwidth and battery power of its nodes [1]. Dynamic Source Routing (DSR), Ad hoc On Demand Distance Vector (AODV), Optimized Link State Routing (OLSR), and Destination Sequence Distance Vector (DSDV) protocols are the most popular MANET routing protocols [1].

Many security techniques are introduced to prevent different MANET attacks [2]. Many of these techniques are directed to protect AODV [4] routing protocol from attacks because it is a popular reactive routing protocol designed for mobile ad hoc network. AODV is self-starting, multi-hop, has low processing and low routing overhead, and suitable for dynamic network changes [3] but it does not take security issues into consideration [5].

In this paper, a new technique is proposed to enhance AODV security. It uses a mobile backbone network to efficiently detect and remove the gray, black, and cooperative black hole attacks based on nodes' trust values. NS2 simulator [6] is used to evaluate the performance of the proposed technique. The simulation results show that, the proposed technique gives minimum routing overhead, minimum delay, and high packet delivery ratio compared with AODV and other routing techniques that are introduced to solve the security issues in AODV algorithm.

The rest of this paper is organized as follows: related work is presented in section 2. The proposed technique is described in section 3. Simulation and comparison results are presented in section 4. Finally, section 5 is a conclusion of the proposed technique.

II. RELATED WORK

This section reviews the AODV routing protocol and its security attacks as well as the advantages and disadvantages of many algorithms that are recently introduced to solve the security issues in AODV.

A. AODV Routing Protocol

The AODV protocol consists of two important phases, Route Discovery and Route Maintenance. In Route Discovery, when a node wants to communicate with another node and there is no valid route in its routing table, it broadcasts a route request packet (RREQ). A node that receives a RREQ for the first time sets up a reverse route to the source node in its routing table. If the node is the destination or has a valid route to the destination, it unicasts a route reply (RREP) along the reverse path back to the source node. Otherwise, it will

increase the hop count in the RREQ by one and forward the RREQ to other nodes. In Route Maintenance phase, neighboring nodes periodically exchange HELLO messages to know its one-hop neighbors. If one node didn't receive a HELLO message from a neighboring node within a certain time interval, the node breaks the routing table information of this neighbor node and sends a Route Error (RERR) message to the nodes on a route with this neighbor.

B. AODV Security Attacks

The work in the current paper focuses on three types of attacks:

Black hole attack [3, 8, 9]: a black hole node is a malicious node that sends a false reply with an apparently valid route to the destination node. It replies every single RREQ with false sequence number, so it acquires the route, and then eavesdrops or drops all data packets that pass through it.

Gray hole attack (selective black hole) [8, 9]: looks like a black hole attack, but a malicious node randomly changes its state between regular node and black hole node. Accordingly, gray hole is harder to be detected by security techniques.

Cooperative black hole attack [9, 10]: two nodes or more in this attack cooperate to gain the path between the source and the destination nodes. When one node gains the path selectively drops or forwards the data packets to one of its cooperating nodes. Cooperation between black hole nodes helps malicious nodes to escape from monitoring techniques.

C. Fighting AODV Security Attacks

There are many techniques presented recently to mitigate security attacks in MANET [1], but this section reviews only some of the techniques that relate to the proposed technique.

Ming-Yang Su [11] presented an intrusion detection system (IDS) to detect and prevent selective black hole attacks. In IDS system, several fixed IDS nodes are distributed and set in sniff mode in order to estimate a suspicious value of a node. The simulation results show that the IDS technique can be used effectively to block the malicious nodes if a proper threshold is set, but IDS technique has some disadvantages: 1) it uses fixed, trusted, and powerful nodes to detect the malicious nodes, which violate the mobility feature of the MANET [12, 13] 2) the scheme suffers from high routing overhead.

The authors in [14] introduced a technique for detecting cooperative black and gray hole attack in MANET using a backbone network of strong nodes established over the ad hoc network. This backbone network monitors the overall traffic in the network with the help of regular untrusted nodes. The disadvantages of this algorithm are: 1) regular nodes can join the backbone based only on their power and location without taking into consideration their reliability and trust 2) the backbone nodes carry an end to end check based on regular nodes' request which can be used by malicious nodes to exhaust backbone resources. 3) The backbone nodes ask normal nodes, which may be malicious to perform monitoring which give deceiving results. 4) Assuming that there is a difference between regular nodes and backbone nodes in terms of power and antenna range which is not suitable. 5) It is not proved that the backbone network is optimal in terms of

minimality and coverage. 6) The technique suffers from high end-to-end delay and high routing overhead. 7) The technique executes an end-to-end check after every transmission of a block of data which is not an optimum solution. 8) The technique assumes that a node has strong neighbors more than malicious nodes, which may not be always satisfied [9, 12, 13].

Also, K. Vishnu, and A. J. Paul [15] presented a mechanism to detect and remove cooperative black and gray holes. It assumes that the network is divided into clusters and a backbone network is present in the MANET [14]. Each backbone node (BBN) knows a valid set of addresses that is used in the network. Only the backbone network in MANET is permitted to select the addresses for non configured hosts [16]. When the source node wants to transmit data, it asks the nearest BBN for non used IP in the network which is called restricted IP (RIP). The source node sends a RREQ for both the destination and the RIP simultaneously. If the source node receives a RREP for the RIP, it means that there is a black hole in that route. The source node sends a few dummy data packets to that destination. When a monitoring node finds that the loss in dummy data packets is more than the normal expected loss at an intermediate node, it informs the source node about this malicious node. Also, the neighbor nodes broadcast an alert message and add this malicious node to the black hole list. This technique has the following drawbacks: 1) regular nodes are assumed to be trusted by default. They participate in the monitoring process and take critical decisions to isolate other nodes, which is not secure [17]. 2) The authors didn't propose simulation results to test the performance of their scheme. 3) The mechanism will fail if malicious nodes keep asking the BBNs for RIP, save RIPs, and stop replying to RIP RREQs. 4) It suffers from routing overhead. 5) It detects black and gray hole nodes depending on the desire of the source node to send data to a destination node.

Authors in [10] presented an enhancement of the AODV to mitigate cooperative black hole attacks by introducing fidelity table wherein every node is assigned a fidelity level that acts as a measure of the reliability of that node. When the destination node receives the data packets, it sends an acknowledgment to the source and the fidelity level of the intermediate node is increased. If the fidelity level of any node drops to zero, it is considered as malicious node and is isolated. The algorithm can mitigate cooperative black hole, but it has many drawbacks [8]: 1) the fidelity tables of nodes are maintained and exchanged periodically among the participating nodes which increase the overhead and the processing delay. 2) Additional overhead and time delay are introduced due to the use of the acknowledgements.

III. THE PROPOSED TECHNIQUE

This paper proposes a technique to enhance AODV security. This technique attempts to detect and remove gray, black and cooperative black hole attacks with the aid of a network of mobile backbone nodes. The proposed technique is divided into four main phases:

1) **Mobile Backbone Network Constructions:** in which, a mobile backbone network is constructed and updated based on nodes trust value

2) **MANET Formation:** in which, new clients join the MANET and the network nodes are grouped to give good performance.

3) **Detection of Malicious Nodes:** in which, two methods are implemented with the aid of the backbone network to detect malicious nodes.

4) **Removal of Malicious Nodes:** this phase starts after detecting malicious nodes; in which, the backbone network isolates the malicious nodes.

These phases will be explained in more details in the next subsections.

B. Mobile Backbone Network Formation

The proposed technique intends to increase the security of AODV depending on the mobile network of secure backbone nodes. This backbone network should be trustable, have dynamic behavior, does not violate the mobility characteristic, structured of the regular MANET nodes, and has good coverage.

To achieve these characteristics, each node in the backbone network maintains two different values:

1) *Monitoring trust value (MTV) for each of its neighbors that represents the reliability of the node.*

2) *Its trust value (TV), which is used to specify its operations and allowed decisions.*

The estimation of the monitoring trust value varies in various introduced techniques. In [11], the estimation of trust value is not adequate since it depends only on the routing control packets and doesn't take dropping data packets into consideration. The techniques that are introduced in [10, 14] estimate the monitoring trust value based on the dropped data packets only but doesn't take into account the routing control packets. Also, they present high overhead and time delay.

The proposed technique in this paper introduces new criteria to estimate the monitoring trust value. The following equation is suggested to estimate the monitoring trust value for node (i):

$$MTV_i = \tanh(C_1 \frac{F_DPs_i}{R_DPs_i}) \tanh(C_2 \frac{R_RREQs_i}{S_RREPs_i}) \quad (1)$$

Where $0 \leq MTV_i \leq 1$, (F_DPs_i) is the number of the forwarded data packets that are not originated from the node, i, (R_DPs_i) is the number of the received data packets that is not targeted to the node, i, (R_RREQs_i) is the number of received route requests to the node, i, (S_RREPs_i) is the number of sent route replies from the node, i, and C_1 and C_2 are constants adjusted experimentally.

To increase the coverage of the backbone network, it should choose new nodes from the neighbors based on their MTVs, power and location to join the backbone network.

The new chosen nodes have lower level than the ones that choose them in the backbone hierarchy. The higher level backbone network nodes assign the lower level ones TVs. The

trust value of the backbone network nodes is estimated using the following suggested equation:

$$TV_i = \frac{1}{L_j} * TV_j * MTV_i, \text{ where } 0 \leq TV \leq 1 \quad (2)$$

Where TV_i is the trust value of the new chosen node, i, TV_j is the trust value of the original backbone node, j. MTV_i is the monitoring trust value of the chosen backbone node, i. L_j is the level of the original backbone node. The backbone network node level is calculated using the following suggested equation:

$$L_i = L_j + 1 \quad (3)$$

Where L_i and L_j are the trust levels of the chosen backbone node, i, and the original backbone node, j, respectively. The highest level in the backbone network hierarchy is one.

The backbone network contains four types of nodes as follows:

1) **Seed Backbone Nodes (SBBNs):** which face the difficulties in the backbone initialization. The mobile backbone network should have high trustable nodes at the start to judge the behavior of the new MANET clients. Accordingly, the backbone network needs to be initialized by powerful trustable mobile seeds before it reaches the autonomous mobile dynamic backbone structure. At least one SBBN is needed to construct the backbone network. SBBNs are distributed in the initialization step to cover the target area. SBBNs have trust value and level equal to one which are the maximum. Each SBBN has a pool of addresses that are used in the network. The SBBNs are the only nodes that have the permission to send addresses to the new nodes that join their clusters. Also, each SBBN monitors other nodes in its cluster to employ alternative backbone node based on its MTV, which is called a backbone node (BBN) and sends it the essential information then, SBBN enters a sleeping mode.

2) **Backbone Nodes (BBNs):** start as regular nodes, then are changed to take the role of SBBN to perform the monitoring function in their clusters and judge the other nodes behavior based on their MTVs. Every BBN employs the highest trusted nearest neighbor node to be its vice backbone node (VBBN) and periodically sends it its control information. To increase the coverage and improve the performance, BBNs can employ other nodes with high MTVs to be capable backbone nodes (CBBNs). There is one BBN in each cluster.

3) **Vice Backbone Nodes (VBBNs):** can take the role of BBN in case of BBN movement or power drop. There is one VBBN in each cluster

4) **Capable Backbone Nodes (CBBNs):** are employed to assist BBNs and to increase the coverage. CBBNs can employ other level of CBBNs.

Each backbone network node (SBBN, BBN, VBBN, CBBN) assigns the new employed backbone node a trust value

and level using equations (2, 3) to specify the operations and allowed decisions for each node in the backbone network

The following steps illustrate the task of initializing the backbone network held by the SBBNs; assuming that initial mobile trustable seeds are equally distributed in the target area, can communicate with each other, know each other locations, contain a pool of addresses, and every SBBN is a seed for a cluster of MANET nodes:

- 1) If SBBN is not in a sleeping mode,
 - a) If SBBN receives newly arrived clients requests to join the most powerful and closest distance SBBN,
 - i. SBBN sends a reply to the client contains a unique address selected randomly from its pool of unused addresses. The process of assigning address to newly arrived clients is described in more details in section (3.5).
 - b) SBBN continuously monitors its clients to judge their performance and sets them monitoring trust values (MTVs).
 - c) For each node, obtain node's MTV,
 - i. If node's MTV is less than experimentally chosen SUSPICIOUS NODE THRESHOLD,
 1. A node is considered suspicious.
 - ii. Else if it finds a regular node that has MTV greater than experimentally chosen BBN THRESHOLD, the closest node to the SBBN, and it is the most powerful,
 1. SBBN employs this node to be the new mobile backbone node (BBN) for this cluster and takes SBBN role.
 2. SBBN sends to it the essential information.
 3. SBBN assigns the BBN's TV.
 4. SBBN starts wake up timer and enters a sleeping mode.
 - d) Start rechecking nodes MTV timer.
 - e) If rechecking timer elapsed,
 - i. Go to step (1.c).
- 2) Else if SBBN is in a sleeping mode,
 - a) If the wake up timer elapsed,
 - i. SBBN wakes up to monitor the backbone network nodes and sets them monitoring trust values (MTVs).
 - ii. For each node, obtain node's MTV,
 1. If the neighbor is BBN and its MTV is less than experimentally BBN THRESHOLD,
 - a. The neighbor status is changed to be a regular one.
 - b. Go to step (1.a).
 2. If the neighbor is VBBN/CBBN and its MTV is less than experimentally VBBN/CBBN node THRESHOLD,
 - a. The neighbor status is changed to be a regular one.
 - b. SBBN informs the BBN.
 3. Else if node's MTV is less than experimentally chosen SUSPICIOUS NODE THRESHOLD,
 - a. A node is considered suspicious.
 - iii. SBBN starts wake up timer and enters a sleeping mode.

The following points illustrate the operations of the backbone network held by the BBNs in every cluster taking into consideration that each cluster has only one BBN:

- 1) BBN takes the role of SBBN or BBN will be a cluster grouping point and the clients are regrouped to join this cluster.
- 2) Start regrouping timer.
- 3) If the regrouping timer elapsed,

- a) Each BBN will be a cluster grouping point and the clients are regrouped to join this cluster. Regrouping process is repeated based on the movement speed.
- b) Start regrouping timer.
- 4) Each BBN adds its neighbors to its MONITORED NODES LIST.
- 5) BBN receives newly arrived clients requests to join the most powerful, closest distance BBN.
 - a) BBN sends a reply to the client contains a unique address selected randomly from its pool of unused addresses. The process of assigning address to newly arrived clients is described in more details in section (3.5).
- 6) Each BBN continuously monitors its neighbors including regular nodes and lower level backbone network nodes to judge their performance and sets them MTVs.
- 7) For each node, obtain its MTV,
 - a) If the neighbor is VBBN/CBBN and its MTV is less than experimentally VBBN, CBBN node THRESHOLD,
 - I) The neighbor status is changed to be regular one
 - II) BBN removes the node's covered addresses from its MONITORED NODES LIST and starts the coverage process.
 - b) Else if node's MTV is less than experimentally chosen SUSPICIOUS NODE THRESHOLD,
 - I) A node is considered suspicious.
 - c) Else if there are no VBBN and a regular node's MTV is greater than experimentally chosen VBBN THRESHOLD, the closest node to the BBN, and it is the most powerful,
 - I) BBN chooses this node to be its vice backbone node.
 - II) BBN assigns VBBN's TV and level.
- 8) Start rechecking nodes MTV timer.
- 9) If the rechecking timer elapsed,
 - a) Go to step 7.
- 10) If there are VBBN,
 - a. Each BBN periodically, based on HELLO message interval, sends the assigned addresses in its cluster and the MONITORED NODES LIST to its VBBN.
- 11) If the BBN suffers a low battery condition and there are VBBN,
 - a. The BBN asks its VBBN to take its IP and role.
 - b. The BBN changes its status to be regular node.
 - c. End.
- 12) If BBN receives information about changing a backbone node status to a regular node,
 - a. BBN removes the node's covered addresses from its MONITORED NODES LIST and starts the coverage process.
- 13) Each BBN periodically, based on HELLO message interval, asks the backbone network nodes in its cluster for their neighbors.
- 14) When BBN receives replies,
 - a. Each BBN add not repeated replies to its MONITORED NODES LIST.
- 15) BBN periodically, based on HELLO message interval, checks its MONITORED NODES LIST.
- 16) If there are assigned addresses in BBN's cluster not in its MONITORED NODES LIST this is an indication that there are unmonitored nodes. In this case, the following is achieved to employ new backbone network nodes which are called CBBNs to monitor the uncovered nodes,
 - a. Inform the backbone network nodes in its cluster with these addresses.
 - b. If BBN finds regular nodes in its neighbors that have MTVs greater than experimentally chosen CBBN THRESHOLD,
 - I) BBN asks if they have these addresses in their neighbor list.
 - II) When BBN receives replies, The BBN,
 - (1) Chooses the one with the highest MTV and

- power to be the new CBBN.
- (2) Adds the new covered address in its MONITORED NODES LIST.
- (3) Assigns the new CBBN TV and level.
- (4) Informs the backbone network with the CBBN's address.
- c. If BBN receives suggestions for CBBNs,
 - I) BBN chooses the one with the highest MTV and power.
 - II) Adds the new covered address in its MONITORED NODES LIST.
 - III) BBN informs the backbone network in its cluster with the CBBN's address.
- 17) Else if there are no assigned addresses in BBN's cluster not in its MONITORED NODES LIST,
 - a. BBN ends the coverage process.

The following steps illustrate the operation of the backbone network held by the VBBN and CBBNs taking into consideration that each cluster can have only one VBBN and more than one CBBN:

- 1) If the VBBN discovers a BBN link failure,
 - a) It takes the IP and the role of BBN.
 - b) End.
- 2) Each VBBN/CBBN continuously monitors its neighbors to judge their performance and sets them monitoring trust values (MTVs).
- 3) For each node, obtain its MTV,
 - a) If the neighbor is CBBN and its MTV is less than experimentally CBBN THRESHOLD,
 - I) The neighbor status is changed to be a regular one and inform the BBN.
 - b) Else if node's MTV is less than experimentally chosen SUSPICIOUS NODE THRESHOLD,
 - I) A node is considered suspicious.
- 4) Start rechecking nodes MTV timer.
- 5) If the rechecking timer elapsed,
 - a) Go to step 3.
- 6) If VBBN/CBBN receives BBN request asks for its neighbors,
 - a) VBBN/CBBN replies with its neighbors.
- 7) If VBBN/CBBN receives request to be changed to a regular node,
 - a) VBBN/CBBN changes its state to be regular node.
- 8) Each VBBN periodically, based on HELLO interval, receives the assigned addresses and the MONITORED NODES LIST from its BBN.
- 9) If VBBN finds that there are assigned addresses in BBN's cluster not in the MONITORED NODES LIST OR If CBBN receives assigned addresses in BBN's cluster and not in BBN's MONITORED NODES LIST,
 - a) If VBBN/CBBN finds regular nodes in its neighbors that have MTVs greater than experimentally chosen CBBN THRESHOLD,
 - I) VBBN/CBBN asks if they have these addresses in their neighbor list.
 - II) When VBBN/CBBN receives replies, The VBBN/CBBN,
 - (1) It sends the BBN a suggestion carries information about the one with the highest MTV and power to be employed as new CBBN.
 - (2) If VBBN/CBBN receives BBN request to employ new CBBN,
 - a. VBBN/CBBN assigns the new CBBN TV and level.
- 10) Else if VBBN finds that there are no assigned addresses in BBN's cluster not in its MONITORED NODES LIST,

- a) VBBN ends the coverage process.

As shown from previous tasks for every backbone network node type, the nodes of the backbone network monitor each other as well as the regular nodes that are located in their transmission range and set them MTVs which represent the reliability of each node in the network. The level of backbone network nodes can be changed based on MTV, power, movement, and coverage. Except the initial seeds, no backbone node is considered trusted forever. Increasing the number of BBNs and CBBNs helps in facing the dynamics of MANET, increases the coverage, increases the reliability, distributes the control, saves the nodes recourses, and speeds up the detection and the removal process.

The construction of the backbone network consumes low overhead because all control information that is exchanged between backbone network nodes is added to the AODV HELLO message as additional fields.

As shown from the discussion, the proposed multi-level backbone network is mobile, dynamic, trusted, powerful, has high coverage, reliable, distributes the control, saves the nodes recourses, and robust can face nodes failure. The backbone network uses multi-hop communication to communicate with each other as well as with regular nodes. Unlike the technique in [11], the proposed technique doesn't use permanent fixed nodes. Also, it is more secure and practical than the introduced backbone in [14] which chooses the backbone nodes based on their power and coverage assuming that all backbone nodes are powerful and trusted by default. The proposed backbone network is constructed and updated based on the nodes trust value in addition to power and coverage. Unlike the other techniques, the proposed technique has not considered any node to be trusted forever including the backbone nodes. Also, the backbone network nodes are the only nodes, that are permitted to monitor and judge the behavior of other nodes, which is considered more secure than the technique that is proposed in [17]. The monitoring process can be used for malicious node detection as well as for backbone construction.

C. MANET Formation

This section describes how new clients join the MANET. The proposed technique uses the BBNs as approximate centers of clusters to facilitate and speed up the communication process. Every BBN has a pool of unique addresses that is used to configure nodes in its cluster. The proposed technique follows the technique that is used in [15, 16] but it modifies the equation that is used by [16] to allocate the range of host addresses as follows:

$$\text{Range of addresses of } BBN_i = i * \text{BaseValue} + n \quad \text{for} \\ 0 \leq n < \text{BaseValue}; 0 \leq i < K \quad (4)$$

Where K is the number of BBNs, and BaseValue is the maximum number of addresses that are supported by every BBN.

Newly arrived clients broadcast requests to BBNs to join the most powerful, closest distance BBN. The clients may be regrouped according to the node's movements. In each cluster, there are one BBN, one VBBN, and CBBNs to cover the cluster area. A lot of clustering techniques for MANET is

discussed in [18]. K-means [19] is one of the simplest algorithms that solves the clustering problem. Accordingly, it is used to group nearby nodes in the proposed technique. The proposed technique tries to keep the BBNs as cluster grouping points even if they are not located exactly in the cluster centers because BBNs are permitted to move randomly.

D. The Detection of black, gray, and cooperative black holes in AODV

Two methods are proposed in the current technique to detect malicious nodes:

1) The first method is based on the monitoring trust values (MTVs) which is estimated using equation (1).

2) The second method is based on baiting the malicious node to reply to requests for route to not existing destination in the MANET.

In the first method, the backbone network periodically checks neighbors MTVs. A node is considered malicious one if its MTV is less than experimentally chosen SUSPICIOUS NODE THRESHOLD. Each node in the backbone network has a suspicious node list. Each entry in this list contains suspicious node ID, discovering nodes TVs, discovering nodes IDs, and suspicious node MTV. The action that is taken by the discovering backbone network node is limited by its trust value.

The following points illustrate the steps executed by the backbone network nodes in the first method:

- 1) The backbone network node checks neighbors MTVs including the other backbone network nodes.
- 2) For each node, obtain next node's MTV,
 - a) If node's MTV is less than experimentally chosen SUSPICIOUS NODE THRESHOLD,
 - i) A node is considered suspicious.
 - ii) If the discovering node TV is greater than experimentally chosen REMOVING NODE THRESHOLD,
 - (1) The discovering node starts the removal process which will be described in detail in section 4.4.
 - iii) Else if the discovering node TV is less than experimentally chosen REMOVING NODE THRESHOLD,
 - (1) The discovering node searches its suspicious node list for the suspicious node ID.
 - (2) If the discovering node does not find the suspicious node ID in its suspicious node list,
 - (a) The discovering node adds an entry contains (discovering node ID, discovering node TV, suspicious node ID, and suspicious node MTV) to its suspicious node list. That entry fields are shown in Figure (4.3).
 - (b) The discovering node informs the backbone network with that entry using additional control fields added to the HELLO message.
 - (3) Else if the discovering node finds an entry of the suspicious node ID in its suspicious node list,
 - (a) If this entry contains only the discovering node which can be happened if the discovering node added this entry before and the suspicious node is not removed yet,

- (i) The discovering node updates its TV and the suspicious node MTV in this entry.
- (ii) The discovering node informs the backbone network nodes.
- (b) Else if this entry contains another discovering nodes including or not including the discovering node which give indication that the discovering node received messages from neighbors confirm that they discover the same suspicious node,
 - (i) The discovering node combines the TVs of all the discovering nodes in the entry including its new TV and calculates combined TV using equation (4.1).
 - (ii) If the combined TV is greater than REMOVING NODE THRESHOLD,
 1. The discovering node starts the removal process.
 - (iii) Else if the combined TV is less than REMOVING NODE THRESHOLD,
 1. If the discovering node is included in the entry,
 - a. The discovering node updates its new TV and the suspicious node MTV.
 2. Else if the discovering node is not included in the entry,
 - a. The discovering node appends its ID, TV, and the suspicious node MTV in the entry.
 3. The discovering node informs the backbone network.
- 3) Set up a timer for rechecking neighbors MTVs.
- 4) If the timer interval elapsed,
 - a) Go to step 1.

The following steps illustrate the operation executed by the backbone network nodes in the first method upon receiving suspicious node entry:

1. The backbone network node receives the suspicious node information.
2. It searches its suspicious node list for the suspicious node ID.
3. If it does not find the suspicious node ID in its suspicious node list,
 - a. It adds the received information as an entry to its suspicious node list.
 - b. It informs the backbone network with that entry.
4. If it finds an entry of the suspicious node ID in its suspicious node list,
 - a. It combines the TVs of all the discovering nodes in the entry with the new received information using equation (5).
 - b. If the combined TV is greater than REMOVING NODE THRESHOLD,
 - i. It starts the removal process.
 - c. If the combined TV is less than REMOVING NODE THRESHOLD,
 - i. It updates the suspicious nodes list entry using the received information.
 - ii. It informs the backbone network.

As stated earlier, in some cases the backbone network nodes need to combine the TVs of all discovering nodes that are recorded in the entry including its new TV. The following equation is used to calculate the combined TV:

$$TV_{combined} = \tanh \sum_{i=1}^n TV_i \quad (5)$$

Where n is the number of the discovering nodes that are indicated in the entry.

As stated earlier, in addition to using the MTV value for detecting malicious nodes, the proposed technique uses another detection method. In this second method, the BBNs periodically perform a special check for malicious node detection. BBNs try to bait the attackers to send RREP to RREQ contains a fake destination address. As stated earlier in section (3.2), every BBN has a pool of disjoint unique addresses that is used to configure its clients in its cluster. This way of address allocation facilitates using the second method for malicious node detection.

The following points illustrate the steps held by the backbone network nodes of the second method:

1. BBN chooses a random unused address called restricted IP address (RIP).
2. BBN uses the AODV HELLO message to send this RIP to the backbone network which is considered as a sign for the backbone network to monitor any nodes that reply the RIP RREQ by RREP message.
3. The backbone network starts to monitor the neighbor nodes for any RREPs to that RIP.
4. BBN sends a RREQ to find a path to this fake destination.
5. If any backbone network node listens a RREP to that RIP RREQ,
 - a. It saves the ID of the node that forwards the reply.
6. If the BBN receives a RREP for RIP RREQ,
 - a. The BBN asks the backbone network for the monitoring information.
 - b. The BBN figures out the source node of the RREP.
 - c. The BBN moves into the removal process that will be described in section (3.4).
7. Set up a timer for baiting again malicious nodes.
8. If the timer interval elapsed,
 - a. Go to step 1.

Using the two proposed methods to detect the malicious nodes, the proposed technique can mitigate black hole, gray hole, and cooperative black hole attacks. The black hole can be caught if it replies the RIP RREQs or if it drops data and sends a lot of RREPs compared to the RREQs. The gray hole attack can be detected by the same algorithm whenever it acts as a black hole node. Also, the proposed technique takes the history of the nodes into consideration when it estimates the MTVs which helps in detecting grayholes. Baiting malicious nodes to reply RIP RREQs can detect and isolate one node of the cooperative blackhole nodes whenever it tries to acquire the route. On the other hand in the monitoring process, if a source node needs to communicate with another node in the network, then the source node initiates the route discovery process by broadcasting RREQ. If one node of the cooperative blackhole nodes succeeds to acquire the route, then the source node starts

to send its data packets. Upon the receipt of data packets, the black hole node starts to forward these data packets to other cooperating nodes. Other cooperating nodes forward these data packets to others and so on until one black hole node drops the data packets. In the proposed technique, the backbone network nodes can detect and isolate the black hole node which drops the data packets. After the isolation of one of the cooperative blackhole nodes, one of the remaining cooperative nodes has to reply RIP RREQs or drop data packets which facilitate its detection process by the proposed algorithm. The detection will continue in such way to catch the cooperative nodes one by one.

It can be shown that the proposed technique is more secure than [14, 15], because BBNs perform the security check periodically not based on a request from regular not trusted nodes, BBNs are the only nodes that perform the monitoring, detect and remove the malicious nodes, know the RIPs and send the RREQs for the RIPs.

The overhead of the proposed technique is lower than the other techniques in [10, 14, 15] because it doesn't use acknowledgments [10], it doesn't maintain and exchange a large amount of control data [10], it doesn't send dummy data [15], and it does not use special control packets to exchange its control information instead, it adds fields to the AODV HELLO message.

E. The Removal of Malicious Nodes

The removal process starts after detecting malicious nodes. The backbone network nodes that have the permission to isolate the malicious node start the removal process by adding the malicious node ID into its black list. Also the discovering node broadcasts the malicious node ID to other nodes in the network using additional control fields added to the HELLO message which is already implemented in AODV [7]. Each node receives the information that is integrated in HELLO message checks that the sender is one of the backbone nodes. After the validation step, the node adds the malicious node ID to its black list and adds the control information into its own HELLO message to redistribute the malicious node ID. Each node in the network ignores route replies (RREPs) and route requests (RREQs) that are received from any node in the black list to isolate the malicious nodes from the network. Also, each node deletes any route in its cache to any node in the black list. If all neighbor nodes around the malicious node do not forward their packets, the malicious node cannot communicate with the other nodes in the MANET and the malicious node is isolated from the network [20].

IV. SIMULATION RESULTS

NS2 simulator [6] is used in this paper to evaluate the performance of the proposed technique compared with other recent techniques such as IDS [11], hash-function [21], and AODV [7]. The simulation results are presented in the next subsections.

A. Comparison with IDS technique

In this subsection, the performance of the proposed technique is compared with Ming-Yang Su technique [11] as well as with the original AODV technique [7].

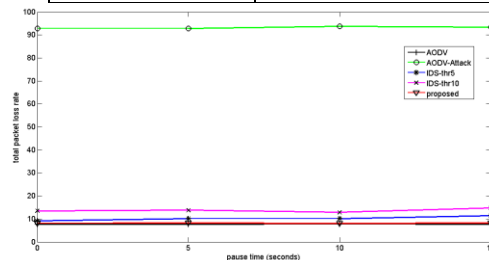
The current comparison is carried out using the same simulation parameters that are used in [11] except that the proposed technique doesn't use extra nodes. The used parameters are listed in Table (I). Random-way-point model [22] is used to allow nodes to move randomly. Each reading in the next figures is the average value resulting from a set of experiments under different scenarios of random movement. In [11], the total packet loss rates are calculated according to the ratio between the number of packets that fail to reach the destinations (missing packets) and the total number of packets that are transmitted from all source nodes of the entire network.

The results of the first comparison between the proposed, IDS, and AODV techniques are shown in Fig. 1. The total packet loss rates of one and two fixed selectively black holes (gray hole) are compared with IDS technique in case of IDS predefined thresholds of 5 and 10. The results are compared also with the ideal not attacked AODV as well as with AODV under attack. As shown from Fig. 1-a, when there is no attack, the mean total packet loss rate for all pause times by AODV is about 7.87%. When there is one fixed selective black hole node the rate by the attacked AODV raises to be about 92.40%. With IDS technique when the threshold value is set to 5, the rate is about 10.05%, and when the threshold is set to 10, the rate is about 13.04%. In the proposed technique the rate is successfully reduced to 8.14%. Fig. 1-b shows the mean total packet loss rate of all pause times when there are two fixed selective black hole nodes. The results are compared with the non attacked AODV in the case of the absence of selective black hole node. When there are no selective black hole nodes, the non attacked AODV gives a rate of about 7.73%, which increases to be about 97.32% when there are two fixed selective black hole nodes. IDS technique gives rate about 11.28% and 14.76% when the threshold values are set to 5, 10 respectively. The rate is successfully reduced to be about 9.83% in the proposed technique. It can be shown from the results that the proposed technique has the lowest mean total loss rate in case of one and two fixed gray hole nodes compared with AODV under attack and IDS. It can be shown that the proposed technique loss rate is very close to the rate of the ideal non attacked AODV.

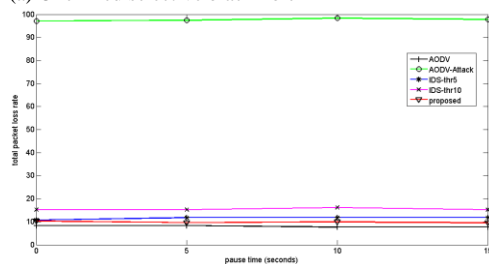
The second comparison between the proposed technique, IDS, and AODV, is carried out when there are one and two random moving selective black hole nodes, the results are illustrated in Fig. 2. As shown from Fig. 2-a, when there is one randomly moving a selective black hole node, the mean total loss rate in AODV under attack is about 86.53%. IDS technique gives rate about 10.29% and 12.55% when the threshold value is set to be 5 and 10 respectively. While in the proposed technique the mean rate is successfully reduced to be about 7.52%, which is close to the ideal non attacked AODV. Also, as shown from Fig. 2-b when there are two randomly moving selective black hole, AODV under attack gives mean total packet loss rate for all pause times about 94.64%, while IDS technique gives rates about 12.03%, and 14.57% with threshold values of 5 and 10 respectively. The rate is successfully reduced to 10.08% in the proposed technique. Also, it can be shown from this comparison that the proposed technique gives the lowest packet loss ratio which is also close to the ideal non attacked AODV ratio.

TABLE I. SIMULATION PARAMETERS

Parameter	Value
Area size	1000 m×1000 m
Normal nodes	50 (distributed and moving randomly)
Connections	20 pairs (40 nodes)
Transmission range	250 m
Traffic type	UDP-CBR (Constant Bit Rate)
Packet size	512 bytes
Mobility	Random-way point model
Maximum speed	20 m/s
Simulation time	500 s
Pause times	0s, 5s, 10s, and 15s
Malicious node(s)	one/two grayhole (fixed/moving)
Traffic rate	5 Kb/second

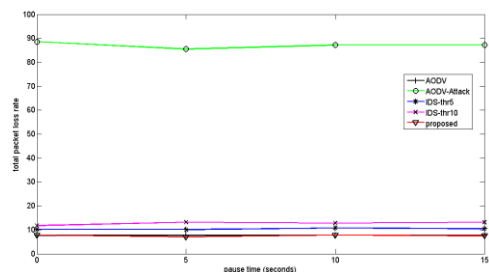


(a) One fixed selective black hole

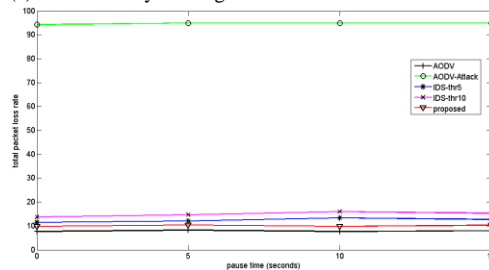


(b) Two fixed selective black holes

Fig. 1. Comparing total packet loss rate in AODV, IDS, and the proposed technique in case of fixed one and two selective black hole (gray hole)



(a) One randomly moving selective black hole



(b) Two randomly moving selective black hole

Fig. 2. Comparing total packet loss rate in AODV, IDS, and the proposed technique in case of randomly moving one and two selective black holes

The authors in [11] give a summary of the true positive rate and false positive rate. A true positive (TP) is defined in [11] as a selective black hole node being correctly judged as a black hole; whereas, a false positive (FP) is a normal node being misjudged as a black hole [11]. The TP rate is defined as the ratio between the number of the TP and the number of black hole nodes while the FP rate is defined as the ratio between the number of the FP and the number of the normal nodes. The results of the comparison between the proposed and IDS techniques are listed Tables (II, III). The results in case of fixed selective black hole(s) are listed in Table (II) and for randomly moving selective black hole(s) are listed in Table (III). As illustrated, the FP and TP of the proposed technique are 0% and 100% respectively in all cases which are better than IDS which gives worst FP rates in some cases. Also, the time of blocking in the proposed technique is better than IDS.

It can be seen from previous comparisons that the proposed technique gives better results than IDS technique and close to the ideal non attacked AODV behavior.

TABLE II. COMPARING TP RATE FP RATE IN IDS, AND THE PROPOSED TECHNIQUE IN CASE OF FIXED 1 AND 2 SELECTIVE BLACK HOLES

(a) TP rate and FP rate for one fixed selective black hole.

Pause time (s)	FP rate			TP rate			Time of blocking (s)		
	IDS (5)	IDS (10)	proposed	IDS(5)	IDS(10)	proposed	IDS (5)	IDS (10)	proposed
0	0(0%)	0(0%)	0(0%)	1(100%)	1(100%)	1(100%)	16.54	21.07	4.65
5	0(0%)	0(0%)	0(0%)	1(100%)	1(100%)	1(100%)	20.13	23.09	5.45
10	0(0%)	0(0%)	0(0%)	1(100%)	1(100%)	1(100%)	21.42	23.08	5.07
15	0(0%)	0(0%)	0(0%)	1(100%)	1(100%)	1(100%)	21.23	21.29	5.75

(b): TP rate and FP rate for two fixed selective black hole.

Pause time (s)	FP rate			TP rate			Time of blocking(s)		
	IDS(5)	IDS (10)	proposed	IDS(5)	IDS(10)	proposed	IDS (5)	IDS (10)	proposed
0	0(0%)	0(0%)	0(0%)	2(100%)	2(100%)	2(100%)	20.60	21.39	5.65
5	0.3(0.6%)	0(0%)	0(0%)	2(100%)	2(100%)	2(100%)	21.68	23.06	5.35
10	0(0%)	0(0%)	0(0%)	2(100%)	2(100%)	2(100%)	21.31	22.76	5.6
15	0.2(0.4%)	0(0%)	0(0%)	2(100%)	2(100%)	2(100%)	21.27	23.03	5.76

TABLE III. COMPARING TP RATE FP RATE IN IDS, AND THE PROPOSED TECHNIQUE IN CASE OF RANDOMLY MOVING 1 AND 2 SELECTIVE BLACK HOLES

(a): TP rate and FP rate for one randomly moving selective black hole.

Pause time (s)	FP rate			TP rate			Time of blocking (s)		
	IDS(5)	IDS (10)	proposed	IDS(5)	IDS (10)	proposed	IDS (5)	IDS (10)	proposed
0	0(0%)	0(0%)	0(0%)	1(100%)	1(100%)	1(100%)	21.08	23.19	5.34
5	0(0%)	0(0%)	0(0%)	1(100%)	1(100%)	1(100%)	21.19	21.99	6.14
10	0.2(0.4%)	0(0%)	0(0%)	1(100%)	1(100%)	1(100%)	21.06	21.24	5.21
15	0(0%)	0(0%)	0(0%)	1(100%)	1(100%)	1(100%)	21.19	21.37	5.45

(b): TP rate and FP rate for two randomly moving selective black hole.

Pause time (s)	FP rate			TP rate			Time of blocking (s)		
	IDS(5)	IDS (10)	proposed	IDS(5)	IDS(10)	proposed	IDS (5)	IDS (10)	proposed
0	0.1(0.2%)	0(0%)	0(0%)	2(100%)	2(100%)	2(100%)	21.24	21.23	5.72
5	0.1(0.2%)	0(0%)	0(0%)	2(100%)	2(100%)	2(100%)	21.27	27.7	5.75
10	0(0%)	0(0%)	0(0%)	2(100%)	2(100%)	2(100%)	21.22	23.58	5.48
15	0(0%)	0(0%)	0(0%)	2(100%)	2(100%)	2(100%)	21.17	21.43	5.85

B. Comparison with Hash-Function Technique

In this subsection, the proposed technique is compared with the technique that is introduced in [21]. The technique uses hash function and message authentication to mitigate black hole attack. The simulation parameters are the same parameters that are used in [21]; these parameters are the same parameters which are listed in Table (I) except that the simulation time is 600 s, the pause times are 0s, 100s, 200s, 300s, 400s, 500s, and 600s, the malicious node is one randomly moving black hole, and the traffic rate is 4 packets/second.

Random-way-point model [22] is used to allow nodes to move randomly. Each reading in the figures is obtained as an average of a set of experiments under different scenarios of random motion.

Authors in [21] use three performance measures; packet delivery ratio, time delay, and normalized control packet overhead. They defined Packet Delivery Ratio as the ratio between the number of data packets successfully delivered to the destinations and the total number of data packets in the network. They defined Time Delay as the difference between the time when the source node broadcasts a RREQ message and the time when the first data packet is received by the destination node. They considered that Normalized Control Packet Overhead is the ratio between the size of all the routing packets and the size of all received data packets.

In the following comparisons, the proposed technique is compared with Hash-Function [21] and AODV [7] techniques

In the first comparison, the packet delivery ratio is obtained for different pause times, the comparison results are shown in Fig. 3. It can be shown that the proposed technique gives the highest packet delivery ratio.

The second comparison is carried out to compare the Time Delay; the results are shown in Fig. 4. It can be shown that the proposed technique introduces the lowest time delay. This comparison is not fair to the proposed technique because the delay is considered for the packets that reach the destination and doesn't take into consideration the packets that will be dropped by the black hole. The delay should be related to the delivery ratio which is not considered by [21].

Fig. 5 shows the results of the third comparison in which the Normalized Control Packet Overhead is compared. As shown, the proposed technique gives the lowest normalized control packet overhead which is close to non attacked AODV.

C. The backbone network security and coverage

The simulation parameters that are used in this subsection are the same parameters that are used in the previous subsection. The first experiment is carried out to see if a malicious node can deceive the backbone network and join it as backbone member. The experiment is carried out with number of malicious nodes equals to 20%, 30%, 40%, 50% of the total number of nodes. The results proved that there is no malicious node could join the backbone network.

In the second experiment, the backbone network coverage is tested and the result is illustrated in Fig. 6. As shown from the figure, by using a low percentage of the backbone nodes, the proposed technique gives high coverage percentage.

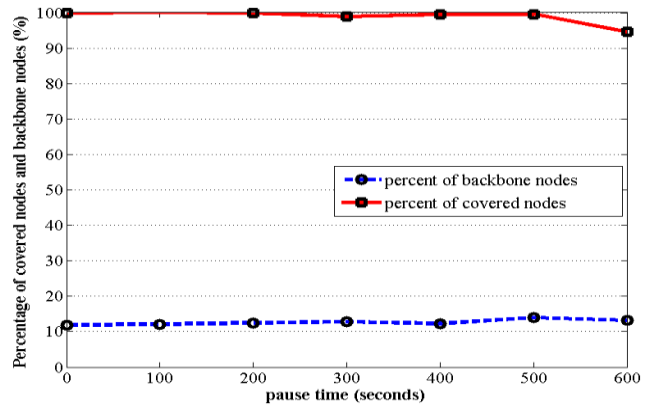


Fig. 6. Backbone network coverage

V. CONCLUSION

A reliable technique is proposed to detect and prevent black, gray, and cooperative black hole attacks in AODV. The proposed technique uses a multi-level mobile backbone network constructed of randomly moving regular MANET nodes chosen based on their trust value, location and power. The backbone network monitors each other as well as regular nodes to estimate monitoring trust value for each node. This value is used as an indicator of malicious behavior. Also, higher level backbone nodes bait the malicious nodes to reply a request for a route to non existing IP. AODV HELLO messages are used to isolate the suspicious node and to exchange the control information. The performance of the proposed technique is compared with AODV, IDS and hash-function techniques. The simulation results show that the backbone network is secure and has high coverage. The proposed technique can highly detect and remove the malicious nodes. It gives the lowest packet loss rate, the lowest end-to-end delay, and the lowest packet overhead.

REFERENCES

- [1] S. Basagni, M. Conti, S. Giordano, and I. Stojmenovic, "Mobile ad hoc networking", Wiley, 2004.
- [2] D. Djenouri, L. Khelladi, and N. Badache, "A survey of security issues in mobile ad hoc networks", IEEE communications surveys, Vol. 7, No. 4, pp. 2-28, 2005.
- [3] B. Kannhavong, H. Nakayama, Y. Nemoto, N. Kato, and A. Jamalipour, "A survey of routing attacks in mobile ad hoc networks", Journal of Wireless communications, IEEE, Vol. 14, No. 5, pp. 85-91, 2007.
- [4] N. Badache, D. Djenouri, and A. Derhab, "Mobility impact on mobile ad hoc routing protocols", ACS/IEEE International Conference on AICCSA, Vol. 3, 2003.
- [5] D. Cerri, and A. Ghioni, "Securing AODV: the A-SAODV secure routing prototype", Communications Magazine, IEEE, Vol. 46, No. 2, pp. 120-125, 2008.
- [6] The Network Simulator ns-2, <http://www.isi.edu/nsnam/ns/>.
- [7] C.E. Perkins, E. Beliding-Royer, S. Das, "Ad hoc on-demand distance vector (AODV) routing", IETF Internet Draft, MANET working group, Jan. 2004.
- [8] S. Kamboj, and M. Dua, "Comparison Study of Various DoS Node Detection Schemes in MANETs", International Journal on Computer Science and Emerging Trends (IJCSSET), Vol. 2, No. 1, pp. 8-15, 2013.
- [9] S. Jain, J. Singhai, and M. Chawla, "A Review Paper on Cooperative Blackhole And Grayhole Attacks in Mobile Ad hoc Networks", International Journal of Ad Hoc, Sensor & Ubiquitous Computing, Vol. 2, No. 3, 2011.

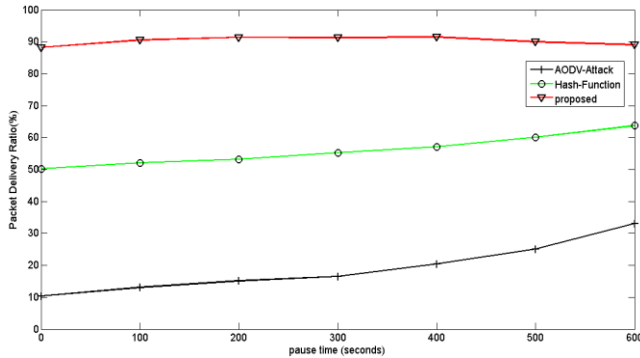


Fig. 3. Total packet delivery ratio for proposed, Hash-Function, and AODV techniques

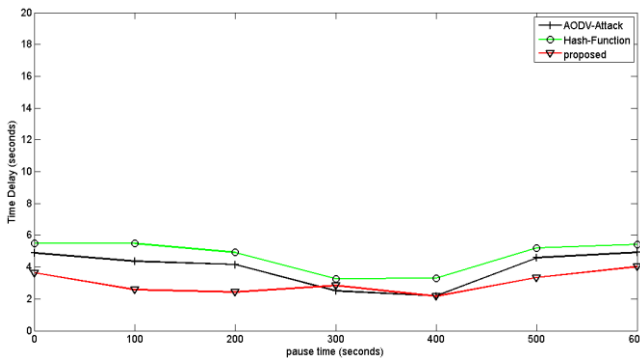


Fig. 4. Time delay for proposed, Hash-Function, and AODV techniques

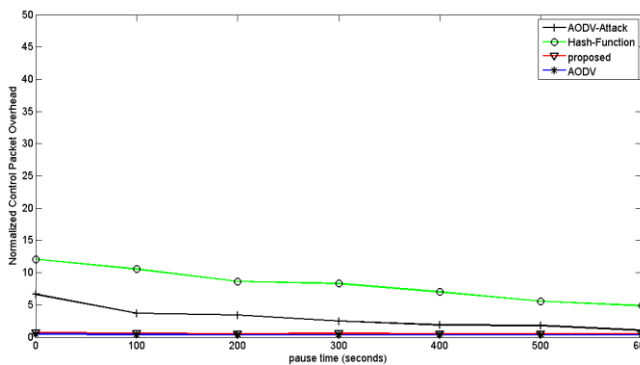


Fig. 5. Normalized Control packet overhead for proposed, Hash-Function, and AODV techniques

- [10] L. Tamilselvan, and V. Sankaranarayanan, "Prevention of co-operative black hole attack in MANET", *Journal of networks*, Vol. 3, No. 5, pp. 13-20, 2008.
- [11] M. Su, "Prevention of selective black hole attacks on mobile ad hoc networks through intrusion detection systems", *Journal of Computer Communications*, Elsevier, Vol. 34, No. 1, pp. 107-117, 2011.
- [12] H. R. Jhaveri, S. J. Patel, and D. C. Jinwala, "Improving route discovery for aodv to prevent blackhole and grayhole attacks in manets", *INFOCOMP Journal of Computer Science*, Vol. 11, No. 1, pp. 1-12, 2012.
- [13] M. B. Jani, and H. Patel, "Mitigation of Blackhole for AODV (Ad hoc On Demand Distance Vector)", *International Journal of Computer Science and Mobile Computing*, Vol. 2, No. 5, pp. 338-345, 2013.
- [14] P. Agrawal, R. K. Ghosh, and S. K. Das, "Cooperative black and gray hole attacks in mobile ad hoc networks", *Proceedings of the ACM 2nd international conference on Ubiquitous information management and communication*, pp. 310-314, 2008.
- [15] K. Vishnu, and A. J. Paul, "Detection and Removal of Cooperative Black/Gray hole attack in Mobile Ad Hoc Networks", *International Journal of Computer Applications*, Vol. 1, No. 22, pp. 38-42, 2010.
- [16] S. Indrasinghe, R. Pereira, and J. Haggerty, "Conflict free address allocation mechanism for mobile ad hoc networks", In *IEEE 21st International Conference*, vol. 1, pp. 852-857, 2007.
- [17] F. H. Tseng, L. D. Chou, and H. C. Chao, "A survey of black hole attacks in wireless mobile ad hoc networks", *Journal of Human-centric Computing and Inf. Sciences*, Springer, Vol. 1, No. 1, pp. 1-16, 2011.
- [18] R. Agarwal, and D. Motwani, "Survey of clustering algorithms for MANET", *International Journal on Computer Science and Engineering*, Vol. 1, No. 2, pp. 98-104, 2009.
- [19] J. MacQueen, "Some methods for classification and analysis of multivariate observations", *fifth Berkeley symposium on mathematical statistics and probability*, Vol. 1, No. 281-297, p. 14, 1967.
- [20] P. Yi, Z. Dai, Y. P. Zhong, and S. Zhang, "Resisting flooding attacks in ad hoc networks", *IEEE International Conference on Information Technology: Coding and Computing (ITCC)*, Vol. 2, pp. 657-662, 2005.
- [21] P. Sachan, and P. M. Khilar, "Authenticated Routing for Ad-Hoc On-Demand Distance Vector Routing Protocol", *Advances in Network Security and Applications*, Springer, pp. 364-373, 2011.
- [22] C. Bettstetter, G. Resta, and P. Santi, "The node distribution of the random waypoint mobility model for wireless ad hoc networks", *IEEE Transactions on Mobile Computing*, Vol. 2, No. 3, pp. 257-269, 2003.

Adoption of e-Government in Pakistan: Demand Perspective

Zulfiqar Haider¹, Chen Shuwen²

Dalian University of Technology,
Faculty of Management and Economics,
No.2 Linggong Road, Ganjingzi District,
Dalian 116023, Liaoning, P.R. China.

Zareen Abbassi³

University of Sindh,
Department of Public Administration
Jamshoro, Pakistan.

Abstract—The reason for this paper is to investigate the variables that empower citizen adoption of e-Government driven organizations in Pakistan, where these offices are at a simple stage. Comprehension citizen's adoption of electronic-government is an essential topic, as the utilization of e-Government has turned into an integral part of administration. Achievement of such activities depends generally on the productive utilization of e-taxpayer supported organizations. Inclusive e-Government is the gateway to the efficiency promised by the Electronic Government. This study utilizes the Unified Theory of Acceptance and Use of Technology (UTAUT) model to inspect the powerful elements of the adoption and utilization of e-Government services in Pakistan from a national point of view. An online survey was led and a factual spellbinding examination was performed on the reactions got from 200 Pakistani nationals. The embraced model can be utilized as a rule for the execution of e-Government services in Pakistan. This study recommends that government ought to run broad publicizing battles to guarantee that individuals are mindful of the services and utilization them. This infers that government ought to place accentuation on expanding familiarity with the services, show the profits of citizens, and empowering confidence in the framework.

Keywords—E-government; adoption; demand perspective; Pakistan

I. INTRODUCTION

This The purpose of this paper is to study the adoption of e-Government in Pakistan through demand perspective. The strategy of e-Government is a fundamental in the modernization of the public sector, which not only allowing the demand of information but also optimize the processes commercially and facilitate communication between the different levels of government[30]. E-Government facilitates the development of public administration activities related to citizens and different companies [25]. E-Government has turned into a mainstream center of government efforts in numerous countries around the globe. It mirrors the expectation for open associations and governments to exploit the correspondences upgrades made conceivable by the ICT insurgency. E government is adopted with the reason for enhancing the services and delivery gave by the government to its citizens [15].

E-government services systems aim to give numerous benefits such as enhancing the processes and operations of government services and upgrading information offering between the government and open [5]. It likewise gives citizens the services in expert way, safely, securely, helpfully,

and with extensive time reserve funds. Nonetheless, the usage of e-Government services is not a basic on-line information procurement, it obliges a profound comprehension of citizens' needs and prerequisites and complete architecture to evade unexpected results [11].

E-government and Internet has rolled out a vital improvement in the entire Pakistani society structure, qualities, society furthermore, the methods for leading business by using the capability of ICT as an instrument in day by day work. In this paper, we distinguish the components that impact Pakistani citizens to acknowledge and utilization e government by applying a corrected UTAUT Model. UTAUT is an exactly accepted model consolidating eight noteworthy models of technology acceptance and their extensions [22]. The outcomes can be utilized by the chiefs and administration creators to progress e-taxpayer driven organizations, and their availability to citizens.

The structure of the paper is as per the following: at to start with, it offers a brief synopsis of past research, a depiction of e-Government in Pakistan, and a diagram of e-Government adoption in developing countries (i.e. South Asia). In the following area, the research methodology and systems are examined. After that, the attained to results are reported in the connection of the related research, alongside a legitimacy and unwavering quality talk [12]. At last, the paper is finished up by inspecting ramifications, constraints and recommendations for future research.

TABLE I. FBR- ONLINE TAX CASES

S/#	Type	Total cases	Selected
1	Income Tax (Corporate)	25,046	1,876
2	Income Tax (Non-Corporate)	840,675	63,050
3	Sales Tax (Corporate)	11,757	1,410
4	Sales Tax (Non-Corporate)	92,455	11,095
5	FED Tax (Corporate)	402	45
6	FED Tax (Non-Corporate)	202	24

Random computer ballot was conducted in respect of six categories i.e. corporate cases of Income Tax, Sales Tax, FED and non-corporate cases of Income Tax, Sales Tax and FED.

The target audience of this paper will be composed citizens or government employees who are paying their taxes online and the issues they are facing in reference to UTAUT model.

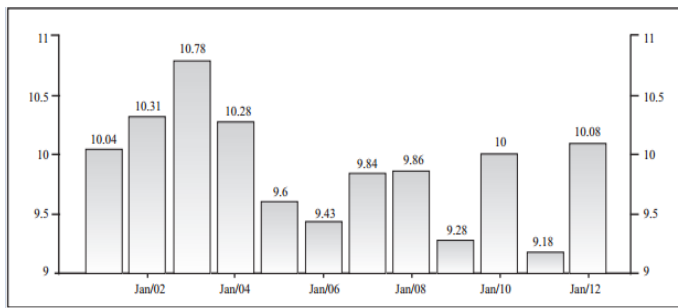


Fig. 1. Online Tax Payers (Obtained from FBR-Pakistan)

UTAUT is one of the most recent developments in the field of general technology acceptance models. This model is used to present a more finish photo of the acceptance process than any past individual models had possessed the capacity to do. Eight models beforehand utilized as a part of the IS writing were converged in an incorporated model, all of which had their starting points in brain research, social science and communications. These models are the TRA, TPB, TAM, TAM2, the Motivational Model of Computer Usage (MM) [12] [21].

The UTAUT holds that four key builds (execution anticipation, exertion hope, social impact, and encouraging conditions) are direct determinants or predictors of usage plan and conduct. Gender, age, experience, and willfulness of utilization are placed to intervene the effect of the four key builds on usage aim and conduct Government

The study of e-Government from the perspective of demand discusses the factors that determine their acceptance, considering those that play an important role in models consumer behavior online will also affect adoption. Most of the citizens in Pakistan are still unaware of online taxpaying system due to the barriers in implementation of e government. Despite this, must be aware of the existence of particularities associated with the incorporation of ICT in the public sector, among them, the complex structure of objectives and Administration performing multiple functions.

Although the public sector enjoyment authority to impose the use of e-Government, their acceptance voluntary lead to an optimized uses [14]. The most widely used model to explain the behavior of Internet consumer is the Technology Acceptance Model (TAM), is an adaptation of the Theory of Reasoned Action (TAR) cantered behavior using new technologies. The TAM replaces beliefs defined in the TAR 2 items denominated usefulness and perceived ease of use, low assuming that these are beliefs that influence the formation of attitudes and, consequently, in the intention and behavior an individual about the use of technology [11]. Meanwhile, the model Diffusion of Innovation (DOI) this is modeled as a process of information collection and reduction of uncertainty with the intention of assessing technology, identifying five elements that can influence it: relative advantage, complexity, compatibility, testability and ease of observation of their benefits [8].

The research question of this study is “What are the fundamental factors and demand perspective that affect the citizens and e-Government adoption in Pakistan?”

The research question can be formed in the accompanying way: RQ1. What is the current circumstance in Pakistan concerning the factors influencing e-taxpayer driven organizations selection and what are the ramifications of this to future organization? This research question was replied by tending to factors of the UTAUT model and directing a survey to portray the current circumstance in Pakistan. Will there be a statistically significant relationship between perceived trust in e-government and behavioral intention to use e-government services? What is the relationship between government policies, strategies and deployment of e-Government in Pakistan? To identify barriers in adoption of e-Government using UTAUT model?

A. Research Aim

First, The aim of this research is to explore the demand-side barriers to e-market in Pakistan Demand Side perspective in public using UTAUT model and the paper presents an in-depth empirical case study of such barriers from a local authority perspective [12]. The aim of this paper is to study the factors that influence the Pakistani citizens for the e-market adoption.

B. Research Method

As already demonstrated, this research is a quantitative study, which develops the UTAUT model to examine the level of change of e-Government appropriation as a result of incorporating the trust build in the model. As it were, the study's center was to show whether trust could enhance the prescient estimation of the UTAUT model, while at the same time testing the unified model's limits, to disclose aim to adopt the e-Government administrations in Pakistan. The research used a correlation research outline with structural comparison displaying (SEM) to focus the degree of connections among the research variables. Also, numerous regression analyses were utilized to anticipate citizens' plan and utilization behavior.

II. LITERATURE REVIEW

This study identifies the determinants of potential users' adoption of e-Government services in a creating nation utilizing a changed form of the UTAUT model. The findings revealed that performance expectancy, effort expectancy, peer influence and facilitating conditions were noteworthy in the adoption of e-government services in Pakistan. These findings are reliable to some degree with those reported in studies led in created countries i.e. [9], [10],[11], [22] and [17], in this way demonstrating that variables, which focus the adoption of e-Government services, identified in created countries could be appropriate in the setting of creating countries. Regardless of the way that the utilization of understudy subjects may have restricted the generalizability of the findings of this study, the research gives helpful experiences into the motivations basic the intentions to utilize e-Government services in creating countries. The likely adoption of e-Government services by understudy subjects is all around anticipated on the premise of the key variables performance expectancy, effort expectancy also, associate influence, and on the premise of facilitating conditions and the directing impacts of Internet experience and sort of scholarly course. Future research could, for instance, incorporate a more extensive scope of members and spread different variables, for example, society what's more, trust. The

use of ICT radically changes the way governments manage and execute their internal processes, and opens the possibility to improve and increase communication channels with citizens [17] [12]. In this paper a conceptualization of e-Government based on its essential features is an integration between its phases and dimensions is performed and considerations to be taken into account for their implementation taking into account the advantages and disadvantages are analyzed.

E-government is seen as “the realization of efficient and effective action by the recipients oriented government, with main emphasis on the citizen, and interaction with these” and its phases (presence, interaction, transaction processing are analyzed and citizen) participation and (external, promotion, internal and relational) dimensions and the interaction between them [13]. These phases and dimensions are not interdependent nor need to complete one to start another. Each has a different purpose and requires different requirements in terms of costs, needs knowledge and level of ICT use [8] [9]. Finally, an analysis of the advantages and disadvantages that must be taken into account to take advantage of early and minimize the latter, during the implementation process of e-Government. 2.1 UTAUT Model

The unified theory of acceptance and use of technology arises from the need to create a common theoretical reference to retake the constructs that other theories and models have proven that were useful in assessing acceptance of technology and realized an evolution in explaining the phenomenon.

The theory constructs studied and formulated a unified integrated them model. Of all the constructs identified four main factors: performance expectancy, effort expectancy, social influence and ease conditions, which are moderated by gender, age, experience and willingness to use; the latter refers to whether the use of technology is voluntary or imposed [41]:

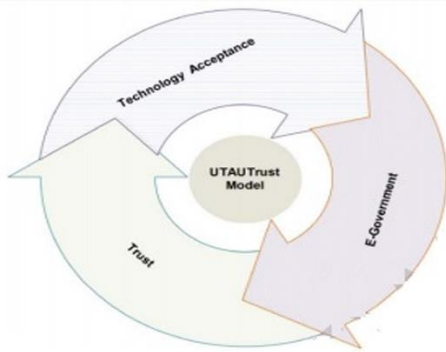


Fig. 2. UTAUT Model

1) Performance Expectation

The degree to which a person believes that using the system will help to make gains in performance. This factor is moderated by gender and age.

2) Expectation of effort

The degree of ease associated with the use of the system. This factor is moderated by gender, age and experience.

3) Social influence

The degree to which a person perceives that other important people believe he should use the system. This factor is

moderated by gender, age, experience & willingness to use [23].

4) Terms of ease

The degree to which a person perceives that the organizational structure and technical infrastructure can support the use of the system. This factor is moderated by age and experience.

The UTAUT model contains five direct determinants of behavioral intention and utilization conduct: (1) performance expectancy, which is "the degree to which an individual accepts that utilizing the system will help him or her to accomplish picks up in employment performance". (2) Effort expectancy, which is "the level of straightforwardness connected with the utilization of the system". (3) Social influence, which is "the degree to which an individual perceives that vital others accept he or she ought to utilize the new system". (4) Facilitating conditions, which is "the degree to which an individual accepts that a hierarchical and technical infrastructure exists to bolster the utilization of the system". And (5) behavioral intention, which is "the individual's subjective likelihood that he or she will perform the conduct in question". In fact, the system of e-Government can successfully replace the need to visit government agencies and reduce the volume of paper records management [44].

It is accounted for that tele density has encountered 6.7 percent development and portable supporter numbers have increment by 10 percent, twofold 2010 figures. Extra versatile infiltration rose to 65.4 percent from 60.4 percent, while broadband client numbers had expanded by 66 percent toward the end of FY2011. Like other creating countries, Pakistan has difficulties, for example, poor IT foundation, low education rates, moderate e-government services improvement, and adoption. As indicated by Almakki (2009), Arab countries have difficulties, for example, the absence of IT base, and additionally social issues. The development of e-Government user-centered requires knowing both the expectations that this has regarding its use, as the elements that favor its adoption. Chan et al. (2010) have proposed the existence of diverse backgrounds of the components UTAUT model in a binding environment. This paper, through the development of a structural equation model using PLS analyzes the effect of such records in a volunteer environment [25] [39]. The results show outcome expectations and the expectations of effort significantly impact on the intention to use the platform of e-Government, and that this influence and facilitating conditions on the use of that platform.

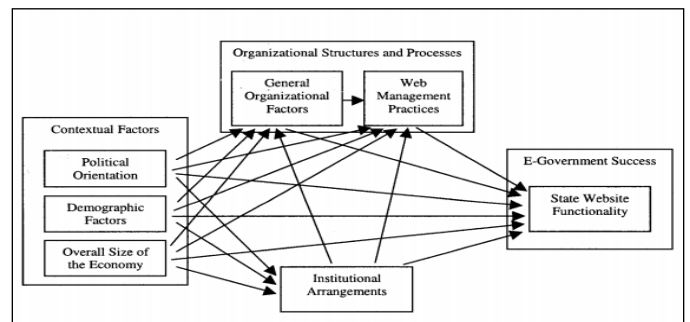


Fig. 3. e-Government adoption model

The levels of e-Government in the Pakistan presents a systematic collection of experiences in which information technologies are used in support of governmental activities that are currently running in the group of countries that have been included in the guide: Bolivia, Chile, Colombia , Ecuador, Peru and Venezuela, Honduras, Nicaragua, Panama, Mexico, Guatemala and Costa Rica [50]. These models of experiences can serve as a basis for developing solutions in other countries, taking into account the importance of the use of ICTs to increase efficiency and effectiveness of public functions, facilitate government-citizen relationship and strengthen national strategies towards promotion of transparency and integrity [14].

Use of Information Technology in Administrations was appreciated by a key element in its modernization, needed to reach the most competitive and dynamic knowledge in the world, capable of sustainable economic growth accompanied by an improvement quantitative and qualitative employment and greater social cohesion [32].

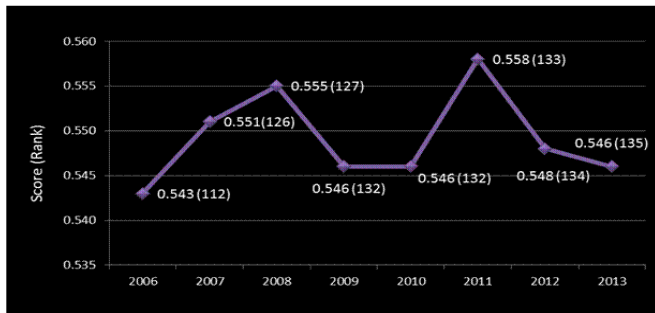


Fig. 4. Recent Changes in E-government

The government agency must provide outside information about itself and the processes that takes place. The level of transparency measures an agency effort to make information available through its website [25]. Transparency is not only reflected in the amount of information, but also in the quality thereof. The five categories are included in the term transparency are:

- 1) *Property: the evidence that the government manages the site and its content.*
- 2) *Contacts / accessibility: how and whom to contact in relation to the activities carried out by the institution.*
- 3) *Information of the institution: its structure and activity area.*
- 4) *Information content: information policy based on the institution.*

There is a vital need that electronic governments across the globe should allow anyone visiting the city website to communicate and interact with employees of the city via the Internet with graphical user interfaces, instant messaging and audio-visual presentations. The government should focus on:

- 1) *The use of ICT, and particularly the Internet.*
- 2) *The use of ICT in all facets of the government organization.*
- 3) *Coordination structures directly by e-government projects and structures to support its action.*

4) *Monitoring the implementation of decisions and recommendations on the program of e-government.*

5) *Monitoring the implementation of e-government projects in view of data and reports prepared by public structures.*

6) *Liaison with the program coordinators of e-government in the various ministries and coordination with them on projects of e-government.*

7) *The development of the technical committee of the communicating administration proposals regarding the administrative, legal and technical issues rose through project monitoring e-government as well as the problems encountered [41].*

8) *Monitoring of projects of international cooperation within the framework of e-government (Welch et al, 2005, p.391).*

A. E-Government in Pakistan

Traditional studies on e-government in the Pakistan focus on observing socioeconomic determinants of local e-Government from an entrepreneurial approach. The results of various researches show that in major municipalities worldwide, all political parties, regardless of their ideology and political stability of government, they focus on promoting the development of a participatory and dynamic e-government. The decline of public trust in government has inspired various proposals for government reform, or market-based approaches for entrepreneurs, in order to improve efficiency and effectiveness, as well as other proposals focused on increasing the participation of citizens in the political process[35][36]. In connection with these recent reforms, in the Pakistan people consider electronic government as a solution to increase in communication of citizens with Public Administration bodies.

TABLE II. PERCENTAGE CHANGE IN ADOPTION OF E-GOVERNMENT

e-Government in Pakistan		
Services	Response %	Response Count
Elections & voting	91.70%	309
Emergency services	85.50%	288
Courts	84.00%	283
Police services	81.30%	274
Economic development	78.30%	264
Road construction	76.30%	257
Public health services	75.10%	253
Corrections	73.90%	249
Building permits	64.40%	217
Parks and recreation	61.70%	208
Children/family services	60.80%	205
Code enforcement	59.90%	202
Animal control	59.60%	201
Street maintenance	56.70%	191
Library	54.60%	184
Welfare services	54.60%	184

There are many frameworks that are being chalked out at the international as far as defining the way e-Government will work and the overall domain it has to follow. The main objective of the framework is to make sure that the efforts that is being put by the countries as far as the implementation of e-Government is concerned does not go wasted and there must be some utility and benefit to it [45] [41]. The approach that is

common in more or less all of these frameworks is that they must be feasible for developing countries and these e-Government services can be utilized in a better manner so make sure that the society and the economy benefits from it [15]. One of the main constituent of all the frameworks is the indicators that are being dished out at the international level (Schmutzer, 2000, pp. 379).

Among the specific online initiatives taken to improve the quality of services, include:

1) *Development of thematic online portals or targeting specific groups, which gather information and transactions related to the subject or group.*

2) *Mechanisms to categories of users to target their requests (e.g. the ability to extract a specialized portal data on enterprises according to their size in order to help small businesses to find more information easily of interest or facilitate access to information by geographic area).*

3) *Use of e-mail lists to communicate information.*

4) *Services that enable individuals to identify users to access information and services tailored to their needs. It may be, for example, access to targeted information or the opportunity to submit a tax return or other forms, or request assistance or submitting comments online.*

Some observers believe that increased access to e-Government allow greater interest, knowledge and policy debate. The Internet allows groups and communities to deliberate in new ways, which can be more effective. Some observers have speculated that the team will direct democracy, the people vote on a wide range of topics. Currently, little evidence shows that this potential is realized. The tendency to simplify and distort information in public discourse is not appeased by changes in the media. Unequal access to the Internet and a wide range of electronic data and communication tools, more or less are divided between people who have studied and people without a strong correlation with income and political participation, creating a digital divide in e-Government despite progress in HCI. Illiteracy and lack of computer skills exacerbate the digital divide in access [27].

Electronic voting is still rare. It covers two types of devices is the electronic voting machines (similar to computers) based on the Internet and to vote by absentee ballot. Various studies reveal that in Europe only a few countries use more or less large scale electronic voting due to below reasons:

1) *More or less of the electorate voting part using electronic devices in the Netherlands, Belgium and Germany.*

2) *Electronic voting is introduced very gradually in Switzerland since 2003.*

3) *Ireland has started to introduce electronic voting in 2002, but suspended its experience.*

4) *In England and Wales, electronic voting is experienced since 2000 during local elections in several communities, but its generalization is not considered.*

5) *In Spain, Italy and Portugal, electronic voting gives rise to the moment of tests without legal value.*

The e-Government Survey 2012 finds that many countries are moving from a unilateral decentralized organizational

model to an integrated, unified, which aims to focus on a single portal services provided by the government to generate greater transparency. Another important point has to do with mobility. The massive development of e-Government initiatives has a multiplier effect in the field of Information Technology. The adoption of technology allows governments and it is public to improve the quality of services to government agencies. Consequently, e-Government portals could become a communication and consultation mechanism immediately, expeditious and low cost in accounting harmonization, thereby encouraging collaboration between states and municipalities to eliminate the drawbacks, shortcomings and to strengthen transparency (Hung et al, 2006, p.97).

One of the things that are very important to understand is that why there is a need for some sort of better convector and indicators that can gauge the efficiency and the overall competency of the e-Government indicators. There is a consensus that the international level that effective policies of the public sector are almost certain to need ICT to make sure that there is certain smoothness as far as operations is concerned (Welch et al, 2005, p.391). Through these indicators, it can be made sure that the overall quality of the service improves and individuals are in a better position to reap some of the benefits of the e-government.

These indicators might also help the government to make sure that all the relevant policies are chalked out in a better manner [36] [11].

Some of the standards are being developed by the international organizations as well as there is some role of academics universities and research institutes that might be aiding in this task (Carter & Bélanger, 2005, p.5). While some of the users of the e-Government advocate the usage of ICT, there are some other measurements services as well that are in place and they are offered through the official website of the government [9] [8].

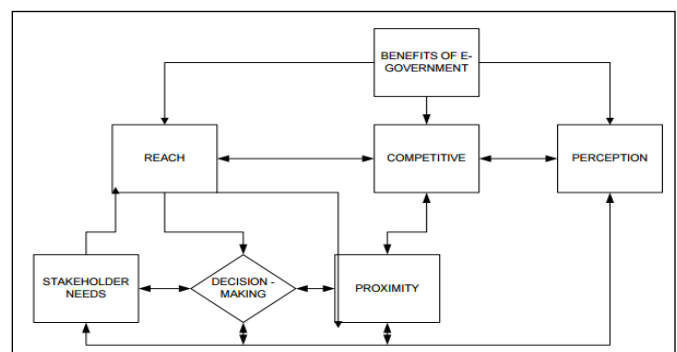


Fig. 5. Benefits of model

III. METHODOLOGIES

As far as the methodologies for the data collection are concerned, some of the surveys that are carried out at the international level and some random web based surveys that are complex in there processing and nature are taken into the consideration. UTAUT proposes four constructs (performance expectancy, effort expectancy, social influence and conditions of the facilities) that influence the intended use and / or use of technology.

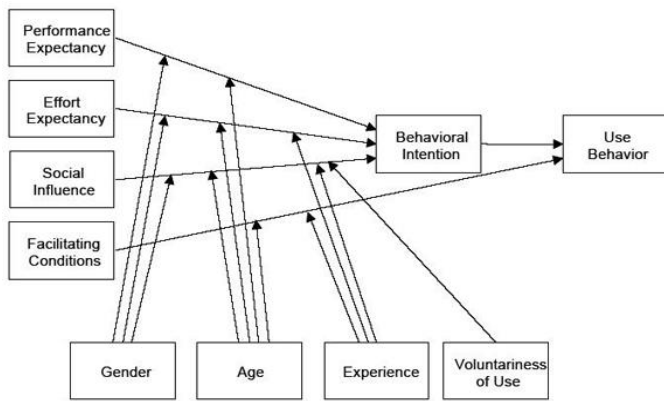


Fig. 6. Study Model (UTAUT)

A. Data Analysis

The goal of the research exhibited in this paper was to study e-Government services in Pakistan from the citizen's perspective. The picked procedure was to utilize the fitting innovation adoption display as a premise for the planning the observational research. The observational information was gathered through an online survey. The largely research procedure is explained as the online survey was conducted for this purpose. The study started with a writing survey and examination. In this stage, pertinent writing on e-Government and related research was gathered and broke down. Alongside that, the different innovation acknowledgement hypotheses also, models were assessed. In this study, quantitative research routines that incorporate polls survey were utilized to direct an interpretive study. The survey was uploaded using internet and utilize. By analyzing, the data collected at different times and on similar issues, this research describes and explains the changes. Secondary data is also used for comparison purposes. Comparisons among the variables can increase the generalization, as well as create additional context. Secondary analysis is comparable with other surveys conducted at different times, creating a unique opportunity to make an empirical description of the long-term changes. The results enjoy more confidence, if repeated in several studies. Secondary data analysis had improved the quality of measurement by extending the scope of the independent variables included in the operationalization of concepts. Using secondary data increases the size of the sample, its representativeness and the number of observations leading to broader applications. Secondary data may be used in the method of triangulation, thereby increasing the accuracy of the results obtained from the analysis of primary data.

B. Framework Creation

The final step of the research approach is to understand the results of research and contextualize them in a framework. The framework will provide considerations for managers seeking to address business challenges in developing countries. The framework will allow the readers to understand research findings, and allow future researcher to improve the framework. UTAUT model is used to understand the implementation of new technology (e-government) in Pakistan.

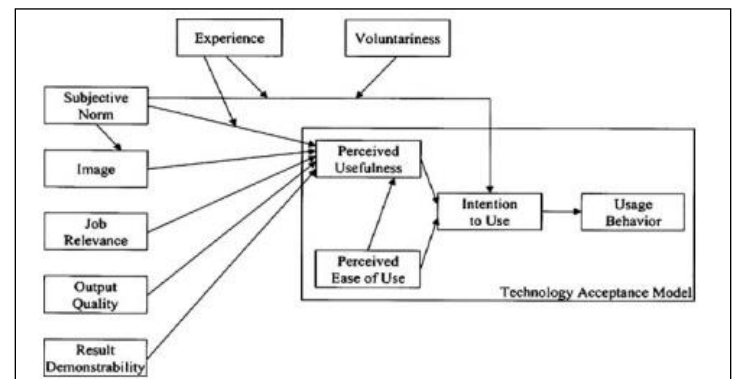


Fig. 7. TAM model (UTAUT Factor)

E-Government indicators are one of the more important policy tools that are in the hand of the Government, they reflect and talk about the overall status and the trends that are being observed as far as policy making is concerned. The collection of the data is one of the compound challenges that are being faced by the Government (Jaeger & Thompson, 2003, p.389).

IV. CHALLENGES

The majority of the developed countries have encountered profits from e-Government administrations; anyhow, there is still much opportunity to get better internationally. Like whatever other advancement, e-Government administrations make various challenges for citizens, and in addition for governments. These challenges incorporate absence of access to e-Government administrations, trust, security concerns, and the advanced separation. In the South Asian area, most gateways and government sites stayed lethargic in 2010 (UN, 2012). In 2010 and 2012, UN e-Government world surveys positioned Pakistan 146th and 156th, respectively. However, as a whole the South Asian locale relapsed in the 2012 survey and stays far beneath the world average [24].

E-Government indicators are classified into four main categories as far as their role and scopes are concerned. These four areas of discussion are:

Use of the ICT is being done by the employees that are working in government institutes. The overall feasibility of ICT as well as some of the government organizations likes the usage of internet. Use of ICT that is being carried out by some of the government organizations. The overall provision to citizens of some country by services by some of the citizens can be evaluated by it.

This has to be kept in mind though that the list is not comprehensive in its nature and it thus provides the starting point to be researched done on this domain [48].

As far as reporting at the international level is concerned, it has to be kept in mind that countries should always be willing to provide the statistical treatment that is enabling some of the government organization to make sure that the metadata is enough to take some necessary decisions about the performance of the e-Government sector (Jaeger & Thompson, 2003, p.389).

There is a statement that is needed to be issues and it takes into account the reference date that has been used as all along the data and use of the different terms that is in the place to make sure that all of them can be addressed.

This is the subset of the general government sector. The general government sector does not take into account some of the public corporations (Hung et al, 2006, p.97). There may be an instance when there will be a government units might be having an entity that is not incorporated as the government entity and still the central government is exercising its control over it, they will be taken as the part of the general government sector.

TABLE III. EXCEL (OUTPUT) SURVEY N=200

Challenges involved in E-Government Adoption		
Variable	Mean (Std. Dev.)	
	High (n=213)	Low (n=1,600)
e-Government score	10.30 (1.52)	4.40 (2.40)
Population	20,088 (20,916)	18,194 (19,676)
Budget	4,426,264 (4.83 e+07)	863,515 (3,601,404)
IT full time employees	2.01 (1.69)	.87 (1.16)
Lack of knowledge	.09 (.29)	.16 (.36)
Privacy issues	.31 (.46)	.27 (.46)
Technology needs	.14 (.34)	.22 (.41)
Security issues	.40 (.49)	.36 (.48)

There is also a scope as far as the expansion of some of the other generally recognized levels of the overall government. The indicators that are brought into the consideration talk about making sure that some of the local and the government units are clearly defined.

This problem cannot be solved without the introduction to the work of public authority's effective modern management techniques that are appropriate to the complexity of the new conditions and requires the system of public administration of openness, of prompt and adequate in relation to the demands of the external environment.

Improving the efficiency of state and municipal government is the main aim of a modern Pakistan administrative reform [27]. Thus, according to the concept of New Public Management, which is the methodological basis of the administrative reform in modern countries, the essence of modern public administration reform is to move the client orientation of the executive authorities, which implies, in particular, the focus of their work to meet the demands of its customers - individuals and organizations [17].

E-Government readiness measures that ability and the willingness of the country and its adaptability to the notion and the implementation of the e-government. The index has been updated time and time again to make sure that it represents the true nature of the governments and also highlights the efforts they are being done by the government in the corresponding period. It has its initiation in 2003 and since then it has made an effort to make sure that it takes into consideration all the major countries that are the part of the agreement (Koh et al, 2008, p.546). The idea is to look at some of the important aspects of the e-government and it talks about and discusses

the quality of the online resources that are being employed by the government as well as the scope of some of the resources.

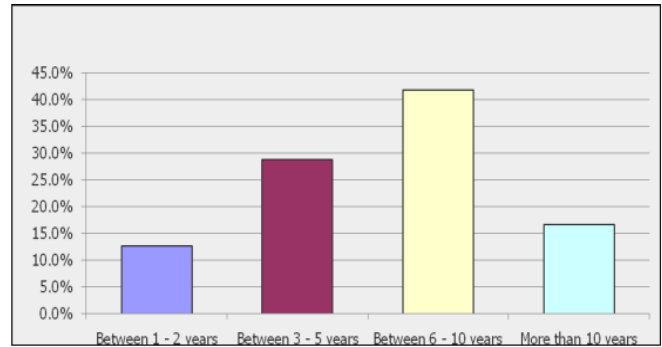


Fig. 8. Adoption of internet among citizens (Statistics Pakistan)

This figure describes about the connectivity that is being seen at the telecommunication level and the overall capability in different period and the quality of the human capital. The weight age that is being given to the efforts that are undertaken by the government but there are some other factors that are also brought into the consideration. The other factors that are being talked about are the level of infrastructure that the country has as well as and the penetration that has been made as far as the operations is concerned is also looked at. There is a close link to the survey and other initiative that is known as E-Participation index [12] [40].

It is clear that the promotion of access to information, the transparency, accountability and anti-corruption in government and public institutions by applying ICT opens opportunities to exercise political and civil rights. This reinforces the democracy and generates a distinct culture that confronts the secrecy, corruption and kidnapping of the public sphere by groups of power. Not only in the public (by increasing their capacity for participation and social control, monitoring or oversight) but also through political leaders and officials public (by increasing levels of awareness of the importance of acting and deciding in favor of social welfare).

In other words, theories relating to e-Government adoption appear to be undeveloped. This paper makes a theoretical contribution to the literature on e-market adoption in an e-government context. It adds to the body of knowledge on institutional theory from the perspective of how organizations respond to external pressure to adopt new technology. In this particular study, local authorities responded to external pressures from central government to adopt e-Government initiatives to engage electronically with suppliers and consumers of services. What the case demonstrates is that the perception of risk, deficits in the knowledge of organizations, trust issues, organizational size, organizational readiness (cognitive and technical), environmental turbulence, competing priorities within organizations. Cultural attitudes within a given sector, lack of clear guidance or guidelines from the imposing institution, the perceived benefits (relative advantage) or consequences (relative disadvantage) of adopting the new technology and the perceived benefits of compliance. With the institutional mandate are all moderating variables (determinant factors) that influence or govern whether organizations within

a given sector adhere to and conform rationally to institutional pressure. Given that the majority of these factors varied between local authorities in the local government sector, dissimilar types of behavior were observed among, but also within, the local authorities.

V. RESULTS

They thus provide that, for the static phase knowledge, the ability to use and perceived functional benefit affect the adoption, while uncertainty, security and the ability perceived use relate directly with confidence the e-Government. For phase interaction, variables that relate they vide intending to adopt e-Government are aware, confidence, ability to use quality information and perceived image. For this phase, perceived uncertainty, perceived safety and perceived ability to use are the variables that have a positive relationship with trust [22]. The UTAUT model raises the expectation of result, the expectative of effort and social influences affect the intended use, while the latter and facilitators conditions determined actual adoption of e-government in Pakistan [4].

TABLE IV. SPSS RESULT- CRONBACH ALPHA α

Constructs	No. of Items	Cronbach Alpha (α)
Performance Expectancy (PE)	5	0.83
Effort Expectancy (EE)	4	0.84
Social Influence (SI)	5	0.77
Facilitating Condition (FC)	3	0.83

The results show that UTAUT is a useful model to explain the intent of the teachers use the combined teaching methodology, showing a predictive power of the set of independent variables on the Behavioral Intention of 35% of the variance ($R^2 = .349, p < 0.000$). The data indicate that the expectation of results ($\beta = .413, p < 0.001$), Facilitators conditions ($\beta = .15, p < 0.001$) and Social Influence ($\beta = .14, p < 0.001$) were determinants behavioral intention to use e-government [21] [37]. The level of Significance is 0.05 ($\alpha = 0.05$). It has to be understood that some of the applications as far as the implementation of the technology is concerned and the thus it has to be kept in mind that the real benefit of the E-Government does not talk about the information and the level of technology but the way technology will be implemented at the organizational level [1] [45].

TABLE V. TEST OF SIGNIFICANCE (SPSS OUTPUT)

Hypotheses/Path	Finding	Conclusion
Research Hypothesis	Beta=0.34	Supported
	Not Significate	Not Supported
	Not Significate	Not Supported

From the survey one thing that could have helped as far as the integration of the e-government into some of the government organizations is that of the strength that is needed to be displayed by the back hand office operations. So there was this need to make sure that back office was the part of the system in a strong manner. This fact was illustrated in the Readiness survey of 2008. Immediately after this report, lot of countries in Northern Europe made sure that they have bring some sort of stability and they have revamped their back office integration task more readily. This is one of the more prime example of how readiness can work wonders as far as the government initiatives are concerned.

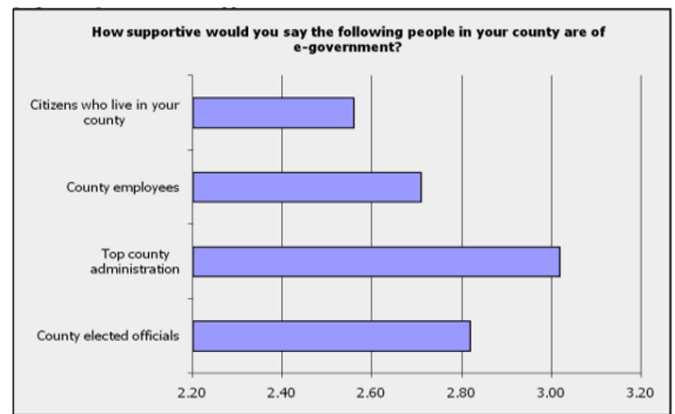


Fig. 9. Effectiveness of e-Government

VI. CONCLUSION

This study applies in a matter of second’s revised UTAUT model on client acceptance and utilization of e-taxpayer supported organizations in KSA. Taking into account the information gathered and the consequences of the examination, it can be presumed that Performance Anticipation, Effort Expectancy, and Facilitating Condition have positive impacts on client expectation to utilize e-taxpayer driven organizations.

Then again, in this study Social Influence was discovered to be inconsequential as far as predicting the behavioral expectation to utilize e-taxpayer supported organizations and its theory was not bolstered. In future work, we would include Trust and website Quality as autonomous variable into our examination model and consider the impacts of other crucial builds of the UTAUT model inside the setting of Pakistani environment. To be more precise and persuading, our work will proceed and new findings will be expected. It can be concluded that the most important goal of the e-government is to make sure that citizens can get the better portfolio of the services and their lives can get enhanced by the overall services that Government can provide to its citizens.

The idea is to make sure that all these services can be delivered to the customer without the citizen having lot of problem and trouble finding them. The transparency is an important aspect as far as operations are concerned as they allow Government to make sure that there is some sense of transparency in the Government operations. Some of the processes that are needed to be done and there are some important certificates can be accessed more readily if E-Government is implemented. E-Government thus can be made to use as a tool that can enhance the life of the citizens and thus achieve better governance. The use of information and technology in all the Government sectors thus will allow the government to see the success their strategies are bringing at the broad way level (Carter & Belanger, 2005, p.5).

In this study, UTAUT model was utilized, as it was recognized as a suitable model based on the writing on e-government selection. By utilizing UTAUT, four builds (i.e. execution anticipation, exertion hope, social impact, and encouraging condition) were connected. The outcomes show that these builds have impact on clients' selection of e-

taxpayer driven organizations in Pakistan. The more clients see e-taxpayer supported organizations as being valuable and straightforward, the more they will mean to utilize such administrations. The outcomes demonstrate that nationals will be all the more ready to embrace.

In order to make sure it happens, there is some need for comparative measure that can improve the performance at the international standards. Government can move forwards in the development of the e-Government and keeping in mind the rising demands of the e-government in the public sector, these policy measures can help the government in a long way. It has to be kept in mind thought that the delivery of the services can be a bit slow at times and thus there are not many Governments that can take care of the necessary capital to take care of all these operational difficulties (Nour et al, 2008, p.461).

The e-Government Survey 2012 finds that many countries are moving from a unilateral decentralized organizational model to an integrated, unified, which aims to focus on a single portal services provided by the government to generate greater transparency. Another important point has to do with mobility.

Future studies are needed to corroborate the results on the properties of the UTAUT of this work and to resolve the limitations presented. One way would involve the review of possible errors in translation that might affect the understanding of some reagents, especially those on the subscales social influence, and facilitating conditions, which showed less internal consistency. Future studies could perform validity tests that allowed to expand and consolidate knowledge about the properties of the instrument in the Pakistani population.

REFERENCES

- [1] Abdallah, S., & Fan, I. S. Framework for e-government assessment in developing countries: case study from Sudan. *Electronic Government, an International Journal*, (2012). 9(2), 158-177.
- [2] Abid, H., & Noreen, U. Ready to E-bank: An exploratory research on adoption of e-banking and e-readiness in customers among commercial banks in Pakistan. *Spider*, (2006). 31(2), 1-31.
- [3] Afsar, S., Mateen, A., & Korbatov, A. Towards the enhancement of electronic democracy: the adoption of reinventing government and electronic government at the municipal level. *J. Agric. Soc. Sci.*, (2005). 1, 133-137.
- [4] Al-Omari, A., & Al-Omari, H. E-government readiness assessment model. *Journal of Computer Science*, (2006). 2(11), pp.841-845. Data taken from <http://asu.edu.jo/Upload/FacultyPub/cddf5964-dd71-47ed-8935-53052a08c5e6.pdf>
- [5] Andersen, T. B.E-Government as an anti-corruption strategy. *Information Economics and Policy*, (2009). 21(3), 201-210.
- [6] Arif, M., & Mahmood, K. The changing role of librarians in the digital world: Adoption of Web 2.0 technologies by Pakistani librarians. *The Electronic Library*, (2012). 30(4), 469-479.
- [7] Bannister, F. The curse of the benchmark: an assessment of the validity and value of e-government comparisons. *International Review of Administrative Sciences*, 73(2), pp.171-188. (2007). Data taken from <http://ras.sagepub.com/content/73/2/171.short>
- [8] Bertot, J. C., Jaeger, P. T., & Grimes, J. M. Using ICTs to create a culture of transparency E-government and social media as openness and anti-corruption tools for societies. *Government information quarterly*, (2010). 27(3), 264-271.
- [9] Blythe, S. E. Pakistan Goes Digital: The Electronic Transactions Ordinance as a Facilitator Growth for E-Commerce. *J. Islamic St. Prac. Int'l L* (2006)., 2, 5.
- [10] Caba Pérez, C., Pedro Rodríguez Bolívar, M., & López Hernández, A. M. e-Government process and incentives for online public financial information. *Online Information Review*, (2008). 32(3), 379-400.
- [11] Carter, L., & Bélanger, F. The utilization of e - government services: citizen trust, innovation and acceptance factors*. *Information Systems Journal*, (2005)15(1), pp.5-25. Data taken from <http://onlinelibrary.wiley.com/doi/10.1111/j.1365-2575.2005.00183.x/full>
- [12] Das, J., DiRienzo, C., & Burbridge, J. Global e-government and the role of trust: A cross country analysis. *Applied Technology Integration in Governmental Organizations: New E-Government Research: New E-Government Research*, (2010). 1.
- [13] Davison, R. M., Wagner, C., & Ma, L. C. From government to e-government: a transition model. *Information Technology & People*, (2005). 18(3), 280-299.
- [14] Dwivedi, Y. K., & Williams, M. D. Demographic influence on UK citizens'e-government adoption. *Electronic Government, an International Journal*, (2008). 5(3), 261-274.
- [15] Dwivedi, Y. K., Khoumbati, K., Williams, M. D., & Lal, B. Factors affecting consumers' behavioural intention to adopt broadband in Pakistan. *Transforming Government: People, Process and Policy*, . (2007). 1(3), 285-297.
- [16] Dwivedi, Y. K., Williams, M. D., Rana, N. P., & Williams, J. Reflecting on e-government research: toward a taxonomy of theories and theoretical constructs. *International Journal of Electronic Government Research*, (2011). 7(4), 64-88.
- [17] Evans, D., & Yen, D. C. E-Government: Evolving relationship of citizens and government, domestic, and international development. *Government information quarterly*, (2006) 23(2), pp.207-235.
- [18] Grönlund, Å. *Electronic government: design, applications and management*. IGI Global, (2002) pp. 15-22
- [19] Gupta, B., Dasgupta, S., & Gupta, A. Adoption of ICT in a government organization in a developing country: An empirical study. *The Journal of Strategic Information Systems*, (2008). 17(2), 140-154.
- [20] Helbig, N. Understanding the complexity of electronic government: Implications from the digital divide literature. *Government Information Quarterly*, (2009) 26(1), pp. 89-97
- [21] Hung, S. Y., Chang, C. M., & Yu, T. J. Determinants of user acceptance of the e-Government services: The case of online tax filing and payment system. *Government Information Quarterly*, (2006). 23(1), pp.97-122. Data taken from <http://www.sciencedirect.com/science/article/pii/S0740624X05000948>
- [22] Irani, Z., Weerakkody, V., Kamal, M., Mohammed Hindi, N., Osman, I. H., Latef Anouze, A., ... & Al-Ayoubi, B. (2012). An analysis of methodologies utilised in e-government research: A user satisfaction perspective. *Journal of Enterprise Information Management*, 25(3), 298-313.
- [23] Jaeger, P. T., & Thompson, K. M. E-government around the world: lessons, challenges, and future directions. *Government Information Quarterly*, (2003). 20(4), pp.389-394. Data taken from http://www.mty.itesm.mx/egap/centros/caep/imagenes/REDIP/2_E_gov_ernment_around_the_world.pdf
- [24] Kamal, M., Weerakkody, V., & Irani, Z. Analyzing the role of stakeholders in the adoption of technology integration solutions in UK local government: An exploratory study. *Government Information Quarterly*, (2011). 28(2), 200-210.
- [25] Koh, C. E., Prybutok, V. R., & Zhang, X. Measuring e-government readiness. *Information & Management*, (2008). 45(8), pp.540-546. Data taken from <http://www.sciencedirect.com/science/article/pii/S0378720608001134>
- [26] Koh, C. E., Prybutok, V. R., Ryan, S., & Ibragimova, B. The importance of strategic readiness in an emerging e-government environment. *Business Process Management Journal*, (2006). 12(1), pp.22-33. Data taken from <http://www.emeraldinsight.com/journals.htm?articleid=1538014&show=abstract>
- [27] Kojima, M., & Johnson, T. Potential for biofuels for transport in developing countries (2005).

- [28] Krishnan, S., & Teo, T. S. Moderating effects of governance on information infrastructure and e - government development. *Journal of the American Society for Information Science and Technology*, (2012). 63(10), 1929-1946.
- [29] Lean, O. K., Zailani, S., Ramayah, T., & Fernando, Y. Factors influencing intention to use e-government services among citizens in Malaysia. *International Journal of Information Management*, (2009). 29(6), 458-475.
- [30] Lenk, K. Electronic government: where are we heading?. In *Electronic Government*. Springer Berlin Heidelberg, (2002) pp. 1-9
- [31] [31] Mujahid, Y. H. (2002). Digital opportunity initiative for Pakistan. *The Electronic Journal of Information Systems in Developing Countries*.
- [32] Nour, M. A., AbdelRahman, A. A., & Fadlalla, A. A context-based integrative framework for e-government initiatives. *Government Information Quarterly*, (2008). 25(3), pp.448-461.
- [33] Nysveen, H., Pedersen, P. E., & Thorbjørnsen, H. Explaining intention to use mobile chat services: moderating effects of gender. *Journal of consumer Marketing*, (2005). 22(5), 247-256.
- [34] Ovais Ahmad, M., Markkula, J., & Oivo, M. Factors affecting e-government adoption in Pakistan: a citizen's perspective. *Transforming Government: People, Process and Policy*, (2013). 7(2), 225-239.
- [35] Paola Torres Maldonado, U., Feroz Khan, G., Moon, J., & Jeung Rho, J. E-learning motivation and educational portal acceptance in developing countries. *Online Information Review*, (2011). 35(1), 66-85.
- [36] Qaisar, N., & Khan, H. G. A. E-Government challenges in public sector: A case study of Pakistan. (2010).
- [37] Rana, N. P., Williams, M. D., Dwivedi, Y. K., & Williams, J. Theories and theoretical models for examining the adoption of e-government services. *E-service Journal*, (2012). 8(2), 26-56.
- [38] Rehman, M., & Esichaikul, V. Factors influencing the adoption of e-government in Pakistan. In *E-Business and E-Government (ICEE)*, 2011 International Conference on (2011). (pp. 1-4). IEEE.
- [39] Rehman, M., Esichaikul, V., & Kamal, M. Factors influencing e-government adoption in Pakistan. *Transforming Government: People, Process and Policy*, (2012). 6(3), 258-282.
- [40] Sahu, G. P. (Ed.). *E-Government Development and Diffusion: Inhibitors and Facilitators of Digital Democracy: Inhibitors and Facilitators of Digital Democracy*. IGI Global. (2009).
- [41] Schmutzer, R. Organizational challenges to the development of electronic government. In *Database and Expert Systems Applications*, 2000. Proceedings. 11th International Workshop on. IEEE, (2000) pp. 379-383
- [42] Seyal, A. H., Noah Abd Rahman, M., & Awg Yussof Hj Awg Mohammad, H. A quantitative analysis of factors contributing electronic data interchange adoption among Bruneian SMEs: A pilot study. *Business Process Management Journal*, (2007). 13(5), 728-746.
- [43] Shafique, F., & Mahmood, K. Indicators of the emerging information society in Pakistan. *Information Development*, (2008). 24(1), 66-78.
- [44] Shareef, M. A., Kumar, U., Kumar, V., & Dwivedi, Y. K. Identifying critical factors for adoption of e-government. *Electronic Government, an International Journal*, (2009). 6(1), 70-96.
- [45] Shin, S. Implementing e-government in developing countries: its unique and common success factors. *American Political Science Association*. (2008).
- [46] Stoltzfus, K., (May). Motivations for implementing e-government: an investigation of the global phenomenon. In *Proceedings of the 2005 national conference on Digital government research* (2005 (pp. 333-338). Digital Government Society of North America.
- [47] Torres, L., Pina, V., & Royo, S. E-government and the transformation of public administrations in EU countries: Beyond NPM or just a second wave of reforms?. *Online Information Review*, (2005). 29(5), 531-553.
- [48] Troshani, I., Jerram, C., & Rao Hill, S. Exploring the public sector adoption of HRIS. *Industrial Management & Data Systems*, (2011). 111(3), 470-488.
- [49] Von Haldenwang, C. Electronic government (e-government) and development. *The European Journal of Development Research*, (2004) 16(2), pp. 417-432
- [50] Wagner, C., Cheung, K., Lee, F., & Ip, R. Enhancing e-government in developing countries: managing knowledge through virtual communities. *The Electronic Journal of Information Systems in Developing Countries*, (2003). P.14.
- [51] Welch, E. W., Hinnant, C. C., & Moon, M. J. Linking citizen satisfaction with e-government and trust in government. *Journal of Public Administration Research and Theory*, (2005). 15(3), pp.371-391. Data taken from <http://jpart.oxfordjournals.org/content/15/3/371.short>
- [52] Wescott, C. G. E - Government in the Asia - pacific region. *Asian Journal of Political Science*, (2001). 9(2), 1-24.
- [53] Zinner Henriksen, H., & Viborg Andersen, K. Electronic records management systems implementation in the Pakistani local government. *Records Management Journal*, 18(1), 40-52. S. Chen, B. Mulgrew, and P. M. Grant, "A clustering technique for digital communications channel equalization using radial basis function networks," *IEEE Trans. on Neural Networks*, (2008). vol. 4, pp. 570-578, July 1993.

AUTHOR PROFILE



Zulfiqar Haider (Syed Zaidi) is working as an Assistant Professor in the Department of Public Administration, University of Sindh, Jamshoro, Pakistan. He is currently pursuing the Ph.D. degree at Dalian University of Technology, Dalian, China.



Prof. Chen Shuwen is working as a Dean & research supervisory duties at the School of public administration, Dalian University of Technology, Dalian, China.



Prof. Dr. Zareen Abbasi is working as a Professor in the Department of Public Administration, University of Sindh, Jamshoro, Pakistan.

Applying Topology-Shape-Metric and FUZZY Genetic Algorithm for Automatic Planar Hierarchical and Orthogonal Graphs

Nahla F.Omran

Department of mathematics
Faculty of science, South Valley University
Qena, Egypt

Sara F. Abd-el ghany

Qena, Egypt

Abstract—The graphs appear in many applications such as computer networks, data networks, and PERT networks, when the network includes a small number of devices, it can be drawn easily by hand, as the number of devices increases, drawing becomes a very difficult task. For this problem we will develop a new method for automatic graph drawing based on two steps, the first is applying the topology –shape –metric that is approaching to orthogonal drawings for the grid and the second step is applying the fuzzy genetic algorithm that is directed, in the topology –shape –metric the final drawing is achieved through three sequential steps: planarization, orthogonalization, and compaction. Each of these steps is responsible for the quality of the final drawing. Then the genetic algorithm applied at the planarization step of the topology-shape-metric to find the geometric position of each vertex to minimize bending in the graph. The developed technique generates a greater number of planar embedding by varying the order of edges' insertion. This is achieved clearly in the Results given in the paper.

Keywords—graph drawing; hierarchical graphs; topology-shape-metric; fuzzy genetic algorithms

I. INTRODUCTION

The expanding use of computers into business, science and the home, making scientists tend to draw diagrams to understand computer software. Graph drawing is a visualization of objects and relations between those objects. The effectiveness of the visualization of a graph is dependent on how efficiently the associated diagram conveys information to the users. The Specific requirements in this application are: Initially we will apply the topology- shape-metric that is divided into three main steps. The first step is the planarization step in this step reduces the number of edge crossings as much as possible.

The second step is the Orthogonalization: The goal of this step is to minimize the number of bends without changing the topology. The third step is the compaction in this step the goal is to minimize the drawing area. The problem of drawing graph was first studied in 1983 by M.R Garey, and D.S. Johnson [1] studied the problem of minimizing the number of edge-crossing. Then, in 1994, Di Battista [2] presented study to produce esthetically pleasing drawings of graphs based on main-cost-flow for both vertical and horizontal edge groups. Later, in 1999 Klau, Petra Mutzel [3] presented an approach based on a branch – and – cut algorithm which computes

optimally labeled orthogonal drawings for compaction and labeling problem. In 2001, Maurizio Patrignani [4] presented study for the complexity of orthogonal compaction based on three problems consist of providing an orthogonal grid drawing, while minimizing the area, the total edge length, or the maximum edge length. In 2002 Markus Eiglsperger, Michael Kaufman [5] present a new compaction algorithm for orthogonal graph drawing with vertices of prescribed size.

Finally, applied the fuzzy genetic algorithm at the planarization step of the TSM. The use of fuzzy logic based techniques for either improving genetic algorithm behavior and modeling genetic algorithm components, the results obtained have been called a Fuzzy genetic algorithm. A FGA may be defined as an ordering sequence of instruction in which some of the instructions or algorithm components may be designed with fuzzy logic- based tools. In 1960[6], the Genetic Algorithms were first described by John Holland and further developed by Holland and his students and colleagues at the University of Michigan in the 1960 and 1970. In 1997, J.Branke, F.Bucher, H.Schmeck [7] use genetic algorithm for undirected graphs. In 1999, D.K.Pratihar, K.Deb, A.Ghosh [8] uses a fuzzy logic and a fuzzy genetic algorithm for the problems with mobile robots. In 2001, I.G. Damousis, K.J. Satsio's, D.P. Labridis, P.S. Dokopoulos [9], combined fuzzy logic and genetic algorithm techniques—application to an electromagnetic field problem. In 2006, P.Kuntz, B.Pinaud, A.Ghosh [10] used a hybrid genetic algorithm to minimizing crossing in hierarchical graphs. In 2007, D.Vrajitoru [11] applied a hybrid genetic algorithm to solve graph drawing problems. In 2011, Bernadette M.M, Gustavo H.D, Frederico G. Guimar, Renato C. M, Petr Ya. E [12] using a fuzzy genetic algorithm for automatic orthogonal graph drawing. In this chapter we will use TSM to minimize the crossing in the graph and then apply the FGA in the second step of TSM to find the geometric position of each vertex to minimize the bends in the graph to produce a graph with good esthetic criteria.

II. THE TOPOLOGY – SHAPE – METRIC TECHNIQUE

The topology-shape-metric was widely discussed and improved in 1998 by I.G. Toll's, G. Di Battista, P. Eades, and R. Tamassia [13]. When applying the topology-shape-metric, the final drawing is produced by applying three consecutive steps: planarization, orthogonalization, and compaction.

A. Planarization

this step determine the topology of the graph , which test if the graph is planar or no. therefore, the goal of this step is to minimize the number of edge crossing as much as possible because the number of crossing affects the understanding of the graph. In 1998, I.G. Toll's, G. Di Battista, P. Eades, R. Tamassia [13] present algorithms that are used to build planar graph.

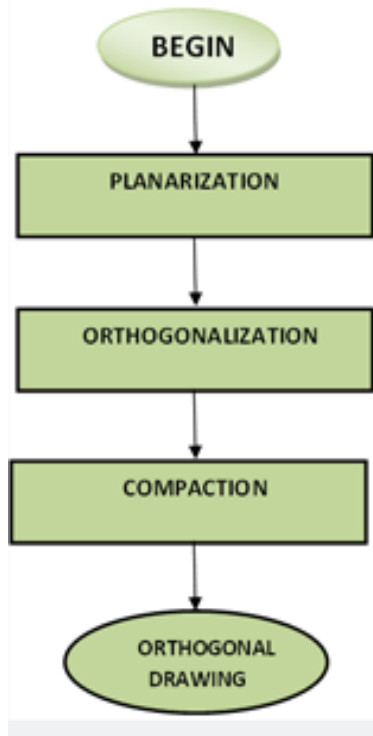


Fig. 1. Block diagram for the topology-shape-metric algorithm [14, 15]

Algorithm (Planarize).

Input: graph G ;

Output: planarization G_+ of G ;

1) Compute a maximal planar subgraph S of the input graph G , and partition the edges into “planar” and “non-planar”, as follows:

a) Start with subgraph G_+ consisting only of the vertices of G , but no edges;

b) For each edge e of G , if the graph obtained by adding e to G_+ is planar, then add e to G_+ and classify e as “planar”, else reject e and classify it as “non planar”.

2) Construct a planar embedding of the planar subgraph G_+ , and the dual graph of S .

3) Add to G_+ the non planar edges, one at a time, each time minimizing the number of crossings. This is done as follows for a non planar edge (u, v) :

a) Find a shortest (least number of edges) path in the dual graph of the current embedding G_+ from the faces incident to u to the faces incident to v ;

b) Add the nonplanar edge and update G_+ as well as its dual graph.

We will apply the planarization step on the graph in Figure2, we will get different planar graph by varying edge insertion to minimize edge crossing.

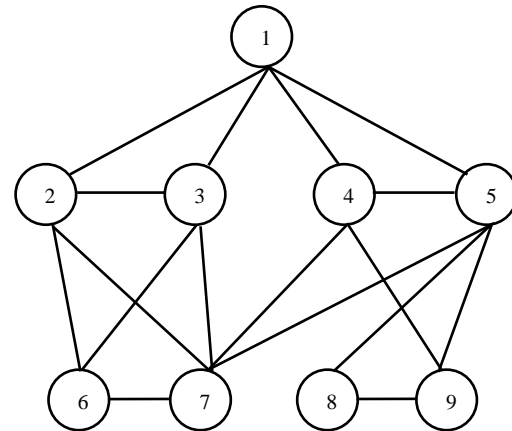


Fig. 2. Example of nonplanar hierarchical graph.

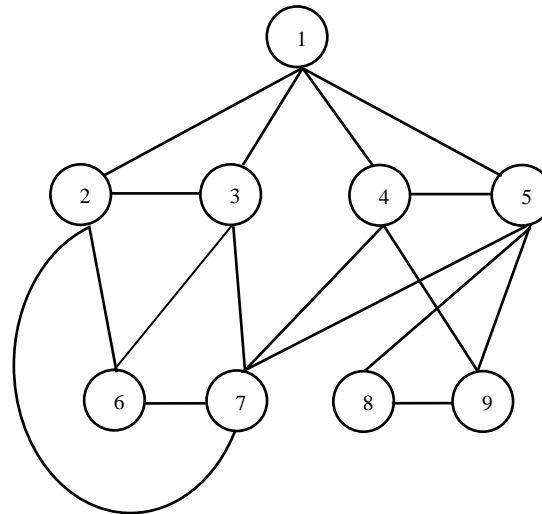


Fig. 3. Nonplanar hierarchical graph with one bend.

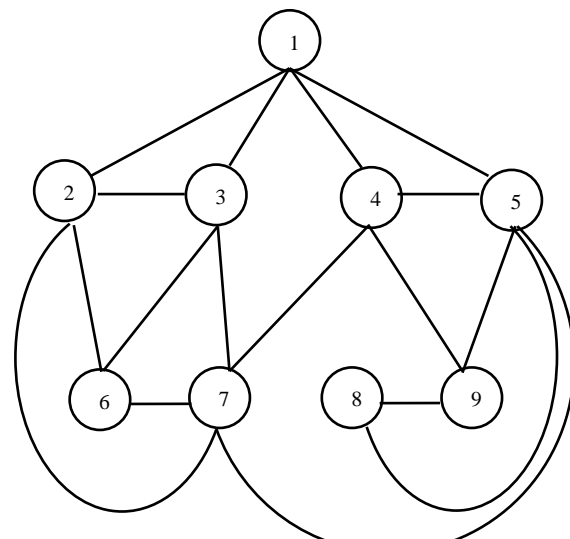


Fig. 4. Planar hierarchical graph with three bends.

The graph in Figure4 is planar but contains three bends, to solve this problem we will apply fuzzy genetic algorithm.

B. Orthogonalization

This step is performed to reshape the drawing to cancel the bends from the graph and the edge become straight line. We will use the Tamassia algorithm [14] computes an orthogonal shape of a planar graph with respect to an input embedding with a minimal number of bends the result of the orthogonalization on the Figure2 is shown in Figure5.

Algorithm. ORTHOGONAL

Input. A biconnected graph G.

Output. An orthogonal drawing of G.

- 1) Compute a st-numbering of G
- 2) Produce a reduced graph G I and modify the st-numbering so that there are no gaps in the st sequence.
- 3) Run Form _pairs on the reduced graph G.
- 4) Place vertices v_1 and v_2 in the same row, if v_2 does not belong to a pair in which it shares a row with another vertex. If v_1 and/or v_2 have degree less than 4, then the placement of v_1 and v_2 might require one or two rows.
- 5) REPEAT
 - a) Consider the next vertex v_i according to Unmarked.
 The st-numbering of G.
 - b) If v_i has already been placed, then go to Step 6.
 - c) If vertex v_i is unassigned, then place v_i in a new row. Connect v_i with each vertex v_j ($j < i$) such that (v_j, v_i) is a directed edge of G. Add as many uncompleted edges as required, depending on v_i 's out degree.
 - d) If vertex v_i is assigned to a pair, then place v_i together with the other vertex in the same pair following the placement rules described above for the specific type of pair.
- 6) UNTIL the only remaining vertex is v_n ,
- 7) Insert v_n , in a new row. If v_n , is of degree 4, then there is an incoming edge that enters v_n , from the top and bends twice. This edge is chosen to be the one that connects to v_n , $_{-1}$.
- 8) Restore the degree 2 vertices of G that were absorbed in Step 2.
- 9) End.

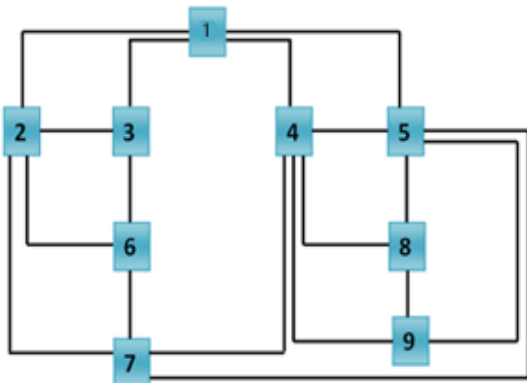


Fig. 5. Orthogonal representation of Figure2

C. Compaction

In this section we will minimize the area of the given orthogonal drawing. The result of this step is shown in Figure6.

Algorithm .planar graph compaction.

Input. $\sqrt{n} \times \sqrt{n}$ bitmap planar graph layout.

Output. $\sqrt{n} \times \sqrt{n}$ bitmap planar graph layout compacted one point to the east.

- 1) identify all points on layout that may possible move to the east. Mark these points to be movable.
- 2) unmark movable points that can cause connectivity violation to be stationary.
- 3) repeat step 2 until no further points is unmarked.
- 4) compact movable points to one point to the east maintaining connectivity.

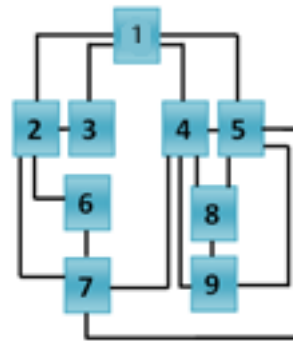


Fig. 6. The result of compaction Figure5

III. PROPOSED FUZZY GENETIC ALGORITHM

We will solve the problem of bending in the graphs by using the FGA to find the geometric position of each vertex. We will applying the FGA at the planarization step to obtain a lot of planar graphs then submit this to orthogonalization and compaction step .The aesthetic criteria takes into consideration:

- 1) The number of crossings f_X in the graph.
- 2) The number of bends f_B in the graph.
- 3) The total sum of the edges' length f_L .

By minimizing all of them we will obtain the optimal graph.

The fitness function $\phi(st,i) = \alpha_1 f_X + \alpha_2 f_B + \alpha_3 f_L$ where $i \in [0, 1]$

We develop the diagram of TSM by adding FGA [15].

Algorithm (TSM-Fuzzy-GA).

Input: graph G;

Output: an optimized planar drawing;

- 1) Generation of the initial population:
N = number of individuals;

a) Generate at random the ordering of edge's insertion from G (represented by an integer permutation);

2) Fitness computation: $i = 0$;
While ($i < N$) do

- a) Submit solution s_i to the planarization step to obtain a planar embedding (Γ_i) and the number of crossings $FX(s_i)$;
 - b) Submit the planar embedding (Γ_i) to the orthogonalization step to obtain the orthogonal representation H , and the number of bends $fB(s_i)$;
 - c) Submit the orthogonal representation H to the compaction step to obtain the final drawing and the total sum of the edges length $fL(s_i)$;
 - d) Calculate the value of the fuzzy membership's μ_{FX} , μ_{fB} , and μ_{fL} ;
 - e) Calculate the fuzzy-max-min aggregation μ_D ;
 - f) $i = i + 1$;
- 3) Record the best individual according to the fitness function;
- 4) Application of the genetic operators for generating the new population. Each crossover operator (PMX or OX) for producing each offspring is selected with equal probability (0.50). The mutation operator to be used (scramble, swap, insert, and invert) for each offspring is also selected at random with equal chance (0.25);
- 5) Application of generational survival selection;
- 6) Go to step 2 until the stop criterion is met;
The results of applying FGA on the Figure2.

First step: in Figure7 we replaced v_2 by v_3 to cancel the bend between v_2 and v_7 .

Second step: in Figure8 we replaced v_8 by v_9 to cancel the bend between v_5 and v_8 .

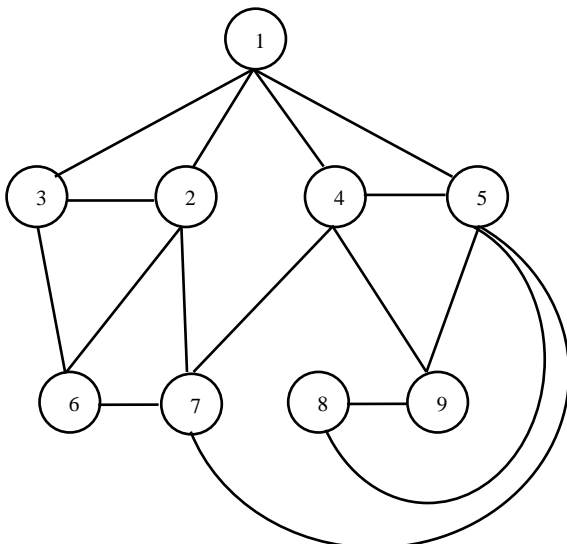


Fig. 7. Planar graph result from applying FGA

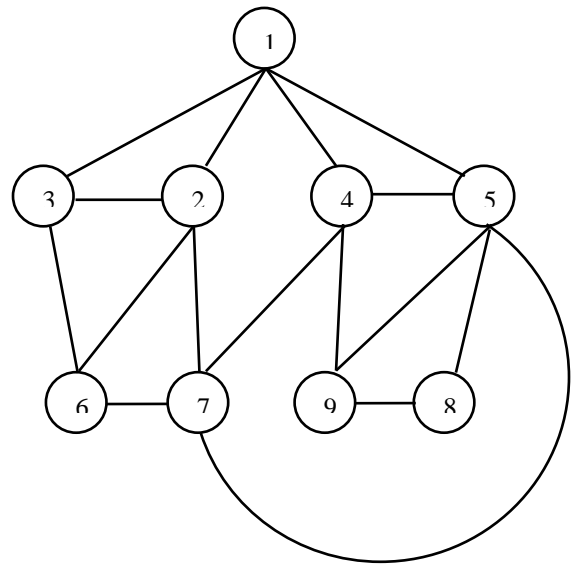


Fig. 8. Best planar graph produced from applying FGA

IV. A GENETIC OPERATOR

in this paper we attempt to solve the problem of bends in the graph by using FGA, the operator used in genetic algorithms to maintain genetic diversity, known as mutation and to combine existing solutions into others, crossover. The main difference between them is that the mutation operators operate on one chromosome, that is, they are unary, while the crossover operators are binary operators. Genetic variation is a necessity for the process of evolution. Genetic operators used in genetic algorithms are analogous to those in the natural world: survival of the fittest, or selection, reproduction (crossover, also called recombination), and mutation.

A. Selection

The Selection operator decides which of the individuals in the population will go into the next generation. This is decided by the fitness value of an individual as calculated by the Fuzzy Fitness Function. At this point the assumption is that a fitness value pertaining to each individual is available.

B. Crossover

Crossover is the most widely used recombination operator. Uniform 1-point crossover has been used. In general, 1-point crossover selects a random cut point and combines the first portion of one parent with the second portion of the other and vice versa to produce two offspring. The individual here consists of an array of cluster numbers. Hence, the main issue in recombination is the renumbering of clusters in the resulting offspring.

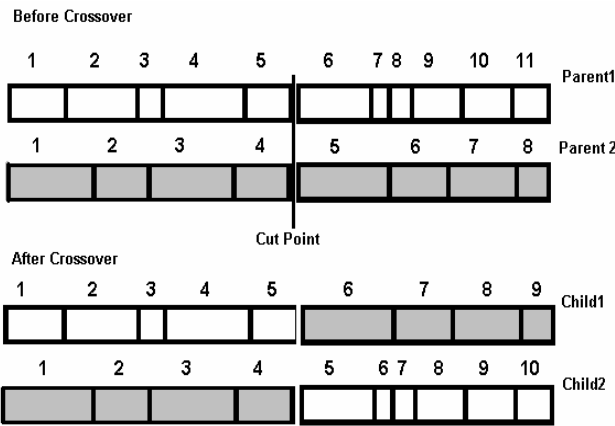


Fig. 9. A Visualization of Crossover

C. Mutation

Mutation is needed to counteract the loss of some potentially useful genetic material during selection and crossover. In an artificial chromosome, this is affected by an Occasional random alteration of the value of a string position. In a binary implementation, a bit value is toggled. In an integer or floating point implementation, a value is changed within an allowed range. This definition cannot be directly applied to the present scenario. A mutation operator that works at the boundaries of clusters has been worked out. For this, the pair-wise fitness between two consecutive data points is found for the whole data set. This has been calculated using the Fuzzy Fitness Function.

1) In each, select randomly, points in the range (1, m_i) where m_i is the number of clusters in the i th individual. The number of clusters to be selected is equal to n mutation as worked out in the previous paragraph. Mutation is applied on the left as well as on the right border of each selected cluster which amounts to $2*n$ mutation borders.

2) for a border point

Two cut-off values (in %) are to be fixed, one below which the pair-wise fitness will be classified as insignificant (lower value ($lval$)) and the other above which the pair-wise fitness will be significant (high value ($hval$)). The actual values will be implementation dependent.

a) Mutation Rule 1: if it's pair-wise fitness with the data point in the neighboring cluster is greater than $hval$ and its pair-wise fitness with the neighboring data point within the same cluster is less than $lval$, and then reallocates the data item to the neighboring cluster.

b) Mutation Rule 2: if it's pair-wise fitness with the data point in the neighboring cluster is greater than $hval$ and also its pair-wise fitness with the neighboring data point within the same cluster is greater than $hval$, then the two clusters can be merged.

c) Mutation Rule 3: if it's pair-wise fitness with the data point in the neighboring cluster is less than $lval$ and its pair-wise fitness with the neighboring data point within the same cluster is also less than $lval$, then the border point can be

made a single-point cluster to which zero fitness value is assigned.

d) Mutation Rule 4: if none of the above conditions apply, then the border is left undisturbed.



Fig. 10. A Visualization of Mutation

V. RESULTS

In this paper, we tested many values for the number of vertices v in the graph, to generate an optimal graph by following the procedure shown in the diagram [16]:

1) We generated graphs by varying the number of vertices V , from 10 to 600 vertices.

2) For each graph in the test set, when we applying the TSM on the graph in Figure 2, at the planarization step the final graph contain three bends shown in Figure 5, but when we applying FGA at the planarization step on the Figure 2, the final graph contains one bend shown in Figure 7.

Table 1 shows the results obtained with the classical topology shape-metric approach.

Table 2 presents the results obtained by the fuzzy genetic algorithm.

TABLE I. RESULTS OBTAINED BY THE CLASSICAL APPROACH TSM

F	f_1	f_B	f_x	v
41	16	3	0	10
87	38	2	1	20
148	64	5	1	30
232	92	6	6	40
441	132	19	24	50
1398	328	49	119	100
2126	632	59	137	150
2757	781	75	194	180
4769	854	172	509	200
5969	1144	197	618	250
9055	2748	173	608	500
9500	2800	198	620	520
10000	3000	200	630	550
10100	3200	195	616	600

TABLE II. RESULTS OBTAINED BY THE FUZZY GENETIC ALGORITHM

Fuzzy genetic algorithm		Test cases			
V-N	stats	f_x	f_i	f_B	μ_D
10-30	Best	0	14	3	1.000
	Average	0	14	3	0.800
20-30	StdDev	0.00	0.52	0.00	0.26
	Best	1	29	1	1.000
30-30	Average	1	30	1	0.675
	StdDev	0.00	0.95	0.00	0.24
40-30	Best	1	48	5	1.000
	Average	1	50	5	0.714
50-30	StdDev	0.00	1.69	0.00	0.25
	Best	5	71	5	0.857
60-30	Average	5	76	6	0.433
	StdDev	0.52	4.37	0.71	0.26
70-30	Best	12	95	10	0.769
	Average	14	100	12	0.446
80-30	StdDev	1.69	7.66	1.96	0.23
	Best	62	250	31	0.917
90-30	Average	70	264	36	0.330
	StdDev	7.89	8.29	3.30	0.31
100-30	Best	86	393	36	1.000
	Average	89	403	38	0.237
110-30	StdDev	5.56	25.33	3.08	0.29
	Best	95	531	51	0.696
120-30	Average	110	536	52	0.321
	StdDev	10.27	35.14	2.17	0.26
130-30	Best	334	687	120	0.698
	Average	342	688	130	0.198
140-30	StdDev	18.57	34.21	5.91	0.26
	Best	427	910	145	0.848
150-30	Average	429	999	152	0.279
	StdDev	12.77	77.81	4.47	0.35
160-30	Best	319	1351	103	1.000
	Average	321	1356	108	0.697
170-30	StdDev	2.67	6.86	6.98	0.41
	Best	310	1387	95	1.100
180-30	Average	303	1493	87	0.699
	StdDev	1.90	6.92	5.00	0.50
190-30	Best	367	1400	163	0.900
	Average	332	1525	145	0.770
200-30	StdDev	20.23	50.72	7.10	0.40
	Best	358	1520	150	1.900
210-30	Average	290	1526	110	0.800
	StdDev	2.29	7.67	7.80	0.56

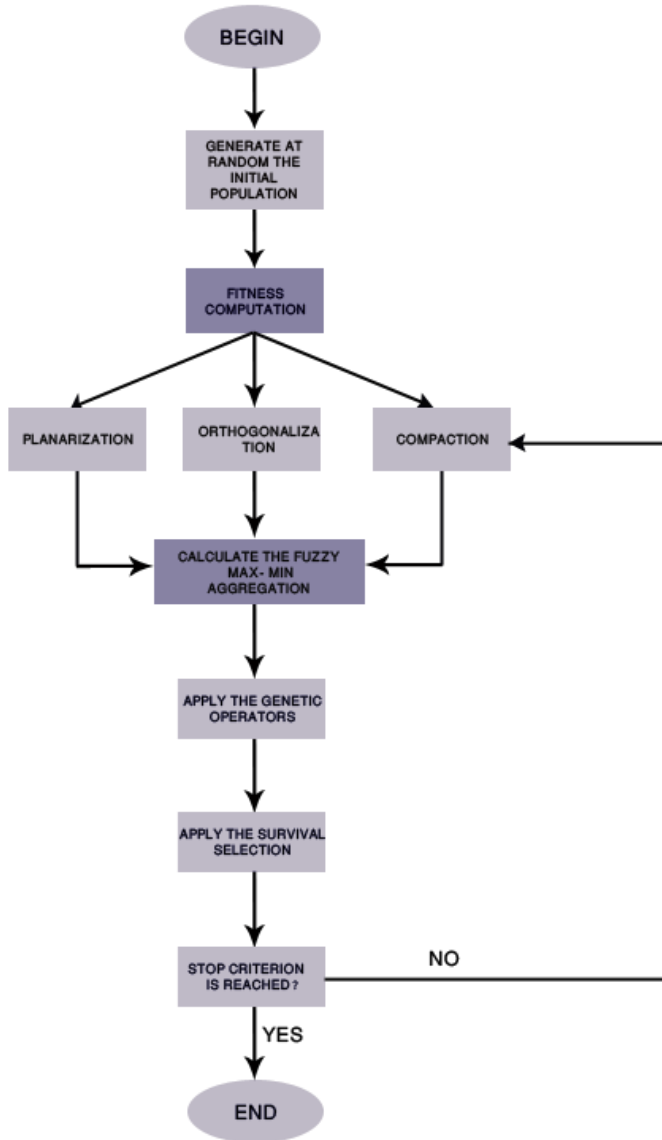


Fig. 11. Block diagram for the TSM-Fuzzy-GA algorithm

VI. CONCLUSIONS

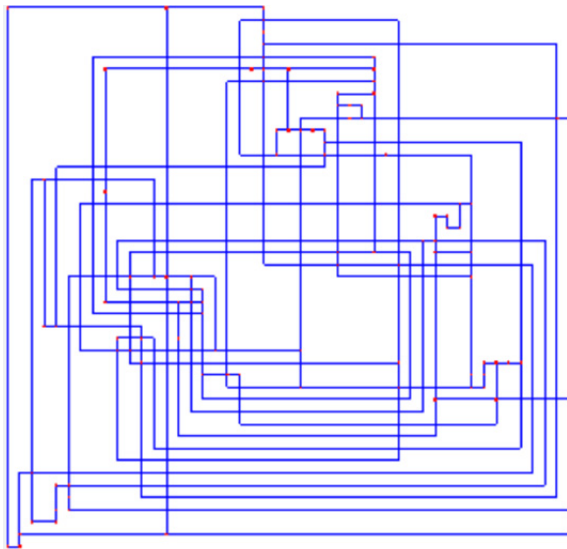
In this paper, we present a new method for automatic orthogonal graph drawing by using fuzzy genetic algorithm at the planarization step of the topology-shape-metric to find the geometric position of each vertex to obtain optimal graphs without crossing and bends.

ACKNOWLEDGMENT

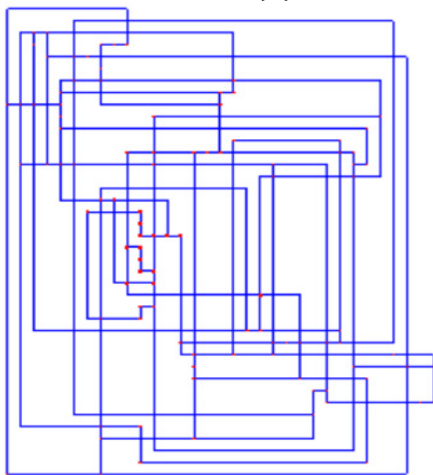
I cannot express my feelings to my family who supported me all over through my way and backed me up during my studies wishing me good luck throughout the preparation of this thesis.

REFERENCES

- [1] M.R. Garey, D.S. Johnson, Crossing number is NP-complete, *SIAM Journal on Algebraic and Discrete Methods* 4 (1983), pp.314-316.
- [2] G. Di Battista, P. Eades, R. Tamassia, I. Toll's, Annotated bibliography on graph drawing algorithms, *Computational Geometry: Theory and Applications* 4 (1994), pp. 240–282.
- [3] G.W. Klau, P. Mutzel, Optimal compaction of orthogonal grid drawings, in: 7th International Integer Programming and Combinatorial Optimization (IPCO) Conference, volume 1610 of Lecture Notes in Computer Science, Springer, 1999.
- [4] M. Patrignani, on the complexity of orthogonal compaction, *Computational Geometry* 19 (2001), pp. 50 – 60.
- [5] M. Eiglsperger, M. Kaufmann, Fast compaction for orthogonal drawings with vertices of prescribed size, in: 9th International Symposium on Graph Drawing (GD 2001), volume 2265 of Lecture Notes in Computer Science, Springer, 2001.
- [6] Holland, J. H. (1992). Genetic algorithms. *Scientific American*, July, 114–116. Forrest, S. (1993). Genetic algorithms: Principles of natural selection applied to computation. *Science*, 261, 872-878.
- [7] J. Branke, F. Bucher, H. Schmeck, Using genetic algorithms for drawing undirected graphs, in: The Third Nordic Workshop on Genetic Algorithms and their Applications, 1997, pp. 200–205.
- [8] D.K. Pratihari, K. Deb, A. Ghosh, Fuzzy-genetic algorithms and time-optimal obstacle-free path generation for mobile robots, *Engineering Optimization* 1 (1999), pp. 117–130.
- [9] P. Kuntz, B. Pinaud, R. Lehn, Minimizing crossings in hierarchical digraphs with a hybridized genetic algorithm, *Journal of Heuristics* 12 (2006) 30–36.
- [10] I.G. Damousis, K.J. Satsios, D.P. Labridis and P.S. Dokopoulos, "Combined fuzzy logic and genetic algorithm techniques—application to an electromagnetic field problem" Elsevier, 2001, pp.371-385.
- [11] D. Vrajitoru, Hybrid multiobjective optimization genetic algorithms for graph drawing, in: Proceedings of the Genetic and Evolutionary Computation Conference (GECCO 07), ACM Press, 2007.
- [12] Bernadette M.M. Neta Gustavo H.D. Arajo, Frederico G. Guimaraes, Renato C. Mesquita, Petr Ya. Ekelc " A fuzzy genetic algorithm for automatic orthogonal graph drawing" journal of SciVerse Science Direct, 2011, pp, 1379-1380.
- [13] I.G. Toll's, G. Di Battista, P. Eades, R. Tamassia, and Graph Drawing: Algorithms for the Visualization Of Graphs, Prentice Hall, 1998.
- [14] R. Tamassia, New layout techniques for entity-relationship diagrams, in: Proceedings of the Fourth International Conference on Entity-Relationship Approach, IEEE Computer Society, Washington, DC, USA, 1985, pp. 304–311.
- [15] T. Lengauer, Combinatorial algorithms for integrated circuit layout, John Wiley & Sons Inc., New York, NY, USA, 1990
- [16] Bernadette M.M. Neta Gustavo H.D. Arajo, Frederico G. Guimaraes, Renato C. Mesquita, Petr Ya. Ekelc " A fuzzy genetic algorithm for automatic orthogonal graph drawing" journal of Sci Verse Science Direct, 2011, p, 1385.



(a)



(b)

Fig. 12. $V = 100$: (a) drawing obtained with topology-shape-metric (b) drawing obtained with the fuzzy genetic algorithm

Performance Enhancement of Scheduling Algorithm in Heterogeneous Distributed Computing Systems

Aida A. NASR

Computer Science & Eng. Dept.,
Faculty of Electronic Engineering,
Menoufia Uni., Menouf 32952,
Egypt

Nirmeen A. EL-BAHNASAWY

Computer Science & Eng. Dept.,
Faculty of Electronic Engineering,
Menoufia Uni., Menouf 32952,
Egypt

Ayman EL-SAYED

Computer Science & Eng. Dept.,
Faculty of Electronic Engineering,
Menoufia Uni., Menouf 32952,
Egypt

Abstract—Efficient task scheduling is essential for obtaining high performance in heterogeneous distributed computing systems. Some algorithms have been proposed for both homogeneous and heterogeneous distributed computing systems. In this paper, a new static scheduling algorithm is proposed called Node Duplication in Critical Path (NDCP) algorithm to schedule the tasks efficiently on the heterogeneous distributed computing systems. The NDCP algorithm focuses on reducing the makespan and provides better performance than the other algorithms in metrics of speedup and efficiency. It consists of two phases, priority phase and processor selection phase. From the theoretical analysis of the NDCP algorithm with other algorithms for a Directed Acyclic Graph (DAG), the better performance is observed.

Keywords—static task scheduling; heterogeneous distributed computing systems; Meta-heuristic algorithms

I. INTRODUCTION

The availability of high-speed networks and diverse sets of resources lead to a platform, called as heterogeneous platform. Such a platform contains interconnected resources with different computing capabilities and different computing speeds. To run an application in this heterogeneous environment, several issues need to be considered such as partitioning the application, scheduling the tasks; etc. It is referred to such a system as Heterogeneous Distributed Computing System (HDCCS). In recent years, HDCCS has emerged as a popular platform to execute computationally intensive applications with diverse computing needs [1].

Task scheduling is of vital importance in HDCCS since a poor task-scheduling algorithm can undo any potential gains from the parallelism presented in the application. In general, the objective of task scheduling is to minimize the completion time of a parallel application by properly mapping the tasks to the processors. There are typically two categories of scheduling models: static and dynamic scheduling. In the static scheduling case, all information regarding the application and computing resources such as execution time, communication cost, data dependency, and synchronization requirement is assumed available a priori. Scheduling is performed before the actual execution of the application [2, 3]. On the other hand, in the dynamic mapping a more realistic assumption is used. Very little a priori knowledge is available about the application and computing resources. Scheduling is done at run-time [4]. In this paper, it is focused on static scheduling. Static scheduling

has three categories: *list-based*, *clustering* and *duplication based*.

List-scheduling algorithms contain two phases: a *task prioritization phase*, and a *machine assignment phase*. In *task prioritization phase*, the algorithms assign a certain priority that is computed, to node in the DAG. In *machine assignment phase*, each task depending on its priority is assigned to machine that minimizes the cost function [5-9]. Examples of list-based algorithms are Heterogeneous Earliest Finish Time (HEFT) and Critical Path on Processor (CPOP) [10]. Another static scheduling category is task duplication based algorithms, in which tasks are duplicated on more than one processor to reduce the waiting time of the dependent tasks. The main idea behind duplication based scheduling is to utilize processor idling time to duplicate predecessor tasks. This may avoid transfer of results from a predecessor, through a communication channel, and may eliminate waiting slots on other processors and reduce the communication overheads [11,12]. An example for duplication algorithms is Heterogeneous Critical Node First (HCNF) and Scalable Task Duplication Based Scheduling (STDS) [13,14].

In this paper, a new algorithm called Node Duplication in Critical Path (NDCP) is developed for static task scheduling for the HDCCS with limited number of processors. The motivation behind this algorithm is to generate the high quality task schedule that is necessary to achieve high performance in HDCCS. The developed algorithm is based on critical path method to give each node a priority, and the duplication algorithm to minimize communication overheads. Finally, idle time is decreased in proposed algorithm.

The remainder of this paper is organized as follows. Section II discusses problem definition. Section III gives an overview of the related work. Section IV presents our developed NDCP algorithm with examples. Section V discusses the results and in section VI, conclusions are given.

II. PROBLEM DEFINITION

Task scheduling for HDCCS is the problem of assigning the tasks of a parallel application to the processors of a HDCCS, which have diverse capabilities, and specifying the start execution time of each task. This must be done in a way that respects the precedence constraints among tasks. An efficient schedule is one that minimizes the total execution time, or the schedule length, of the parallel application [15-23].

The models of HDCS [24] and the model of application to be considered in this work can be described as follows. By using DAG, the parallel application is represented. DAG is defined by the tuple (T,E), where T is a set of n tasks and E is a set of e edges. Each $t_i \in T$ represents a task in the parallel application, which in turn is a set of instructions that must be executed sequentially in the same processor without interruption. Each edge $(t_i, t_j) \in E$ represents a precedence constraint, such that the execution of $t_j \in T$ cannot be started before $t_i \in T$ finishes its execution. If $(t_i, t_j) \in E$, then t_i is a parent of t_j and t_j is a child of t_i . A task with no parents is called an entry task t_{entry} , and a task with no children is called an exit task t_{exit} . Each edge $(t_i, t_j) \in E$ has a value that represents the estimated inter-task communication cost required to pass data from the parent task t_i to the child task t_j . Because tasks might need data from their parent tasks, a task can start execution on a processor only when all data required from its parents become available to that processor; at that time the task is marked as ready. The speed of the inter-processor communication network is assumed to be much lower than the speed of the intra-processor bus. Therefore, when two tasks are scheduled on the same processor the communication cost between these tasks can be ignored. The HDCS is represented by a set P of m processors that have diverse capabilities. The $n \times m$ computation cost matrix C stores the execution costs of tasks. Each element $c_{i,j} \in C$ represents the estimated execution time of task t_i on processor p_j . Precise calculation of the running times of the tasks on the processors is unfeasible before running the application. All processors in the HDCS are assumed to be fully connected. Communications between processors occur via independent communication units; this allows for concurrent execution of computation of tasks and communications between processors. After scheduling all the tasks of a parallel application on the processors of a HDCS, the schedule length is defined as the longest finish time of the HDCS. Fig. 1 presents an example of a parallel application consisting of five tasks and a HDCS with two processors, where the application is represented as a DAG and the execution costs estimated for the five tasks on the HDCS are shown as a computation cost matrix.

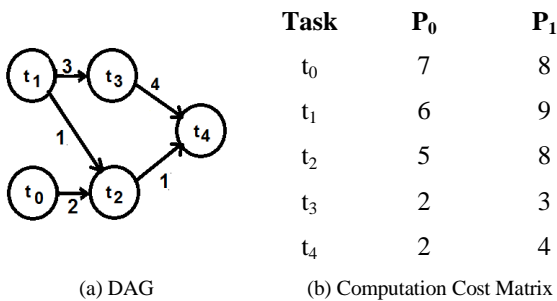


Fig. 1. Example of a DAG and Computation Cost Matrix

Definition (1) Critical Path (CP): CP of a DAG is the longest path from the entry task to the exit task in the graph.

Definition (2) $EST(t_i, P_j)$ [10]: Denotes the Earliest Start Time of a task t_i on a processor P_j and is defined as shown in Equation 1.

$$EST(t_i, P_j) = \max\{T_{Available}(P_j), \max\{AFT(t_k) + c_{k,i}\}\} \quad (1)$$

Where $T_{Available}(P_j)$ is the earliest time at which processor P_j is ready. $AFT(t_k)$ is the Actual Finish Time of a task t_k (where t_k is the parent of task t_i and $k=1, 2, \dots, n$) on the processor P_j . $c_{k,i}$ is the communication cost from task t_k to task t_i , $c_{k,i}$ equal zero if the predecessor task t_k is assigned to processor P_j . For the entry task, $EST(t_{entry}, P_j) = 0$.

Definition (3) $EFT(t_i, P_j)$ [10]: Denotes the Earliest Finish Time of a task t_i on a processor P_j and is defined as shown in Equation 2.

$$EFT(t_i, P_j) = EST(t_i, P_j) + w_{i,j} \quad (2)$$

Which is the Earliest Start Time of a task t_i on a processor P_j plus the computational cost $w_{i,j}$ of t_i on a processor P_j .

Definition (4) Data Ready Time (DRT): is the idle time waited by a t_i on processor p_j .

Definition (5) Maximum Parent (MP): maximum parent of task t_i is a parent task t_k such that the value of $EFT(t_k, p_m) + c(t_k, t_i)$ is the largest among all t_i 's parent tasks.

Definition (6) Very Important Task (VIT): is the task that belongs to the critical path of DAG.

III. RELATED WORK

In this section, it is given an overview of some algorithms, specifically list-based scheduling algorithms.

A. Heterogeneous Earliest Finish Time

The HEFT algorithm executes in two phases: a task-prioritizing phase and processor selection phase [10]. In task prioritizing phase, the algorithm selects the task with the highest upward rank at each step. Upward rank is given by Equation 3.

$$Rank_u = \bar{w}_i + \max_{n_j \in succ(n_i)} (\bar{c}_{i,j} + rank_u(n_j)) \quad (3)$$

Where $succ(n_i)$ is the set of immediate successors of task n_i , $\bar{c}_{i,j}$ is the average communication cost of edge (i,j) , and \bar{w}_i is the average computation cost of task n_i . In processor selection phase, the selected task is assigned to the processor which minimizes its earliest finish time with an insertion-based approach. The algorithm has an $O(n^2p)$ time complexity for n nodes and p processors.

B. Critical Path On Processor Algorithm

The CPOP algorithm consists of two phases: prioritizing phase and processor selection phase [10]. In task prioritizing phase, the algorithm selects the task with the highest of upward rank + downward rank value at each step. Downward rank can be calculated by Equation 4.

$$Rank_d(n_i) = \max_{n_j \in pred(n_i)} (\bar{c}_{j,i} + \bar{w}_j + rank_d(n_j)) \quad (4)$$

Where $pred(n_i)$ is the set of immediate predecessors of task n_i . The algorithm targets scheduling of all critical tasks (i.e., tasks on the critical path of the DAG) onto a single processor, in which the critical tasks are executed in minimum time as possible. If the selected task is noncritical, the processor selection phase is based on earliest execution time with insertion-based scheduling. Like HEFT algorithm, the CPOP

algorithm has an $O(n^2p)$ time complexity for n nodes and p processors.

C. Path-based Heuristic Task Scheduling Algorithm

The PHTS algorithm is proposed for a bounded number of heterogeneous processors consisting of three phases namely, a path-prioritizing phase, task selection phase, and processor selection phase [25]. Path prioritizing phase for computing the priorities for all possible paths. Each path is assigned by a value called rank(p_j), is given by Equation 5.

$$Rank(p_j) = \sum_{t_i \in p_j} \bar{w}_i + \overline{c_{i,succ(t_i)}} \text{-----}(5)$$

Where \bar{w}_i is the average computation cost of a task t_i . It is computed by $\bar{w}_i = \sum_{j=1}^m w_{i,j} / m$, and $\overline{c_{i,succ(t_i)}}$ is the communication cost of edge from task t_i to its successor, if exists.

In task selection phase, the algorithm selects the unscheduled tasks from the paths in the sorted path list. During the task selection, the algorithm applies the following conditions on each task:

- The task should not be scheduled earlier.
- The task has no parents or its parents are scheduled.

Finally, the algorithm apply the processor selection phase like HEFT algorithm. The algorithm has an $O(n^2p)$ time complexity for n nodes and p processors.

D. Highest Communicated Path of Task Algorithm

HCPT algorithm consists of three phases called, level sorting phase, task prioritizing phase and processor selection phase. In level sorting phase, the given DAG is traversed in a top-down fashion to sort tasks in each level. In task prioritizing phase, the HCPT algorithm computes the task priority by using the rank value as shown in Equation 6.

$$Rank(t_i) = MCP(t_i) + \max_{t_j \in succ(t_i)} (\bar{c}_{i,j} + Rank(t_j)) \text{-----}(6)$$

Where MCP(t_i) refers to Mean Communication of Parent tasks. It is computed by Equation 7.

$$MCP(t_i) = (\sum_{j=1}^n C_{j,i}) / y \text{-----}(7)$$

Where y is the number of parent tasks. Finally, the algorithm apply the processor selection phase like HEFT and PHTS algorithms [8].

IV. OUR SCHEDULING ALGORITHM

The Node Duplication in Critical Path (NDCP) algorithm is developed for static task scheduling algorithm for HDACS with limited number of processors. This algorithm based on Critical Path Merge [26] (CPM) technique and task duplication technique.

Any algorithm applying the list scheduling technique has the freedom to define the two criteria: the priority scheme for the nodes and the choice criterion for the processor. Fig. 2 and Fig. 3 show steps of NDCP algorithm. It consists of two phases namely, priority phase and processor selection phase.

A. Priority phase

In this paper, NDCP algorithm modified into priority scheme, where gives the priority to the path instead of the node.

```

Schedule_Task( $t_i$ )
Begin
For each processor  $P_j$  in the processor set ( $P_j \in Q$ ) do
    Compute  $EFT(t_i, P_j)$  value
End for
Assign  $t_i$  to the  $p_j$  that minimizes  $EFT$ 
If  $t_i$  is VIT
{
If  $DRT(t_i, p_j) > w(MP, p_j)$ 
If  $EST(t_i, p_j) > EFT(MP, p_j)$ 
{
    Duplicate  $MP$  on  $P_j$  without violate the dependency constraints
    Update  $EFT$  of  $t_i$  on  $p_j$ 
}
}
End
    
```

Fig. 2. Schedule Task Function

```

Set the computation cost of tasks & communication cost of edges
While there are tasks in given DAG do
    Compute the Critical Path
        using  $CP_x = \text{Max}(\sum_1^n w(t_i)_{max} + \sum \overline{c_{i,succ(t_i)}})$ 
    Put critical path in Critical Path List (CPL)
    Remove critical path from the DAG
    Update the DAG
End While
For each path  $CP_x$  in CPL
For each task  $t_i$  in  $CP_x$ 
If  $t_i$  has no parents or all parents are scheduled then
{
    Call Schedule_Task( $t_i$ )
For each Waited Task  $t_w$  in WL
If all parents of  $t_w$  are scheduled then
{
    Call Schedule_Task( $t_w$ )
    Remove  $t_w$  from WL
}
}
End for
Else
    Put task  $t_i$  in Waited List (WL)
End for
End for
    
```

Fig. 3. Node Duplication in Critical Path Algorithm (NDCP)

The NDCP algorithm computes the critical path of the DAG using Equation 8.

$$CP = \text{Max}\{\sum_1^b w(t_i)_{max} + \sum c_{i,succ(t_i)}\} \text{-----}(8)$$

Where $w(t_i)_{max}$ is the maximum computation of task t_i . b is the number of CP tasks. $c_{i,succ(t_i)}$ is the communication between t_i and its successor, where t_i and its successor belong to the same critical path. Then the algorithm removes this critical path.

After updating the DAG, the algorithm computes next critical path and so on. The NDCP computes a critical path using the largest weight for each task t_i at slowest processor p_j . It sorts all critical paths into critical path list CPL in descending order.

B. Processor selection phase

This phase consists of two stages: processor stage and duplication test stage. In processor stage, NDCP algorithm selects a CP_x from CPL, and then it selects task t_i from CP_x . If t_i has no parents or all parents are scheduled, the algorithm calls *Scheduled_Task* function (as shown in Fig. 2). In *Schedule_Task* function, the NDCP algorithm calculates EFT of task t_i by Equation 2 for each processor, and selects the processor that has a minimum EFT to assign the task. With high performance algorithms, some processors are idle during the execution of the application because of DRT. If DRT is enough to duplicate MP, the execution time of the parallel application could be reduced [11]. So, the algorithm applies task duplication to reduce the makespan. *Schedule_Task* function executes also stage of duplication test. The algorithm test, if DRT of task t_i is more than the weight of MP on the same processor p_j , the algorithm duplicates the MP on p_j and updates EFT of task t_i . The duplication stage is applied on VIT only. This must be done without violating the precedence constrains among tasks. If the task has parents without scheduled, the algorithm puts this task in waiting list W_L to be ready. Once all parents are scheduled, the algorithm selects the task from W_L to schedule. It also removes that task from W_L and continues. Using W_L guarantees scheduling all of important tasks early. A case study is taken into account as following.

Case Study: Considering the application DAG shown in Fig. 4, Table 1 shows the computation matrix. The generated schedule along with stepwise trace of the HCPT algorithm and NDCP algorithm are shown in Fig. 5. With applying task duplication, DRT of tasks decreases. So the schedule length with task duplication decreases. PHTS, HEFT and CPOP algorithms were also applied on sample DAG 1, and the results respectively were 47,47, 57. From Fig. 5 it is clear that our algorithm is outperforms the others because it marks 43 units. So the scheduling performance is enhanced. It is noted that, the NDCP algorithm applies task duplication to decrease the communication overhead by using idle time in scheduling. The NDCP algorithm applies the task duplication on VIT only, because the task that belongs to the critical path is critical task. So, if EFT of VIT decreases the schedule length of application will decrease. When the algorithm duplicates a task, it decreases DRT of its childs and decreases also EFT. This leads to good utilization of processors in the system.

TABLE I. COMPUTATION MATRIX

Task	P0	P1
T ₁	10	7
T ₂	8	5
T ₃	5	2
T ₄	15	10
T ₅	12	14
T ₆	3	7
T ₇	8	4
T ₈	4	3

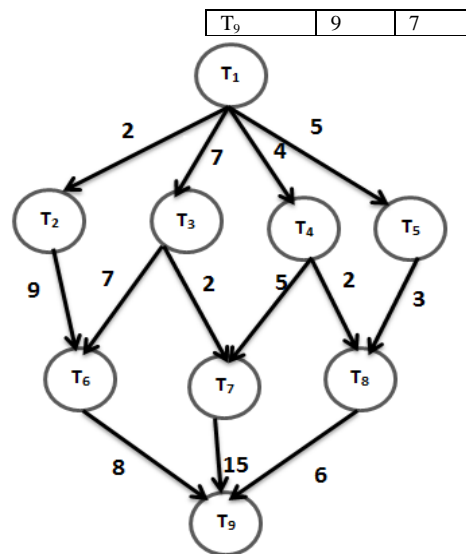


Fig. 4. Sample DAG 1

V. RESULTS AND DISCUSSIONS

A. Simulation Environment

To evaluate the performance of our developed NDCP algorithm, a simulator had been built using visual C#.NET 4.0 on machine with:

- Processor: Intel(R) Core(TM) i3 CPU M 350 @2.27GHz.
- Installed memory RAM: 4.00 GB.
- System type: window 7, 64-bit.

To test the performance of NDCP algorithm with the other algorithms a set of randomly generated graphs is created by varying a set of parameters that determines the characteristics of the generated DAGs.

These parameters are described as follows:

- DAG size: n : The number of tasks in the DAG.
- Density:

It is used "sameprob" method to generate the DAG [27]. Let A denote a task connection matrix with elements $a(i,j)$, where $0 \leq i \leq n$, and $0 \leq j \leq n$, represent the task number (t_0 is the entry dummy node and t_n is the exit dummy node). When $a(i,j)=1$, t_i precedes task t_j , when $a(i,j)=0$, t_i and t_j are independent of each other. In the "sameprob" edge connection method, $a(i,j)$ is determined by independent random values defined as follows:

$P[a(i,j)=1] \leq prob$ for $1 \leq i < j \leq n$ and $P[a(i,j)=0] > prob$ for $1 \leq i < j \leq n$, $P[a(i,j)=0]=1$ if $i \geq j$, where $prob$ indicates the probability that there exists an edge (precedence constraint) between t_i and t_j .

With six different numbers of processors varying from 8, 16, 32, 64 and 80 processors. For each number of processors, six different DAG sizes have been generated varying from 40, 60, 80,100,120 and 150 tasks. In each experiment, the

probability p is assigned from the corresponding set given below:

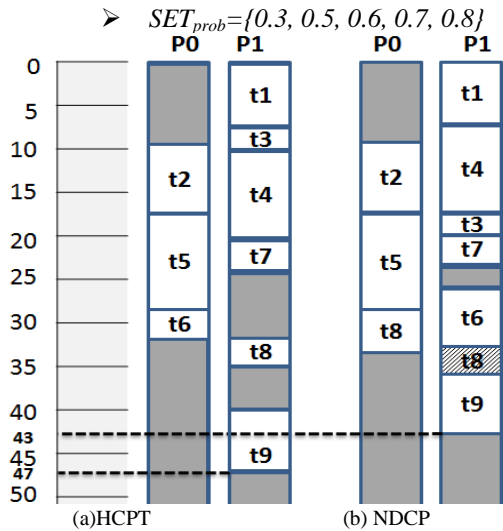


Fig. 5. The Schedules Generated by Different Algorithms

HCPT is applied (which is the best performance algorithm compared with HEFT, PHTS, CPOP), PHTS (which is the best performance algorithm compared with HEFT) and NDCP algorithms on Standard Task Graph Set (STG) (a kind of benchmark for evaluation of multiprocessor scheduling algorithms) [28]. The results of PHTS, HCPT and NDCP respectively were 254, 230 and 205 units on random task graph 50/tmp/50/rand0005.stg. In addition, the algorithms are applied on random task graph 50/tmp/50/rand0000.stg, and the results of PHTS, HCPT and NDCP were 88, 80 and 76 units respectively, from the results, it is noted that our algorithm outperforms other algorithms compared in performance.

B. Comparison Metrics and Results

The comparison metrics are schedule length, speedup, efficiency, and time complexity.

1) Schedule Length

Schedule length is the maximum finish time of the exit task in the scheduled DAG [26]. The main function of task scheduling is minimizing an application time, so schedule length is the important metric to measure performance of task scheduling algorithm. The NDCP algorithm used critical path to detect task priority, because the critical path contains a very important tasks. The NDCP algorithm computes the first critical path to get rid the critical tasks then it computes the next critical path (after updating DAG) to get rid the next critical tasks and so on. It deals with the DAG, after computing a critical path, as a new DAG with new critical path. The NDCP algorithm uses also task duplication to reduce DRT of the successors, and it could reduce the overall time of application. The algorithm duplicates MP of VIT only. Therefore, the NDCP algorithm is more efficient than other algorithms. This appeared from Fig. 6 to Fig. 10. Figures show scheduling length versus number of tasks with varying number of processors 8, 16, 32, 64 and 80. Performance ratio in schedule length is 11%.

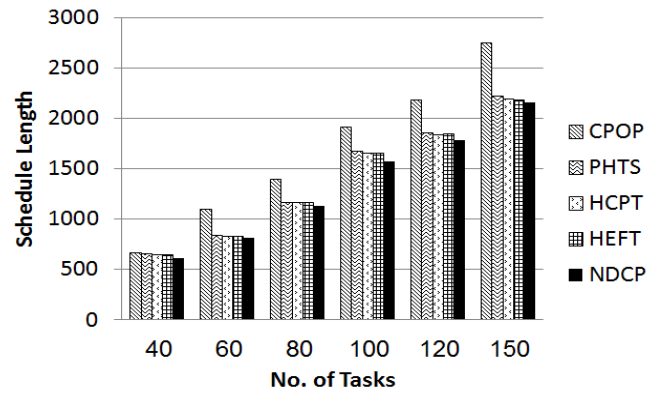


Fig. 6. Schedule Length at 8 Processors

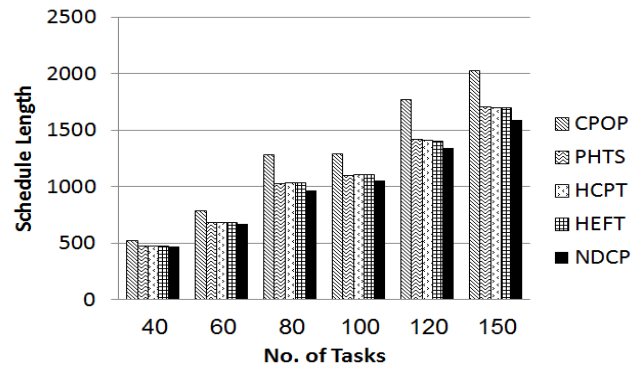


Fig. 7. Schedule Length at 16 Processors

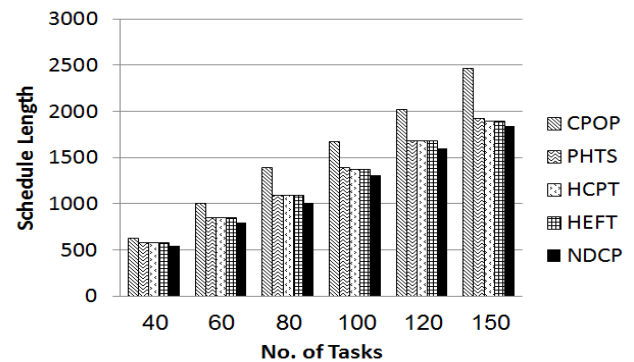


Fig. 8. Schedule Length at 32 Processors

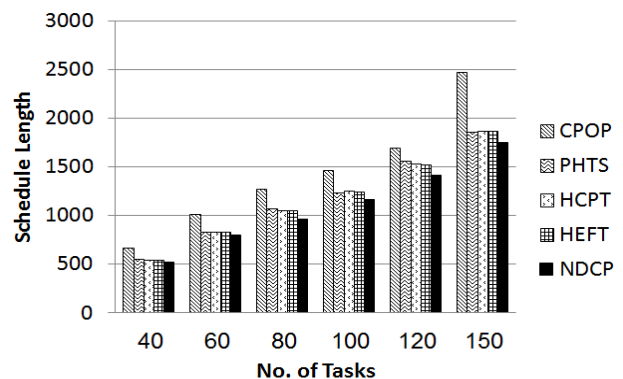


Fig. 9. Schedule Length at 64 Processors

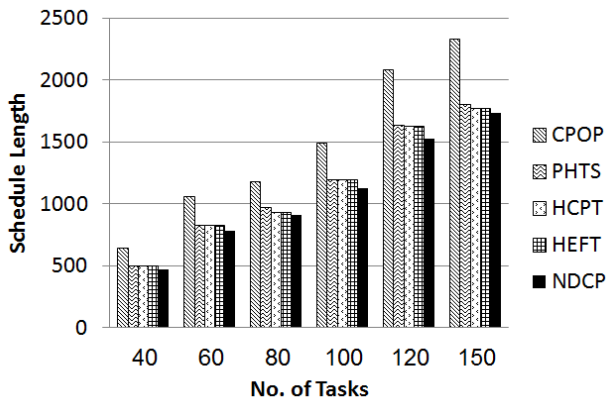


Fig. 10. Schedule Length at 80 Processors

2) Speedup

Speedup of a schedule is defined as the ratio of the schedule length obtained by assigning all tasks to the fastest processor, to the schedule length of application [24]. The speedup is given by Equation 9.

$$Speedup = \frac{Min_{j \in P} [\sum_{i \in V} w(i, j)]}{SL} \quad \text{----- (9)}$$

Where $w(i, j)$ means the weight of task t_i on processor p_j and SL means the schedule length. Speedup is a good measure for the execution of an application program on a parallel system. Due to minimize schedule length, all processors have finished tasks execution earlier and speedup of *NDCP* algorithm increases. The results of the comparative study according to the speedup parameter have been presented from Fig. 11 to Fig. 16. According to the results, performance ratio of speedup is calculated as 10.5%.

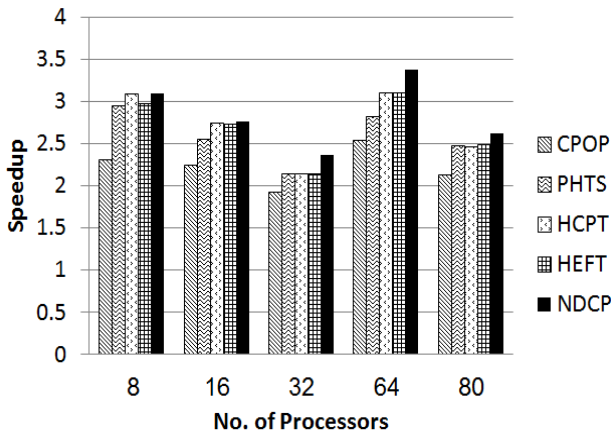


Fig. 11. Speedup with 40 Tasks

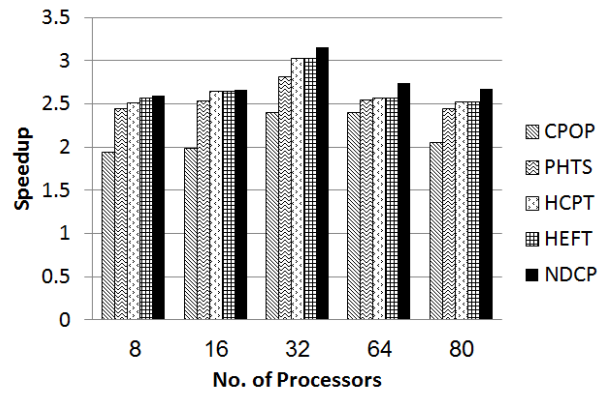


Fig. 12. Speedup with 60 Tasks

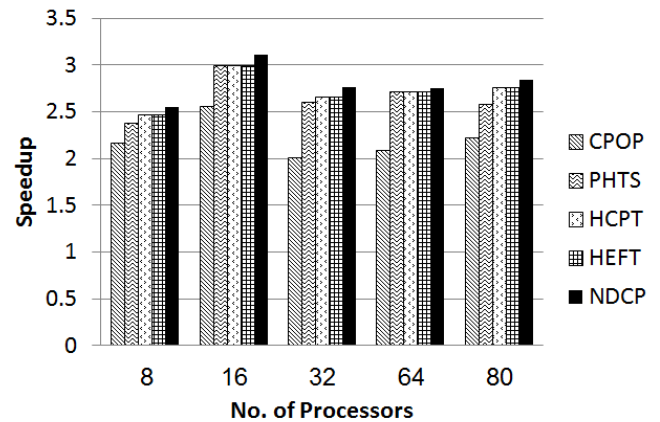


Fig. 13. Speedup with 80 Tasks

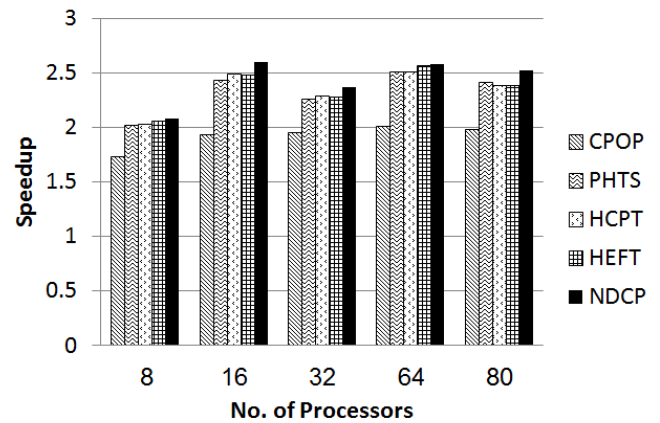


Fig. 14. Speedup with 100 Tasks

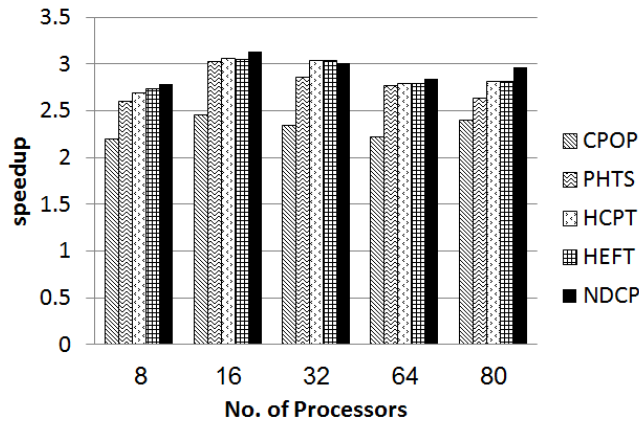


Fig. 15. Speedup with 120 Tasks

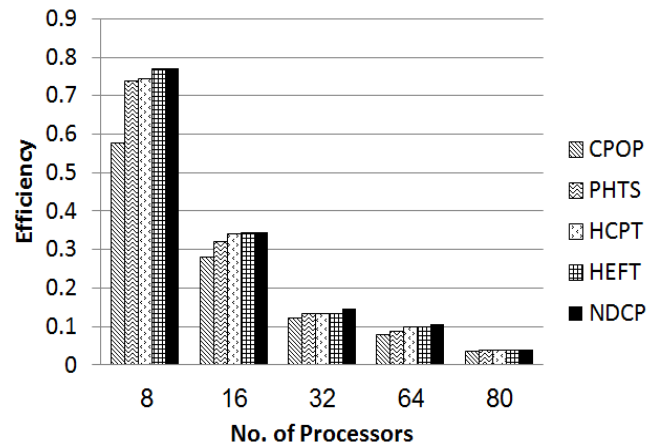


Fig. 17. Efficiency with 40 Tasks

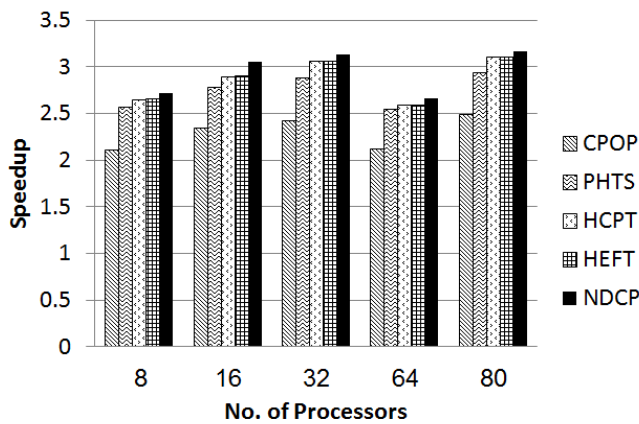


Fig. 16. Speedup with 150 Tasks

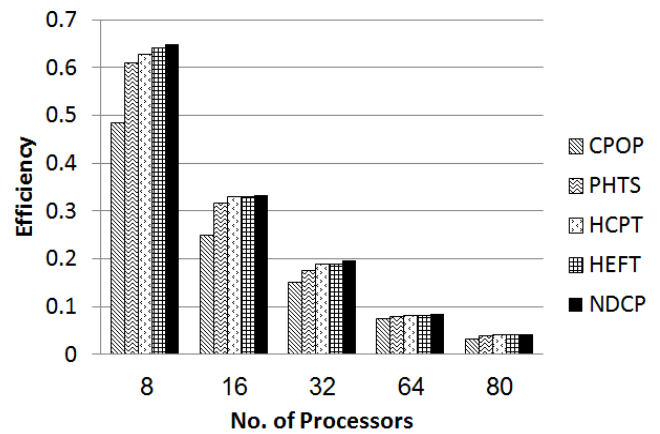


Fig. 18. Efficiency with 60 Tasks

3) Efficiency

Efficiency is the speedup divided by the number of processors used [24]. The efficiency is described in Equation 10.

$$Efficiency = \frac{Speedup}{number\ of\ processors\ used} \quad \text{-----(10)}$$

Using task duplication involves the largest number of parallel computers and makes balance between them. Efficiency is an indication to what percentage of a processors time is being spent in useful computation. So efficiency of the *NDCP* algorithm outperforms efficiency of the other algorithms. From Fig. 17 to Fig. 22, figures show efficiency of the *NDCP* algorithm compare with *HEFT*, *CPop*, *PHTS* and *HCPT* algorithms. The performance ratio in efficiency which has been achieved by *NDCP* algorithm is 9.3%. According to efficiency parameter, our proposed *NDCP* algorithm achieves better performance than the other algorithms.

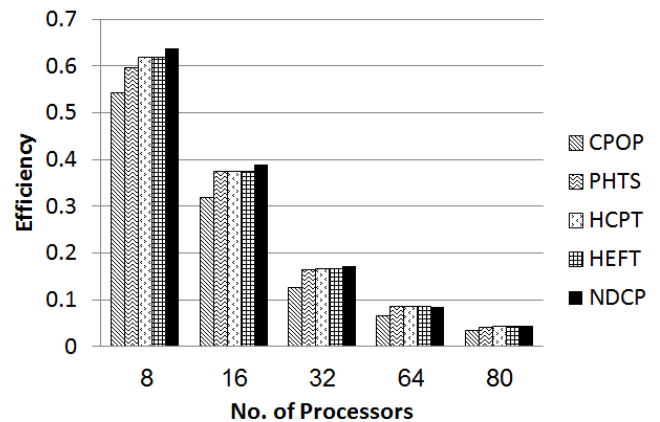


Fig. 19. Efficiency with 80 Tasks

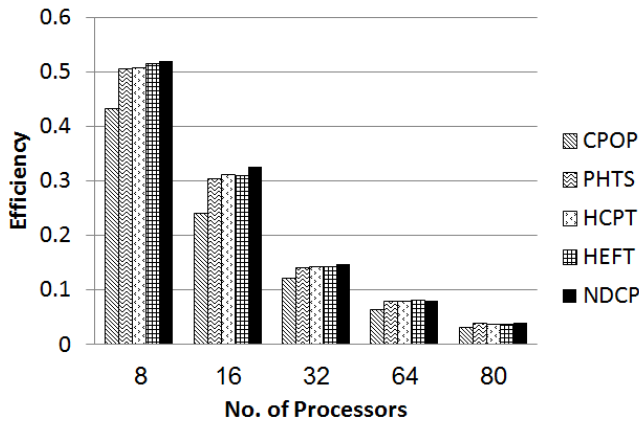


Fig. 20. Efficiency with 100 Tasks

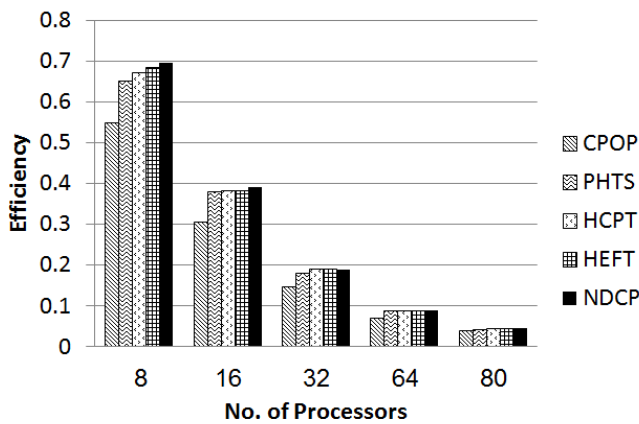


Fig. 21. Efficiency with 120 Tasks

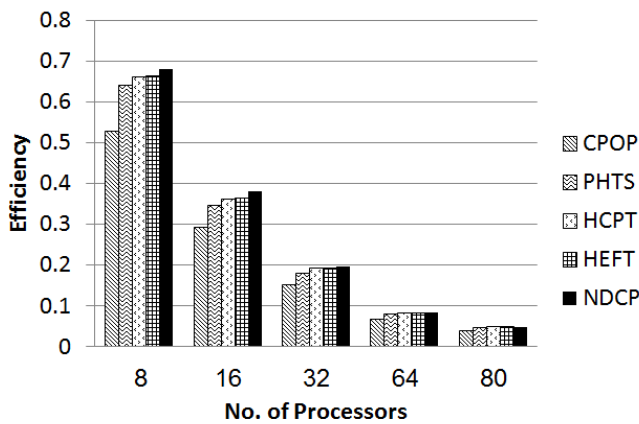


Fig. 22. Efficiency with 150 Tasks

4) Time Complexity

Time complexity is the amount of time taken to assign every task to specific processor according to specific priority. The *NDCP* algorithm has an $O(cp+w*p*n)$ time complexity for *cp* number of critical paths, *w* number of waited tasks $w < n$, *n* number of tasks and *p* number of processors.

Our algorithm may approximate time complexity into $O(wpn)$. From Table II; it is noted that, task duplication algorithms have high time complexity. But *NDCP* algorithm has the lowest time complexity because The *NDCP* algorithm tests task duplication, if there is an idle time at specific processor not at all processors. The algorithm assigns the task firstly then examines, if there is enough idle time before the task to duplicate its parent or not so, our algorithm has the lowest time complexity for task duplication. This makes the *NDCP* algorithm outperformance the other algorithms.

TABLE II. TIME COMPLEXITY OF SOME ALGORITHMS

Algorithm	Complexity	Use Duplication
HEFT	$O(n^2p)$	No
CPOP	$O(n^2p)$	No
PHTS	$O(n^2p)$	No
HCPT	$O(n^2p)$	No
HCNF	$O(n^2 \log n)$	Yes
NDCP	$O(wpn)$	Yes

VI. CONCLUSIONS

In this paper, a new scheduling algorithm has been presented for heterogeneous distributed computing systems (*HDCS*) to enhancement scheduling performance. This algorithm based on *Critical Path Merge (CPM)* technique and task duplication technique. The *NDCP* algorithm duplicate MP for VIT only. The performance analysis showed that the proposed *NDCP* algorithm has better performance than *HCPT*, *PHTS*, *HEFT* and *CPOP* algorithms. According to the simulation results, it is found that the *NDCP* algorithm is better than the other algorithms in terms of schedule length, speedup and efficiency. The *NDCP* algorithm also has the lowest time complexity $O(wpn)$. Performance improvement ratio in schedule length is 11%, performance improvement ratio in speedup is 10.5% and performance improvement ratio in efficiency is 9.3%. In addition, the algorithms are applied on Standard Task Graph *STG* as a benchmark, and it is observed that *NDCP* algorithm is more efficient than the other algorithms.

REFERENCES

- [1] N. Wahid and N. Wafa, "A New DAG Scheduling Algorithm for Heterogeneous Platforms," in 2nd IEEE International Conference on Parallel Distributed and Grid Computing (PDGC). IEEE, pp.114-119, 2012.
- [2] H. Yang, P. Lee and C. Chung, "Improving static task scheduling in heterogeneous and homogeneous computing systems," IEEE International Conference on Parallel Processing, ICPP, pp. 45-45, 2007.
- [3] H. Izakian, A. Abraham and V. Snasel, "Comparison of heuristics for scheduling independent tasks on heterogeneous distributed environments," in Proc. of the International Joint Conference on Computational Sciences and Optimization, IEEE, vol. 1, pp. 8-12, 2009.
- [4] Sheikh, H.F., Ahmad, I., "Dynamic task graph scheduling on multicore processor for performance, energy, and temperature optimization," IEEE, Published in: Green Computing Conference (IGCC), Page: 1-6, June 2013.
- [5] E. Ilavarasan and P. Thambidurai, "Low Complexity Performance Effective Task Scheduling Algorithm for Heterogeneous Computing Environments," Journal of Computer Sciences, 3(2):94-103, 2007.

- [6] H. Arabnejad and J. Barbosa, "List scheduling algorithm for heterogeneous systems by an optimistic cost table," IEEE Transactions on Parallel and Distributed Systems, , no. 99, pp. 1–1, 2013.
- [7] R. Eswari and S. Nickolas, "Expected completion time-based scheduling algorithm for heterogeneous processors", in Proc. International Conf. Information Communication and Management, IPCSIT vol.16 2011, pp.72-77.
- [8] A. A. Nasr, N. A. El-bahnasawy and A. El-sayed, "Task scheduling optimization in heterogeneous distributed systems", International Journal of Computer Applications (0975 – 8887), Volume 107, No 4, December 2014.
- [9] A. A. Nasr, N. A. El-bahnasawy and A. El-sayed, "Task scheduling algorithm for high performance heterogeneous distributed systems", International Journal of Computer Applications (0975 – 8887) , Volume 110 – No. 16, January 2015.
- [10] H.Topcuoglu, S. Hariri, and M.Y.Wu, "Performance-effective and low-complexity task scheduling for heterogeneous computing", IEEE Trans. Parallel and Distributed Systems, Vol. 13, No.3, pp. 260- 274, March 2002.
- [11] I. Ahmad, and Y. Kwok, "A New approach to scheduling parallel programs using task duplication," Proc. International Conf. Parallel Processing, Vol.2, pp. 47-51, 1994
- [12] J. Mei and K Li, "Energy-Aware scheduling algorithm with duplication on heterogeneous computing systems," Publish in: Grid Computing (GRID), ACM/IEEE 13th International Conference, Page: 122 -129, Sept. 2012.
- [13] Baskiyar, S.; SaiRanga, P.C., "Scheduling directed a-cyclic task graphs on heterogeneous network of workstations to minimize schedule length," 2003 International Conference on Parallel Processing Workshops, pp.97,103, 6-9 Oct. 2003.
- [14] S. Darbha and D. P. Agrawal, "A task duplication based scalable scheduling algorithm for distributed memory systems". J. Parallel Distrib. Comput, Vol. 46, PP. 15-27, 1997.
- [15] Y. Xu, K. Li, T. T. Khac and M. Qiu, "A multiple priority queuing genetic algorithm for task scheduling on heterogeneous computing systems," IEEE 14th International Conference on High Performance Computing and Communications, pp. 639-646, 2012.
- [16] M. Gallet, L. Marchal and F. Vivien, "Efficient scheduling of task graph collections on heterogeneous resources," IPDPS 2009-Proceeding of the IEEE International Parallel and Distributed Processing Symposium, pp.1-11, 2009.
- [17] C. Hui, " A high efficient task scheduling algorithm based on heterogeneous multi-core processor", IEEE, Database Technology and Application (DBTA) Pages 1-4, Nov. 2010.
- [18] Y. Kang and Y. Lin, "a recursive algorithm for scheduling of tasks in a heterogeneous distributed environment," IEEE, Biomedical Engineering and information (BMEI), Pages 2099-2103, Vol:4, Oct. 2011.
- [19] Rahman M., Venugopal S. and Buyya R., "A dynamic critical path algorithm for scheduling scientific workflow applications on global grids," 3rd IEEE International Conference on e-Science and Grid Computing, pp. 35-42, 2007.
- [20] Al Na'mneh, R.A., Darabkh, K.A., "A new genetic-based algorithm for scheduling static tasks in homogeneous parallel system," Published in: Robotics, Biomimetics, and Intelligent Computational Systems (ROBIONETICS), 2013 IEEE International Conference, Page: 46-50, Nov. 2013.
- [21] Thambidurai P. and Mahilmanan R., "Performance effective task scheduling algorithm for heterogeneous computing system," IEEE Proceedings of the 4th International Symposium on Parallel and Distributed Computing, pp. 28-39, 2005.
- [22] R. Bajaj and D. P. Agrawal, "Improving scheduling of tasks in a heterogeneous environment," IEEE Transactions Parallel Distributed System, vol. 15, pp. 107–118, February 2004.
- [23] M. Ehsan, M. Sajjad, H. Altaf, N. Muhammad and A. Shoukat, "SDBATS: A novel algorithm for task scheduling in heterogeneous computing systems," IEEE International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum, pp.43-53, 2013.
- [24] N. A. Bahnasawy, M. A. Koutb, M. Mosa and F. Omara, "A new algorithm for static task scheduling for heterogeneous distributed computing systems," African Journal of Mathematics and Computer Science Research Vol. 4(6), pp. 221-234, June 2011.
- [25] R. Eswari and S. Nickolas, "Path-based heuristic task scheduling algorithm for heterogeneous distributed computing systems", International Conference on Advances in Recent Technologies in Communication and Computing, 2010.
- [26] K. S. Manoj and T. Rajesh, "A survey on scheduling of parallel programs in heterogeneous systems," International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 1, Issue 8, October 2012.
- [27] V. A. F. Almeida, I. M. M Vasconcelos, J. N. C. Áraabe and D. A. Menascé. " Using Random Task Graphs to Investigate the Potential Benefits of Heterogeneity in Parallel Systems", Proc. Supercomputing '92, pp. 683-691 (1992).
- [28] <http://www.kasahara.elec.waseda.ac.jp/schedule/index.html>.

Numerical Evaluation of the Effect of Gradient on Reflection Coefficient of Continuously Graded Layer

Ahmed Markou, Hassan Nounah

Laboratory of Metrology and Information Processing, Department of Physics
Faculty of Sciences, Ibnou Zohr University
Agadir, Morocco

Abstract—This paper presents a numerical model, based on transfer matrix method, for modeling the propagation of surface acoustic waves at the interface formed by the coupling liquid and a continuously inhomogeneous thin layer on a semi-infinite substrate. The tow-dimensional spectrum of reflection coefficient computed by this model, allows determining the modes which propagate at the studied interface, this model treats different profile of gradients, and the numerical results obtained show that the reflection coefficient is sensitive to the variation of these gradients.

Keywords—surface acoustic waves; nondestructive testing; functionally graded materials; dispersion curve; Lamb modes

I. INTRODUCTION

The functionally graded materials are recently developed in Japan for to be used as thermal barrier materials for aerospace structures and fusion reactors (high temperature applications), and now they are developed for the general use in different engineering applications. The FGMs are made from different of material constituents, where their properties such as elastic constants and density are varying according to the spatial coordinates [1]. The profile of gradient in physical properties obtained is governed by the procedure of fabrications [2][3]. For modeling FGMs, tow approaches are possible. The first assumed that the gradient is piecewise varying, and the FGM is slicing into finite homogeneous layers [4][5]. The second approach assumed that the gradient is varying continuously between the tow basic material properties [6]. Due to their complex structure, the characterization of FGMs poses a great challenge. The surface acoustic wave is widely used for characterizing the profile of gradient near to their surface [7] because surface acoustic wave (SAW) propagation, is strongly dependent to the local properties of the materials [8]. The good Knowledge of the reflection coefficient is necessary to determine and analyze the reflected or transmitted modes propagating at the interface of studied structures [9].

In inhomogeneous medium, the equation governing the propagation of elastic wave, is a system of three second order differential equations with no constant coefficients, for the displacement field, in which involved spatial derivatives of the elasticity coefficients, these type of equations can be solved analytically for specific profile of gradients [10][11][12].

In the present article, we have used the transfer matrix method [13][14] and the theory developed by L.M.Brekhovskikh [15] to compute the reflection coefficient

and the dispersion curves of different numerical profiles of gradient simulating an FGM structure. The numerical sample considers the transversal cut near to the surface of titanium plate after the oxidation process, where the titanium layer presents a continuous gradient in its elastic properties crossing its depth. The inhomogeneous area is divided into some elementary layers with the same thickness. The number of elementary layers is selected such that the error rate on the velocity of the Rayleigh mode is less than one percent in the frequency range.

II. THEORY AND METHOD

The geometry of the problem is illustrated in the Fig.1. At each layer the elastic properties are constants, and then, the continuous gradient is replaced by piecewise constant functions. The minimal slicing to ensure reliable results and easy calculations are determined. (Fig.2)

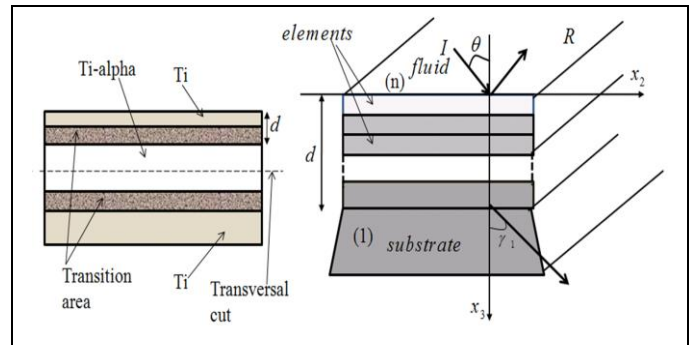


Fig. 1. Geometry of graded layer on semi-infinite substrate

The gradient variations are in the direction (ox_3), and then, all properties of the material are only depending to the x_3 coordinate. Taken account of the boundary conditions, the displacement -stress vector at each tow adjacent layers can be written as following [15]:

$$\xi^m(x_3) = A^m(x_3)\xi^{m-1}(x_3) \quad (1)$$

Where ξ and A are, respectively, the displacement-stress vector and transfer matrix of the layer m , such as:

$\xi^m = \begin{bmatrix} u_{11}^m & u_{33}^m & \sigma_{33}^m & \sigma_{13}^m \end{bmatrix}$ And $A^m = a_+^m (a_-^m)^{-1}$, these tow matrices are obtained for the same layer m (at its limits, upper and lower interface), they contain the same properties of

the layer m , but they differ by the term which contains the expression $\exp(jkx_3)$ [15].

In the general case, the displacement-stress vector has six components and the matrix A is at sixth order [16]. But according to the theory developed by L.M.Brekhovskikh, the calculation of the reflection coefficient requires only fourth components of displacement-stress vector [15].

Step by step, using (1), and taken account the boundary conditions at the interface separating two successive layers, it is possible to express the displacement-stress vector at the interface liquid/first elementary layer ($x_3=0$) as a function of that at the interface last elementary layer/ substrate ($x_3=d$):

$$\xi^n = A^n A^{n-1} \dots A^1 \xi^1 \quad (2)$$

And then, the transfer matrix of the continuous graded layer is as following (n is the number of elementary layers):

$$A = \prod_n^1 A^m \quad (3)$$

At the interface liquid/layer the shear component σ_{13}^n is null (the liquid is considered perfect, its shear modulus is then null), and then from (2) and (3), we can deduce the following equation:

$$A_{41}u_{11}^1 + A_{42}u_{33}^1 + A_{43}\sigma_{33}^1 + A_{44}\sigma_{13}^1 = \sigma_{13}^n = 0 \quad (4)$$

By using (2) and (4), we deduce the following system of equations:

$$\begin{aligned} u_{33}^n &= M_{22}u_{33}^1 + M_{23}\sigma_{33}^1 + M_{24}\sigma_{13}^1 \\ \sigma_{33}^n &= M_{32}u_{33}^1 + M_{33}\sigma_{33}^1 + M_{34}\sigma_{13}^1 \end{aligned} \quad (5)$$

With:

$$M_{pi} = A_{pi} - \frac{A_{4i} - A_{p1}}{A_{41}}, \quad p=2, 3 \text{ and } i=2, 3, 4.$$

At the two interfaces liquid/first elementary layer and last elementary layer/substrate, the normal displacement and stress components can be written as following [14]:

$$\begin{aligned} u_{33}^n &= j\alpha_n \phi_{inc} (1 - R) \\ \sigma_{33}^n &= -\omega^2 \rho_n \phi_{inc} (1 + R) \end{aligned} \quad (6)$$

ϕ_{inc} and R (quotient between amplitudes of incident and reflected waves) and ρ_n and α_n are respectively, the amplitude of incident acoustic wave and reflection coefficient and density and wave number in coupling liquid. ω is the angular frequency of acoustic wave. θ is the incident angle of acoustic wave. θ_1 and γ_1 are the refraction angles of transmitted waves in the substrate. d is the thickness of graded layer.

$$\begin{aligned} u_{11}^1 &= j(W k_L^1 \sin \theta + P k_T^1 \cos \gamma_1) \\ u_{33}^1 &= j(-W k_L^1 \cos \theta_1 + P k_L^1 \sin \theta) \\ \sigma_{33}^1 &= -\omega^2 \rho_1 (W \cos 2\gamma_1 - P \sin 2\gamma_1) \\ \sigma_{13}^1 &= \omega^2 \rho_1 (P \cos 2\gamma_1 + W \frac{C_{T1}^2}{C_{L1}^2} \sin 2\theta_1) \end{aligned} \quad (7)$$

C_{L1} and C_{T1} are the velocities of longitudinal and shear waves in the substrate. P and W (amplitudes of acoustic waves in the substrate) are two unknowns, and only the rate $\frac{P}{W}$ is necessary to determine the reflection coefficient. This quotient can be deduced from (4) [15].

Finally, R is the reflection coefficient, it is a function of incident angle and frequency. By using (4) and (5) and (6) and (7), its expression can be written in the following form:

$$R(\omega, \theta) = \frac{\frac{\sigma_{33}^n}{\omega u_{33}^n} - j\omega \frac{\rho_n}{\alpha_n}}{\frac{\sigma_{33}^n}{\omega u_{33}^n} + j\omega \frac{\rho_n}{\alpha_n}} \quad (8)$$

The tangent hyperbolic profile of longitudinal and transversal velocities is given by the following formula:

$$C = C_0 + (C_d - C_0) \left[\frac{\tanh\left(a(x_3 - \frac{d}{2})\right) - \tanh\left(-a(\frac{d}{2})\right)}{\tanh\left(a\frac{d}{2}\right) - \tanh\left(-a\frac{d}{2}\right)} \right] \quad (9)$$

C_0 and C_d are the ultrasonic velocities in the surface of layer and in the substrate. a is a given parameter.

III. NUMERICAL RESULTS AND DISCUSSION

The ultrasonic velocities in substrate of Titanium alpha-case and in Titanium layer [18] and the numerical data used in simulations are regrouped in the table I. The coupling liquid surmounted this structure is the water.

TABLE I. INPUT DATA USED IN SIMULATIONS

	V_L (m/s)	V_T (m/s)	ρ (kg/m ³)	Thickness(μ m)
Titanium alpha-c	6660	3553	4460	-
Titanium	6060	3230	4460	100
Water	1500	-	1000	-

A. Stabilization of the velocity of Rayleigh mode

To determine the number of elementary layers sufficient to ensure the accuracy of calculations, the frequency is fixed at 150 MHz and the input number of elementary layers is increasing until the stabilization of velocity of Rayleigh mode.

In the Titanium, the wavelength of Rayleigh at the frequency of 150 MHz is $\lambda_R=20\mu m$, which is less than thickness of the graded layer ($d=100\mu m$). About twelve layers the velocity of Rayleigh mode (for linear profile) becomes constant and stabilizes at the 3020.30 m/s and the relative error is null (Fig. 2). For example, at teen layers the velocity of Rayleigh mode is 3010.10 m/s and the error is about 0.34%. For a graded aluminum layer on the substrate of silica, for to have an error less than 1%, it should about twenty layers [4][5][17]. This difference can be explained by the significant difference between the properties of the graded layer and those of substrate, in contrast to the Titanium alpha and Titanium where their properties are very close and the step of variation of the profile is weak, then, the continuity is assured only for several elementary layers.

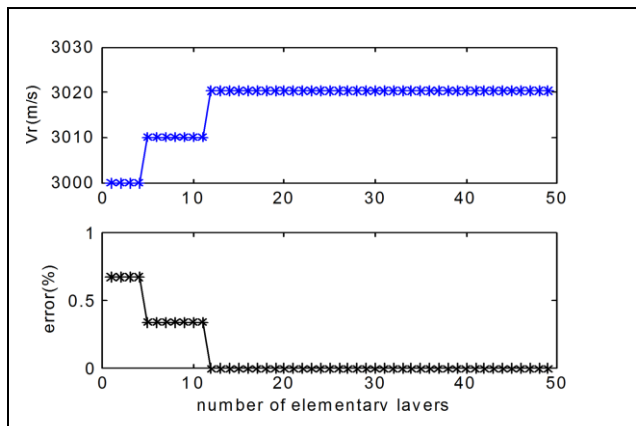


Fig. 2. Variation of the velocity of Rayleigh mode with elementary layers

B. Reflection coefficient and dispersion curve

The reflection coefficient in (8) is computed for each frequency at each incident angle. The length of vectors which contain frequencies and incident angles are respectively choosing 401 and 451 elements. The step between tow successive frequencies is fixed at $\Delta f=0.375\text{MHz}$ in the frequency range of [0 150MHz]. For the incident angles the step is $\Delta\theta=0.1089$ degree for the incident angle range of [0 50 degree]. The reflection coefficient is then a matrix of 401×451 elements. We have remarked that the image of the phase of this matrix (or its imaginary or its real part) gives the dispersion of the generalized Lamb modes reflected by the studied structure (fig. 3 and Fig.4).

We have remarked that the counter of the image in Fig. 3 can clearly show all generalized modes reflected in the coupling liquid (Fig.4), this method gives the dispersion curve from reflection coefficient without solving (4).

The image in Fig.3 shows a series of modes which propagate at the interface of studied structure. Tow types of acoustic surface modes are present:

The Rayleigh and pre-Rayleigh modes (sezawa modes):

The first mode is the Rayleigh mode (Rayleigh wave), its velocity decrease, with frequency, from 4020 m/s to reach the asymptotic value of 3020.30 m/s at high frequencies. The velocity of first Sezawa mode decrease from the 6612 m/s (the longitudinal velocity in substrate is 6660 m/s) to the

asymptotic value of 3342 m/s (the transversal velocity in layer is 3230 m/s). Other modes (sezawa modes) vary between 5600 m/s and the asymptotic velocity of 3342 m/s at high frequencies.

Higher order modes: which decrease rapidly from high velocities to the velocity of longitudinal mode in the layer (the asymptotic value of 6100 m/s) (Fig.4).

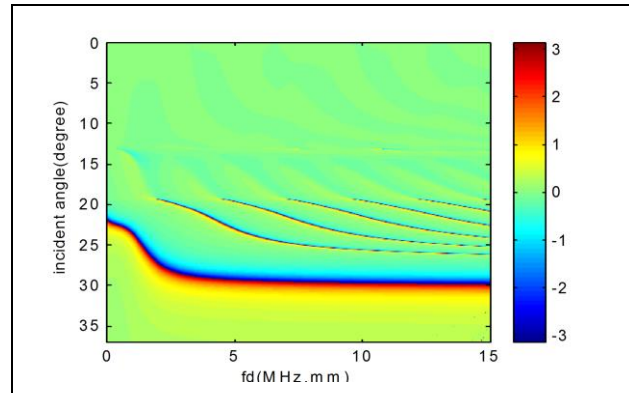


Fig. 3. Dispersion image for linear profile-phase of the matrix R

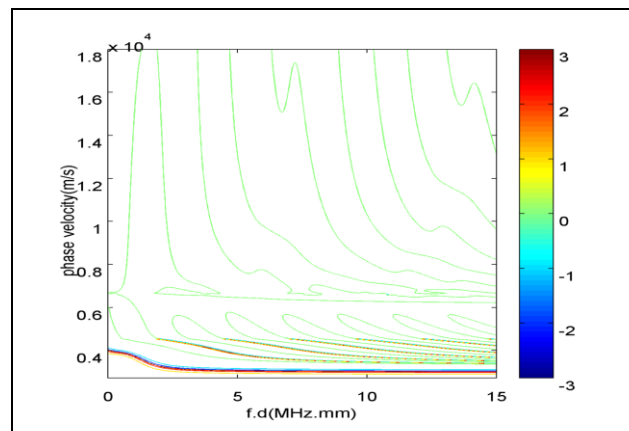


Fig. 4. Contour of the image in Fig. 3

At low frequencies ($f=0.375\text{MHz}$, Fig.5) and at the incident angle of $\theta=22$ degree, only the Rayleigh mode is existing, its acoustic energy is totally reflected by the interface liquid/graded layer ($R=1$). When the frequency increases, many modes appear at different incident angles (Fig.6). Their velocities can be determined by using the following formula:

$$v = \frac{v_{liq}}{\sin \theta_c} \tag{10}$$

v_{liq} is the longitudinal velocity in the coupling liquid and θ_c is the critical incident angle. This angle can be determined from the phase of the reflection coefficient, it corresponds to the pick figured in the phase of the reflection coefficient (Fig.5).

The phase velocity of the Rayleigh mode and Sezawa mode is determined by using (10). At each frequency, the critical angles are determined by using an appropriate algorithm. The dispersion curve of these tow modes is presented in Fig.7.

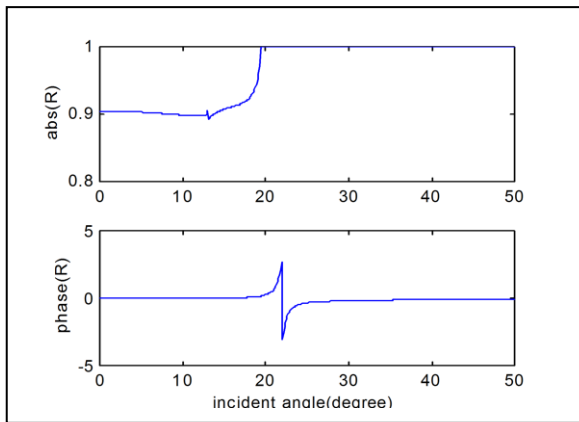


Fig. 5. Modulus and phase of the RC at $f=0.375\text{MHz}$

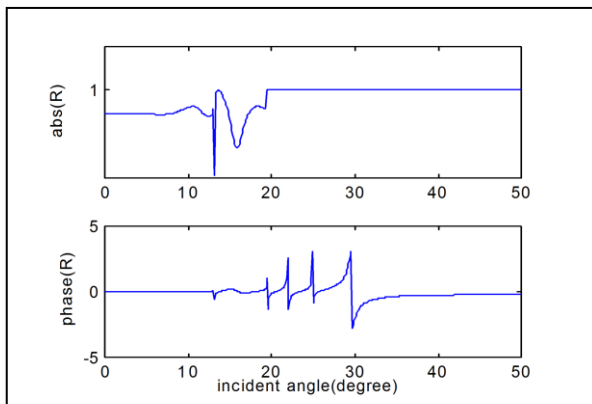


Fig. 6. Modulus and phase of the RC at $f=74.62\text{ MHz}$

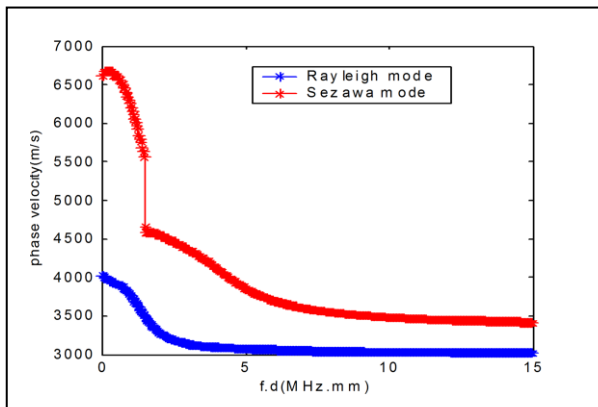


Fig. 7. Dispersion curve of Rayleigh and sezawa modes

C. effect of the gradient on reflection coefficient

In the case of the inverse problem i.e. when the reflection coefficient and dispersion curve are known and elastic properties of the studied structure are unknown, in this case, the complete characterization of continuously graded profile is very delicate.

However, the reflection coefficient is the good indicator of the profile of heterogeneity of graded materials.

The modulus of reflection coefficient for different profiles is presented in the following figures:

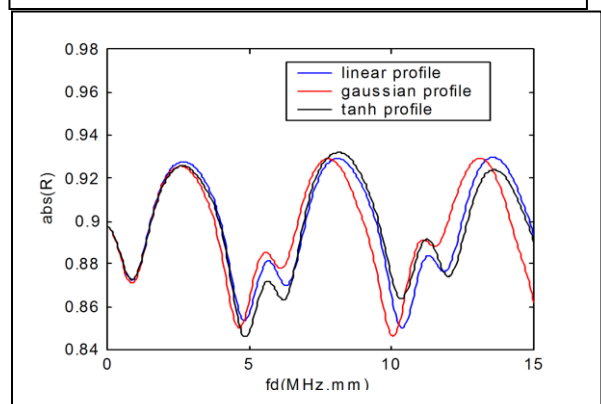
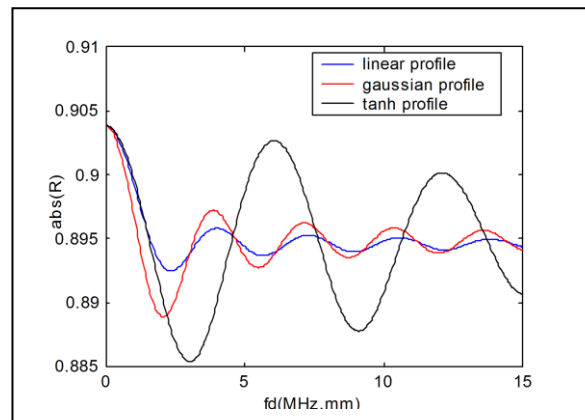


Fig. 9. Modulus of the reflection coefficient at $\theta=11\text{ degree}$

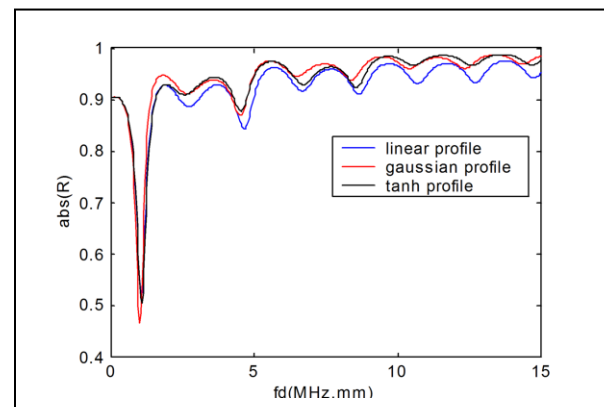


Fig. 10. Modulus of the reflection coefficient at $\theta=14\text{ degree}$

From the Fig.8 and Fig.9 and Fig.10, we can observe the influence of the profile of heterogeneity on the spectrum of reflection coefficient and we note the following remarks:

- The resonance in the reflection coefficient.
- The reflection coefficient is sensitive to the shape of gradient at certain frequency range and incident angles.
- At normal incidence, the reflection coefficient is more sensitive to the profile of gradient (Fig.8).
- At low frequencies (0 to 20MHz), the reflection coefficient is insensitive to the nature of the gradient.

Interpretations:

- At normal incidence, the acoustic wave propagates in the direction of the gradient, the vertical interference with the heterogeneity occurs.
- At low frequencies, the wavelength is important compared to the thickness of graded area, the whole of acoustic energy is located in the substrate.
- At high frequencies, the wavelength is weak compared to the thickness of the heterogeneous area, the acoustic wave is sensitive to the spatial variation of gradient.

D. Study of Rayleigh and Sezawa modes

For exciting only these tow first modes, the incident angle is fixed at the critical angle of $\theta=26$ degree, which correspond to the velocity of 3421 m/s for the both modes (Fig.11).

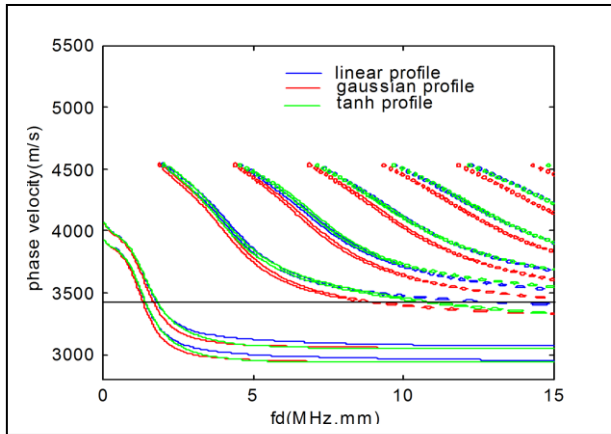


Fig. 11. Dispersion curve for Rayleigh and Sezawa modes-real part of the R

At this incident angle, the modulus of reflection coefficient is equal to one in all frequency range (Fig.5 and Fig.6). For evaluating the influence of the profile of the gradient, the phase of reflection coefficient gives the best results.

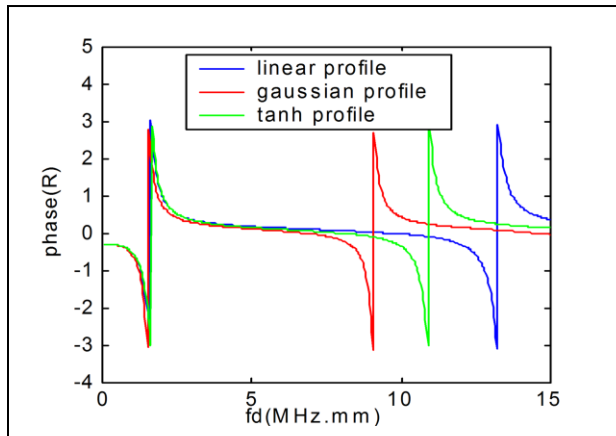


Fig. 12. Phase of the RC at $\theta=26$ degree for Rayleigh and sezawa modes

From the Fig.12, we can observe the influence of the nature of the gradient on the phase of reflection coefficient for the Sezawa mode. The frequency of existence of this mode depends to the shape of gradients. The Rayleigh mode dose

not influenced by the profile of gradient (fig.12). The results are regrouping in table II:

TABLE II. FREQUENCY OF EXCITATION OF THE TOW FIRST MODES

	gaussian pr.	linear pr.	tanh pr.
Rayleigh mode			
f(MHz)	15.375	16.125	16.50
$\lambda(\mu\text{m})$	222.55	212.20	207.38
Sezawa mode			
f(MHz)	91.125	132.75	109.5
$\lambda(\mu\text{m})$	37.550	27.776	31.249

The second mode (or sezawa mode) is excited at high frequency ($\lambda < d$) that's why the phase of reflection coefficient of this mode is more sensitive to the profile of gradients (table II).

At the incident angle of $\theta=26$ degree, the wavelength of the Rayleigh mode is important than thickness of graded layer, the acoustic wave is not affected by the heterogeneous area ($\lambda > d$) (Fig12), (table.II).

When the incident angle increases, the Rayleigh mode (first mode) is excited at high frequencies and becomes sensitive to the profiles ($\lambda < d$) (Fig.13) (Table.III).

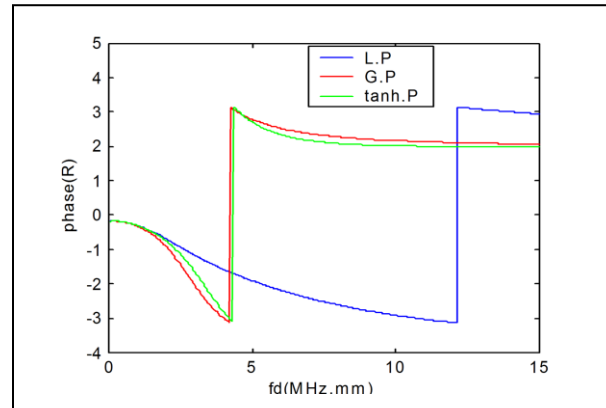


Fig. 13. Phase of the RC for Rayleigh mode at $\theta=30$ degree

TABLE III. FREQUENCY OF EXCITATION OF RAYLEIGH MODE

	gaussian pr.	linear pr.	tanh pr.
Rayleigh mode			
f(MHz)	49.875	121.500	48.375
$\lambda(\mu\text{m})$	60.557	24.858	62.435

IV. CONCLUSION

This numerical study, based on transfer matrix method, of the propagation of acoustic wave in continuously graded thin layer on semi-infinite substrate, aims to understand the behavior of acoustic wave at the interface of such structure, in order to discover an efficient tool for characterizing the gradient in elastic properties present near to the surface of these structures.

We used the model described in this article to compute the two-dimensional spectrum of the reflection coefficient which is a good indicator of the heterogeneity of materials. At normal incidence and at high frequencies, the reflection coefficient is more sensitive to the gradient. For the Rayleigh and Sezawa modes, the phase of the reflection coefficient is best for characterizing the profile of heterogeneity at high frequencies. The frequency of excitation of these modes is influenced by the shape of gradient.

The gradient has little influence on the velocity of Rayleigh and Sezawa modes and this due to the nature of propagation of these modes, which propagate perpendicularly to direction of the heterogeneity, the vertical decreasing of the displacement field is very rapid, and then, the vertical interferences are very weak.

The choice of the incident angle and the frequency is mandatory for ensuring best characterization of FGMs.

REFERENCES

- [1] Suresh, S., M. A., "Fundamentals of functionally graded materials", 1998.
- [2] S.Y. Sung, Y.J. Kim, "Alpha-case formation mechanism on titanium investment castings", *Materials Science and Engineering A* 405 (2005) 173–177.
- [3] J. F. Groves and H. N. G. Wadley, "Functionally graded materials synthesis via low vacuum directed vapor deposition", *Composites Part B* 28B(1997), 57-69.
- [4] H.Nounah, B.Cros, and J.Attal, "Theoretical approach of the characterization of gradients in elastic properties by acoustic microscopy", *Eur. Phys. J. AP* 5, 221-226, 1999.
- [5] H. Nounah, B. Cros and J. Attal, "Modelling of gradients in elastic properties for acoustic microscopy characterizations", *Eur. Phys. J. AP* 6 (3) 237 (1999).
- [6] Shuvalov A. L., Le Clezio E., Feuillard G., "The state-vector formalism and the Peano-series solution for modelling guided waves in functionally graded anisotropic piezoelectric plates", *Int. J. Eng. Sci.*, 46, 929–47, 2008.
- [7] A.Atalar, H.Koymen and L.Degertekin, "Characterization of layered materials by the lamb wave lens", *proc. of ultrasonic symposium*, New York, 1989.
- [8] J. Kushibiki, T. Ishikawa, and N.Chubachi, "Cut-off characteristic of leaky sezawa and pseudo-sezawa wave modes for thin-film characterization", *Appl. Lett.*57(19), 5 November 1990.
- [9] D.L.Folds, C.D.Loggins, "Transmission and reflection of ultrasonic waves in layered media", *J.Acoust.Soc.Am.*, 62, 1102-1109, 1977.
- [10] Liu, J. Y., Tsai, S. H., Wang, C. C., and Chu, C. R., "Acoustic wave reflection from a rough seabed with a continuously varying sediment layer over-lying an elastic basement", *Journal of Sound and Vibration* 275, 3-5 (2004),739-755.
- [11] Chiu, T.-C., and Erdogan, F., "One-dimensional wave propagation in a functionally graded elastic medium", *Journal of Sound and Vibration* 222, 3(1990), 453-487.
- [12] Robins, A. J., "Reflection of plane acoustic-waves from a layer of varying density", *Journal of the Acoustical Society of America* 87, 4 (1990), 1546-1552.
- [13] W.T.Thomson, "Transmission of elastic waves through a stratified solid medium", *J.Appl.phys.*, 21, 89-93, 1950.
- [14] N.A. Haskell, "The dispersion of surface waves on multilayered media", *Bull. Seismol. Soc. Am.* 43, 377–393, 1953.
- [15] L.M. Brekhovskikh, "Waves in layered media", academic press, New York, 1960.
- [16] Adnan H. Nayfeh, Wael G. Abdelrahman, and Peter B. Nagyc, "Analyses of axisymmetric waves in layered piezoelectric rods and their composites", *J. Acoust. Soc. Am.* 108 (4), October 2000.
- [17] H. Nounah, "Modélisation et caractérisation des matériaux à gradient de propriétés mécaniques par des méthodes microacoustiques", phd thesis, Montpellier II, France, 1995.
- [18] C. Baron, "le developemen en serie de peano du matricant pour l'étude de la propagation des ondes élastiques en milieux à propriétés continûment variables", Phd thesis, France, 2005.

Automatic Ferrite Content Measurement based on Image Analysis and Pattern Classification

Hafiz Muhammad Tanveer
Center for High Energy Physics (CHEP)
University of the Punjab
Lahore, Pakistan

Waleed Asif
Computer Science Department
Qarshi University
Lahore, Pakistan

Hafiz Muhammad Tahir Mustafa
Electrical Engineering Department, College of Engineering
& Emerging Technology (CEET), University of the Punjab
Lahore, Pakistan

Munir Ahmad
Institute of Nuclear Medicine and Oncology (INMOL)
Lahore, Pakistan

Muhammad Anjum Javed, Maqsood Ahmad
Center for High Energy Physics (CHEP)
University of the Punjab

Abstract—The existing manual point counting technique for ferrite content measurement is a difficult time consuming method which has limited accuracy due to limited human perception and error induced by points on boundaries of grid spacing. In this paper, we present a novel algorithm, based on image analysis and pattern classification, to evaluate the volume fraction of ferrite in microstructure containing ferrite and austenite. The prime focus of the proposed algorithm is to solve the problem of ferrite content measurement using automatic binary classification approach. Classification of image data into two distinct classes, using optimum threshold finding method, is the key idea behind the new algorithm. Automation of the process to measure the ferrite content and to speed up specimen's testing procedure is the main feature of the newly developed algorithm. Improved performance index by reducing error sources is reflected from obtained results and validated through the comparison with a well-known method of Ohtsu.

Keywords—Pattern classification; Decision threshold; Machine learning; Microstructure

I. INTRODUCTION

In the perspective of materials science and engineering, the microstructure (microscopic image) of specimen is the most considerable entity. The study of microstructure correlates the properties of material with microstructure. Quantitative measures of micrographs determine the specific characteristics of microstructures.

The measurement of different metallic components present in specimen, by employing microstructure to evaluate percentage composition, is an important aspect of quantitative metallography. Vital role of quantitative metallography is very important in the discipline of materials science and engineering. It provides the basis to develop an appropriate mathematical model by considering the relations between processes, mechanical properties and microstructures.

Considering these points, the vital role of quantitative metallography is very obvious in the discipline of materials science and engineering [1-3].

Ferrite content measurement is a significant parameter which determines the mechanical strength of material. Experimental techniques for ferrite content measurement are broadly classified as destructive techniques and non-destructive techniques [4]. Destructive techniques refer to the measurements from microstructures. On the other hand, non-destructive techniques are directly applicable to the material's specimen for ferrite content measurement. There is no need to acquire fine microstructure of specimen for ferrite content measurement procedure. In destructive techniques, measurements are taken from microstructures, while in non-destructive techniques specimen is directly employed for measurements [5]. Manual point counting and Image analysis are examples of destructive techniques while magnetic method (Magne-Gage instrument) and magnetic induction method (ferritescope or vibrating sample magnetometer, VSM) falls under the category of non-destructive techniques for ferrite content measurement[1;4].

Ferritescope is a device which measures the ferrite contents from material's specimen directly instead of microstructure. The principle of magnetic induction is applied to measure the ferrite contents by employing ferritescope[6]. Magnetic portion of specimen interacts with the magnetic field generated by coil and induced voltage measures the proportion of ferrite content in second coil. The correct identification of ferrite and non-ferrite structures is a particular advantage of magnetic induction technique. The magnetic permeability of steel plays an important role for the measurement of ferrite content because the amount of ferrite present in specimen tends to correlate with the permeability of steel. This permeability specifies the ferrite content of material being analyzed, when calibrated with standards having known ferrite

content. Digital read-out dial or calibrated dial provides the measure of ferrite content [4]. Ferritescope is very costly device, so in spite of having high accuracy, this method is not cost effective and rarely used.

Manual point counting method is based on stereological principle in which a square grid of particular dimensions is superimposed over the specimen's microstructure. This square grid is moved systematically through the specified number of different fields to cover specimen's surface area. This method gives an impartial statistical estimation of volume fraction corresponding to ferrite or any identifiable constituent in microstructure. There are few sources of error associated with manual point counting method such as points on grid boundaries and human perception limitation which are responsible for low accuracy of obtained results. The American Society for Testing Materials (ASTM E562) [7] standard describes the detailed procedure for the measurement of volume fraction of ferrite.

Image analyzer is a device capable of measuring ferrite content measurement from specimen's microstructure directly. The accuracy of desired results is strictly dependent on the acquired image quality. In the presence of bad contrast or analyzed features have a discontinuous outline in microstructure, the automatic measurements of image analyzer provide inaccurate results [8].

The microstructure under experiment contains the information of two metallic components. One of them is Austenite and other one is Ferrite [9]. The histogram of microstructure has two distinct regions. First region corresponds to Ferrite component and second region indicates the presence of Austenite. This pattern of histogram in microstructure attracts the attention to solve the problem of separating each metallic component by volume. Pattern classification based on Image data analysis is the possible approach towards efficient solution for ferrite content measurement. Image data analysis based on histogram pattern is the first step in classifier design. Optimum decision threshold between two regions separates one class of data from other [7;10-13].

The analysis and implementation of systems that are able to learn pattern from data is the subject of machine learning. Supervised machine learning and un-supervised machine learning are two primary machine learning techniques. In supervised learning, data analysis and prior knowledge of class label are used to design a classifier whereas finding pattern in data without any prior knowledge is the main focus of un-supervised learning [14;15]. Supervised learning leads to classification, whereas un-supervised learning ends up with clustering (similar groups in data). In variety of situations, linear and non-linear classifiers are employed to separate data into different classes. Linear classifier is most suitable option in situations where decision boundary may be marked as hyper-plane among different classes. In complex data distributions, inter mixing of data from different classes is quite random, and not possible to separate data into different classes by passing hyper-plane as decision threshold. Non-linear classifier deals such type of complex data distributions [7;15]. The Pattern classification deals with the identification

and separation of data into different classes. Feature selection or extraction plays an important role in determining the performance of classifier.

Otsu's method provides a global threshold value to convert the grayscale images into binary images. The existence of two distinct and separable categories of image pixels is an essential condition for this technique to work. As a consequence, the image histogram must follow bi-modal pattern. The threshold value corresponds to the optimal decision boundary for classification of image pixels. Image pixels having values below this threshold belongs to one class and pixels with values above threshold represent other class. These two classes describe the two distinct phases present in image. The percentage of each phase is evaluated quite easily after classification of pixels.

There exists no method in literature for automatic ferrite content measurement from microstructure by using image analysis and pattern classification. Up to our knowledge, present research fills this gap and provides an efficient solution for ferrite content measurement [13;15-17]. The GPF algorithm based on image processing and pattern classification has the ability to automatically measure the volume fraction of ferrite. This method will enhance the specimen's testing speed to a great extent. Machine learning nature of GPF algorithm in this paper has minimized the errors of manual point counting technique. The proposed technique will probably replace the tedious manual point counting method [17].

This paper is organized as follows. Section II explains the methodology and experimental details. The analysis and discussion on obtained results are included in section III. Finally, the section IV deals with the conclusion and possible future directions to extend the present research along with the potential application areas.

II. EXPERIMENTAL DETAILS

First of all, specimen is prepared by following some etching standard. The heat treatment and cooling method assigns a unique color to each metallic component in specimen. Microstructure (microscopic image) of specimen is taken with some suitable magnification index (200 or 500 times) to visualize inner details clearly. In the present study the specimen contains two metallic components (ferrite and austenite) and correspondingly two phases in microstructure. Each phase having distinct color and particular range of pixel intensities. The binary classifier is designed to separate ferrite content from specimen's microstructure. Analysis of image histogram suggests the linear classifier with optimal decision threshold to solve problem. MATLAB software is an efficient tool to program the functionality of linear classifier.

The process of evaluating the ferrite content measurement in a specimen having two phases is categorized in analysis and classification modules. The function $f(x, y)$ describes the discrete-time version of specimen's microstructure with spatial resolution of $(M \times N)$,

$$f(x, y) = \sum_{j=-N/2}^{N/2} \sum_{i=-M/2}^{M/2} f(i, j) \delta(x - i, y - j) \quad (1)$$

In the quantization of 2-D discrete time signal, the standard of 256 gray levels (0-255) is most appropriate to

assign intensity level of each image pixel. One byte (8-bits) is required to store the information of one pixel in gray scale image. For analysis purpose, RGB image is converted to grayscale image to find the range of pixel intensities corresponding to each phase.

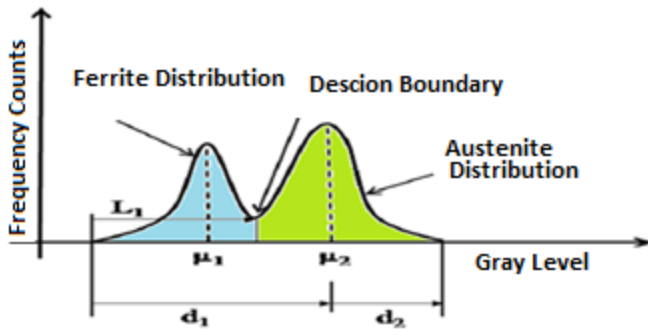


Fig. 1. Description of proposed algorithm with various parameter definitions

For instance, the distribution of data in Figure 1 describes two distinct regions before and after decision boundary. The region before decision boundary indicates the presence of ferrite contents and after that austenite part. Distribution of data in Figure 1 allows the application of Gaussian kernel to each component. Statistical relations for normal distribution are applicable to each component of distribution. Gray levels corresponding to two data peaks describe the mean locations (μ_1 and μ_2) of two Gaussian kernels. In existing situation, the decision boundary is approximately at $\mu_1 + \sigma_1$ from mean (μ_1) of first Gaussian kernel and at $\mu_2 - \sigma_2$ from data peak (μ_2) of second Gaussian kernel. The standard deviation σ from mean (μ) covers approximately 84.4% of area under distribution curve. Decision threshold is the intersection of two Gaussian kernels, the lowest pixel count with respective gray level between two optimized data peaks. This optimum threshold minimizes the classification error to a negligible value. The reason behind that lies in the fact that this optimized threshold is actually the starting point of second distribution and last point of first distribution. The proposed algorithm first determines the maximum data peak (maximum pixel count with respective gray level) of whole image.

$$y_1 = \Phi[g(f)] , \quad 0 \leq f \leq 255 \quad (3)$$

In above equation, Φ is the function that provides the maximum pixel count of whole image histogram with corresponding gray level. Suppose y_1, f_1, n_1 and $N = (f_1 - n_1)$ are maximum pixel count, corresponding gray level, gray level at non-zero starting pixel count and total number of data samples (data peaks) between f_1 and n_1 in histogram respectively. The maximum data peak may fall in any distribution. In case, second distribution contains maximum data peak, the process of finding the data peak in first distribution is quite straight forward. When maximum data peak falls in first distribution, reversal in the order of distributions needs to be considered to make the process identical to above mentioned. Recursive window is applied

from data peak in second distribution by considering two consecutive data samples (peaks) in single step. The process continues, until non-zero starting point of first distribution reaches. This is an indication to the completion of first iteration.

Image histogram gives information of relative frequency of pixels at particular gray level. The mathematical formulation of histogram is represented by function $g(f)$.

$$g(f) = n_f , \quad 0 \leq f \leq 255 \quad (2)$$

Histogram pattern of microstructure shows two distinct and linearly separable regions. The existence of two distinct regions provides evidence to the two metallic components in specimen's microstructure. One region corresponds to the existence of ferrite and other to austenite. The primary function of linear classifier is to find optimal decision boundary. Ferrite class label is assigned to pixel intensity range below decision boundary. The range of pixel intensity above decision threshold is assigned austenite class labels. Analysis of microstructure gives an identification of different phases, and classification separates metallic components in terms of pixel count with specific intensity range. The scope of present research is to separate ferrite content by volume in microstructure having two phases based on pattern classification. The beauty of proposed algorithm is to transform the complex and lengthy procedure for measurement of metallic components to a simple arithmetic problem. Finding data peaks in each distribution and decision boundary in between provide the required solution of data classification problem. Minimum pixel count between two data peaks is the optimum decision threshold and the most important parameter for the classification of image data.

The number of data samples reduces to one half after each iteration, and this procedure ends up with two optimized data peaks with corresponding gray levels of both distribution parts. One of these optimized data peaks refers to the mean of first Gaussian kernel and other to the second Gaussian kernel. Decision threshold provides the pixel intensity range for ferrite content measurement. On the basis of this information, volume fraction of ferrite is determined by comparing the pixel counts in ferrite region to the total number of pixels.

$$g_i^m(q_j)|_{j=1}^i = g^{m-1}(f_p) * H_{u \rightarrow u-1}|_{u=f_1-2k}^i \quad (4)$$

(* Operator defines recursive window for the selection of greater sample value in the comparison of every two samples till the completion of iteration).

$$\frac{N}{2} \leq i \leq 2 , \quad n_1 \leq f_p \leq f_1 , \quad 0 \leq k \leq i , \quad 1 \leq m \leq n \quad (5)$$

The vector q_j contains gray levels at ascending order index, after the completion of each iteration, the length of this vector reduces to one half. Window function $H_{u \rightarrow u-1}$ compares two data samples in one step and selects the highest of two. The result will be the reduction of data peaks to one half for next iteration. Algorithm flow chart is shown in Figure 2.

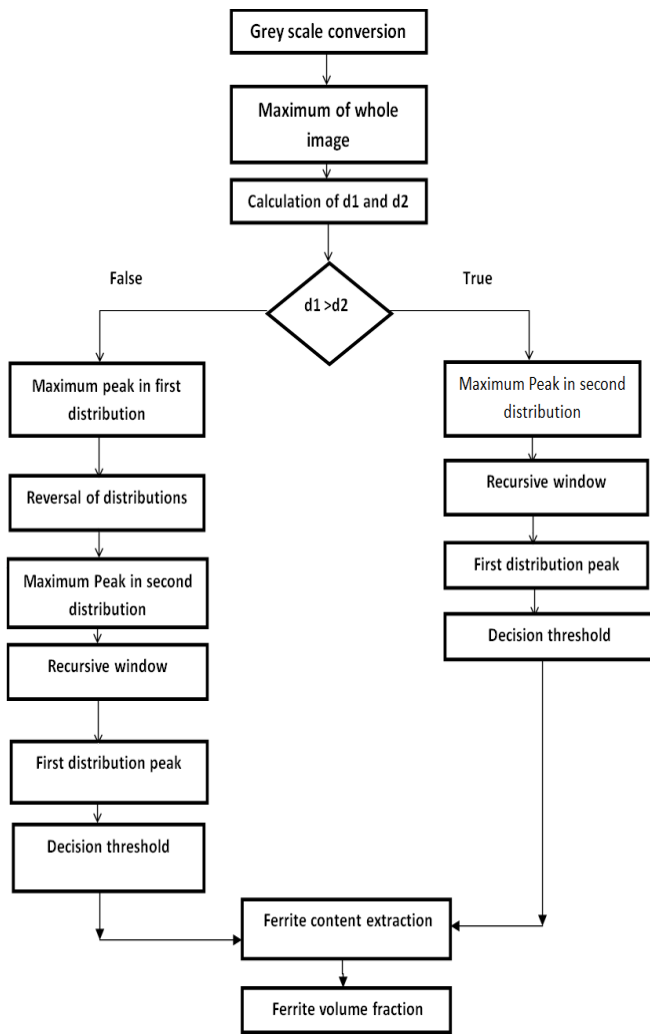


Fig. 2. Algorithm flow chart of the proposed algorithm to calculate ferrite volume fraction in a microstructure

We select four microstructures (Sample 1, 2, 3 and 4) having variations in histogram pattern. First microstructure (Sample 1) contains histogram in which both regions of distribution are very similar to Gaussian kernel. The histogram pattern in second microstructure (Sample 2) indicates Gaussian trend in first distribution part, but second distribution part doesn't obey Gaussian pattern in second half. In third microstructure (Sample 3), the situation is approximately same as described in Sample 2. Worst scenario is being presented having lot of irregularities in both distribution parts in fourth and last microstructure (Sample 4). Figure 3 shows four samples (microstructure images) used to measure ferrite volume ratio.

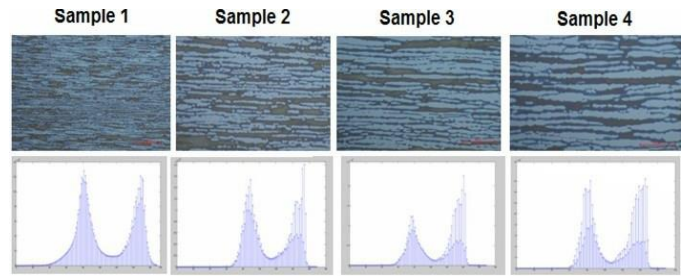


Fig. 3. Four different microstructure samples used to evaluate performance of various algorithms for the measurement of ferrite volume fraction. The samples were selected based on the spread of histogram degrading from an almost Gaussian curve

III. RESULTS AND DISCUSSIONS

The chi-square goodness-of-fit (GOF) test is employed to know that data samples came from a population with specific distribution. Information of model (distribution) for which decision is required and sufficient number of data samples are the limitations of chi-square GOF test. In this case, there is a need to test that data samples follow Gaussian distribution. This hypothesis is tested by applying chi-square GOF test, in which chi-square statistic for the given data samples is compared against some critical value. The degree of freedom and significance level (0.05 default) are important factors to decide some critical value. In case of hypothesis acceptance, the Gaussian kernel is applied on the distribution of data and all the standard mathematical relations for normal distribution are valid here.

The chi-square statistic value is less than critical value ($X_{statistic}^2 < X_{critical}^2$) for different microstructure images used in the experimental procedure. The values of degree of freedom and level of significance are 6 and 0.1 respectively. This test validates the GOF for Gaussian distribution. Gaussian kernel is applied to approximate the distribution of data corresponding to both phases present in microstructure. The GPF algorithm works accurately in this situation and gives reliable results. These results include mean, variance, standard deviation associated to the distribution of each phase and most important one is the optimal decision boundary for data classification.

For manual point counting (MPC) procedure, 30 equal sized fields were placed on each sample image in order to measure ferrite content to cover the whole image and pixels were counted for their gray level classification manually. Figure 4 presents data for Sample 1 for all 30 fields evaluated. Table 1 bears values for various parameters for each sample evaluated using MPC algorithm. In all four samples ferrite volume fraction evaluated as percentage of the ferrite content present are very close to each other with relatively low error.

TABLE I. FERRITE VOLUME FRACTION MEASUREMENT USING MPC WITH 30 FIELDS PER MICROSTRUCTURE

Sample No.	Total Fields Counted	Phase Vol. Fraction (%)	Standard Deviation	95% CI
1	30	40.1	7.7	2.9
2	30	47.1	7.6	2.9
3	30	44.5	10.1	3.8
4	30	44.5	8.8	3.3

TABLE II. DECISION BOUNDARY CALCULATION BY SPC AND GPF ALGORITHMS

Sample No.	μ_1	σ_1	SPC ($\mu_1 + \sigma_1$)	GPF (lowest pixel count)
1	82	30.1	112.1	116
2	90	17.7	107.7	116
3	79	21.5	100.5	109
4	85	18.6	103.6	108

TABLE III. FERRITE VOLUME FRACTION MEASUREMENT USING PROPOSED ALGORITHM

Sample No.	Pixel Intensity Range	Ferrite Volume Fraction
1	13-189	56.59
2	56-169	53.03
3	36-157	44.82
4	45-160	48.53

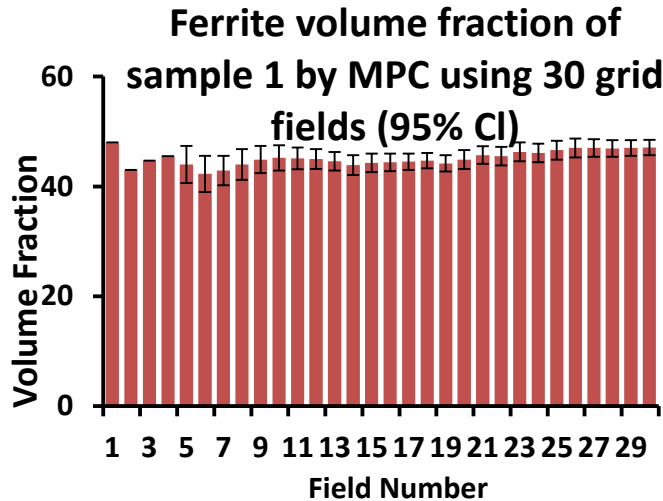


Fig. 4. Calculation of ferrite volume fraction of sample no. 1 by manual point counting using 30 fields per microstructure with relative error bars

For statistical point counting (SPC) algorithm, Gaussian kernel is applied on each distribution of histogram to obtain parameters (μ , σ). Decision boundary is approximately at the spread of σ_1 from mean μ_1 of first distribution. Ferrite content measurement is provided by considering pixel count for the range of gray levels below decision boundary. Classification boundary by applying Gaussian kernel is close to the result of proposed algorithm. This closeness in results provides mathematical support of Gaussian kernel to GPF algorithm.

Table 2 shows the parameters (μ_1 , σ_1) of first distribution part representing ferrite content region, and decision boundary values evaluated based on statistical estimation and proposed GPF algorithm. Pixel intensity range for ferrite content measurement in various microstructures is provided in Table 3.

The ferrite volume fraction by considering optimized decision boundary based on proposed algorithm for different microstructures are shown in Table 2. Several microstructures having highly symmetrical to worst possible scenario have been considered to check the validity of GPF algorithm. In the presence of irregularities, where second distribution doesn't obey Gaussian pattern, the GPF algorithm works well and provide quite satisfactory results.

The results obtained by applying GPF algorithm are approximately same to the one achieved by applying Gaussian kernel to each component of distribution. Sources of errors like points on grid boundaries and limited human perception have been resolved in the considerations of GPF algorithm. Thereby, results provided by GPF algorithm are more accurate and reliable than manual point counting method. The distance between first non-zero pixel count with corresponding gray level and optimal decision boundary L_1 indicates the range of pixel intensities for ferrite content measurement. Total pixel count for pixel intensity range L_1 indicates the presence of ferrite content measurement in microstructure having two phases. Application of Gaussian kernel on both distribution parts provides approximately the same result. The closeness of results obtained by applying Gaussian kernel and GPF algorithm provides mathematical support to measurement process.

If we carefully observe the Table1, we can clearly see that Table1 shows the results by manually applying the grid, field by field on the microstructures. Measurements of thirty fields are recorded, the volume fraction of ferrite in each field with 95% confidence interval and relative accuracy has been considered to average out the result. In microstructure (sample 3), ferrite volume fraction by applying manual point counting and GPF algorithm are very close to each other. On the other hand, there is a difference of approximately 16% in case of microstructure (sample 1). The reason behind this big difference lies in the fact that acquired image doesn't fulfill the required essential conditions for analysis. This difference will reduce to an acceptable level in the presence of conditions such as color contrast between different phases, clear grain boundaries and noise free focus. In sample 4 and sample 2, the difference in volume fraction is approximately 4% and 6% respectively.

The comparison between Otsu's method and proposed GPF algorithm is shown in table 4. Threshold values and percentage of ferrite content measurement for different microstructures are obtained by applying these two algorithms. The results obtained by employing two different techniques are very close to each other. Also the results obtained by GPF algorithm are very close to MPC results. This indicates the validity and reliability of results provided on the basis of GPF algorithm.

TABLE IV. COMPARISON BETWEEN OTSU'S METHOD AND GPF ALGORITHM

Sample No.	Otsu's Method		GPF Algorithm	
	Threshold	Ferrite Vol. Fraction (%)	Threshold	Ferrite Vol. Fraction (%)
1	113	57.65	116	56.59
2	118	62.65	116	53.03
3	110	49.25	109	44.82
4	110	56.36	108	48.53

The approach of GPF algorithm presented in this paper considers the whole image, pixel by pixel for the classification of all the data into two classes by deciding the optimized decision boundary. The results obtained in two techniques are shown in Table 3, but GPF algorithm adds the factor of accuracy by removing doubts and approximations associated with manual point counting measurement technique.

IV. CONCLUSION

Automation of ferrite content measurement based on image analysis and pattern classification is much faster and more accurate in comparison to the conventional manual point counting technique. The selection of optimal decision boundary minimizes classification errors and adds factor of accuracy to acquired results in relatively less time. Limitations associated with image analyzer and error sources regarding human perception in manual point counting have been resolved in machine learning algorithm for pattern classification. The crux of this research is to increase the efficiency of manual point counting for ferrite content measurement process by introducing the automatic computerized method to perform the same task in a very short span of time.

The present research deals with the separation of two phases in a microstructure image. Support Vector Machine (SVM) is a well-known binary classifier, the comparison of GPF algorithm with SVM may be one of the possible future directions to extend the present research activity. There is another dimension to proceed with the application of GPF algorithm i.e. generalization of GPF algorithm to work with microstructure images having more than two phases is another promising future direction. There is a definite need to analyze in great detail; the data present in the microstructure images with more than two phases. In this way important information of hidden pattern in the data can also be acquired for analysis.

This research may prove beneficial for the modern metallurgical industry all around the world in terms of accuracy and time saving. Material quality is related to the percentage of ferrite content in the material; GPF algorithm is able to automatically measure the ferrite content in the material more accurately in a minimum time and with much less chances of errors. Instead of a simple histogram analysis, segmentation may also be used for error reduction analysis. Computational efficiency with mathematical simplicity of analysis is another area to work in future.

The application area of this newly developed algorithm is not limited to metallurgy and material sciences. This method for binary classification will work reasonably well in the field of medical diagnosis and as a test method in factories to qualify a sample as pass or fail. This method may also be applicable to check an item for a specific qualitative property using the microstructure image of that particular item.

REFERENCES

- [1] R.T.DeHoff, E. H. Aigeltinger, Experimental Quantitative Microscopy with Special Applications to Sintering. In: J. Hirschhorn, K. Roll, editors. Advanced Experimental Techniques in Powder Metallurgy, Springer US; 1970, 81-137.
- [2] E. Underwood, Quantitative Stereology for Microstructural Analysis. In: J. McCall, W. Mueller, editors. Microstructural Analysis, Springer US; 1973, 35-66.
- [3] Practical Applications of Quantitative Metallography ASTM, 1983.
- [4] J. Niagaj, Au. Mazur, Review of methods for measurement of ferrite content in high alloyed steels and their welded joints. Welding International 2012;28:345-353.
- [5] F.D. Huang, L. Froyen, Quantitative analysis of microstructure in metals with computer assistance. Non-destructive Testing and Ultrasonics 2001;6:
- [6] GCd. Souza, JM. Pardal, SSM. Tavares, MdPC. Fonseca, JLF. Martins, EPd. Moura, I. CardoteFilho. Evaluation of phases proportions in welded joints of duplex stainless steel by non-destructive testing. Soldag.insp. 2013; 18:158-168.
- [7] L. Devroye. Automatic Pattern Recognition: A Study of the Probability of Error. IEEE Transactions on Pattern Analysis and Machine Intelligence 1988; 10:530-543.
- [8] JE. Hilliard, JW. Cahn. An Evaluation of Procedures in Quantitative Metallography for Volume-Fraction Analysis. In: The Selected Works of W. John. Cahn, John Wiley & Sons, Inc.; 1998, 63-64.
- [9] AW. Brewer, RL. Moment. Techniques and Standards for Measuring Ferrite in Austenitic Stainless Steel Welds. Welding Journal 1976;55:159s-164s.
- [10] JC. Russ. The Image Processing Handbook, Sixth Edition Taylor & Francis, 2011.
- [11] P. Osmond, H. Petitgand, JR. Philippe. Use of image processing in iron and steel making industry. Rev.Metall. 2000;2:187-196.
- [12] F. Keinosuke. Introduction to statistical pattern recognition (2nd ed.) Academic Press Professional, Inc., 1990.
- [13] PA. Deuyver, J. Kittler. Pattern Recognition: A Statistical Approach Prentice-Hall, 1982.
- [14] L. Chulhee, DA. Landgrebe. Feature extraction based on decision boundaries. Pattern Analysis and Machine Intelligence, IEEE Transactions on 1993;15:388-400.
- [15] L. Kanal. Patterns in pattern recognition: 1968-1974. Information Theory, IEEE Transactions on 1974;20:697-722.
- [16] G. Nagy. State of the art in pattern recognition. Proceedings of the IEEE 1968; 56:836-863.
- [17] S. Salzberg. Programs for Machine Learning by J. Ross Quinlan. Morgan Kaufmann Publishers, Inc., 1993. Mach Learn 1994; 16:235-240.

Development of Eye-Blink and Face Corpora for Research in Human Computer Interaction

Emmanuel Jadesola Adejoke.

Dept. of Computer science
Bingham University
Nassarawa, Nigeria

Ibiyemi Tunji Samuel

Dept. of Electrical Engineering
University of Ilorin
Ilorin, Nigeria

Abstract—A major requirement in face recognition research and coded voluntary eye-blink based sign language communication research is a robust face and eye-blink image corpora. The effectiveness, confidence level, and acceptability level of developed algorithms for face recognition and eye-blink based sign language communication depends largely on availability of relevant corpora in these fields of international standard. The wave of security challenges with attendant wanton destruction to lives and properties particularly in our country makes deployment of appropriate information technology to curb it imperative. Hence, the motivation of this work in provision of face and eye-blink image corpora with local contents to serve as input dataset to our developed face recognition authenticated driven coded eye-blink triggered actionable alert.

Keywords—Face Recognition; Coded Voluntary Eye Blink; Sign Language Communication; Authenticated driven; corpora

I. INTRODUCTION

In recent years, there have been efforts to augment traditional Human Computer interfaces with intelligent interfaces that allow users to interact with the computer more naturally and effectively. The goal is to develop HCI that are more man-centered in communication than computer-centered.[1] These are to respond to users natural communication cues such as gestures, facial expressions, gaze directions etc. and have been implemented in systems such as face recognition systems, eye blink recognition system, lip reading systems, voice recognition and translation systems and sign language recognition systems.

Although there are existing reliable methods of biometric personal identification such as finger print analysis, retina or iris scan they are yet to gain general acceptance from the public [2]. Facial recognition is chosen because it is a more passive and non-intrusive system of authentication also it is an identification technique that is closer to the way human beings recognize each other. With the current insecurity in worldwide a non-intrusive authentication system such as face recognition will be desirable where the people involved may not even be conscious of been captured.

Eye blink as a sign language, has for ages been a mode of communication for human beings, where information of different codes are being passed from one person to another. For security or emergency issues a real time eye blink communication would be a very reliable mode of communication for quick responses. It is reported that the eyes are directly connected to brain [3], they are the last part of our

body on which we can lose control in man. For some persons, who are suffering from a brain-stem stroke, neuro-motor disability or due to any accident, the eyes are the only option for communication with the world. Blink of their eyes can be converted to vocabulary for such persons. But continuous monitoring and understanding is required to understand and communicate with them properly. This is a bit difficult for human to understand the meaning of those blinks continuously [4].

It is noticed that world-wide there are many computer users who, due to their physical condition, are not able to use their hands or any part of body. Most of them have good control of their eyes and therefore communication with the world is done solely using the movement of their eyes or eye lids. These eyes blink or eye lids movement can be replaced with mouse click functionality. Implementing this man centered communication mode(eye blink communication) to a computer system for responses in these types of situations is very timely. Hence the development of the face and eye blink corpora. Many face databases has been created and made available for use and research purposes but most if not all of them are for the white coloured man. In this paper is presented corpora for the black coloured man both of frontal still faces for face recognition and dynamic video streams for eye blink recognition is presented.

In this proposed system, a cctv colour camera captures video stream digital images fed into a module for detecting face in the input image and crop out the face image.

Due to variations in the size and illumination effects on the cropped faces both size (geometric) and illumination normalization is carried out on the cropped face. The normalized face is then passed to face recognition unit and the eye-blink recognition unit simultaneously.

On detection and decoding of eye blinks, the desired action is performed if and only if the person making the blinks is authorized to communicate with the machine. Authorization is based on face recognition. The actionable operation supported in this work is automatic telephone dialing of some selected phone numbers relaying pre-recorded 'Save-Our-Soul' message by person under security threat or by an incapacitated medical patient In the section two of this paper the data collection experiment is discussed, the section three describes the corpora creation process and how it can be accessed, and section four is the conclusion of the paper.

II. RELATED WORKS

A summary of some existing face databases is given in [5], where the authors created a database of black faces used to benchmark algorithms developed for skin segmentation, face development and recognition under real life situation. The researchers present still frontal view black face database collected under controlled conditions for researches that require two dimension still face images. In [6] RPI ISL presents seven eye databases used primarily in eye detection, eye tracking and eye blink recognition. Existing eye blink databases capture only single eye blinks but this collection contains both single and double eye blinks to make room for more robust and extensive researches in eye blink recognition.[7] Presents the ZJU Eye blink video database which consists of 80 video clips of 20 individuals in the avi format. This was used in the development of an eye detector and eye blink detector using AdaBoost tools and recorded a 99.34% rate of eye detection and a 96% accuracy of eye blink detection. This newly created eye blink video corpora contains 90 video streams of four minutes each capturing both involuntary eye blinks and prompted single and double eye blinks for a wider scope of research in these area.[8]The BioID Face Database is a real world condition database with varying illumination background and face size. The dataset consists of 1521 gray level images with a resolution of 384x286 pixels. Each one shows the frontal view of a face of one out of 23 different test persons. For comparison reasons the set also contains manually set eye positions. The images are stored in single files using the portable gray map (pgm) data format. The database is published to give all researchers working in the area of face detection the possibility to compare the quality of their face detection algorithms with others. A new corpora is being created to have black faces databases relevant to the international community to aid more researches in the fields of face detection, face recognition and eye blink recognition.

[9]The AT&T laboratory created a database of faces that contains a set of face images taken between April 1992 and April 1994 at the lab. The database was used in the context of a face recognition project carried out in collaboration with the Speech, Vision and Robotics Group of the Cambridge University Engineering Department. There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open / closed eyes, smiling / not smiling) and facial details (glasses / no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The files are in PGM format, the size of each image is 92x112 pixels, with 256 grey levels per pixel. The images are organised in 40 directories (one for each subject), which have names of the form sX, where X indicates the subject number (between 1 and 40). In each of these directories, there are ten different images of that subject, which have names of the form Y.pgm, where Y is the image number for that subject (between 1 and 10). The database can be retrieved from http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/at_t_faces.tar.Z as a 4.5Mbyte compressed tar file or from [\[t_faces.zip\]\(#\) as a ZIP file of similar size. The database was utilized in \[10\].](http://www.cl.cam.ac.uk/Research/DTG/attarchive/pub/data/at</p></div><div data-bbox=)

III. DATA COLLECTION EXPERIMENTS

The corpora consist of only Africans between the ages of eight and forty five. A total of 123 male and 80 female frontal images were captured. It consists of mainly adults, only twelve children were captured, this is to ensure a large age variance. The pictures and streams were captured without glasses; only eight people were captured with plain glasses on. Of the eighty females captured, six of them had veils covering their hair to their bodies. The video streams are a total of four minutes each of static frontal recordings of involuntary and voluntary single and double eye blinks. The facial expression for both the pictures and the video is neutral. The pictures are stored in the BMP format, while the videos are stored in the AVI format. The picture frames from the video streams are stored in the BMP format.

A. Data Collection Process

A video recording cam-coder was used to capture the data for both still pictures and video clips. The choice of a cam-coder over a web cam was to enable the researches get good quality pictures and video streams. A JVC camera model GZ-MG 275 was used with f-stop f/1.4, exposure time 1/25 seconds, ISO speed ISO-200 and EXIF version 0221.

The consideration of the site setting was for the camera to be four feet in front of the person to be captured with a white background. Where illumination was not sufficient, flood lights were made available. Figure 1 Depicts a picture of the data capture studio setup.

B. Black Eye-Blink and Face Corpora (BEFC)

In the BEFC a total of two hundred and three (203) people were captured. One hundred and twenty three (123) males and eighty (80) females. Five frontal view pictures were captured of each of people in jpeg format at a dimension of 1632 by 1224, horizontal and vertical resolution of 72dpi each, resolution unit of 2, bit dept 24 and color representation of



Fig. 1. Picture of Data Capture Studio Setup

sRGB. These were later converted into bmp format to ensure we have a loss less file. The bmp files from the pictures are of size 1632 by 1224 pixels each to be down sized to 640 by 480 pixels for a more reduce storage size and processing time. Each file is 900Kb on disk and 24 bpp(bit dept) each. Figure 2 shows sample faces in the face corpora.

The video clips were made for four minutes each. They are frontal view video clips, the first two minutes was to capture involuntary eye blinks, in the third minute, prompted single eye blinks were taken and in the fourth minute prompted double eye blinks were taken. The video clips were taken in the .MOD format and later converted to the avi format because decoding program is easily done with the avi format. The .MOD files had a frame dimension of 720 by 576, frame rate of 25fps, data rate 7700kbps and total bit rate 7956 kbps. The avi format has frames are of the dimension 720 by 576 pixels, data rate 1041 Kbps(kilobytes per second) , total bit rate of 1169 Kbps, frame rates 25 frames per second, bit rates 128 kbps, and of an average size of 33.7 Mb. A total of 90 videos clips were taken, 50 males and 40 females. The video clips were latter split into frames of bmp format of size 1.18Mb each, dimension 720 by 572, 24 bit colour. Eight of the people were captured with glasses to see how sensitive the system will be with glasses.

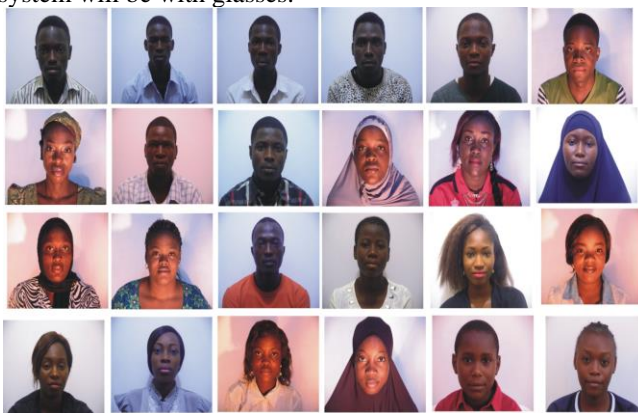


Fig. 2. Sample Faces in the Black Face Corpora

C. Quantity and Quality of Collected Data

The corpora stores one thousand one hundred and fifteen (1015) frontal view face pictures both of jpeg and bmp formats, ninety (90) video clips for four minutes each in the avi format and ninety folders of the frames of each video clip in the bmp format. The pictures are sharp, bright and distinct. There are some shadows at the background; the white background assumed various shades depending on the dominant color of the cloth of the person.

IV. CREATION AND ACCESS OF CORPORA

A. Description Of The Corpora Creation and Access Process

The corpora are organised of four folders, the first folder contains 1165 pictures in the jpeg format, the second contains 1165 pictures in the bmp format, the third contains 90 video clips of four minutes each and the fourth contains 90 folder of the frames of each video clip in the bmp format.

The pictures have names of the form Fac_x_t. file extension. Where x stands for number numbering from 001-223 and t stands for image number 1 to 5 for each subject. Figure 3 shows a sample of named pictures. The video streams have names of the form MOVxy.file extension, where x numbers from 1 to 90 and y is the number corresponding to the number of the picture of the person in the face database.

While the frame folders have names of the form MOVx , where x numbers from 1 to 90. . The corpora can be accessed by contacting the first authors, the website where it can be uploaded is still under development and once it is ready it will be made public.

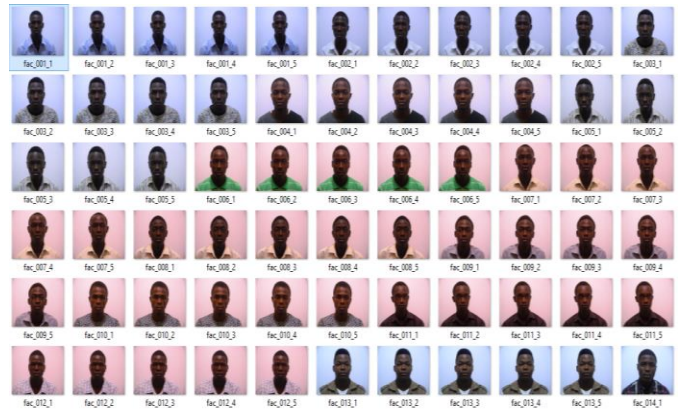


Fig. 3. Arrangement and Naming in The Black Face Corpora

V. CONCLUSION

This paper has discussed a new black face and eye blink corpora created for the development of a face recognition enabled by eye blink recognition system. The corpora are available for research as a contribution to the fields of HCI and bioinformatics for academic and non-commercial use especially for black faces. The system that will be developed will be applicable in the security system of the nation and also in the health sector

REFERENCES

- [1] R. T. Narmadha , T. Mythili, R.T. Nivetha, “ Real Time HCI Using Eye Blink Detection”. *Internationa Journal Of Computer Science And Mobile Computing*, Vol. 3, Issue 1 2014
- [2] M. Meenakshi, “ Real-Time facial recognition system- Design, implementation and validation”. *Journal of signal processing theory and applications 2013*.
- [3] J.A. Emmanuel and T.S.Ibiyemi, “A Framework for a Face Recognition Enabled Eye Blink communication”, in Book of Abstract, 1st International Conference on Sustainable Development (ICSD 2015), Baze University, Abuja Nigeria.
- [4] S. Priya, Parmar and Nehal Chitaliya, “Detect Eye blink using Motion analysis Method”, *International Journal of Emerging Technologies in Computational and Applied Sciences (IJETCAS)*, 2013
- [5] J.A. Ojo and S.A. Adeniran, “ Colour Face Image Database for Skin Segmentation, Face detection, Recognition and Tracking for Black Faces Under Real Life Situations”, *International Journal of Image procesing(IJIP)*, vol 4 Issue 6, 2011
- [6] http://www.ecse.rpi.edu/~cvrl/database/ISL_IR_Eye_Database.htm
- [7] Gang Pan, Lin Sun, Zhaohui Wu, Shihong Lao, Eyeblink-based Anti-spoofing in Face Recognition from a Generic Webcam, *The 11th IEEE International Conference on Computer Vision (ICCV'07)*, Rio de Janeiro, Brazil, 2011.
- [8] <https://www.bioid.com/About/BioID-Face-Database>
- [9] www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html
- [10] Ferdinando Samaria, Andy Harter, Parameterisation of a Stochastic Model for Human Face Identification, *Proceedings of 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota FL, December 1994.

Comparison Fractal Color Image Compression using YIQ and YUV Color Model

Eman A. Al-Hilo

Physics Department/ College of Education for Girls
Kufa University/ Iraq

Rusul Zehwar

Physics Department/ College of Education for Girls
Kufa University/ Iraq

Abstract—The principle of fractal image coding is that the image converges to a stable image by iterating the contractive transformations on an arbitrary initial image. This algorithm Partition the image into a number of range blocks and domain blocks. For every range block, the best matching domain block is searched for among all domain blocks by performing a set of transformations on each block. For color image compression, the Fractal coding is applied on different planes of color image independently by treating each plane as gray level image. The coordinate systems used for color image are RGB, YIQ and YUV. To encode a color image the main idea is to divide the image into its three different layers or components (RGB, YIQ and YUV). It is then possible to compress each of these layers separately, handle each of the layers as an independent image. In this paper the data of the color component (R,G,B) are transformed two times in two program separately, ones for YIQ and other for YUV color space. The results show that using (YUV) color space is more useful and efficient than using YIQ in fractal image compression, where PSNR increase 0.1% , CR increase 0.31% and ET decrease 2.321%.

Keywords—Compression image; fractal color image compression and iterated function system

I. INTRODUCTION

Colors are important for human for communicating with the daily encountered objects as well as his species, these colors should be represented formally and numerically within a mathematical formula so it can be projected on device computer storage and applications, this mathematical representation is known as color model that can hold the color space, by the means of color's primary components (Red, Green, and Blue) the computer can visualizes what the human does in hue and lightness. Most of these techniques reduce the redundancies between color components by transforming the color primaries into a decorrelated color model such as YUV and YIQ. [1]

The YIQ color space has been used in NTSC (National Television System Committee) color TV system and has been employed in USA, Canada, Japan and Korea. The Y stands for luminance components, and I and Q stand for chrominance information for representing color. [2]

RGB values can be transformed to YIQ color space using equation (1), while YIQ values can be converted back to RGB using equation (2). [3,4]

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.144 \\ 0.596 & -0.247 & -0.322 \\ 0.212 & -0.523 & 0.311 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \dots\dots\dots(1)$$

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 0.956 & 0.621 \\ 1.000 & -0.272 & -0.647 \\ 1.000 & -1.105 & 1.702 \end{bmatrix} \cdot \begin{bmatrix} Y \\ I \\ Q \end{bmatrix} \dots\dots\dots(2)$$

The YUV color space defines is a term of one luminance (Y channel) and two chrominance components (UV channels).U represents the color difference between blue signal and luminance (B–Y) and V represents the difference between red and luminance (R–Y). RGB values can be transformed to YUV color space using equation (3), while YUV values can be converted back to RGB using equation (4). [3,4]

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \dots\dots\dots(3)$$

This is the inverse matrix to get the RGB components out of the YUV color:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} 1.000 & 0.000 & 1.140 \\ 1.000 & -0.395 & -0.581 \\ 1.000 & 2.032 & 0.000 \end{bmatrix} \cdot \begin{bmatrix} Y \\ U \\ V \end{bmatrix} \dots\dots\dots(4)$$

Many researches study the color image compression. Porat [2001] study the effect of intercolor correlation on color image compression. He has been compare color compression techniques that take advantage of this inter color correlation. Most of these techniques reduce the redundancies between color components by transforming the color primaries into a de correlated color space, such as YIQ or YUV. [5]

Kasambe and Patel [2007] they have to provide a method that uses genetic algorithms to speed up. The coordinate systems used for color image are RGB, YIQ and YUV. The time required for the account in fractal image compression with acceptable Image quality for color and gray scale images. [2]

Al-Hilo [2007] has studied speeding fractal color image compression by moment feature includes converting the RGB model system to YUV model and minimize of the rang of U and V because of most of the image data are concentrated in the range of Y [6]

II. MATCHING PROCESS

After the generation of the range and domain pools; one takes each range block listed in the range pool and map it with all the domain blocks listed in the domain pool. At each mapping instance one determines the mapping coefficients, i.e., scale (s) and offset (o), which are called the IFS coefficient. These parameters (s) and (o) are determined by applying the least sum X^2 of square errors between r_i' and r_i according to following equation: [7]

$$x^2 = \sum_{i=0}^{n-1} (r_i' - r_i)^2 \dots\dots\dots(5)$$

The minimum of x^2 occurs when:

$$\frac{\partial x^2}{\partial y} = 0 \text{ and } \frac{\partial x^2}{\partial o} = 0 \dots\dots\dots(6)$$

Substituting equation (5) in (6) and using equations (7) one gets:

$$S = \frac{n \sum_{i=0}^{n-1} r_i d_i - \sum_{i=0}^{n-1} r_i \sum_{i=0}^{n-1} d_i}{n \sum_{i=0}^{n-1} d_i^2 - (\sum_{i=0}^{n-1} d_i)^2} \dots\dots\dots(7)$$

$$O = \frac{\sum_{i=0}^{n-1} r_i \sum_{i=0}^{n-1} d_i^2 - \sum_{i=0}^{n-1} r_i d_i \sum_{i=0}^{n-1} d_i}{n \sum_{i=0}^{n-1} d_i^2 - (\sum_{i=0}^{n-1} d_i)^2} \dots\dots\dots(8)$$

$$x^2 = \frac{1}{n} [\sum_{i=0}^{n-1} r_i^2 + s \sum_{i=0}^{n-1} d_i^2 - 2 \sum_{i=0}^{n-1} r_i d_i + 2o \sum_{i=0}^{n-1} d_i + o(no - 2 \sum_{i=0}^{n-1} r_i)] \dots\dots\dots(9)$$

where, d_i is the i^{th} pixel value of the matched domain block.

r_i is the i^{th} pixel value of the range block.

n is the number of pixels in each block (i.e. the block size).

III. ENCODING TECHNIQUE

The implementation encoding method could be summarized by the following steps:

- 1) Load BMP image and put it in (R,G,B) arrays (i.e., three 2D arrays).
- 2) Convert (R,G,B) arrays to (Y,I,Q) arrays in first program and convert to (Y,U,V) in second program.
- 3) Down sample (I and Q) to quarter of its original size in first program and to (U and V) in second program.
- 4) For each component (i.e., the original Y, and the down sampled (I,Q) for first program and (the original Y, and the down sampled (U,V) for second program) do:
 - a) Establish the range array
 - b) Down sample the range image to produce the domain array.
 - c) Partitioning:
 - (1) The range array must be partitioned into non-overlapping fixed blocks, to generate the range blocks (r_1, \dots, r_n).

- (2) The domain must be partitioned into overlapping blocks, using specific step size, to generate the domain blocks (d_1, \dots, d_n). They should have the same size of range blocks.

d) Searching:

- (1) Pick up a domain block from the domain pool.
- (2) Perform one of the isometric mappings.
- (3) Calculate the scale (s) and offset (o) coefficient using equations (7-8).
- (4) Apply the following condition to bound the value of (s) and offset (o) coefficient:

If $s < s_{min}$ then $s = s_{min}$

Else if $s > s_{max}$ then $s = s_{max}$

If $o < o_{min}$ then $o = o_{min}$

Else if $o > o_{max}$ then $o = o_{max}$

- (5) Quantize the value (s) and offset (o) using equations (7-8).
- (6) Compute the approximation error (χ^2) using equation (9).
- (7) After the computation of IFS code and the sum of error (χ^2) of the matching between the range and the tested domain block., the (χ^2) is compared with registered minimum error (χ^2_{min}); such that:
If $\chi^2 < (\chi^2_{min})$ then

$$s_{opt} = i_s; o_{opt} = i_o, \chi^2_{min} = \chi^2$$

PosI=domain block index

Sym=symmetry index

End if

- (8) If $\chi^2_{min} < \epsilon$ then the search across the domain blocks is stopped, and the registered domain block is considered as the best matched block
- (9) Repeat steps (4) to (10) for all symmetry states of the tested domain block.
- (10) Repeat steps (3) to (11) for all the domain blocks listed in the domain pool.
- (11) The output is the set of IFS parameters (i.e., $i_s, i_o, posI, Sym$) which should be registered as a set of fractal coding parameters for the tested range block.
- (12) Repeat steps (1) to (12) for all range blocks listed in the range pool
- (13) Store all IFS mapping parameters as an array of record. The length of this array is equal to the number of range blocks in the range pool.

IV. DECODING TECHNIQUE

The decoding process by YIQ and YUV model can be summarized in the following steps:-

1) *Generating arbitrary domain pool. The domain pool could be initialized as a blank image or a piece of image extracted from any available image.*

2) *The values of the indices of (i_s) and (i_o) for each range block should be mapped to reconstruct the quantized values of the scale (s_q) and offset (o_q) coefficients. This step is called the dequantization step.*

3) *Choosing the number of possible iterations, and the threshold value of the mean square error (TMSE). At each iteration the following steps are performed:*

a) *For each range block one determines the coordinates (x_d, y_d), of the best matched domain, from the IFS parameters ($posI$), in order to extract the domain block (d) from the arbitrary domain image.*

b) *For each range block, its approximation r_i' is obtained by multiplying the corresponding best matched domain block (d) by the scale value (s_q) and adding to the result the offset value (o_q), according to equation.:*

$$r_i' \approx sd_i + o \dots \dots \dots (10)$$

c) *The generated block is transformed (rotated, reflected, or both) according to its corresponding IFS symmetry parameter value (Sym).*

d) *The generated r_i' block is placed in its position in the decoded image array (range image).*

e) *Checking whether there is another range block, if yes then steps (b,c,d) are repeated.*

f) *Down sampling the reconstructed image (range pool) in order to produce the domain pool by using the averaging (or integer) sampling.*

g) *Calculating the mean square error (MSE) between the reconstructed range and the previous reconstructed range image. If the MSE is greater than (TMSE) value then the iteration continued and the above steps (a-f) should repeated; this iteration is continued till reaching the attractor state (i.e., the newly reconstructed range image is very similar to the previous reconstructed image), otherwise the iteration continue till reaching the predefined maximum number of iterations (m), in our program used (m=20).*

4) *The above (steps 3a-3g) should be implemented upon the three components (Y,I,Q in first program and Y,U,V to second program) to produce the attractor of each component.*

5) *Converting the reconstructed (YIQ color components in first program and YUV in second program) to RGB components sing the inverse (YIQ and YUV) transform equations respectively (2,4).*

6) *Calculating the fidelity criteria for each RGB component, then determines the overall value of the fidelity criteria for the RGB reconstructed image.*

V. TESTS RESULTS

This work is carried out in Visual Basic 6.0 version on Laptop (hp): intel (R) Core (TM) i5-2430M CUP @ 2.40 GHz Processor, 64-bit Operating System and 6.00 GB RAM. To evaluate the performance of the established colour FIC system by YIQ and YUV model, the proposed system has been tested using Lena image (256x256 pixel, 24bits) as test image. These tests explore the effect of the following coding parameters on the compression performance parameters of the established system of YIQ color space and YUV color space:

A. Maximum and Minimum Scale Tests

This set of tests was conducted to study the effect of MinScale, and MaxScale on the compression performance parameters of the reconstructed image in YIQ and YUV models. Table (1) and (2) show the effects of this test on compression performance parameters for YIQ and YUV respectively.

TABLE I. EFFECTS OF MAXSCALE AND MINSKALE ON THE COMPRESSION PERFORMANCE PARAMETERS USING (YIQ) COLOR MODEL

MaxScale	MinScale	PSNR	CR	ET(sec)
1	-1	30.85	8.73	54.65
	-1.5	30.94	8.73	54.34
	-2	30.94	8.73	54.26
	-2.5	30.92	8.73	54.46
	-3	30.90	8.73	53.96
2	-1	31.02	8.73	54.99
	-1.5	31.05	8.73	54.17
	-2	31.03	8.73	54.36
	-2.5	31.02	8.73	53.99
	-3	31.01	8.73	54.97
3	-1	31.04	8.73	54.25
	-1.5	31.04	8.73	54.19
	-2	31.05	8.73	54.76
	-2.5	31.04	8.73	56.56
	-3	30.98	8.73	55.27

TABLE II. EFFECTS OF MAXSCALE AND MINSKALE ON THE COMPRESSION PERFORMANCE PARAMETERS USING (YUV) COLOR MODEL

MaxScale	MinScale	PSNR	CR	ET(sec)
1	-1	30.93	9.04	58.59
	-1.5	31.02	9.04	57.98
	-2	31.02	9.04	57.72
	-2.5	31.02	9.04	58.33
	-3	31.02	9.04	57.52
2	-1	31.11	9.04	57.75
	-1.5	31.14	9.04	58.10
	-2	31.15	9.04	57.50
	-2.5	31.15	9.04	57.87
	-3	31.12	9.04	57.91
3	-1	31.11	9.04	58.83
	-1.5	31.13	9.04	58.41
	-2	31.14	9.04	57.64
	-2.5	31.13	9.04	55.33
	-3	31.13	9.04	58.33

1) Step Size Tests

In this set of tests the effect of the shift StepSize parameter is studied. Figures (1) and (2) show the effects of this test using YIQ and YUV models respectively.

Original image	StepSize (1)	StepSize (2)	StepSize (3)	StepSize (4)
				
PSNR	31.5	31.05	30.71	30.48
CR	8.73	8.73	8.73	8.73
ET	208.23	53.45	24.08	14.20

Fig. 1. Effect of StepSize parameter on the compression performance parameters using (YIQ)

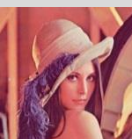

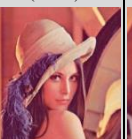
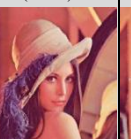
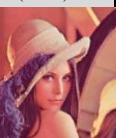
Original image	DomSize (128x128)	DomSize (64x64)	DomSize (32x32)	DomSize (16x16)
				
PSNR	31.05	30.49	29.49	27.80
CR	8.73	8.73	8.73	8.73
ET	53.75	13.24	3.48	1.14

Fig. 5. The effects of DomSize on the compression performance parameters using YIQ





Original image	StepSize (1)	StepSize (2)	StepSize (3)	StepSize (4)
				
PSNR	31.60	31.15	30.83	30.57
CR	9.04	9.04	9.04	9.04
ET	223.88	54.68	24.69	14.46

Fig. 2. Effect of StepSize parameter on the compression performance parameters using (YUV)

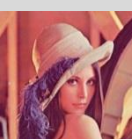
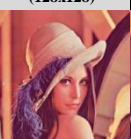
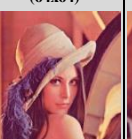
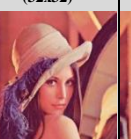
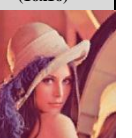
Original image	DomSize (128x128)	DomSize (64x64)	DomSize (32x32)	DomSize (16x16)
				
PSNR	31.15	30.55	29.95	27.85
CR	9.04	9.04	9.04	9.04
ET	54.85	13.68	4.56	1.23

Fig. 6. The effects of DomSize on the compression performance parameters using YUV

2) Block Size Tests

This set of conducted tests aimed to investigate the effects of the BlockSize parameter on the compression performance parameters. Figures (3) and (4) show the effects of this test using YIQ and YUV models respectively.

Original image	BlockSize (2x2)	BlockSize (4x4)	BlockSize (8x8)	BlockSize(16x16)
				
PSNR	34.32	31.05	27.15	23.68
CR	2.04	8.73	37.46	161.68
ET	30.68	53.57	31.27	24.17

Fig. 3. Effects of BlockSize parameter on the compression performance parameters using (YIQ)

Original image	BlockSize (2x2)	BlockSize (4x4)	BlockSize (8x8)	BlockSize(16x16)
				
PSNR	28.80	31.15	27.19	23.68
CR	2.11	9.04	38.89	168.32
ET	101.70	57.62	31.34	23.59

Fig. 4. Effects of BlockSize parameter on the compression performance parameters using (YUV)

3) Domain Size Tests

This set of tests was performed to define the effect of the DomSize on the compression parameters. Figures (5) and (6) show the effects of this test using YIQ and YUV models respectively

4) Permissible Error Value (ϵ_o) Tests

This set of tests was performed to study the effects of permissible error level (ϵ_o) on the compression performance parameters. Tables (3) and (4) show the effects of this test using YIQ and YUV models respectively

TABLE III. EFFECTS OF PERMISSIBLE ERROR VALUE (E) ON THE COMPRESSION PERFORMANCE PARAMETERS USING (YIQ)

ϵ_o	PSNR	CR	ET(sec)
0.1	31.048	8.73	54.72
0.2	31.048	8.73	55.36
0.3	31.048	8.73	54.49
0.4	31.047	8.73	54.27
0.5	31.048	8.73	54.27
0.6	31.047	8.73	53.62
0.7	31.045	8.73	55.33
0.8	31.044	8.73	55.00
0.9	31.041	8.73	51.67
1	31.033	8.73	50.32
2	30.826	8.73	33.21
3	30.367	8.73	20
4	29.959	8.73	14.22
5	29.534	8.73	10.10

TABLE IV. SHOWS THE DIFFERENT BETWEEN YUV, YIQ MODEL FOR LENA IMAGE

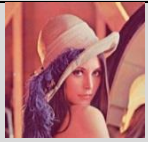
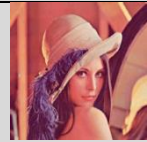
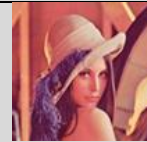


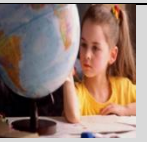
Original image	YIQ	YUV	
			Percentage Error %
PSNR	31.05	31.15	
CR	8.73	9.04	Increase 0.31%
ET	53.75	57.62	increase 2.23%

TABLE V. SHOWS THE DIFFERENT BETWEEN YUV, YIQ MODEL FOR SUN CHILD IMAGE

Original image	YIQ	YUV	
			Percentage Error %
PSNR	35.2	36.32	
CR	8.73	9.04	Increase 0.31%
ET	51.48	55.14	Increase 3.66%

VI. CONCLUSIONS

The results indicate that using YUV model is more efficient than YIQ model as in following:

- 1) The PSNR is increase 0.10% in YUV model than in (YIQ) model.
- 2) The CR is increase 0.31% than in YUV model than in (YIQ) model.
- 3) The ET parameter in the (YUV) is increase 2.23% than in (YIQ) model.

Table (5) and (6) summarizes the results for Lena and sun child image respectively.

TABLE VI. EFFECTS OF PERMISSIBLE ERROR VALUE (E) ON THE COMPRESSION PERFORMANCE PARAMETERS USING (YUV)

ϵ_0	PSNR	CR	ET(sec)
0.1	31.1466	9.04	53.68
0.2	31.1465	9.04	54.95
0.3	31.145	9.04	54.41
0.4	31.144	9.04	54.26
0.5	31.140	9.04	56.67
0.6	31.142	9.04	55.59
0.7	31.140	9.04	54.54
0.8	31.136	9.04	57.74
0.9	31.133	9.04	51.00
1	31.120	9.04	50.72
2	30.862	9.04	33.42
3	30.151	9.04	20.19
4	29.540	9.04	14.19
5	28.994	9.04	5.05

REFERENCES

- [1] Ibraheem, N., Hasan, M., Khan, R., Mishra, P., " *Understanding Color Models: A Review*", ARPN Journal of Science and Technology, VOL. 2, NO. 3, April 2012.
- [2] Kasambe p.v. and Patel M., " *Modified Fractal Image Compression Using Genetic Algorithms*", Mumbai, India, Vol. 1, 134, 2007.
- [3] Ford A., Roberts A., " *Color Space Conversions*", August 11, 1998(b).
- [4] González J. M. C., Rodríguez M. A. V., Pulido J. A. G., and Pérez J. M. S., " *Digital Signal Processing*", Available online 23 October 2009.
- [5] Porat M., " *The Effect of Inter Color Correlation on Color Image Compression*", M.Sc., Electrical Engineering, January 2001.
- [6] Al-Hilo, E.A., " *Speeding-up Fractal Colored Image Compression using Moments Features*", PhD Thesis, College of Science, AL-Mustansiriyah University, Baghdad, 2007.
- [7] Fisher, Y., " *Fractal Image Compression Theory and Application*", Book, University of California, Institute for Nonlinear Science, Springer-Verlay, New York, Inc, 1995.

A 'Cognitive Driving Framework' for Collision Avoidance in Autonomous Vehicles

Alan J. Hamlet

Department of Mechanical and
Aerospace Engineering
University of Florida
Gainesville, Florida, USA

Carl D. Crane

Department of Mechanical and
Aerospace Engineering
University of Florida
Gainesville, Florida, USA

Abstract—The Cognitive Driving Framework is a novel method for forecasting the future states of a multi-agent system that takes into consideration both the intentions of the agents as well as their beliefs about the environment. This is particularly useful for autonomous vehicles operating in an urban environment. The algorithm maintains a posterior probability distribution over agent intents and beliefs in order to more accurately forecast their future behavior. This allows an agent navigating the environment to recognize dangerous situations earlier and more accurately than competing algorithms, therefore allowing the agent take actions in order to prevent collisions. This paper presents the Cognitive Driving Framework in detail and describes its application to intersection navigation for autonomous vehicles. The effects of different parameter choices on the performance of the algorithm are analyzed and experiments are conducted demonstrating the ability of the algorithm to predict and prevent automobile collisions caused by human error in multiple intersection navigation scenarios. The results are compared to the performance of prevailing methods; namely reactionary planning and constant velocity forecasting.

Keywords—Multi-agent systems; autonomous vehicles; intent prediction; non-linear filtering; Bayesian filtering;

I. INTRODUCTION

The potential safety and convenience benefits that autonomous vehicles can provide to our society are myriad. The World Health Organization reported that in 2010, 1.24 million people died due to road vehicle accidents.¹ In addition to the potential of reducing this massive loss of life, autonomous vehicles have shown promise in increasing vehicle efficiency and convenience for drivers [1]–[3].

The vast majority of current autonomous vehicle architectures employ a reactionary response to changes in the environment. These systems require very frequent and rapid re-planning in order to avoid dynamic obstacles. Another intuitive approach is to have the autonomous vehicle predict where the dynamic obstacles are going to be in order to plan a path. One popular approach to making this prediction is to assume the dynamic obstacle continues to move in a straight line at its current velocity, as is done in the 'velocity obstacle' literature [4], [5]. Cornell's autonomous vehicle, Skynet, uses this type of approach by using an extended Kalman filter to track dynamic obstacles and then calculates the 'time to

collision' assuming a constant speed and heading [6]. This approach does not take into account the control decisions made by the dynamic obstacle that affect its trajectory, as is the case for pedestrians and other vehicles.

Some research has begun to incorporate the intentions of the dynamic obstacle in order to more intelligently predict its future position. Some methods used to predict intent are hidden Markov models [7], [8], Markov decision processes [9], and Gaussian processes or mixture models [10], [11]. These methods attempt to model trajectories and classify the dynamic obstacles' motion according to the corresponding intent. While this body of research is a step toward realizing more intelligent vehicles that truly understand their environment, it fails to consider how the obstacles' understanding of the environment will affect its future state.

The aforementioned types of planners work sufficiently well for navigating in urban environments where other vehicles are driving safely, but widespread adoption of autonomous vehicles will take time and human driven vehicles will remain on the roads for many years. With the presence of non-autonomous vehicles, the potential for accidents caused by human error will persist. According to a research study by the National Highway Traffic Safety Administration, 93% of traffic accidents were caused by human error.² Autonomous vehicles need to be able to operate alongside human drivers and prevent these potential collisions caused by human error. Reactive planners will often fail to recognize these dangerous situations in time to prevent a collision. Using constant velocity forecasting can result in overly cautious driving, due to frequent false predictions of dangerous situations.

In this research, both the intent and the belief of a dynamic obstacle are considered when modeling the future states of the obstacle. This is beneficial for situations in which a dynamic obstacle, e.g. a pedestrian or another vehicle, may have an incorrect belief about the environment. For example, an obstacle vehicle trying to merge into traffic may believe it has more space than it actually does or it may not see an oncoming vehicle due to occlusions or driver error. In these situations, just knowing the driver's intent does not suffice since for the same intent she may yield or begin to merge depending on her belief.

¹http://www.who.int/gho/road_safety/mortality/traffic_deaths_number/en

²<http://www.nhtsa.gov/people/injury/research/udashortrpt/background.html>

In this paper, the dynamics of a multiple-vehicle system are modeled as a dynamic Bayesian network (DBN). Obstacle vehicles' actions are dependent on both their intent and their belief of the surrounding environment. This idea is similar to that proposed in [12], but in this paper the problem is not formulated as a Markov decision process as to avoid discretization of the state space. This is required in order to achieve the resolution necessary for the autonomous vehicle domain. Inference is performed over the network using a particle filter to jointly estimate the vehicle's intent and belief. Future vehicle states are then forecast using Monte Carlo simulation and the probability of a future collision is calculated. Simulation results show that this method of joint inference allows an autonomous vehicle to predict a collision with enough time to take evasive action.

The remainder of this paper is structured as follows. In section 2, an overview of the cognitive driving framework is given. The manner of representing the system state and dynamics is described. In section 3, the process for formulating the problem as inference over a dynamic Bayesian network is explained. The structure of the DBN is detailed and the method of performing joint inference over the network using particle filtering is discussed. At the end of section 3, forecasting the future state of the system using Monte Carlo simulation is explained. Next, in section 4, a detailed analysis is given on how different parameter choices affect the performance of the algorithm. Then simulation results demonstrating the accuracy of the proposed method are presented. Finally, concluding remarks are given and future research directions are discussed in section 5.

II. THE COGNITIVE DRIVING FRAMEWORK

This section provides an overview of the cognitive driving framework by describing how the state of an intersection environment with multiple vehicles is represented and by defining the form of the system dynamics.

In the cognitive driving framework, or CDF, the system consists of two vehicles, the obstacle vehicle and the ego vehicle, in a known environment. The joint state of the two vehicles is called the system pose and is represented as

$$\mathbf{S}_t = \begin{bmatrix} {}^1\mathbf{x}_t \\ {}^2\mathbf{x}_t \end{bmatrix}, \quad (1)$$

where a superscript 1 denotes the obstacle vehicle and a superscript 2 denotes the ego vehicle. In this paper, the term 'ego vehicle' refers to the vehicle that is trying to predict the intent of the obstacle vehicle.

In order to provide a general algorithm, the system dynamics are assumed to be nonlinear and of the form

$$\mathbf{S}_{t+1} = \begin{bmatrix} {}^1\mathbf{x}_{t+1} \\ {}^2\mathbf{x}_{t+1} \end{bmatrix} = \begin{bmatrix} f({}^1\mathbf{x}_t, {}^1\mathbf{u}_t, {}^1\boldsymbol{\nu}_t) \\ f({}^2\mathbf{x}_t, {}^2\mathbf{u}_t, {}^2\boldsymbol{\nu}_t) \end{bmatrix}, \quad (2)$$

where ${}^i\mathbf{u}_t$ is the control input and ${}^i\boldsymbol{\nu}_t$ is the process noise for vehicle i at time t . The controller for the autonomous vehicle running the CDF (the ego vehicle) is assumed to be of the form

$${}^2\mathbf{u}_t = h({}^2\mathbf{x}_t, {}^1\mathbf{x}_t, {}^2I), \quad (3)$$

where the arguments to the nonlinear function $h()$ are the vehicle's own state, the state of the obstacle vehicle, and the

intent of the ego vehicle, respectively, at time t . The intent variable, iI , represents the current behavior the vehicle is trying to execute (e.g. turn left or go straight through the intersection). Here the controller, $h()$, is both highly nonlinear and discontinuous as it is a function of both continuous and discrete variables. The nonlinearities arise not only from the piecewise nature due the discrete intent variable, but also from the nonlinear kinematics of the system and the nonlinear dependence on the obstacle vehicle state.

The controller for the obstacle vehicle is modeled similarly as

$${}^1\mathbf{u}_t = h({}^1\mathbf{x}_t, \mathbf{B}_t, {}^1I). \quad (4)$$

The difference here is that the obstacle vehicle is not assumed to have exact knowledge of the ego vehicle's state. Instead, the obstacle vehicle's controller operates on the assumed state of the ego vehicle, the *belief*, \mathbf{B}_t . It should be noted that in this context the belief is simply a point, not a distribution or density as sometimes used in the literature. If the ego vehicle has not been observed by the obstacle vehicle, then $\mathbf{B}_t = \emptyset$. The obstacle updates its belief according to the equations

$$\mathbf{B}_{t+1} = g(\mathbf{B}_t, \mathbf{O}_{t+1}) \quad (5)$$

$$\mathbf{O}_t = k(\mathbf{S}_t, \beta, \mathbf{e}_t), \quad (6)$$

where \mathbf{O}_t is the obstacle vehicle's observation at time t , and β is a parameter that represents the probability of the obstacle vehicle observing the ego vehicle at any given discrete time step. The observation noise, \mathbf{e}_t , is normally distributed with a mean of zero. Given \mathbf{B}_t and \mathbf{O}_{t+1} , \mathbf{B}_{t+1} updates deterministically. The observation model, $k()$, determines from the system pose if the ego vehicle is in the obstacle vehicle's *isovist*: the volume of space with line of sight visibility from the obstacle vehicle's pose. If the ego vehicle is occluded by other vehicles or buildings, it will not be in the obstacle vehicle's isovist, and thus $\mathbf{O}_t = \emptyset$. If the ego vehicle is in the obstacle vehicle's isovist, then the obstacle vehicle will make a noisy observation of the ego vehicle's pose with probability β .

The goal of the cognitive driving framework is to allow the ego vehicle to predict the future states of the obstacle vehicle using this model in order to prevent collisions. This is done by performing online inference of the obstacle vehicle's belief and intent, \mathbf{B}_t and I_t . The following section details the procedure for performing this joint inference and prediction.

III. FILTERING AND FORECASTING

In this section, the model outlined in the previous section is formulated as a dynamic Bayesian network. How online inference is performed using a particle filter is described and a procedure for using Monte Carlo simulation for forecasting future system states is presented.

A. Dynamic Bayesian Network

The cognitive driving framework uses a dynamic Bayesian network to capture the dependencies between the random variables in the CDF system dynamics. A Bayesian network is a directed acyclic probabilistic graphical model that is used to represent a set of random variables and their conditional dependencies. A *dynamic* Bayesian network is a Bayesian network

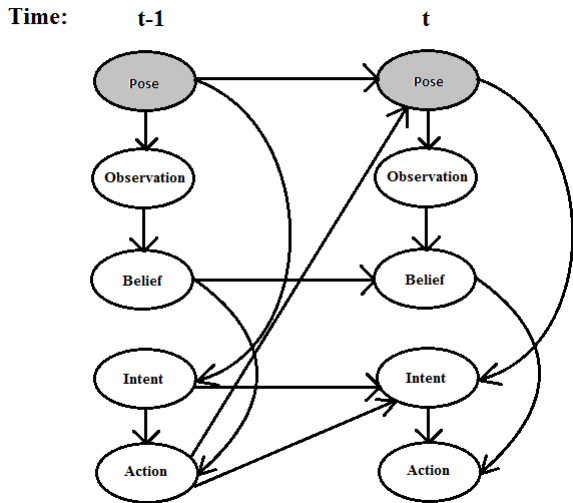


Fig. 1: The structure of the DBN used in the cognitive driving framework.

which relates the variables to each other over sequential time steps. In literature, dynamic Bayesian networks are sometimes referred to as *two-time-slice Bayesian networks* because at any point in time t , the value of a variable in the network can be calculated from the prior value (at time $t-1$) and the independent variables [13]. Kalman filter models and Hidden Markov Models are special cases of DBN's. In Kalman filter models, both the system dynamics and the measurements are assumed to be linear Gaussian. In hidden Markov models, the dynamics and measurements both have discrete distributions. DBN's make no assumptions about the form of the dynamics or measurements and allow the hidden state of the system to be factored into separate variables so the structure of the dependencies between the variables can be exploited.

The structure of the DBN used in the CDF is depicted in figure 1. This graphical model reflects the dependencies given by the equations in section II. The gray nodes in the graph denote the variable known by the ego vehicle, the system pose, S_t , as given in equation 1. In some contexts, because the value of this variable is provided to the ego vehicle by its sensors, it is called the observation. In this work, the observation, O_t , refers to the obstacle vehicle's noisy measurement of the ego vehicle's pose, 2x_t .

Between time-slices, the variables in the DBN flow temporally from left to right and within a time-slice they flow (more-or-less) from top to bottom. The system pose affects the obstacle vehicle's observation which in turn determines the obstacle vehicle's belief. The obstacle vehicle's intent and belief of the system pose inform the obstacle vehicle's controller. The joint actions of the two vehicles result stochastically in the next system pose. Without loss of generality, the intent of the obstacle vehicle is assumed to be constant throughout an episode.

B. Filtering

Now that the two-vehicle system dynamics are represented as a DBN, a method of filtering needs to be implemented

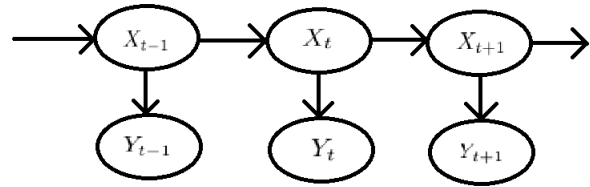


Fig. 2: Simplified model of the CDF using the joint DBN state variable X_t .

in order to perform online inference of the obstacle vehicle's belief and intent. By combining equations 2 through 6 we can represent the DBN state and dynamics, respectively, as

$$X_t = [S_t, B_t, I]^T \quad (7)$$

$$X_{t+T} \sim P(X_{t+T}|X_{t+T-1}) \quad (8)$$

This condenses the DBN in figure 1 to that shown in figure 2, which is the typical representation for filtering problems. The variable Y_t represents the measurement, which in this study is the system pose, S_t .

The system dynamics are highly non-linear, as shown in section II. Instead of linearizing the dynamics at the expense of accuracy of the estimation, a non-linear Monte Carlo based filtering method was employed. Monte Carlo (MC) methods are ideally suited for the current application due to their ability to model highly non-linear systems with multimodal, non-Gaussian distributions [14]. In this study, the DBN state is composed of both continuous and discrete variables, representing discrete intention hypotheses making traditional linearization methods such as the Extended or Unscented Kalman Filter unsuitable for this application.

Particle filters are sequential Monte Carlo methods that maintain an estimate of the posterior distribution of the system state as a set of particles. This non-parametric representation is capable of representing arbitrarily complex distributions as long as a large enough particle set is used. Each particle is initialized according to the a priori distribution and is propagated through the noisy system dynamics. The particles are then resampled according to the particles' importance weights. The importance weight of a particle is proportional to the likelihood of the particle generating the measurement, $Y_t = S_t$. In this study, the likelihood is represented as a Gaussian distribution centered around the measured system pose. The weight of each particle is proportional to the probability of the system pose of the particle given the measured system pose, as shown in the equation below.

$$w_t^{[m]} \propto P(X_t^{[m]}|S_t) \sim \mathcal{N}(S_t, \Sigma). \quad (9)$$

$$X_t^{[m]} \sim P(X_t^{[m]}|X_{t-1}^{[m]}) \quad (10)$$

Σ is the variance of the Gaussian likelihood function. A superscript $[m]$ denotes that the variable corresponds to the m^{th} particle. The weights are normalized such that they sum to one.

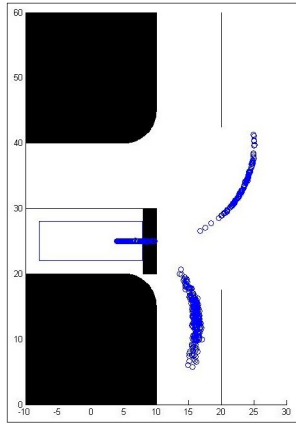


Fig. 3: The distribution of particles using Monte Carlo simulation for a look-ahead time of 2.5 seconds.

Once the importance weights are calculated, the particles are re-sampled in order to move the posterior distribution toward the region of the state space that matches the measurement. The technique of *stratified* or *low-variance* sampling is employed to reduce computational complexity [15]. The resulting set of particles is an approximation to the actual state of the system. As the number of particles, M , approaches infinity, the particle set converges to the true distribution of the state.

It is worthy of note that the version of the Cognitive Driving Framework presented here does not require that the obstacle vehicle controller (equation 4) be known; in fact, nor is it required that the system dynamics (equation 2) are known. All that is required for the CDF is that a simulation of the DBN system dynamics is available, i.e. that samples can be generated from the distribution $P(\mathbf{X}_{t+1}|\mathbf{X}_t)$.

C. Forecasting

Using the particle filtering method described in the previous subsection, an online estimate of the system state can be maintained. To predict and anticipate collisions in the future, though, the future state of the system must be estimated. To accomplish this, the CDF uses Monte Carlo simulation to propagate the particle set representing the current system state, \mathbf{X}_t , forward in time, creating a new particle set, \mathbf{X}_{t+T} , that approximates the state of the system at some time $t+T$ in the future. This is done by recursively sampling the DBN system dynamics given in equation 8. The probability of a collision can then be calculated from this new particle set by simply determining the percentage of particles in the set that represent a collision state.

An example future state distribution is shown in figure 3. Here, a vehicle is at a stop sign at a T-intersection and has the option to turn left or right. If it is turning left, it may also choose to yield if there is oncoming traffic. In the figure, the set of particles for a look-ahead time of 2.5 seconds is plotted. Three distinct modes can be seen in the distribution corresponding to the three potential behaviors of the vehicle: turning left, turning right, and yielding. This situation will

```

1: Algorithm Cognitive_Driving_Framework( $\chi_{t-1}, \mathbf{S}_t$ )
2:  $\chi_t = \bar{\chi}_t = \emptyset$ 
3: for  $m = 1$  to  $M$  do
4:    $\bar{\mathbf{X}}_t^{[m]} \sim P(\mathbf{X}_t^{[m]}|\mathbf{X}_{t-1}^{[m]})$ 
5:    $w_t^{[m]} = P(\bar{\mathbf{X}}_t^{[m]}|\mathbf{S}_t)$ 
6:    $\bar{\chi}_t = \bar{\chi}_t + \langle \mathbf{X}_t^{[m]}, w_t^{[m]} \rangle$ 
7:  $\chi_t = \text{Resample}(\bar{\chi}_t, w_t)$ 
8:  $\chi_{t+T} = \chi_t$ 
9:  $\text{collisions} = 0$ 
10: for  $m = 1$  to  $M$  do
11:   for  $\tau = 1$  to  $T$  do
12:      $\chi_{t+\tau}^{[m]} \sim P(\chi_{t+\tau}^{[m]}|\chi_{t+\tau-1}^{[m]})$ 
13:      $\text{collisions} = \text{collisions} + \text{CheckCollision}(\chi_{t+\tau}^{[m]})$ 
14:  $P(\text{collision}) = \text{collisions}/M$ 
15: if  $P(\text{collision}) \geq \text{Threshold}$  then
16:    $\text{EmergencyStop}$ 
17: return  $\chi_t$ 

```

Fig. 4: Pseudo-code overview of an update for the Cognitive Driving Framework algorithm.

be discussed and analyzed in more detail in the experimental results, section IV.

The cognitive driving framework algorithm is given in figure 4. In lines 2 through 5 the particle filter update is performed on the particle set, χ . Line 2 samples the next state from the DBN dynamics and line 3 sets the weight for the new sample based on the measured system pose. After this is done for all the particles, this weighted particle set, $\bar{\chi}_t$, is resampled according to the weights in order to move the distribution of the particles toward the measurement, as shown in line 5. In line 6, the particle set representing the future state of the system, χ_{t+T} is initialized to be equal to the set representing the current state, χ_t . Line 7 initially sets the number of particles in a collision state to zero. Lines 8 through 11 recursively propagate each particle one at a time through the system dynamics to obtain samples of the system state T time steps in the future. After each sample is propagated through the dynamics T times, it is checked to see if it is in a collision state in line 11, and the number of particles in a collision state is counted. Line 12 then calculates the probability of a collision as the number of particles in a collision state divided by the total number of particles. If the probability of a collision is higher than the set threshold, then an emergency braking maneuver is triggered in line 14. Otherwise, the algorithm just returns the particle set representing the current system state, χ_t .

There are multiple parameters used in this algorithm, namely, the look-ahead, T , the threshold, Threshold , and the number of particles, M . The selection of values for these parameters and their influence on the performance of the algorithm is discussed in section IV-B.

IV. EXPERIMENTAL RESULTS

In order to demonstrate the ability of the CDF to perform joint inference on the intent and belief of an obstacle vehicle as well as forecast the future state of the multi-agent system, two

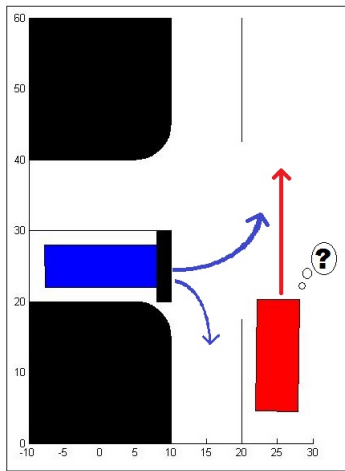


Fig. 5: A close up image of the simulated intersection environment for scenario one. The ego vehicle (red) attempts to predict the intent of the obstacle vehicle (blue).

simulated experiments were conducted. For comparison, the same experiments were also conducted using both a reactive planner and a constant velocity planner. This section first describes the simulation set up for the two tested scenarios. Next, a detailed analysis is performed to determine the optimal choice of parameters as well as examine the effects of parameter choice on the performance of the CDF algorithm. Lastly, the simulation results using all three methods are presented and discussed.

A. Simulation Setup

The CDF was tested using two simulated T-intersection scenarios. Scenario one is depicted in figure 5. The autonomous vehicle (the ego vehicle, in red) has the right of way. An obstacle vehicle (blue) is stopped at a stop sign at the intersection and can either turn left into the ego vehicle's lane or turn right. It is desirable for the ego vehicle to predict not just the intention of the obstacle vehicle to turn left, but whether the obstacle vehicle is going to turn left in front of the ego vehicle or if it is going to yield.

Scenario two takes place in the same intersection environment, but the ego vehicle is now traveling south (toward the bottom of the figure) through the intersection. The obstacle vehicle is driving north toward the intersection and has the option to proceed straight through the intersection or to turn left. Again, it is desirable to predict not just whether the obstacle vehicle is going to turn left or go straight, but whether the obstacle vehicle is going to turn left in front of the ego vehicle, yield to the ego vehicle, or go straight.

The simulation uses a bicycle kinematic model for the vehicles as described in [10]. The pose for a vehicle from equation 1 is given by the four dimensional vector

$$\mathbf{x}_t = \begin{bmatrix} x_t \\ y_t \\ \theta_t \\ v_t \end{bmatrix}, \quad (11)$$

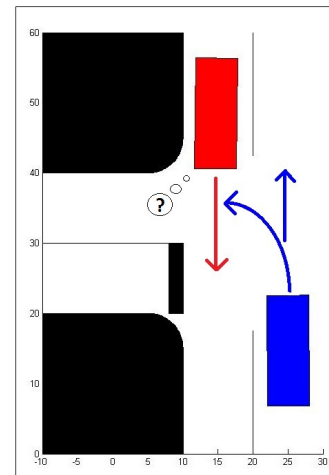


Fig. 6: A close up image of the simulated intersection environment for scenario two. The ego vehicle (red) attempts to predict the intent of the obstacle vehicle (blue).

where x_t and y_t are the position of the center of the rear axle in the ground plane, θ_t is the vehicle's orientation, and v_t is the speed of the vehicle, all at time t . The superscript indicating which vehicle the pose corresponds to has been left off here for clarity. The dynamics from equation 2 are then given by

$$\mathbf{x}_{t+1} = f(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) = \begin{bmatrix} x_t + v_t \cdot \Delta t \cdot \cos \theta_t \\ y_t + v_t \cdot \Delta t \cdot \sin \theta_t \\ \theta_t + \frac{v_t \cdot \Delta t}{l} \cdot \tan ({}_2u_t + {}_2\nu_t) \\ v_t + ({}_1u_t + {}_1\nu_t) \cdot \Delta t \end{bmatrix} \quad (12)$$

where the elements of the two dimensional control input are acceleration, ${}_1u_t$, and steering angle, ${}_2u_t$. The two components of the process noise, ${}_1\nu_t$ and ${}_2\nu_t$, are both zero mean Gaussian noise affecting the realization of the controller's commanded acceleration and steering angle, respectively. The parameter l is the wheelbase of the vehicle. In these simulations, a time step, Δt , of 0.1 seconds is used.

The controller used in this simulation is a piecewise function that is composed of a different path following controller for each intent, I . A hand tuned finite state machine determines whether the vehicle should yield to the other vehicle or if it is clear to proceed. As shown in equations 3 and 4, the ego vehicle determines its control input based on the known poses of both vehicles, while the obstacle vehicle only has access to its own pose and a noisy estimate of the ego vehicle's pose.

B. Parameter Analysis

In this section, the affect of varying the parameter values in the CDF algorithm are analyzed. In particular, the look-ahead (how far into the future to forecast the system state) and the threshold (the collision probability at which evasive action should be triggered) are examined.

When analyzing the performance of the CDF algorithm, two key metrics were considered: the percentage of imminent collisions avoided and the number of false positive predictions of an imminent collision. Both of these metrics need to be considered when selecting values for the look-ahead and the

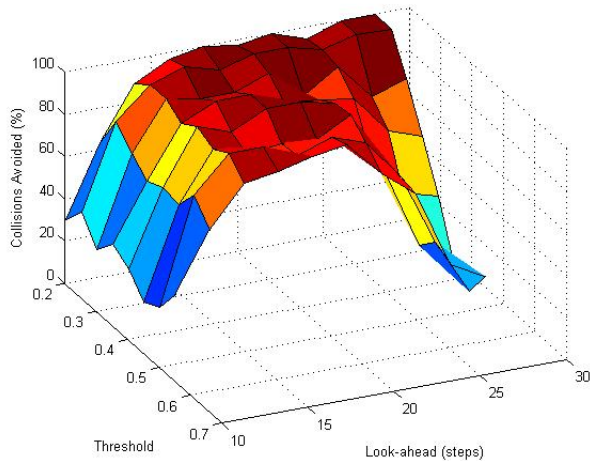


Fig. 7: How CDF performance varies with different values for the look-ahead and the threshold.

threshold because the parameters that result in good performance on one metric cause poor performance according to the other. The optimal selection of parameters will simultaneously minimize both the number of collisions and the number of false positive collision predictions.

In order to characterize the effects of different parameter values on the performance of the CDF algorithm, a series of simulations were run on a range of values for both the look-ahead and the threshold. For each $(T, Threshold)$ pair, 100 simulations were run on scenario one of the T-intersection navigation problem.

Figure 7 shows how the percentage of imminent collisions avoided varied with the parameter settings. Figure 8 shows how the number of false positive imminent collision predictions varied with different parameter values. As one would expect, lower values for the threshold correspond to a greater number of collisions avoided but also correspond to a greater number of false positive predictions. The influence of the look-ahead is less obvious, as poor performance occurs at look-ahead values that are both too small or too large. For small values of the look-ahead, there is not enough time to take evasive action to prevent the collision by the time it is recognized. On the other hand, for larger values of the look-ahead, the covariance of the future state distribution grows very large making it more difficult to recognize situations where collisions are imminent.

The selection of the parameter values can be seen as a 'cautiousness/aggressiveness' setting for the autonomous vehicle. For some parameter settings (i.e. low threshold values), the vehicle will behave very cautiously, avoiding 100% of imminent collisions but also frequently braking unnecessarily when collisions are falsely predicted. Alternatively, the settings can be tuned so that the vehicle will not brake until it is nearly certain the obstacle vehicle is going to cause a collision. In this study, the collision avoidance metric was weighted more heavily and parameter values were chosen that had a reasonably low number of false positive predictions. Ultimately the look-ahead was chosen to be 1.6 seconds (16 time steps) and the threshold was set to a collision probability of 0.35.

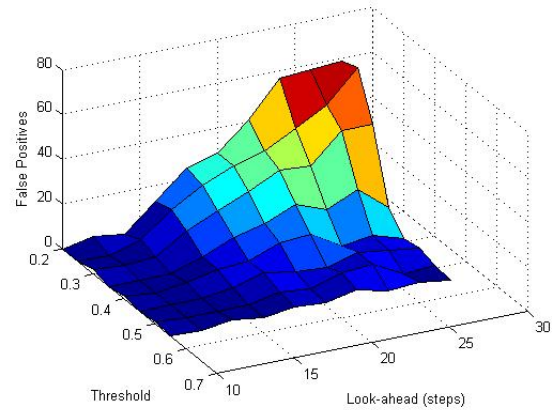


Fig. 8: Number of false positive collision predictions with different values for the look-ahead and the threshold.

The number of particles, M , used to approximate the current and future state distributions was chosen heuristically. As the number of particles increases, so does the accuracy of the approximation, but this improved accuracy comes at the cost of increased computational expense. Therefore, it is desirable to use as few particles as possible while still maintaining a sufficiently accurate posterior distribution. It was observed in this study that as few particles as 100 could be used to approximate the posterior distribution with acceptable convergence results.

C. Results and Analysis

Experiments were performed to compare the CDF to both a purely reactionary planner and a constant velocity planner. These two methods are commonly used in moving object tracking and collision avoidance. Reactive planners continually check to see if the planned path is still clear, if it is not, the robot will either stop or plan a new path avoiding the obstacle that is blocking the current path. Reactive planners are extremely simple but are only effective for slow moving robots using very frequent update rates. Constant velocity planners are the most common type of planner where the robot tracks moving objects' positions and velocities. Collisions are predicted by assuming the robot and the obstacle will maintain a constant velocity. The constant velocity trajectories are checked to see if a collision state will occur in the future. This technique often results in more intelligent trajectories than reactive planning, but also often results in frequent false positive collision predictions and performs poorly when the constant velocity assumption is violated.

Two experiments were performed, corresponding to the two scenarios presented in section IV-A. In both experiments, the CDF algorithm calculates the probability of a collision at a look-ahead of 1.6 seconds in the future, or 16 time steps ahead at simulation rate of 10 Hz. If the probability of a collision exceeds the threshold of 0.35, the ego vehicle then brakes at the maximum rate in an attempt to avoid the collision. The maximum rate of deceleration used in the experiments was 16 ft/s^2 , which is reasonable for a passenger vehicle traveling on a road surface with a moderate coefficient of friction.

TABLE I: Simulation results using the cognitive driving framework in scenario one.

Scenario	Prediction			Total
	Cutoff	Yield	Right	
Cutoff	385	1	18	404
Yield	2	312	15	329
Right	0	0	267	267
				1000

When testing the reactive planner, this braking maneuver was initiated when any part of the obstacle vehicle entered the ego vehicle's lane. When testing the constant velocity forecaster, the braking maneuver was initiated when the planner reported a collision state at a look-ahead of 1.6 seconds, the same as used with the CDF. The constant velocity trajectories were assumed to be deterministic. Each planner was run at a rate of 10 Hz.

The first experiment performed was scenario one as described in section IV-A. In this experiment, the ego vehicle is driving north through the intersection at a speed of about 30 miles per hour (48 km/h). The obstacle vehicle is stopped at the stop sign and randomly chooses to turn right, with probability 0.25, or left, with probability 0.75. Based on the obstacle vehicle's stochastic observations, it sometimes falsely believes the intersection is clear and turns in front of the ego vehicle, causing a collision to be imminent. A collision is considered 'imminent' if the system state is such that if the ego vehicle does not take preventative measures, a collision will result. The parameter β in equation 6 was set to 0.05. This selection of β corresponds to about a 40 percent chance of the obstacle vehicle observing the ego vehicle within the first second of simulation.

Table I details the simulation results for the first experiment using the CDF. The simulation was run for a total of 1000 episodes and the obstacle vehicle cutoff the ego vehicle a total of 404 times. In this context, 'cutoff' means that if the ego vehicle were to keep its speed constant and not take preventative measures, a collision would result. The results show that the CDF was able to recognize 385 out of 404 imminent collisions and only 2 benign situations were mistaken as cutoff situations.

For comparison, the simulation was then run using the reactive planner and the constant velocity planner, again for 1000 episodes each. Table II compares the performance of all three methods. The CDF was able to avoid 95% of the imminent collisions caused by the obstacle vehicle, while the reactive planner was only able to avoid 47% of the imminent collisions. The constant velocity planner was able to prevent almost all of the imminent collisions at 98%, but at the cost of a very high false positive rate of 43%. Here, false positive means that at some point, the planner believed a collision to be imminent and initiated the emergency braking maneuver when in fact a collision was not imminent. The CDF only had a false positive rate of 9.4% and since the CDF continuously updates its online estimate of the obstacle vehicle's intent, it quickly recognizes when the obstacle vehicle is actually turning right and aborts the braking maneuver.

TABLE II: Comparison of simulation results for scenario one.

	Collisions Imminent	Collisions Occurred	False Positives	Collisions Avoided
CDF	404	22	9.4 %	94.6%
Reactive	404	213	0%	47.3%
Velocity	390	7	43.3%	98.2%

The second experiment performed was scenario two as described in section IV-A. In this experiment, the ego vehicle is driving south through the intersection, again at a speed of about 30 miles per hour, while the obstacle vehicle is heading north toward the intersection. The obstacle vehicle randomly chooses to go straight, with probability 0.25, or left, with probability 0.75. As with the first experiment, if the obstacle vehicle's intent is to turn left, it may cutoff the ego vehicle depending on it's belief. This scenario is more difficult to recognize than scenario one since the obstacle vehicle is moving at a much higher speed.

The results for the CDF algorithm on this experiment are given in table III and the comparison between all three methods is given in table IV. It can be seen that the CDF only had one false negative classification, recognizing 196 out of 197 imminent collisions before they occurred, 94% of which were able to be avoided. The reactive planner performed very poorly, only avoiding 2.5% of imminent collisions. This low collision prevention rate is due to the fact that the reactive planner is unable to detect the dangerous situation with enough time to take evasive action. The constant velocity planner also performed poorly, only preventing 8.6% of collisions. The CDF did have a significant number of false positive classifications, though, due to recognizing that the obstacle vehicle's intent is to turn left but not recognizing that the obstacle vehicle is going to yield. Similarly to scenario one, this results in initiating the braking maneuver and then aborting once the ego vehicle recognizes that the obstacle vehicle is yielding.

After testing these three different planners on two different intersection navigation scenarios, it is clear that the CDF is better at optimizing the trade off between avoiding collisions and minimizing false positive classifications (overly cautious driving) for a variety of situations. The main advantage to the reactive planner is that it has a very low false positive rate (0% for the two scenarios tested here), but it performs very poorly at preventing collisions. Both of these facts are a result of the planner not recognizing a collision as imminent until it is nearly about to occur, so the planner is very confident that the collision is indeed imminent, but there is not enough time to prevent the collision at typical driving speeds. The constant velocity planner performed well on scenario one, with a collision avoidance rate 3.6% higher than the CDF, but had nearly 5 times as many false positive classifications. Furthermore, the constant velocity planner performed very poorly on scenario two, only preventing 8.6% of imminent collisions. This is caused by the planner's failure to anticipate the driver's turning action due to the constant velocity assumption.

The cognitive driving framework, on the other hand, was able to prevent about 94% of collisions in both scenarios. At the same time, the CDF had a fairly low false positive

TABLE III: Simulation results using the cognitive driving framework in scenario two.

Scenario	Prediction			Total
	Cutoff	Yield	Right	
Cutoff	196	0	1	197
Yield	273	279	5	557
Right	1	0	245	246
				1000

TABLE IV: Comparison of simulation results for scenario two.

	Collisions Imminent	Collisions Occurred	False Positives	Collisions Avoided
CDF	197	12	34.1%	93.9%
Reactive	200	195	0%	2.5%
Velocity	234	214	8.0%	8.6%

prediction rate with only 9.4% false positive predictions in scenario one and 34% in scenario two. As discussed in section IV-B, by adjusting the parameters in the CDF algorithm, the 'cautiousness' of the autonomous vehicle can be tuned to users' preferences. Additionally, in future work, a fuzzy logic controller can be used to determine when and how much the ego vehicle should brake in order to better optimize collision avoidance and user comfort.

V. CONCLUSION AND FUTURE WORK

This paper presented the cognitive driving framework, a method for joint inference of the intent and belief of an obstacle vehicle in an intersection navigation scenario. The goal of the CDF is to allow an autonomous vehicle to predict when a potentially hazardous situation is about to occur early enough to allow the autonomous vehicle to take evasive action to prevent a collision. The formulation of the problem as a dynamic Bayesian network was presented. A non-linear filtering method was proposed using a particle filter to estimate the posterior distribution of the state of the DBN. Monte Carlo simulation is used to estimate the future state distribution and calculate the probability of a collision. Finally, the accuracy of the estimation method was demonstrated by simulating two intersection navigation scenarios where an obstacle vehicle cuts off the autonomous vehicle. The simulation results show that the CDF is able to predict and prevent 94% of imminent collisions in two different intersection navigation scenarios. For comparison, the same simulations were run using a purely reactionary planner and a constant velocity planner. The results show that the reactive planner prevented only 47% and 3% of imminent collisions on the two scenarios. The constant velocity planner performed well on the first scenario, preventing 98% of collisions, but only prevented 9% of collisions in the second scenario.

This work has some natural extensions that should be explored. Adding additional vehicles to the intersection environment would lead to some interesting challenges that the CDF should be evaluated on, such as how the algorithm handles occlusions and how computation time scales with the

number of vehicles. The framework could be strengthened by relaxing the assumption that the vehicle poses are known. Additionally, the authors are implementing the cognitive driving framework on an autonomous vehicle platform in order to test the algorithm in an actual intersection navigation scenario.

REFERENCES

- [1] R. Bishop, "Intelligent vehicle applications worldwide," *Intelligent Systems and their Applications, IEEE*, vol. 15, no. 1, pp. 78–81, 2000.
- [2] T. C. Folsom, "Social ramifications of autonomous urban land vehicles," in *IEEE International Symposium on Technology and Society*, 2011.
- [3] Z. Juan, J. Wu, and M. McDonald, "Socio-economic impact assessment of intelligent transport systems," *Tsinghua Science & Technology*, vol. 11, no. 3, pp. 339–350, 2006.
- [4] J. Alonso-Mora, A. Breitenmoser, P. Beardsley, and R. Siegwart, "Reciprocal collision avoidance for multiple car-like robots," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 360–366.
- [5] B. Kluge and E. Prassler, "Recursive agent modeling with probabilistic velocity obstacles for mobile robot navigation among humans," in *Autonomous Navigation in Dynamic Environments*. Springer, 2007, pp. 121–134.
- [6] I. Miller, M. Campbell, D. Huttenlocher, F.-R. Kline, A. Nathan, S. Lupashin, J. Catlin, B. Schimpf, P. Moran, N. Zych *et al.*, "Team cornell's skynet: Robust perception and planning in an urban environment," *Journal of Field Robotics*, vol. 25, no. 8, pp. 493–527, 2008.
- [7] D. Vasquez, T. Fraichard, and C. Laugier, "Incremental learning of statistical motion patterns with growing hidden markov models," *Intelligent Transportation Systems, IEEE Transactions on*, vol. 10, no. 3, pp. 403–416, 2009.
- [8] R. Kelley, A. Tavakkoli, C. King, M. Nicolescu, M. Nicolescu, and G. Bebis, "Understanding human intentions via hidden markov models in autonomous mobile robots," in *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*. ACM, 2008, pp. 367–374.
- [9] T. Bandyopadhyay, C. Z. Jie, D. Hsu, M. H. Ang Jr, D. Rus, and E. Frazzoli, "Intention-aware pedestrian avoidance," in *Experimental Robotics*. Springer, 2013, pp. 963–977.
- [10] F. Havlak and M. Campbell, "Discrete and continuous, probabilistic anticipation for autonomous robots in urban environments," *Robotics, IEEE Transactions on*, vol. 30, no. 2, pp. 461–474, 2014.
- [11] D. Ellis, E. Sommerlade, and I. Reid, "Modelling pedestrian trajectory patterns with gaussian processes," in *Computer Vision Workshops (ICCV Workshops), 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 1229–1234.
- [12] C. L. Baker, R. R. Saxe, and J. B. Tenenbaum, "Bayesian theory of mind: Modeling joint belief-desire attribution," in *Proceedings of the thirty-second annual conference of the cognitive science society*, 2011, pp. 2469–2474.
- [13] U. Lerner, R. Parr, D. Koller, G. Biswas *et al.*, "Bayesian fault detection and diagnosis in dynamic systems," in *AAAI/IAAI*, 2000, pp. 531–537.
- [14] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, 2002.
- [15] G. Kitagawa, "Monte carlo filter and smoother for non-gaussian non-linear state space models," *Journal of computational and graphical statistics*, vol. 5, no. 1, pp. 1–25, 1996.

Denoising CT Images using wavelet transform

Lubna Gabralla

Faculty of Computer Science & Information Technology
Sudan University of Science & Technology
Khartoum, Sudan

Marwan Zaroug

Department of Computer Science,
College of science and arts of Baljurashi, Albaha University
Albaha, Kingdom of Saudi Arabia

Hela Mahersia

Department of Computer Science,
College of science and arts of Baljurashi, Albaha University
Albaha, Kingdom of Saudi Arabia

Abstract— Image denoising is one of the most significant tasks especially in medical image processing, where the original images are of poor quality due the noises and artifacts introduces by the acquisition systems. In this paper, we propose a new image denoising scheme by modifying the wavelet coefficients using soft-thresholding method, we present a comparative study of different wavelet denoising techniques for CT images and we discuss the obtained results. The denoising process rejects noise by thresholding in the wavelet domain. The performance is evaluated using Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE). Finally, Gaussian filter provides better PSNR and lower MSE values. Hence, we conclude that this filter is an efficient one for preprocessing medical images.

Keywords— Computed Tomography; Discrete wavelet transform; Lung cancer; Thresholding

I. INTRODUCTION

Despite the advances in oncological care, lung cancer remains the largest cause of death both worldwide and within the Kingdom of Saudi Arabia with an overall 5-year survival rate of only 15%. According to GLOBOCAN 2012 [7][8][23], lung cancer accounts 2 million deaths annually.

Recently, In KSA, the prevalence of lung cancer has increased significantly in the recent years; this is, mainly attributed to the increased incidence of smoking among men and students. The survival of patients is closely correlated to the stage of the detected lung cancer. Obviously, early detection of cancerous pulmonary nodules should improve a patient's chances for survival.

Computed tomography (CT) is the most commonly used diagnosis technique for detecting small pulmonary nodules because of its sensitivity and its ability to visualize a complete three-dimensional structure of the human thorax. It basically uses x-rays to obtain structural and functional information about the human body. An example of lung CT images is given in Figure 1.

However, the CT image quality is influenced by the radiation dose since it increases with the significant amount of radiation dose. Unfortunately, this increases the amount of x-rays being absorbed by the human body and increases the chances of cancer [15]. On the contrary, we need to reduce the radiation dose and this leads to noisy CT images. The presence of noise gives spotted images with blurred appearance [1].



Fig. 1: An example of lung CT image from ELCAP database

Thus, recovering an original image from noisy image remains a challenging problem that has received an increasing attention in recent years [15]. The recovering can be accomplished by image denoising, a process of estimating the original image from an image that has been contaminated by noise degradation [1]. Different methods have been proposed in literature for denoising lung CT images [18][23]. In [9], the authors proposed a low-pass Gaussian filter to improve the original CT images. This filter was also used in the work of Gurcan et al. [12], Lin and Yan [16], and Lin et al. [17]. Gaussian smoothing filters were employed by Pu et al. [20], Wei et al. [24], Gori et al. [10], and Retico et al. [21] to eliminate the image artifacts. In [14], Kim et al. used median filters to reduce noise. Also, 3D multi-scale filters were used in [22] to enhance lung nodules.

These traditional methods are restricted to the analysis of spatial interactions over relatively small neighborhoods on a single scale [19]. Nevertheless, other methods based on multi-resolution analyses and wavelet transforms, become more effective because of their capability to capture the signal details in different scales [15][19].

However, to our knowledge, only few works in literature considered the wavelet approach to denoise CT lung images [2][3]. In [2], the authors propose a fusion algorithm based on wavelet transform and canny operator to detect image edges, which may reduce the noise and obtain the continuous and distinct edges, whereas in [3], the authors combine Curvelet transformation with Monte-Carlo algorithm. Firstly, CT image's Curvelet decomposition is processed, then, Monte-Carlo algorithm is used to estimate high frequency coefficients.

In this paper, we propose an efficient noise reduction technique for CT images using wavelet-based thresholding. The proposed technique consists of two different stages of processing, wavelet transformation and thresholding.

This paper is organized as follows: In Section 2, a brief introduction to wavelet transform is given. The proposed denoising system is explained in Section 3. In Section 4, experimental results for various type of noise are discussed in detail. Finally, concluding remarks are given in Section 5.

II. DISCRETE WAVELET TRANSFORM

Wavelets are functions generated from one single function Ψ by dilations and translations. The basic idea of the wavelet transform is to represent any arbitrary function as a superposition of wavelets. Any such superposition decomposes the given function into different scale levels where each level is further decomposed with a resolution adapted to that level [19]. The translated and dilated wavelet functions derived from the mother wavelet Ψ are given by equation (1):

$$\Psi_{a,b} = \frac{1}{\sqrt{a}} \Psi\left(\frac{t-b}{a}\right) \text{ with } a \neq 0 \quad (1)$$

Where a is the scale coefficient and b is the translation coefficient. Thus, the wavelet transform of the signal $x(t)$ is given by:

$$C_{a,b} = \int_{-\infty}^{+\infty} x(t) \cdot \Psi_{a,b}(t) dt \quad (2)$$

Where the function $\Psi_{a,b}$ must be square integrable and must have compact support.

By applying DWT to an image, the image is decomposed into four sub-bands as shown in Fig. 1(a). The sub-bands labeled LH1, HL1 and HH1 represent the finest scale wavelet coefficients, also called detail images while the sub-band LL1 corresponds to coarse level coefficients, also called approximation image. To obtain the next coarse level of wavelet coefficients, the sub-band LL1 alone is further decomposed and sampled. This results in two level wavelet

decomposition as shown in Fig. 1(b). To obtain the next decomposition, LL2 will be used. This process continues until the final scale is reached.

The basic principle of denoising by wavelets was first proposed by E.-L Donoho in [4-6], where he thresholds the wavelet coefficients to zero if their values are below a certain threshold.

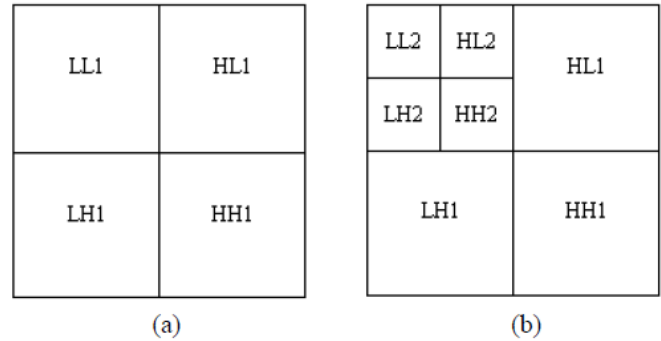


Fig. 2: the 2 level discrete wavelet decomposition for an image, from [19]

According to the noise model, a variety of threshold choosing methods can be mainly divided into four threshold selection rules [11] that are as follows:

A. Universal Thresholding

Universal threshold is the default method and yields the largest threshold. This type of global thresholding method was developed by [4] and the threshold value is given in equation (3) as :

$$T_{Donoho} = \hat{\sigma} \sqrt{2 \log n} \quad (3)$$

Where n is the sample size and $\hat{\sigma}$ is an estimate of the noise level σ .

B. Heursure Thresholding

Mixed rule is a mixture of the two previous rules: Rigrsure and universal threshold. First step calculates the variables A and B according to the system of Eq. (4)

$$\begin{cases} A = \frac{\sum_{i=1}^n |x_i|^2 - n}{n} \\ B = \sqrt{\frac{1}{n} \left[\frac{\log n}{\log 2} \right]^3} \end{cases} \quad (4)$$

If A is less than B the universal form threshold is as Eq. (3) is used, else threshold selection rule based on Rigrsure is adopted. A and B are defined by [11].

C. Minimax Thresholding

A fixed threshold selected to obtain minimum of maximum performance for mean square error against an ideal procedure.

The minimax principle is used in statistics in order to find a good estimator. The algorithm of the threshold selection is :

$$T = 0.3936 + 0.1829 \frac{\log n}{\log 2} \quad (5)$$

D. Hard and Soft thresholding

There are two thresholding methods common used the hard-threshold function which selects all wavelet coefficients that are greater than the given threshold T and sets the others to zero as shown in Eq. (6) below:

$$x_H = \begin{cases} 0 & \text{if } |x| \leq T \\ x & \text{if } |x| > T \end{cases} \quad (6)$$

The other popular alternative is the soft-threshold function [13] (also called the shrinkage function) which shrinks the wavelet coefficients by T towards zero. This type of threshold is defined by Eq. (7)

$$x_S = \begin{cases} 0 & \text{if } |x| \leq T \\ \text{sgn}(x)(|x| - T) & \text{if } |x| > T \end{cases} \quad (7)$$

In the next part, we calculate a new threshold function that we test for both soft and hard denoising algorithms.

III. PROPOSED DENOISING TECHNIQUE FOR CT IMAGES

We modify the thresh value given by equation (3) by adding a corrective term that takes into account the variation in the computed tomography images. The new threshold function T_{new} is calculated using the following equation:

$$T_{new} = \hat{\sigma} \sqrt{2 \log n - 8 \left(\frac{\sigma^2(x)}{\mu(x)} \right)} \quad (8)$$

Where n is the sample size, $\hat{\sigma}$ is an estimate of the noise level σ , σ^2 and μ are respectively, the variance of the and the mean of the input noisy image.

Then we apply this new thresh T_{new} in both equations (6) and (7).

The proposed denoising scheme is explained in Figure 3.

IV. RESULTS AND DISCUSSION

In our experiments, we have taken several gray scale images taken from ELCAP database, each of size 512 × 512. Noises tested in this work are the gaussian noise and the speckle noise. The noise levels are taken as 10, 20, 30 and 50. The wavelets tested in our experiments are: Daubechies, Haar, Symlet and Coifflet.

The results are compared with the universal threshold using both soft and hard threshold rules. The objective quality of the reconstructed image is measured by peak signal to noise ratio. The results are given in Table 1, Table 2, Figure 5. and Figure 6. It is evident from Table 1 and Table 2. that our proposed scheme outperforms the universal thresholding algorithm specially for lung CT images, for all values of noise levels considered in experiments. The robustness of the modified proposed thresh over the universal one proposed in [4] can be proved from Figure 4.

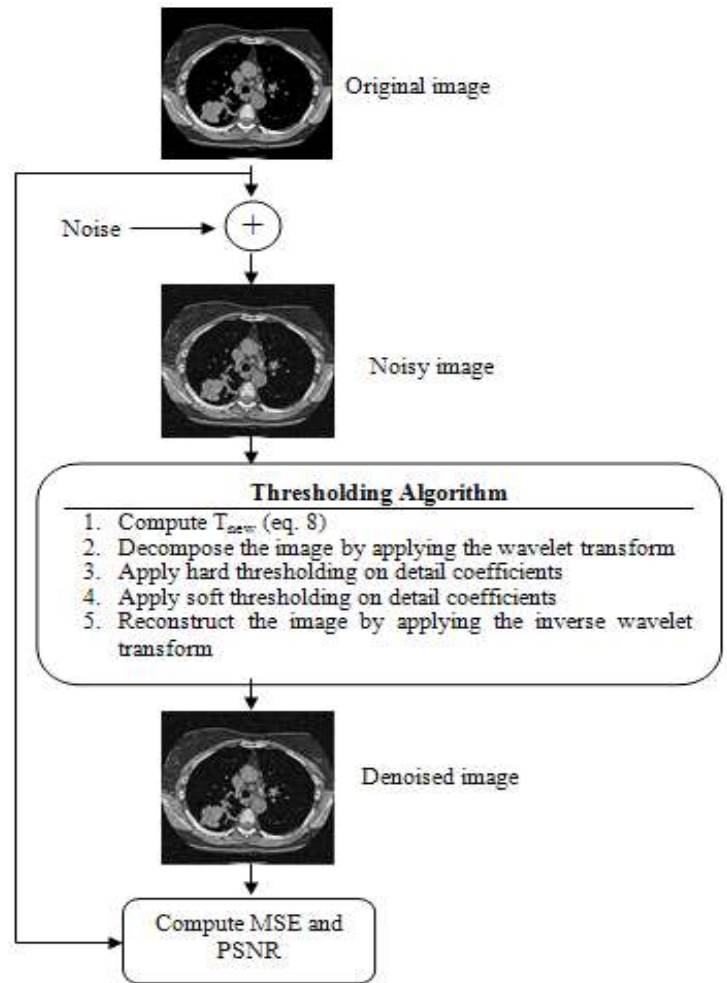


Fig.3 : Proposed denoising scheme

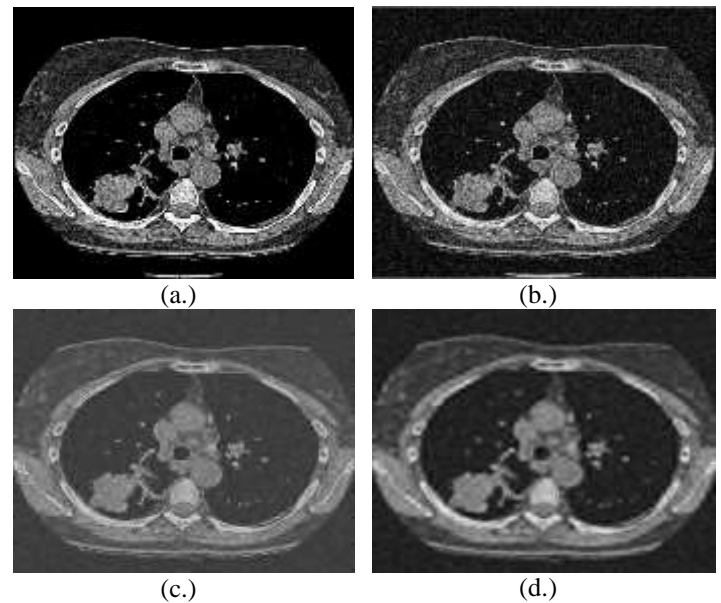


Fig. 4: (a) Original CT image (b) noisy image with $\sigma = 50$ (c) de-noised using hard thresholding with our proposed thresh, and (d) de-noised using soft thresholding with our proposed thresh

TABLE I. RESULTS OBTAINED FOR THE GAUSSIAN NOISE

Noise Levels	Wavelet	Wavelet level	Results (PSNR in dB)			
			Soft denoising		Hard denoising	
			T_{Donoho}	T_{new}	T_{Donoho}	T_{new}
10	Haar	2	21,43	23,32	24,86	27,28
	Haar	3	21,07	23,01	24,71	27,19
	Haar	4	20,93	22,89	24,68	27,17
	Db4	2	21,54	23,36	24,75	27,09
	Db4	3	21,12	23,00	24,55	26,96
	Db4	4	20,96	22,86	24,51	26,93
	Symlet 4	2	21,59	23,41	24,81	27,11
	Symlet 4	3	21,20	23,07	24,61	26,98
	Symlet 4	4	21,05	22,93	24,57	26,95
	Coiflet 4	2	21,60	23,41	24,82	27,12
	Coiflet 4	3	21,19	23,06	24,62	26,97
	Coiflet 4	4	21,04	22,92	24,56	26,94
20	Haar	2	18,47	19,68	20,17	21,84
	Haar	3	17,99	19,30	19,95	21,71
	Haar	4	17,77	19,12	19,90	21,67
	Db4	2	18,73	19,85	20,25	21,89
	Db4	3	18,18	19,41	19,98	21,71
	Db4	4	17,93	19,21	19,90	21,66
	Symlet 4	2	18,74	19,89	20,30	21,95
	Symlet 4	3	18,23	19,48	20,07	21,79
	Symlet 4	4	17,99	19,29	19,99	21,74
	Coiflet 4	2	18,77	19,91	20,33	21,96
	Coiflet 4	3	18,23	19,47	20,07	21,78
	Coiflet 4	4	17,99	19,29	19,98	21,73
30	Haar	2	17,24	18,03	18,23	19,40
	Haar	3	16,68	17,59	17,95	19,22
	Haar	4	16,38	17,37	17,85	19,18
	Db4	2	17,58	18,28	18,38	19,51
	Db4	3	16,93	17,78	18,03	19,29
	Db4	4	16,59	17,53	17,90	19,21
	Symlet 4	2	17,58	18,29	18,42	19,54
	Symlet 4	3	16,97	17,83	18,12	19,37
	Symlet 4	4	16,65	17,60	18,00	19,30
	Coiflet 4	2	17,61	18,32	18,41	19,56
	Coiflet 4	3	16,98	17,84	18,07	19,36
	Coiflet 4	4	16,66	17,60	17,95	19,29
50	Haar	2	16,03	16,32	16,28	16,93
	Haar	3	15,39	15,84	15,89	16,73
	Haar	4	14,95	15,52	15,71	16,63
	Db4	2	16,39	16,65	16,62	17,13
	Db4	3	15,70	16,10	16,16	16,86
	Db4	4	15,24	15,76	15,95	16,74
	Symlet 4	2	16,39	16,66	16,62	17,13
	Symlet 4	3	15,71	16,14	16,21	16,91
	Symlet 4	4	15,27	15,81	16,02	16,80
	Coiflet 4	2	16,42	16,69	16,65	17,15
	Coiflet 4	3	15,75	16,16	16,22	16,91
	Coiflet 4	4	15,32	15,83	16,01	16,78

TABLE II. RESULTS OBTAINED FOR THE SPECKLE NOISE

Noise Levels	Wavelet	Wavelet level	Results (PSNR in dB)			
			Soft denoising		Hard denoising	
			T_{Donoho}	T_{new}	T_{Donoho}	T_{new}
10	Haar	2	21,60	23,60	25,42	28,54
	Haar	3	21,23	23,27	25,23	28,39
	Haar	4	21,10	23,16	25,19	28,35
	Db4	2	21,71	23,63	25,28	28,25
	Db4	3	21,28	23,25	25,05	28,04
	Db4	4	21,12	23,12	24,99	27,99
	Symlet 4	2	21,76	23,68	25,35	28,27
	Symlet 4	3	21,35	23,31	25,10	28,06
	Symlet 4	4	21,21	23,19	25,05	28,01
	Coiflet 4	2	21,77	23,69	25,33	28,25
	Coiflet 4	3	21,35	23,31	25,09	28,04
	Coiflet 4	4	21,20	23,18	25,02	27,99
20	Haar	2	18,73	20,02	20,63	22,73
	Haar	3	18,22	19,59	20,35	22,52
	Haar	4	18,01	19,43	20,28	22,47
	Db4	2	18,99	20,20	20,70	22,73
	Db4	3	18,39	19,70	20,36	22,47
	Db4	4	18,15	19,51	20,27	22,40
	Symlet 4	2	19,01	20,24	20,77	22,82
	Symlet 4	3	18,46	19,78	20,47	22,56
	Symlet 4	4	18,23	19,60	20,38	22,50
	Coiflet 4	2	19,04	20,26	20,77	22,82
	Coiflet 4	3	18,46	19,78	20,43	22,55
	Coiflet 4	4	18,23	19,59	20,34	22,48
30	Haar	2	17,69	18,55	18,81	20,30
	Haar	3	17,05	18,03	18,42	20,02
	Haar	4	16,75	17,81	18,32	19,95
	Db4	2	18,04	18,82	18,99	20,39
	Db4	3	17,29	18,21	18,53	20,05
	Db4	4	16,96	17,96	18,39	19,96
	Symlet 4	2	18,04	18,84	19,01	20,47
	Symlet 4	3	17,34	18,28	18,61	20,17
	Symlet 4	4	17,02	18,04	18,48	20,07
	Coiflet 4	2	18,08	18,88	19,02	20,48
	Coiflet 4	3	17,35	18,28	18,59	20,15
	Coiflet 4	4	17,03	18,04	18,46	20,05
50	Haar	2	16,95	17,34	17,29	18,20
	Haar	3	16,09	16,65	16,72	17,79
	Haar	4	15,62	16,31	16,50	17,66
	Db4	2	17,40	17,73	17,72	18,44
	Db4	3	16,43	16,92	17,00	17,94
	Db4	4	15,91	16,54	16,75	17,77
	Symlet 4	2	17,40	17,73	17,74	18,43
	Symlet 4	3	16,45	16,97	17,08	18,00
	Symlet 4	4	15,96	16,60	16,85	17,84
	Coiflet 4	2	17,44	17,77	17,79	18,47
	Coiflet 4	3	16,48	16,98	17,10	17,98
	Coiflet 4	4	16,00	16,62	16,83	17,82

PSNR of different denoising methods for a gaussian noise

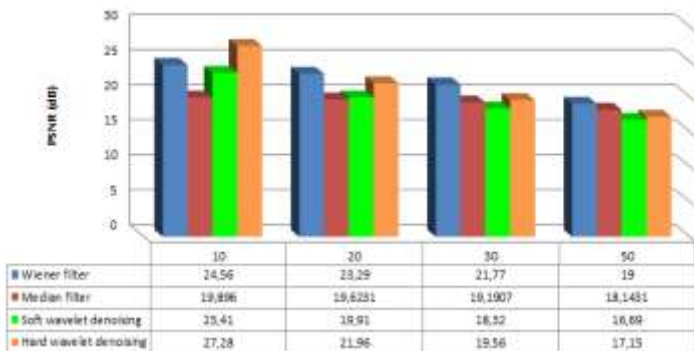


Fig. 5: Different denoising scheme for a Gaussian noise

PSNR of different denoising methods for a speckle noise

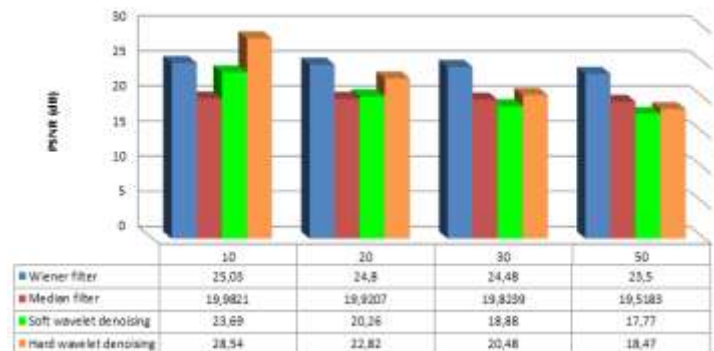


Fig. 6: Different denoising scheme for a Speckle noise

Table 3 gives an idea about the results obtained using wavelets for denoising CT lung images. The applied noises are Gaussian taken with two levels, 10 and 20. We can see that our results, in term of PSNR, are similar to those given by Bhadauria and Singh [2].

TABLE III. DIFFERENT RESULTS PRESENTED IN LITERATURE

	PSNR (db)		Used database
	$\sigma = 10$	$\sigma = 20$	
Results given in [2]	29.36	25.56	LIDC
Our results	27.12	21.96	ELCAP

V. CONCLUSION

In this paper, we have proposed a new denoising scheme that removes the noise significantly and performs the universal denoising scheme in terms of PSNR for all values of noise level. The comparative result shows that the proposed threshold value found to be better than universal threshold. The comparative PSNR value of the proposed threshold improves with increase in the noise level. In this sense our threshold value is an important contribution to the choice of the threshold to remove the noise from the image using wavelets.

ACKNOWLEDGMENT

We would like to thank Al BAHA University for supporting this work and providing the research funds.

REFERENCES

- [1] Ali, S.-A., S. Vathsal, and L. Lal kishore, An Efficient Denoising Technique for CT Images using Windowbased Multi-Wavelet Transformation and Thresholding, *European Journal of Scientific Research*, Vol. 48, No. 2, pp. 315-325, 2010.
- [2] Bhadauria, H.- S. and A. Singh, Wavelet and Canny Based Edge Detection Method for Noisy Lung CT Image, *International Journal of Emerging Technology and Advanced Engineering*, Vol. 3, No. 5, pp. 776-780, 2013.
- [3] Deng, J. L. Haiyun, and Hao Wu, A CT Image Denoise Method Using Curvelet Transform, *Communication Systems and Information Technology*, Vol. 100, pp. 681-687, 2011.
- [4] Donoho D.-L. and I.-M. Johnstone, Ideal Spatial Adaptation via Wavelet Shrinkage, *Biometrika*, Vol. 81, No. 3, pp. 425-455, 1994.
- [5] Donoho D.-L., De-Noising by Soft-Thresholding, *IEEE trans. On Information Theory*, Vol. 41, No. 3, pp. 613-627, 1995.
- [6] Donoho D.-L. and I.-M. Johnstone, Adapting to Unknown Smoothness via Wavelet Shrinkage," *Journal of American Statistical Association*, Vol. 90, No. 432, pp. 1200-1224, 1995.
- [7] Dubey, S. and C.-A. Powell, Update in lung cancer 2008, *Am. J. Respir. Crit. Care Med.*, Vol. 179, No. 10, pp. 860-868, 2009.

- [8] Ferlay, J., I. Soerjomataram, M. Ervik, R. Dikshit, S. Eser, C. Mathers, M. Rebelo, D.M. Parkin, D., Forman and F. Bray, Cancer Incidence and Mortality Worldwide in 2012, Report of the International Agency for Research on Cancer, [http:// globocan.iarc.fr](http://globocan.iarc.fr).
- [9] Garnavi, R., A. Baraani-Dastjerdi, H. Abrishami Moghaddam, M. Giti and A.-A. Rad, new segmentation method for lung HRCT images, In Proc. of the International Conference on Digital Image Computing: Techniques and Applications, DICTA '05, pp. 52, 2005.
- [10] Gori, I., R. Bellotti and P. Cerello, Lung nodule detection in screening computed tomography, In Proc. of the IEEE Nuclear Science Symposium Conference Record, pp. 3489 – 3491, San Diego, 2006.
- [11] Guo D.-f., W.-H. Zhu, Z.-M. Gao, and J.-Q. Zhang, A study of wavelet thresholding denoising, In Proc. of the 5th International Conference on Signal Processing WCCC-ICSP'2000, pp. 329-332, 2000.
- [12] Gurcan, M.N., B. Sahiner, and N. Petrick, Lung nodule detection on thoracic computed tomography images: preliminary evaluation of a computer-aided diagnosis system. *Med. Phys.* Vol. 29, pp. 2552-2558, 2002.
- [13] Joy J., S. Peter, and N. John, Denoising Using Soft Thresholding, *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, Vol. 2, pp. 1027-1031, 2013.
- [14] Kim, H., T. Nakashima, and Y. Itai, Automatic detection of ground glass opacity from the thoracic MDCT images by using density features, In Proc. of the International Conference on Control, Automation and Systems, pp. 1274-1277, Seoul, 2007.
- [15] Lee S.-L.-A., A.-Z. Kouzani, and E.-J. Hu, Automated detection of lung nodules in computed tomography images: a review, *Machine Vision and Applications*, Vol. 23, pp. 151-163, 2012.
- [16] Lin, D.-T. and C.-R Yan, Lung nodules identification rules extraction with neural fuzzy network, In proc of the 9th International Conference of Information Processing (ICONIP), Vol. 4, pp. 2049 – 2053, Singapore, 2002.
- [17] Lin, D.-T., C.-R. Yan, and W.-T. Chen, Autonomous detection of pulmonary nodules on CT images with a neural network-based fuzzy system, *Comput. Med. Imaging Graph.* Vol. 29, pp. 447-458, 2005.
- [18] Mahersia, H., L. Gabralla and M. Zaroug, Lung Cancer Detection on CT scan images: A Review on the analysis Techniques, *International Journal of Advanced Research in Artificial Intelligence*, Vol. 4, No. 4, pp. 38-45, April 2015.
- [19] Mahersia, H., K. Hamrouni, and N. Ellouze, Wavelet texture analysis and fuzzy C-mean classification for image segmentation, In Proc. of the 9th Maghrebien Conference on Information Technologie MCSEAI06, Agadir, Maroc, 2006.
- [20] Pu, J., J. Roos, and C.-A. Yi, Adaptive border marching algorithm: automatic lung segmentation on chest CT images, *Comput. Med. Imaging Graph.*, Vol. 32, pp. 452-462, 2008.
- [21] Retico, A., P. Delogu, and M.-E. Fantacci, Lung nodule detection in low-dose and thin-slice computed tomography, *Comput. Biol. Med.*, Vol. 38, pp. 525-534, 2008.
- [22] Senthil-Kumar T.-K. and E.-N. Ganesh, Proposed Technique for Accurate Detection/Segmentation of Lung Nodules using Spline Wavelet Techniques, *Int J Biomed Sci.*, Vol. 9, No. 1, pp. 9-17, 2013.
- [23] Vijaya G. and A. Suhasini, An Adaptive Preprocessing of Lung CT Images with Various Filters for Better Enhancement, *Academic Journal of Cancer Research*, Vol. 7, No. 3, pp. 179-184, 2014.
- [24] Wei, G.-Q., L. Fan and J. Qian, Automatic detection of nodules attached to vessels in lung CT by volume projection analysis, *Medical Image Computing and Computer-assisted Intervention, MICCAI*, Vol. 2488, pp. 746-752, 2002.