

Ajith Abraham
Jaime Lloret Mauri
John F. Buford
Junichi Suzuki
Sabu M. Thampi (Eds.)

Communications in Computer and Information Science

192

Advances in Computing and Communications

First International Conference, ACC 2011
Kochi, India, July 2011
Proceedings, Part III

Part 3

 Springer

Ajith Abraham Jaime Lloret Mauri
John F. Buford Junichi Suzuki
Sabu M. Thampi (Eds.)

Advances in Computing and Communications

First International Conference, ACC 2011
Kochi, India, July 22-24, 2011
Proceedings, Part III

Volume Editors

Ajith Abraham
Machine Intelligence Research Labs (MIR Labs)
Auburn, WA, USA
E-mail: ajith.abraham@ieee.org

Jaime Lloret Mauri
Polytechnic University of Valencia
Valencia, Spain
E-mail: jlloret@dcom.upv.es

John F. Buford
Avaya Labs Research
Basking Ridge, NJ, USA
E-mail: john.buford@gmail.com

Junichi Suzuki
University of Massachusetts
Boston, MA, USA
E-mail: jxs@acm.org

Sabu M. Thampi
Rajagiri School of Engineering and Technology
Kochi, India
E-mail: smthampi@acm.org

ISSN 1865-0929
ISBN 978-3-642-22719-6
DOI 10.1007/978-3-642-22720-2
Springer Heidelberg Dordrecht London New York

e-ISSN 1865-0937
e-ISBN 978-3-642-22720-2

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.2, I.4, I.5, C.2, F.1, H.3

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The First International Conference on Advances in Computing and Communications (ACC 2011) was held in Kochi during July 22–24, 2011. ACC 2011 was organized by Rajagiri School of Engineering & Technology (RSET) in association with the Association of Computing Machinery (ACM)- SIGWEB, Machine Intelligence Research Labs (MIR Labs), International Society for Computers and Their Applications, Inc. (ISCA), All India Council for Technical Education (AICTE), Indira Gandhi National Open University (IGNOU), Kerala State Council for Science, Technology and Environment (KSCSTE), Computer Society of India (CSI)- Div IV and Cochin Chapter, The Institution of Electronics and Telecommunication Engineers (IETE), The Institution of Engineers (India) and Project Management Institute (PMI), Trivandrum, Kerala Chapter. Established in 2001, RSET is a premier professional institution striving for holistic excellence in education to mould young, vibrant engineers.

ACC 2011 was a three-day conference which provided an opportunity to bring together students, researchers and practitioners from both academia and industry. ACC 2011 was focused on advances in computing and communications and it attracted many local and international delegates, presenting a balanced mixture of intellects from the East and from the West. ACC 2011 received 592 research papers from 38 countries including Albania, Algeria, Bangladesh, Brazil, Canada, Colombia, Cyprus, Czech Republic, Denmark, Ecuador, Egypt, France, Germany, India, Indonesia, Iran, Ireland, Italy, Korea, Kuwait, Malaysia, Morocco, New Zealand, P.R. China, Pakistan, Rwanda, Saudi Arabia, Singapore, South Africa, Spain, Sri Lanka, Sweden, Taiwan, The Netherlands, Tunisia, UK, and USA. This clearly reflects the truly international stature of ACC 2011. All papers were rigorously reviewed internationally by an expert technical review committee comprising more than 300 members. The conference had a peer-reviewed program of technical sessions, workshops, tutorials, and demonstration sessions.

There were several people that deserve appreciation and gratitude for helping in the realization of this conference. We would like to thank the Program Committee members and additional reviewers for their hard work in reviewing papers carefully and rigorously. After careful discussions, the Program Committee selected 234 papers (acceptance rate: 39.53%) for presentation at the conference. We would also like to thank the authors for having revised their papers to address the comments and suggestions by the referees.

The conference program was enriched by the outstanding invited talks by Ajith Abraham, Subir Saha, Narayan C. Debnath, Abhijit Mitra, K. Chandra Sekaran, K. Subramanian, Sudip Misra, K.R. Srivathsan, Jaydip Sen, Joyati Debnath and Junichi Suzuki. We believe that ACC 2011 delivered a high-quality, stimulating and enlightening technical program. The tutorials covered topics of

great interest to the cyber forensics and cloud computing communities. The tutorial by Avinash Srinivasan provided an overview of the forensically important artifacts left behind on a MAC computer. In his tutorial on “Network Forensics,” Bhadran provided an introduction to network forensics, packet capture and analysis techniques, and a discussion on various RNA tools. The tutorial on Next-Generation Cloud Computing by Pethuru Raj focused on enabling technologies in cloud computing.

The ACC 2011 conference program also included five workshops: International Workshop on Multimedia Streaming (MultiStreams 2011), Second International Workshop on Trust Management in P2P Systems (IWTMP2PS 2011), International Workshop on Cloud Computing: Architecture, Algorithms and Applications (CloudComp 2011), International Workshop on Identity: Security, Management and Applications (ID2011) and International Workshop on Applications of Signal Processing (I-WASP 2011). We thank all the workshop organizers as well as the Workshop Chair, El-Sayed El-Alfy, for their accomplishment to bring out prosperous workshops. We would like to express our gratitude to the Tutorial Chairs Patrick Seeling, Jaydeep Sen, K.S. Mathew, and Roksana Boreli and Demo Chairs Amitava Mukherjee, Bhadran V.K., and Janardhanan P.S. for their timely expertise in reviewing the proposals. Moreover, we thank Publication Chairs Pruet Boonma, Sajid Hussain and Hiroshi Wada for their kind help in editing the proceedings. The large participation in ACC2011 would not have been possible without the Publicity Co-chairs Victor Govindaswamy, Arun Saha and Biju Paul.

The proceedings of ACC 2011 are organized into four volumes. We hope that you will find these proceedings to be a valuable resource in your professional, research, and educational activities whether you are a student, academic, researcher, or a practicing professional.

July 2011

Ajith Abraham
Jaime Lloret Mauri
John F. Buford
Junichi Suzuki
Sabu M. Thampi

Organization

ACC 2011 was jointly organized by the Department of Computer Science and Engineering and Department of Information Technology, Rajagiri School of Engineering and Technology (RSET), Kochi, India, in cooperation with ACM/SIGWEB.

Organizing Committee

Chief Patrons

Fr. Jose Alex CMI	Manager, RSET
Fr. Antony Kariyil CMI	Director, RSET

Patron

J. Isaac, Principal	RSET
---------------------	------

Advisory Committee

A. Krishna Menon	RSET
A.C. Mathai	RSET
Fr. Varghese Panthalookaran	RSET
Karthikeyan Chittayil	RSET
Vinod Kumar, P.B.	RSET
Biju Abraham	
Narayamparambil	RSET
Kuttyamma A.J.	RSET
Asha Panicker	RSET
K. Rajendra Varmah	RSET
P.R. Madhava Panicker	RSET
Liza Annie Joseph	RSET
Varkey Philip	RSET
Fr. Joel George Pullolil	RSET
R. Ajayakumar Varma	KSCSTE
K. Poullose Jacob	Cochin University of Science & Technology
H.R. Mohan, Chairman	Div IV, Computer Society of India (CSI)
Soman S.P., Chairman	Computer Society of India (CSI), Cochin Chapter
S. Radhakrishnan, Chairman	Kerala State Centre, The Institution of Engineers (India)

Steering Committee

John F. Buford	Avaya Labs Research, USA
Rajkumar Buyya	University of Melbourne, Australia
Mukesh Singhai	University of Kentucky, USA
John Strassner	Pohang University of Science and Technology, Republic of Korea
Junichi Suzuki	University of Massachusetts, Boston, USA
Ramakrishna Kappagantu	IEEE India Council
Achuthsankar S. Nair	Centre for Bioinformatics, Trivandrum, India

Conference Chair

Sabu M. Thampi	Rajagiri School of Engineering and Technology, India
----------------	---

ACC 2011 Program Committee Chairs

General Co-chairs

Ajith Abraham	Machine Intelligence Research Labs, Europe
Chandra Sekaran K.	National Institute of Technology Karnataka, India
Waleed W. Smari	University of Dayton, Ohio, USA

Program Co-chairs

Jaime Lloret Mauri	Polytechnic University of Valencia, Spain
Thorsten Strufe	Darmstadt University of Technology, Germany
Gregorio Martinez	University of Murcia, Spain

Special Sessions and Workshops Co-chairs

El-Sayed El-Alfy	King Fahd University of Petroleum and Minerals, Saudi Arabia
Silvio Bortoleto	Positivo University, Brazil

Tutorial Co-chairs

Patrick Seeling	University of Wisconsin - Stevens Point, USA
Jaydeep Sen	Tata Consultancy Services, Calcutta, India
K.S. Mathew	Rajagiri School of Engineering and Technology, India
Roksana Boreli	National ICT Australia Ltd., Australia

Demo Co-chairs

Amitava Mukherjee
Bhadran V.K.

IBM Global Business Services, India
Centre for Development of Advanced
Computing, Trivandrum, India

Janardhanan P.S.

Rajagiri School of Engineering and Technology,
India

Publicity Co-chairs

Victor Govindaswamy
Arun Saha
Biju Paul

Texas A&M University, USA
Fujitsu Network Communications, USA
Rajagiri School of Engineering and Technology,
India

Publication Co-chairs

Pruet Boonma
Sajid Hussain
Hiroshi Wada

Chiang Mai University, Thailand
Fisk University, USA
University of New South Wales, Australia

ACC 2011 Technical Program Committee

A. Hafid

Network Research Lab, University of Montreal,
Canada

Abdallah Shami

The University of Western Ontario, Canada

Abdelhafid Abouaissa

University of Haute Alsace, France

Abdelmalik Bachir

Imperial College London, UK

Abdelouahid Derhab

CERIST, Algeria

Abhijit Mitra

Indian Institute of Technology Guwahati, India

Adão Silva

University of Aveiro, Portugal

Adel Ali

University Technology Malaysia

Ahmed Mehaoua

University of Paris Descartes, France

Ai-Chun Pang

National Taiwan University, Taiwan

Ajay Gupta

Western Michigan University, USA

Alberto Dainotti

University of Naples "Federico II", Italy

Alessandro Leonardi

University of Catania, Italy

Alex Galis

University College London, UK

Alexey Vinel

Saint Petersburg Institute, Russia

Ali Abedi

University of Maine, USA

Alicia Triviño Cabrera

Universidad de Málaga, Spain

Alireza Behbahani

University of California, Irvine, USA

Alois Ferscha

University of Linz, Austria

Al-Sakib Khan Pathan

International Islamic University, Malaysia

Amar Prakash Azad

INRIA, France

Amirhossein Alimohammad

University of Alberta, Canada

Amit Agarwal

Indian Institute of Technology, Roorkee, India

Amitava Mukherjee	IBM Global Business Services, India
Anand Prasad	NEC Corporation, Japan
Andreas Maeder	NEC Laboratories Europe, Germany
Ankur Gupta	Model Institute of Engineering and Technology, India
Antonio Coronato	ICAR-CNR, Naples, Italy
Antonio Pescapé	University of Naples Federico II, Italy
António Rodrigues	IT / Instituto Superior Técnico, Portugal
Anura P. Jayasumana	Colorado State University, USA
Arnab Bhattacharya	Indian Institute of Technology, Kanpur, India
Arun Saha	Fujitsu Network Communications, USA
Arvind Swaminathan	Qualcomm, USA
Ashley Thomas	Secureworks Inc., USA
Ashraf Elnagar	Sharjah University, UAE
Ashraf Mahmoud	KFUPM, Saudi Arabia
Ashwani Singh	Navtel Systems, France
Athanasios Vasilakos	University of Western Macedonia, Greece
Atilio Gameiro	Telecommunications Institute/Aveiro University, Portugal
Aydin Sezgin	Ulm University, Germany
Ayman Assra	McGill University, Canada
Aytac Azgin	Georgia Institute of Technology, USA
B. Sundar Rajan	Indian Institute of Science, India
Babu A.V.	National Institute of Technology, Calicut, India
Babu B.V.	BITS-Pilani, Rajasthan, India
Babu Raj E.	Sun College of Engineering and Technology, India
Balagangadhar G. Bathula	Columbia University, USA
Borhanuddin Mohd. Ali	Universiti Putra Malaysia
Brijendra Kumar Joshi	Military College, Indore, India
Bruno Crispo	Università di Trento, Italy
C.-F. Cheng	National Chiao Tung University, Taiwan
Chang Wu Yu	Chung Hua University, Taiwan
Charalampos Tsimenidis	Newcastle University, UK
Chih-Cheng Tseng	National Ilan University, Taiwan
Chi-Hsiang Yeh	Queen's University, Canada
Chitra Babu	SSN College of Engineering, Chennai, India
Chittaranjan Hota	BITS Hyderabad Campus, India
Chonho Lee	Nanyang Technological University, Singapore
Christian Callegari	University of Pisa, Italy
Christos Chrysoulas	Technological Educational Institute, Greece
Chuan-Ching Sue	National Cheng Kung University, Taiwan
Chung Shue Chen	TREC, INRIA, France

Chun-I. Fan	National Sun Yat-sen University, Taiwan
Chutima Prommak	Suranaree University of Technology, Thailand
Dali Wei	Jiangsu Tianze Infoindustry Company Ltd, P.R. China
Danda B. Rawat	Old Dominion University, USA
Daniele Tarchi	University of Bologna, Italy
Davide Adami	CNIT Pisa Research Unit, University of Pisa, Italy
Deepak Garg	Thapar University, India
Demin Wang	Microsoft Inc., USA
Dennis Pfisterer	University of Lübeck, Germany
Deyun Gao	Beijing Jiaotong University, P.R. China
Dharma Agrawal	University of Cincinnati, USA
Dhiman Barman	Juniper Networks, USA
Di Jin	General Motors, USA
Dimitrios Katsaros	University of Thessaly, Greece
Dimitrios Vergados	National Technical University of Athens, Greece
Dirk Pesch	Cork Institute of Technology, Ireland
Djamel Sadok	Federal University of Pernambuco, Brazil
Eduardo Cerqueira	Federal University of Para (UFPA), Brazil
Eduardo Souto	Federal University of Amazonas, Brazil
Edward Au	Huawei Technologies, P.R. China
Egemen Cetinkaya	University of Kansas, USA
Elizabeth Sherly	IITM-Kerala, India
El-Sayed El-Alfy	King Fahd University, Saudi Arabia
Emad A. Felemban	Umm Al Qura University, Saudi Arabia
Eric Renault	TELECOM & Management SudParis, France
Errol Lloyd	University of Delaware, USA
Ertan Onur	Delft University of Technology, The Netherlands
Faouzi Bader	CTTC, Spain
Faouzi Kamoun	WTS, UAE
Fernando Velez	University of Beira Interior, Portugal
Filipe Cardoso	ESTSetubal/Polytechnic Institute of Setubal, Portugal
Florian Doetzer	ASKON ConsultingGroup, Germany
Francesco Quaglia	Sapienza Università di Roma, Italy
Francine Krief	University of Bordeaux, France
Frank Yeong-Sung Lin	National Taiwan University, Taiwan
Gianluigi Ferrari	University of Parma, Italy
Giuseppe Ruggeri	University "Mediterranea" of Reggio Calabria, Italy
Grzegorz Danilewicz	Poznan University of Technology, Poland
Guang-Hua Yang	The University of Hong Kong, Hong Kong
Guo Bin	Institut Telecom SudParis, France

Hadi Otrok	Khalifa University, UAE
Hamid Mcheick	Université du Québec à Chicoutimi, Canada
Harry Skianis	University of the Aegean, Greece
Hicham Khalife	ENSEIRB-LaBRI, France
Himal Suraweera	Singapore University of Technology and Design, Singapore
Hiroshi Wada	University of New South Wales, Australia
Hong-Hsu Yen	Shih-Hsin University, Taiwan
Hongli Xu	University of Science and Technology of China, P.R. China
Houcine Hassan	Technical University of Valencia, Spain
Hsuan-Jung Su	National Taiwan University, Taiwan
Huaiyu Dai	NC State University, USA
Huey-Ing Liu	Fu-Jen Catholic University, Taiwan
Hung-Keng Pung	National University of Singapore
Hung-Yu Wei	NTU, Taiwan
Ian Glover	University of Strathclyde, UK
Ian Wells	Swansea Metropolitan University, UK
Ibrahim Develi	Erciyes University, Turkey
Ibrahim El rube	AAST, Egypt
Ibrahim Habib	City University of New York, USA
Ibrahim Korpeoglu	Bilkent University, Turkey
Ilja Radusch	Technische Universität Berlin, Germany
Ilka Miloucheva	Media Technology Research, Germany
Imad Elhajj	American University of Beirut, Lebanon
Ivan Ganchev	University of Limerick, Ireland
Iwan Adhicandra	The University of Pisa, Italy
Jalel Ben-othman	University of Versailles, France
Jane-Hwa Huang	National Chi Nan University, Taiwan
Jaydeep Sen	Tata Consultancy Services, Calcutta, India
Jiankun Hu	RMIT University, Australia
Jie Yang	Cisco Systems, USA
Jiping Xiong	Zhejiang Normal University of China
José de Souza	Federal University of Ceará, Brazil
Jose Moreira	IBM T.J. Watson Research Center, USA
Ju Wang	Virginia State University, USA
Juan-Carlos Cano	Technical University of Valencia, Spain
Judith Kelner	Federal University of Pernambuco, Brazil
Julien Laganier	Juniper Networks Inc., USA
Jussi Haapola	University of Oulu, Finland
K. Komathy	Easwari Engineering College, Chennai, India
Ka Lok Hung	The Hong Kong University, Hong Kong
Ka Lok Man	Xi'an Jiaotong-Liverpool University, China
Kaddar Lamia	University of Versailles Saint Quentin, France
Kainam Thomas	Hong Kong Polytechnic University

Kais Mnif	High Institute of Electronics and Communications of Sfax, Tunisia
Kang Yong Lee	ETRI, Korea
Katia Bortoleto	Positivo University, Brazil
Kejie Lu	University of Puerto Rico at Mayaguez, USA
Kemal Tepe	University of Windsor, Canada
Khalifa Hettak	Communications Research Centre (CRC), Canada
Khushboo Shah	Altusystems Corp, USA
Kotecha K.	Institute of Technology, Nirma University, India
Kpatcha Bayarou	Fraunhofer Institute, Germany
Kumar Padmanabh	General Motors, India
Kyriakos Manousakis	Telcordia Technologies, USA
Kyung Sup Kwak	Inha University, Korea
Li Zhao	Microsoft Corporation, USA
Li-Chun Wang	National Chiao Tung University, Taiwan
Lin Du	Technicolor Research and Innovation Beijing, P.R. China
Liza A. Latiff	University Technology Malaysia
Luca Scalia	University of Palermo, Italy
M Ayoub Khan	C-DAC, Noida, India
Maaruf Ali	Oxford Brookes University, UK
Madhu Kumar S.D.	National Institute of Technology, Calicut, India
Madhu Nair	University of Kerala, India
Madhumita Chatterjee	Indian Institute of Technology Bombay, India
Mahamod Ismail	Universiti Kebangsaan Malaysia
Mahmoud Al-Qutayri	Khalifa University, UAE
Manimaran Govindarasu	Iowa State University, USA
Marcelo Segatto	Federal University of Espírito Santo, France
Maria Ganzha	University of Gdansk, Poland
Marilia Curado	University of Coimbra, Portugal
Mario Fanelli	DEIS, University of Bologna, Italy
Mariofanna Milanova	University of Arkansas at Little Rock, USA
Mariusz Glabowski	Poznan University of Technology, Poland
Mariusz Zal	Poznan University of Technology, Poland
Masato Saito	University of the Ryukyus, Japan
Massimiliano Comisso	University of Trieste, Italy
Massimiliano Laddomada	Texas A&M University-Texarkana, USA
Matthias R. Brust	University of Central Florida, USA
Mehrzad Biguesh	Queen's University, Canada
Michael Alexander	Scaledinfra Technologies GmbH, Austria
Michael Hempel	University of Nebraska - Lincoln, USA
Michael Lauer	Vanille-Media, Germany
Ming Xia	NICT, Japan
Ming Xiao	Royal Institute of Technology, Sweden
Mohamed Ali Kaafar	INRIA, France

Mohamed Cheriet	Ecole de Technologie Superieure, Canada
Mohamed Eltoweissy	Pacific Northwest National Laboratory, USA
Mohamed Hamdi	Carthage University, Tunisia
Mohamed Moustafa	Akhbar El Yom Academy, Egypt
Mohammad Banat	Jordan University of Science and Technology, Jordan
Mohammad Hayajneh	UAEU, UAE
Mohammed Misbahuddin	C-DAC, India
Mustafa Badaroglu	IMEC, Belgium
Naceur Malouch	Université Pierre et Marie Curie, France
Nakjung Choi, Alcatel-Lucent	Bell-Labs, Seoul, Korea
Namje Park	Jeju University, South Korea
Natarajan Meghanathan	Jackson State University, USA
Neeli Prasad	Center for TeleInFrastructure (CTIF), Denmark
Nen-Fu Huang	National Tsing Hua University, Taiwan
Nikola Zogovic	University of Belgrade, Serbia
Nikolaos Pantazis	Technological Educational Institution of Athens, Greece
Nilanjan Banerjee	IBM Research, India
Niloy Ganguly	Indian Institute of Technology, Kharagpur, India
Pablo Corral González	University Miguel Hernández, Spain
Patrick Seeling	University of Wisconsin - Stevens Point, USA
Paulo R.L. Gondim	University of Brasília, Brazil
Peter Bertok	Royal Melbourne Institute of Technology (RMIT), Australia
Phan Cong-Vinh	London South Bank University, UK
Pingyi Fan	Tsinghua University, P.R. China
Piotr Zwierzykowski	Poznan University of Technology, Poland
Pascal Lorenz	University of Haute Alsace, France
Pruet Boonma	Chiang Mai University, Thailand
Punam Bedi	University of Delhi, India
Qinghai Gao	Atheros Communications Inc., USA
Rahul Khanna	Intel, USA
Rajendra Akerkar	Western Norway Research Institute, Norway
Raul Santos	University of Colima, Mexico
Ravishankar Iyer	Intel Corp, USA
Regina Araujo	Federal University of Sao Carlos, Brazil
Renjie Huang	Washington State University, USA
Ricardo Lent	Imperial College London, UK
Rio G. L. D'Souza	St. Joseph Engineering College, Mangalore, India
Roberto Pagliari	University of California, Irvine, USA
Roberto Verdone	WiLab, University of Bologna, Italy
Roksana Boreli	National ICT Australia Ltd., Australia

Ronny Yongho Kim	Kyungil University, Korea
Ruay-Shiung Chang	National Dong Hwa University, Taiwan
Ruidong Li	NICT, Japan
S. Ali Ghorashi	Shahid Beheshti University, Iran
Sahar Ghazal	University of Versailles, France
Said Souhli	Ericsson, Sweden
Sajid Hussain	Fisk University, USA
Salah Bourennane	Ecole Centrale Marseille, France
Salman Abdul Moiz	CDAC, Bangalore, India
Sameh Elnikety	Microsoft Research, USA
Sanjay H.A.	Nitte Meenakshi Institute, Bangalore, India
Sathish Rajasekhar	RMIT University, Australia
Sergey Andreev	Tampere University of Technology, Finland
Seshan Srirangarajan	Nanyang Technological University, Singapore
Seyed (Reza) Zekavat	Michigan Technological University, USA
Sghaier Guizani	UAE University, UAE
Shancang Li	School of Engineering, Swansea University, UK
Shi Xiao	Nanyang Technological University, Singapore
Siby Abraham	University of Mumbai, India
Silvio Bortoleto	Positivo University, Brazil
Simon Pietro Romano	University of Naples Federico II, Italy
Somayajulu D. V. L. N.	National Institute of Technology Warangal, India
Song Guo	The University of British Columbia, Canada
Song Lin	University of California, Riverside, USA
Soumya Sen	University of Pennsylvania, USA
Stefano Ferretti	University of Bologna, Italy
Stefano Giordano	University of Pisa, Italy
Stefano Pesic	Cisco Systems, Italy
Stefano Tomasin	University of Padova, Italy
Stefanos Gritzalis	University of the Aegean, Greece
Steven Gordon	Thammasat University, Thailand
Suat Ozdemir	Gazi University, Turkey
Subir Saha	Nokia Siemens Networks, India
Subramanian K.	Advanced Center for Informatics and Innovative Learning, IGNOU, India
Sudarshan T.S.B.	Amrita Vishwa Vidyapeetham, Bangalore, India
Sugam Sharma	Iowa State University, USA
Surekha Mariam Varghese	M.A. College of Engineering, India
T. Aaron Gulliver	University of Victoria, Canada
Tao Jiang	Huazhong University of Science and Technology, P.R. China
Tarek Bejaoui	Mediatron Lab., Carthage University, Tunisia
Tarun Joshi	University of Cincinnati, USA
Theodore Stergiou	Intracom Telecom, UK

Thienne Johnson	University of Arizona, USA
Thomas Chen	Swansea University, UK
Tsern-Huei Lee	National Chiao Tung University, Taiwan
Usman Javaid	Vodafone Group, UK
Vamsi Paruchuri	University of Central Arkansas, USA
Vana Kalogeraki	University of California, Riverside, USA
Vehbi Cagri Gungor	Bahcesehir University, Turkey
Velmurugan Ayyadurai	University of Surrey, UK
Vicent Cholvi	Universitat Jaume I, Spain
Victor Govindaswamy	Texas A&M University, USA
Vijaya Kumar B.P.	Reva Institute of Technology and Management, Bangalore, India
Viji E Chenthamarakshan	IBM T.J. Watson Research Center in New York, USA
Vino D.S. Kingston	Hewlett-Packard, USA
Vinod Chandra S.S.	College of Engineering Thiruvananthapuram, India
Vivek Jain	Robert Bosch LLC, USA
Vivek Singh	Banaras Hindu University, India
Vladimir Kropotov	D-Link Russia, Russia
Wael M El-Medany	University of Bahrain, Kingdom of Bahrain
Waslon Lopes	UFCEG - Federal University of Campina Grande, Brazil
Wei Yu	Towson University, USA
Wei-Chieh Ke	National Tsing Hua University, Taiwan
Wendong Xiao	Institute for Infocomm Research, Singapore
Xiang-Gen Xia	University of Delaware, USA
Xiaodong Wang	Qualcomm, USA
Xiaoguang Niu	Wuhan University, P.R. China
Xiaoqi Jia	Institute of Software, Chinese Academy of Sciences, P.R. China
Xinbing Wang	Shanghai Jiaotong University, P.R. China
Xu Shao	Institute for Infocomm Research, Singapore
Xueping Wang	Fudan University, P.R. China
Yacine Atif	UAE University, UAE
Yali Liu	University of California, Davis, USA
Yang Li	Chinese Academy of Sciences, P.R. China
Yassine Bouslimani	University of Moncton, Canada
Ye Zhu	Cleveland State University, USA
Yi Zhou	Texas A&M University, USA
Yifan Yu	France Telecom R&D Beijing, P.R. China
Yong Wang	University of Nebraska-Lincoln, USA
Youngseok Lee	Chungnam National University, Korea
Youssef SAID	Tunisie Telecom/Sys'Com Lab,ENIT, Tunisia
Yuan-Cheng Lai	Information Management, NTUST, Taiwan
Yuh-Ren Tsai	National Tsing Hua University, Taiwan

Yu-Kai Huang	Quanta Research Institute, Taiwan
Yusuf Ozturk	San Diego State University, USA
Zaher Aghbari	University of Sharjah, UAE
Zbigniew Dziong	University of Quebec, Canada
Zhang Jin	Beijing Normal University, P.R. China
Zhenghao Zhang	Florida State University, USA
Zhenzhen Ye	iBasis, Inc., USA
Zhihua Cui	Taiyuan University of Science and Technology, China
Zhili Sun	University of Surrey, UK
Zhong Zhou	University of Connecticut, USA
Zia Saquib	C-DAC, Mumbai, India

ACC 2011 Additional Reviewers

Akshay Vashist	Telcordia Technologies, USA
Alessandro Testa	University of Naples Federico II, Italy
Amitava	Academy of Technology, India
Ammar Rashid	Auckland University of Technology, New Zealand
Anand	MITS, India
Bjoern W. Schuller	Technical University, Germany
Chi-Ming Wong	Jinwen University of Science and Technology, Taiwan
Danish Faizan	NIC-INDIA, India
Fatos Xhafa	UPC, Barcelona Tech, Spain
Hooman Tahayori	Ryerson University, Canada
John Jose	IIT Madras, India
Jyoti Singh	Academy of Technology, India
Koushik	West Bengal University of Technology, India
Long Zheng	University of Aizu, Japan
Manpreet Singh	M.M. Engineering College, India
Maria Striki	Telcordia Technologies, Piscataway, USA
Mohamad Zoinol Abidin	Universiti Teknikal Malaysia Melaka, Malaysia
Mohamed Dahmane	University of Montreal, Canada
Mohd Helmy Abd Wahab	Universiti Tun Hussein Onn Malaysia, Malaysia
Mohd Riduan Bin Ahmad	Universiti Teknikal Malaysia Melaka, Malaysia
Mohd Sadiq	Jamia Millia Islamia, India
Mudhakar Srivatsa	IBM T.J. Watson Research Center, USA
Nan Yang	CSIRO, Australia
Nurulnadwan Aziz Aziz	Universiti Teknologi MARA, Malaysia

Pooya Taheri	University of Alberta, Canada
R.C. Wang	NTTU, Taiwan
Roman Yampolskiy	University of Louisville, USA
Shuang Tian	The University of Sydney, Australia
Syed Abbas Ali	Ajman University of Science & Technology, UAE
Velayutham	Adhiparasakthi Engineering College, Melmaruvathur, India
Yeong-Luh Ueng	National Tsing Hua University, Taiwan

International Workshop on Identity: Security, Management and Applications (ID 2011)

General Chairs

Paul Rodrigues (CTO, WSS, India)	Hindustan University, India
H.R. Vishwakarma (Secretary, Computer Society of India)	VIT University, India

Program Chairs

P. Krishna Reddy Sundar K.S.	IIIT, Hyderabad, India Education & Research, Infosys Technologies Limited, India
Srinivasa Ragavan S. Venkatachalam	Intel Inc, USA Jawaharlal Nehru Technological University, India

Organizing Chair

Madhan Kumar Srinivasan	Education & Research, Infosys Technologies Limited, India
-------------------------	--

Organizing Co-chairs

Abhi Saran	London South Bank University, UK
Anireddy Niranjana Reddy	University of Glamorgan, UK
Revathy Madhan Kumar	Education & Research, Infosys Technologies Limited, India

Technical Program Committee

Arjan Durresi	Indiana University Purdue University Indianapolis, USA
Arun Sivanandham	Infosys Technologies Limited, India
Avinash Srinivasan	Bloomsburg University, USA
Bezawada Bruhadeshwar	IIIT, Hyderabad, India
Bhaskara Reddy AV	Infosys Technologies Limited, India
Bipin Indurkha	IIIT, Hyderabad, India

C. Sunil Kumar	Jawaharlal Nehru Technological University, India
Chandrabali Karmakar	Infosys Technologies Limited, India
Farooq Anjum	On-Ramp Wireless, USA
Gudipati Kalyan Kumar	Excellence India, India
Hamid Sharif	University of Nebraska-Lincoln, USA
Hui Chen	Virginia State University, USA
Jie Li	University of Tsukuba, Japan
Kalaiselvam	Infineon Technologies, Germany
Lau Lung	UFSC, Brazil
Lukas Ruf	Consecom AG, Switzerland
Manik Lal Das	Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), India
Manimaran Govindarasu	Iowa State University, USA
Narendra Ahuja	University of Illinois, USA
Omar	University of Jordan, Jordan
Pradeep Kumar T.S.	Infosys Technologies Limited, India
Pradeepa	Wipro Technologies, India
Rajiv Tripathi	NIT, Allahabad, India
Rakesh Chithuluri	Oracle, India
Sanjay Chaudhary	Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), India
Santosh Pasuladi	Jawaharlal Nehru Technological University, India
Satheesh Kumar Varma	IIIT, Pune, India
Saurabh Barjatiya	IIIT, Hyderabad, India
Sreekumar Vobugari	Education & Research, Infosys Technologies Limited, India
Suthershan Vairavel	CTS, India
Tarun Rao	Infosys Technologies Limited, India
Thomas Little	Boston University, USA
Tim Strayer	BBN Technologies, USA
V. Balamurugan	IBM, India
Vasudeva Varma	IIIT, Hyderabad, India
Vinod Babu	Giesecke & Devrient, Germany
Yonghe Liu	UT Arlington, USA

International Workshop on Applications of Signal Processing (I-WASP 2011)

Workshop Organizers

Jaison Jacob	Rajagiri School of Engineering and Technology, India
Sreeraj K.P.	Rajagiri School of Engineering and Technology, India
Rithu James	Rajagiri School of Engineering and Technology, India

Technical Program Committee

A. Vinod	NTU, Singapore
Aggelos Katsaggelos	Northwestern University, USA
Bing Li	University of Virginia, USA
Carlos Gonzalez	University of Castilla-La Mancha, Spain
Damon Chandler	Oklahoma State University, USA
Egon L. van den Broek	University of Twente, The Netherlands
Feng Wu	Microsoft Research Asia, P.R. China
Hakan Johansson	University of Linköping, Sweden
Joaquim Filipe	EST-Setubal, Portugal
Lotfi Senahdji	Université de Rennes 1, France
Reyer Zwiggelkaar	Aberystwyth University, UK
Xianghua Xie	Swansea University, UK
Yoshikazu Miyanaga	Hokkaido University, Japan

International Workshop on Cloud Computing: Architecture, Algorithms and Applications (CloudComp 2011)

Workshop Organizers

Binu A.	Cochin University of Science and Technology, India
Biju Paul	Rajagiri School of Engineering and Technology, India
Sabu M. Thampi	Rajagiri School of Engineering and Technology, India

Technical Program Committee

Antonio Puliafito	University of Messina, Italy
Bob Callaway	IBM, USA
Chee Shin Yeo	Institute of High-Performance Computing, Singapore
Chin-Sean Sum	National Institute of Information and Communications Technology, Japan
Ching-Hsien Hsu	Chung Hua University, Taiwan
Drissa Houatra	Orange Labs, France
Deepak Unnikrishnan	University of Massachusetts, USA
Jie Song	Northeastern University, P.R. China
Salah Sharieh	McMaster University, Canada
Francesco Longo	Università di Messina, Italy
Fabienne Anhalt	Ecole Normale Supérieure de Lyon-INRIA, France
Gaurav Somani	LNMIIT, Jaipur, India
Haibing Guan	Shanghai Jiao Tong University, P.R. China
Hongbo Jiang	Huazhong University of Science and Technology, P.R. China
Hongkai Xiong	Shanghai Jiao Tong University, P.R. China
Hui Zhang	Nec Laboratories America, USA
Itai Zilbershtein	Avaya, Israel
Jens Nimis	University of Applied Sciences, Germany
Jie Song	Software College, Northeastern University, China

Jorge Carapinha	PT Inovação S.A. Telecom Group, Portugal
Junyi Wang	National Institute of Information and Communications Technology, Japan
K. Chandra Sekaran	NITK, India
Kai Zheng	IBM China Research Lab, P.R. China
Krishna Sankar	Cisco Systems, USA
Laurent Amanton	Havre University, France
Luca Caviglione	National Research Council (CNR), Italy
Lukas Ruf	Consecom AG, Switzerland
Massimiliano Rak	Second University of Naples, Italy
Pallab Datta	IBM Almaden Research Center, USA
Pascale Vicat-Blanc Primet	INRIA, France
Prabu Dorairaj	NetApp Inc, India
Shivani Sud	Intel Labs, USA
Shuicheng Yan	National University of Singapore, Singapore
Siani Pearson	HP Labs, UK
Simon Koo	University of San Diego, USA
Srikumar Venugopal	UNSW, Australia
Stephan Kopf	University of Mannheim, Germany
Thomas Sandholm	Hewlett-Packard Laboratories, USA
Umberto Villano	University of Sannio, Italy
Vipin Chaudhary	University at Buffalo, USA
Yaozu Dong	Intel Corporation, P.R. China
Zhou Lan	National Institute of Information and Communications Technology, Japan

International Workshop on Multimedia Streaming (MultiStreams 2011)

Program Chairs

Pascal Lorenz	University of Haute Alsace, France
Fan Ye	IBM T.J. Watson Research Center, USA
Trung Q. Duong	Blekinge Institute of Technology, Sweden

Technical Program Committee

Guangjie Han	Hohai University, P.R. China
Alex Canovas	Polytechnic University of Valencia, Spain
Brent Lagesse	Oak Ridge National Laboratory, USA
Chung Shue Chen	INRIA-ENS, France
Debasis Giri	Haldia Institute of Technology, India
Mario Montagud	Universidad Politécnic de Valencia, Spain
Doreen Miriam	Anna University, India
Duduku V. Viswacheda	University Malaysia Sabah, Malaysia
Elsa Macías López	University of Las Palmas de Gran Canaria, Spain
Eugénia Bernardino	Polytechnic Institute of Leiria, Portugal
Fernando Boronat	Instituto de Investigación para la Gestión Integrada de Zonas Costeras, Spain
Jen-Wen Ding	National Kaohsiung University of Applied Sciences, Taiwan
Joel Rodrigues IT	University of Beira Interior, Portugal
Jo-Yew Tham	A*STAR Institute for Infocomm Research, Singapore
Marcelo Atenas	Universidad Politecnica de Valencia, Spain
Jorge Bernabé	University of Murcia, Poland
Bao Vo Nguyen	Posts and Telecommunications Institute of Technology, Vietnam
Hans-Juergen Zepernick	Blekinge Institute of Technology, Sweden
Jose Maria Alcaraz Calero	University of Murcia, Spain
Juan Marin Perez	University of Murcia, Spain
Lei Shu	Osaka University, Japan
Lexing Xie	The Australian National University, Australia
Marc Gilg	University of Haute-Alsace, France
Miguel Garcia	Polytechnic University of Valencia, Spain
Mohd Riduan Bin Ahmad	Universiti Teknikal Malaysia, Malaysia

Phan Cong-Vinh

Alvaro Suárez-Sarmiento

Song Guo

Tin-Yu Wu

Zhangbing Zhou

Zuqing Zhu

Juan M. Sánchez

Choong Seon Hong

London South Bank University, UK

University of Las Palmas de Gran Canaria,
Spain

University of British Columbia, Canada

Tamkang University, Taiwan

Institut Telecom & Management SudParis,
France

Cisco System, USA

University of Extremadura, Spain

Kyung Hee University, Korea

Second International Workshop on Trust Management in P2P Systems (IWTMP2PS 2011)

Program Chairs

Visvasuresh Victor

Govindaswamy

Jack Hu

Sabu M. Thampi

Texas A&M University-Texarkana, USA

Microsoft, USA

Rajagiri School of Engineering and Technology,
India

Technical Program Committee

Haiguang

Ioannis Anagnostopoulos

Farag Azzedin

Fudan University, P.R. China

University of the Aegean, Greece

King Fahd University of Petroleum & Minerals,
Saudi Arabia

Roksana Boreli

Yann Busnel

Juan-Carlos Cano

Phan Cong-Vinh

Jianguo Ding

Markus Fiedler

Deepak Garg

Felix Gomez Marmol

Paulo Gondim

Steven Gordon

Ankur Gupta

National ICT Australia, Australia

University of Nantes, France

Universidad Politecnica de Valencia, Spain

London South Bank University, UK

University of Luxembourg, Luxemburg

Blekinge Institute of Technology, Sweden

Thapar University, Patiala, India

University of Murcia, Spain

Universidade de Brasilia, Brazil

Thammasat University, Thailand

Model Institute of Engineering and Technology,
India

Houcine Hassan

Yifeng He

Michael Hempel

Salman Abdul Moiz

Guimin Huang

Universidad Politecnica de Valencia, Spain

Ryerson University, Canada

University of Nebraska-Lincoln, USA

CDAC, India

Guilin University of Electronic Technology,
P.R. China

Renjie Huang

Benoit Hudzia

Helge Janicke

Washington State University, USA

SAP Research, UK

De Montfort University, UK

Mohamed Ali Kaafar	INRIA, France
Eleni Koutrouli	National University of Athens, Greece
Stefan Kraxberger	Graz University of Technology, Austria
Jonathan Loo	Middlesex University, UK
Marjan Naderan	Amirkabir University of Technology, Iran
Lourdes Penalver	Valencia Polytechnic University, Spain
Elvira Popescu	UCV, Romania
Guangzhi Qu	Oakland University, USA
Aneel Rahim	COMSATS Institute of Information Technology, Pakistan
Yonglin Ren	SITE, University of Ottawa, Canada
Andreas Riener	University of Linz, Austria
Samir Saklikar	RSA, Security Division of EMC, India
Thomas Schmidt	HAW Hamburg (DE), Germany
Fangyang Shen	Northern New Mexico College, USA
Thorsten Strufe	TU Darmstadt, Germany
Sudarshan Tsb	Amrita School of Engineering, India
Demin Wang	Microsoft, USA
Fatos Xhafa	UPC, Barcelona, Spain
Jiping Xiong	Zhejiang Normal University, P.R. China
Chang Wu Yu	Chung Hua University, Taiwan

Table of Contents – Part III

Security, Trust and Privacy

Chaotic Integrity Check Value	1
<i>Prathuri Jhansi Rani and S. Durga Bhavani</i>	
An In-Depth Analysis of the Epitome of Online Stealth: Keyloggers; and Their Countermeasures	10
<i>Kalpa Vishnani, Alwyn Roshan Pais, and Radhesh Mohandas</i>	
Cancelable Biometrics for Better Security and Privacy in Biometric Systems	20
<i>Sanjay Ganesh Kanade, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi</i>	
Advanced Clustering Based Intrusion Detection (ACID) Algorithm	35
<i>Samarjeet Borah, Debaditya Chakravorty, Chandan Chawhan, and Aritra Saha</i>	
Measuring the Deployment Hiccups of DNSSEC	44
<i>Vasilis Pappas and Angelos D. Keromytis</i>	

Sensor Networks

Self-organizing MAC Protocol Switching for Performance Adaptation in Wireless Sensor Networks	54
<i>Fan Yu and Subir Biswas</i>	
DFDNM: A Distributed Fault Detection and Node Management Scheme for Wireless Sensor Network	68
<i>Indrajit Banerjee, Prasenzit Chanak, Biplab Kumar Sikdar, and Hafizur Rahaman</i>	
An Optimized Reduced Energy Consumption (OREC) Algorithm for Routing in Wireless Sensor Networks	82
<i>Joydeep Banerjee, Swarup Kumar Mitra, Pradipta Ghosh, and Mrinal Kanti Naskar</i>	
Improving Energy Efficiency of Underwater Acoustic Sensor Networks Using Transmission Power Control: A Cross-Layer Approach	93
<i>Sumi A. Samad, S.K. Shenoy, and G. Santhosh Kumar</i>	
A Collaborative, Secure and Energy Efficient Intrusion Detection Method for Homogeneous WSN	102
<i>T. Mohamed Mubarak, Syed Abdul Sattar, Appa Rao, and M. Sajitha</i>	

An Efficient and Hybrid Key Management Scheme for Three Tier Wireless Sensor Networks Using LU Matrix	111
<i>Manivannan Dorai Pandian, Ezhilarasi Rajapackiyam, P. Neelamegam, and Anuj Kumar Rai</i>	
Grey System Theory-Based Energy Map Construction for Wireless Sensor Networks	122
<i>Vivek Katiyar, Narottam Chand, and Surender Soni</i>	
An Entropic Approach to Data Aggregation with Divergence Measure Based Clustering in Sensor Network	132
<i>Adwitiya Sinha and D.K. Lobiya</i>	
Energy Efficient Routing Protocols for Wireless Sensor Networks Using Spatial Correlation Based Collaborative Medium Access Control	143
<i>A. Rajeswari and P.T. Kalaivaani</i>	
Signal and Image Processing	
Palmprint Authentication by Phase Congruency Features	157
<i>Jyoti Malik, G. Sainarayanan, and Ratna Dahiya</i>	
Design and Implementation of 3D DWT for 4D Image Based Noninvasive Surgery	168
<i>P.X. Shajan, N.J.R. Muniraj, and John T. Abraham</i>	
Stratified SIFT Matching for Human Iris Recognition	178
<i>Sambit Bakshi, Hunny Mehrotra, and Banshidhar Majhi</i>	
Quality Index Based Face Recognition under Varying Illumination Conditions	185
<i>K.T. Dilna and T.D. Senthilkumar</i>	
Noise Adaptive Weighted Switching Median Filter for Removing High Density Impulse Noise	193
<i>Madhu S. Nair and P.M. Ameera Mol</i>	
SMT-8036 Based Implementation of Secured Software Defined Radio System for Adaptive Modulation Technique	205
<i>Sudhanshu Menta, Surbhi Sharma, and Rajesh Khanna</i>	
Abstraction of Exudates in Color Fundus Images	213
<i>Richu Paul and S. Vasanthi</i>	
A Histogram Adaptation for Contrast Enhancement	221
<i>Lisha Thomas and K. Santhi</i>	
Evaluating the Performance of a Speech Recognition Based System	230
<i>Vinod Kumar Pandey and Sunil Kumar Kopparapu</i>	

Unified Approach in Food Quality Evaluation Using Machine Vision	239
<i>Rohit R. Parmar, Kavindra R. Jain, and Chintan K. Modi</i>	
Mach-Zehnder Interferometer Based All-Optical Peres Gate	249
<i>G.K. Maity, J.N. Roy, and S.P. Maity</i>	
Using PSO in Image Hiding Scheme Based on LSB Substitution	259
<i>Punam Bedi, Roli Bansal, and Priti Sehgal</i>	
Matrix Embedding Using Random Linear Codes and Its Steganalysis	269
<i>P. Harmya, S.S. Anju, Noopa Jagadeesh, and Aishwarya Nandakumar</i>	
An Efficient Directional Weighted Median Switching Filter for Impulse Noise Removal in Medical Images	276
<i>Madhu S. Nair and J. Reji</i>	
An Improved Handwritten Text Line Segmentation Technique	289
<i>M. Mohammadi, S.S. Mozaffari Chanijani, V.N. Manjunath Aradhya, and G.H. Kumar</i>	
Skew Estimation for Unconstrained Handwritten Documents	297
<i>V.N. Manjunath Aradhya, C. Naveena, and S.K. Niranjan</i>	
Recognition of Simple and Conjunct Handwritten Malayalam Characters Using LCPA Algorithm	304
<i>M. Abdul Rahiman and M.S. Rajasree</i>	
A Fuzzy Genetic Approach to Impulse Noise Removal	315
<i>K.K. Anisha and M. Wilscy</i>	
Chain Code Histogram Based Facial Image Feature Extraction under Degraded Conditions	326
<i>Soyuj Kumar Sahoo, Jitendra Jain, and S.R. Mahadeva Prasanna</i>	
Object Recognition Based on Fuzzy Morphological Polynomial Signal Representation	334
<i>Chin-Pan Huang, Ping S. Huang, Chaur-Heh Hsieh, and Tsorng-Lin Chia</i>	
Face Detection for Skin-Toned Images Using Signature Functions	342
<i>H.C. VijayLakshmi and Sudarshan PatilKulkarni</i>	
Recurrent Neural Network Based Phoneme Recognition Incorporating Articulatory Dynamic Parameters	349
<i>Mohammed Rokibul Alam Kotwal, Foyzul Hassan, Md. Mahabubul Alam, Abdur Rahman Khan Jehad, Md. Arifuzzaman, and Mohammad Nurul Huda</i>	

Gaussian Noise and Haar Wavelet Transform Image Compression on Transmission of Dermatological Images	357
<i>Kamil Dimililer and Cemal Kavalcioglu</i>	
Image Processing Techniques for Glaucoma Detection	365
<i>Mishra Madhusudhan, Nath Malay, S.R. Nirmala, and Dandapat Samerendra</i>	
Transmitter Preprocessing Assisted MIMO SDMA Systems over Frequency-Selective Channels	374
<i>Shriram Swaminathan, Suraj Krishnan, and Prabagarane Nagaradjane</i>	
A Fuzzy Neuro Clustering Based Vector Quantization for Face Recognition	383
<i>Elizabeth B. Varghese and M. Wilscy</i>	
3D Face Recognition Using Orientation Maps	396
<i>B.H. Shekar, N. Harivinod, and M. Sharmila Kumari</i>	
The Optimal Wavelet for Speech Compression	406
<i>Shijo M. Joseph and P. Babu Anto</i>	
Skew Angle Estimation and Correction for Noisy Document Images	415
<i>M. Manomathi and S. Chitrakala</i>	
Face Recognition Using ATP Feature Set under Difficult Lighting Conditions	425
<i>Lincy Thomas and Komathy Karuppanan</i>	
Classification of Mammogram Images Using Discrete Wavelet Transformations	435
<i>K.K. Rajkumar and G. Raju</i>	
Optimized Trace Transform Based Feature Extraction Architecture for CBIR	444
<i>Meena S. Maralappanavar, K. Pramod, and K. Linganagouda</i>	
Multi-algorithm Fusion for Speech Emotion Recognition	452
<i>Gyanendra K. Verma, U.S. Tiwary, and Shaishav Agrawal</i>	
Combining Chain-Code and Fourier Descriptors for Fingerprint Matching	460
<i>C.Z. Geevar and P. Sojan Lal</i>	
Facial Emotion Recognition Using Different Multi-resolution Transforms	469
<i>Gyanendra K. Verma, U.S. Tiwary, and Mahendra K. Rai</i>	

Robust Watermarking through Spatially Disjoint Transformations	478
<i>Reena Gunjan, Saurabh Maheshwari, Vijay Lazmi, and M.S. Gaur</i>	

Soft Computing Techniques

An ANN Based Approach to Calculate Robotic Fingers Positions	488
<i>Ankit Chaudhary, J.L. Raheja, Kunal Singal, and Shekhar Raheja</i>	

Word Classification Using Neural Network	497
<i>A. Muthamizh Selvan and R. Rajesh</i>	

Very Short Term Wind Power Forecasting Using PSO-Neural Network Hybrid System	503
<i>E. Pratheepraj, Anuj Abraham, S.N. Deepa, and V. Yuvaraj</i>	

A Class of Recurrent Neural Network (RNN) Architectures with SOM for Estimating MIMO Channels	512
<i>Kandarpa Kumar Sarma and Abhijit Mitra</i>	

An Efficient Framework for Prediction in Healthcare Data Using Soft Computing Techniques	522
<i>Veena H. Bhat, Prasanth G. Rao, S. Krishna, P. Deepa Shenoy, K.R. Venugopal, and L.M. Patnaik</i>	

Process Oriented Guided Inquiry Learning for Soft Computing	533
<i>Clifton Kussmaul</i>	

A Modified and Efficient Shuffled Frog Leaping Algorithm (MSFLA) for Unsupervised Data Clustering	543
<i>Suresh Chittineni, Dinesh Godavarthi, A.N.S. Pradeep, Suresh Chandra Satapathy, and P.V.G.D. Prasad Reddy</i>	

Neighborhood Search Assisted Particle Swarm Optimization (NPSO) Algorithm for Partitional Data Clustering Problems	552
<i>R. Karthi, C. Rajendran, and K. Rameshkumar</i>	

System Software

Portability in Incremental Compilers	562
<i>P.R. Mahalingam and C. Unnikrishnan</i>	

Test Case Optimization Using Artificial Bee Colony Algorithm	570
<i>Adi Srikanth, Nandakishore J. Kulkarni, K. Venkat Naveen, Puneet Singh, and Praveen Ranjan Srivastava</i>	

Vehicular Communications Networks

Adaptive Power Allocation in CI/MC-CDMA System Using Genetic Algorithms	580
<i>Santi P. Maity and Sumanta Hati</i>	
PHY Abstraction for MIMO Based OFDM Systems	590
<i>Tarun kumar Juluru and Anitha Sheela Kankacharla</i>	
Level Crossing Rate in Land Mobile Satellite Channel with a Modified Nakagami-Lognormal Distribution	601
<i>Sayantan Hazra and Abhijit Mitra</i>	
Cache Invalidation for Location Dependent and Independent Data in IVANETS.	609
<i>Anurag Singh, Narottam Chand, and Lalit Kr Awasthi</i>	
VCAR: Scalable and Adaptive Data Dissemination for VANET	615
<i>K. Naveen and Komathy Karuppanan</i>	
Efficient Distributed Group Authentication Protocol for Vehicular Ad Hoc Network	624
<i>Priya Karunanithi and Komathy Karuppanan</i>	
Opportunistic Dissemination for Accident Management Using Vehicular Networks	634
<i>R. Namritha and Komathy Karuppanan</i>	
Machine Learning Approach for Multiple Misbehavior Detection in VANET	644
<i>Jyoti Grover, Nitesh Kumar Prajapati, Vijay Laxmi, and Manoj Singh Gaur</i>	
Improved Position-Based Routing in Vehicular Ad Hoc Networks Using P-DIR Method	654
<i>Ram Shringar Raw and D.K. Lobiyal</i>	
IMS and Presence Service Integration on Intelligent Transportation Systems for Future Services	664
<i>Andrés Garcia, José Santa, Antonio Moragón, and Antonio Farnando Gómez-Skarmeta</i>	
Author Index	677

Chaotic Integrity Check Value

Prathuri Jhansi Rani and S. Durga Bhavani

Department of Computer & Information Sciences,
University of Hyderabad, Hyderabad, India

Abstract. Chaotic cryptography is slowly emerging as a subfield of cryptography. Many encryption algorithms, secure hash functions and random number generators have been proposed in the literature which are based on well-known chaotic functions. Chaotic keyed hash functions are proposed in the literature but have not been analysed for integrity check value purpose in the literature. We propose a keyed chaotic hash function based on parametrized family of logistic and tent maps and develop a message authentication code (MAC) which outputs a 128 bit message digest(MD).The keyed hash function is designed such that it is resistant to preimage and second preimage attacks. The chaotic hash functions proposed in the literature use a multitude of chaotic maps and we show in this paper that using two chaotic maps judiciously achieves a secure keyed hash function. The proposed keyed hash function is analysed for its sensitivity to the two secret keys that of initial value as well as the parameter value of the family of functions. Experiments in which as the secret keys are infinitesimally changed, the hash value obtained is shown to have nearly 50% of the bits different from the original MD. Further similar experiments are repeated with MD truncated to the first 96 bits, which is the default length for authentication data field in IPSec authentication header and encapsulating security payload. Once again the confusion is shown to be very close to 50%.

1 Introduction

A keyed hash function or message authentication code(MAC) is a family $h_k : k \in K$ of hash functions, where K is a key space of h [1]. A hash function is a function that takes a variable-length input string and converts it to a fixed-length(smaller) output string called hash value or message digest. $h : (0, 1)^* \rightarrow (0, 1)^n$ is such that h satisfies the three security properties: **collision resistance**, **preimage** and **second preimage resistance** [2]. Integrity check value(ICV) is an authentication field in IPSec Authentication Header and Encapsulating security payload. The ICV is a truncated version of message authentication code(MAC) produced by MAC algorithm HMAC-MD5-95 or HMAC-SHA1-96 [3]. The term ICV is referred as Keyed Hash function or Message Authentication Code. The full HMAC value is calculated and first 96 bits are considered, which is the default length for the authentication data field in IPSec Authentication Header and Encapsulating security payload. Recent investigations reveal that several well-known methods such as MD5, SHA1 and RIPEMD too are not immune to

collisions [4,5]. Hence ICV algorithm needs to be reinvestigated in the light of new algorithms emerging out of chaotic cryptography.

Richard Taylor proposed an integrity check value algorithm for stream ciphers [7]. It is well-known that keyed hash functions are directly useful to be used as MAC. Chaotic keyed hash functions have been proposed in the literature by Zhang et al. based on n th-order chaotic system [10]. A novel approach to keyed hash function construction that uses a chaotic neural network has been proposed by Xiao et al. [12]. These algorithms then can be theoretically used as integrity check value algorithms. There is no explicit evaluation of hash function algorithms to be used as integrity check value algorithms in the literature. The chaotic hash functions proposed in the literature use a multitude of chaotic maps and we show in this paper that using two chaotic maps judiciously achieves keyed hash function with less computational complexity. We propose a chaotic hash function and conduct a systematic analysis of its incorporation into integrity check value algorithm. We assess how the truncation of a MD of length 128 to say 96 bits influences its performance and report our results.

2 Proposed Hash Function

We propose a one-way hash function based on seminal paper of Baptista's symmetric encryption algorithm [8]. Baptista's encryption algorithm maps each character a of the message to the number of iterations n that the logistic map takes to reach ϵ -intervals I_a . We modify Baptista's encryption algorithm to build a hashing function. Divide the input bit sequence into 8 blocks, B_0, B_1, \dots, B_7 each block having k bytes, $k \in N$. Choose secret value $0 < x_0 < 1$, $\lambda \in (3.8, 4.0)$ the control parameter of the logistic equation $f_\lambda(x) = \lambda x(1-x)$ and choose intervals of I_m contained in $(0,1)$ for each m . Initialise $x(0) = x_0$.

Algorithm 1.

```

1: for  $B_i = 0$  to 7 do
2:   for byte  $m = 0$  to  $k - 1$  of the message of block  $B_i$  do
3:     Iterate the logistic map until  $f^{n_m}(x(m)) \in I_m$ .
4:     Re-Set  $x(m + 1) = f^{n_m}(x(m))$ 
5:   end for
6:   Assign  $h_i = n$ , for the block  $B_i$  /*Thus the block gets compressed by 16 bits.*/
7: end for
8:  $h(P) = (h_0h_1h_2h_3h_4h_5h_6h_7)$  /* The final hash value is obtained by concatenating  $h_i$ 's*/
```

Note that in the algorithm, hash value h_i that emerges for each block B_i is n_k where

$$f^{n_{k-1}}(\dots f^{n_2}(f^{n_1}(f^{n_0}(x_0)))) \in I_{k-1}$$

n_{k-1} is the integer number of iterations that corresponds to the $(k-1)$ -th byte.

It can be seen that this scheme is not secure from preimage attack. For a given hash value y , one can produce an input (preimage) x such that $h(x) = y$ as follows. Suppose $y = n$ say, find a character a such that $f^n(x_0) \in I_a$. So trivially a one character preimage exists for the given y . Assume that the minimum length of the message is greater than 1. First note that

$$f^{n_{k-1}}(\dots f^{n_2}(f^{n_1}(f^{n_0}(x_0)))) = f^{n_0+n_1+\dots+n_{k-1}}(x_0).$$

Choose any sequence of integers that ends in n , such as $n_0n_1\dots n_{k-1}n$. We can construct the preimage of y as follows: Find character p_0 where $f^{n_0}(x_0) \in I_{p_0}$. Now reset x_0 as $f^{n_0}(x_0)$. Then find p_1 such that $f^{n_0+n_1}(x_0) \in I_{p_1}$ and so on. Finally we get p_{k-1} such that

$$f^{n_0+n_1+\dots+n_{k-1}}(x_0) \in I_{n_{k-1}}.$$

Then the required preimage for $y = n$ is $p_0p_1\dots p_{k-1}$. Following the above scheme, second preimages can be found. Since a collision resistant hash function is second preimage resistant, the hash function will be automatically prone to collisions. [9] Hash function needs to be strengthened by choosing a large n as well as with perturbation of initial value.

3 Strengthening the Proposed Hash Function

In this section we rectify the vulnerability of the hash function using two schemes *Perturb_{IC}* and *Vote_{PRNG}*.

3.1 Scheme A

- *Perturb_{IC}* : *Perturb_{IC}* requires perturbing x_0 using character of the message at each step. This scheme is clearly depicted in Figure [1] with the following notation. Let $byte_i = a_i$ then $A_i = ASCII(a_i)$ and $I_i = I_{a_i}$.

3.2 Scheme B

- *Vote_{PRNG}* : It was also found in the experiments the small integers that get generated as hash values in the above Algorithm [1] lead to collisions. We take a vote by a pseudo-random number generator (PRNG), a one-dimensional chaotic tent map *PRNG_{Tent}* to decide n which is depicted in the Figure [1]

Tent Map

$$f(x) = \begin{cases} 2x & \text{if } 0 \leq x < \frac{1}{2} \\ -2x + 2 & \text{if } \frac{1}{2} < x \leq 1 \end{cases}$$

The tent map is iterated and the PRNG algorithm outputs the iteration number at which the value falls within a prefixed ϵ -interval. Since the tent map is almost a linear map, the computations are inexpensive as compared to other non-linear chaotic maps. Good results are obtained when the PRNG is tested for randomness using NIST(National institute of standards and technology) test suite.

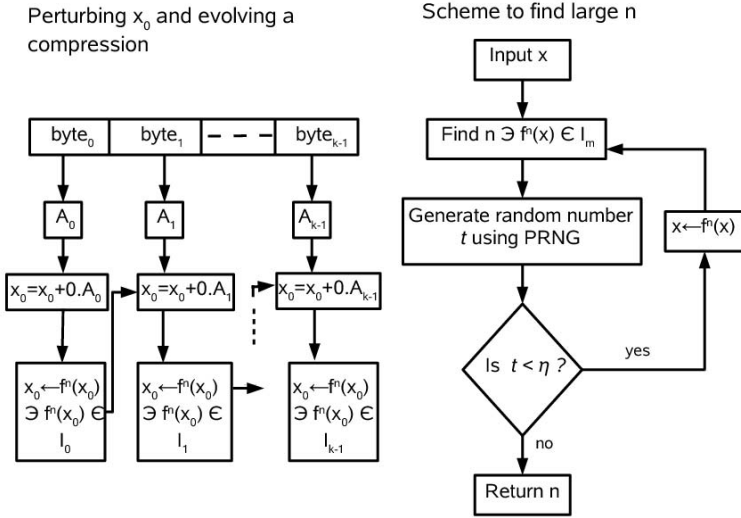


Fig. 1. Perturbing x_0 and evolving a compression function and Scheme to find large n

4 Results and Analysis

4.1 Checking for Collisions in the Whole Space

A large number of different input messages say $P_1, P_2, P_3, \dots, P_N$ of length 800 bits are chosen randomly to test for collisions. Hash values are computed for 3 lakh messages which is roughly 2^{21} for this test. Results of the experiments show that no two input messages are compressed to an identical message digest. Then to test for applicability to integrity check value, we truncate the MD to its first 96-bits and repeat the collision experiment. It is found that after truncation too no two hash values are identical to each other. The hash values obtained for a sample of five input messages is shown in the Table [1](#)

4.2 Sensitivity of Proposed Keyed Hash Function with Respect to Input Message

Let P be the input message and P' denote the message in which i th bit is toggled. Then $d_H(h(P), h(P'))$ denotes the number of bits changed in output with a 1 bit change in the input. The minimum hamming distance,

$$d_{min} = \min_{1 \leq i \leq N} d_H(h(P), h(P'))$$

and maximum hamming distance ,

$$d_{max} = \max_{1 \leq i \leq N} d_H(h(P), h(P'))$$

are computed. Further, the average d_{avg} and the standard deviation d_{std} of the distribution of Hamming distances over N trials are computed to get a

Table 1. MD generated for the input messages $P_1, P_2, P_3, \dots, P_5$

Input	Message Digest
P1	010111110010111000111011010011010001111011110100 1000010101110000110001111000101111011010001001101
P2	111110101001000111111010001100001011010011000010 111110000101010110110101100011111010010001001001
P3	000111001101110111011011010001111110111000010001 000001001001111000001010100100111110100101101011
P4	101010000010000010101100010100000001011100100011 100010011011110010000111111000011100001011011110
P5	1000011110101011111110100000110001111110101110010 11011010001110111111110110101011100011101010110

Table 2. Collision resistance analysis for the proposed keyed hash function

N	256	512	1024	2048	4096	8192	10000
d_{min}	49	47	44	44	44	43	43
d_{max}	75	76	80	80	83	84	84
d_{avg}	63.05	63.03	63.10	63.03	63.21	63.06	63.03
d_{std}	4.92	4.92	4.93	4.92	4.94	4.93	4.92

Table 3. Statistical comparison of changed bit number

	This Paper	Ref. [11]	Ref. [12]	Ref. [13]	MD5
d_{avg}	63.03	64.01	63.85	63.91	64.03
d_{std}	4.92	4.47	4.52	4.36	4.42

better idea of the distribution. The messages obtained by changing one bit in the original message exhibit hash values that have, on average, nearly 50% of the bits different from the original Message Digest.

The Figure 2 shows that the proposed algorithm exhibits desirable security with the number of changed bits due to a 1-bit toggle in plain text being 63 in the 128 bit message digest which is very close to the ideal value of 50% probability. Further all the results achieved are comparable to those of the recent work as shown in the Table 3. The space in which the tests are conducted is large enough to indicate that the values obtained by d_{avg} etc lie close to the true values of the distribution.

Comparison with MD5: Another experiment in which the number of identical bytes that occur in the same positions is calculated. Figure 3 shows the distribution of the number of ASCII characters with the same value at the same location of the proposed function and that of MD5 by toggling a bit in a sample of 10 input messages. This shows that the proposed function is on par with MD5. This is only a sample experiment and in the Figure 6 the results are shown when run on a large number of input messages.

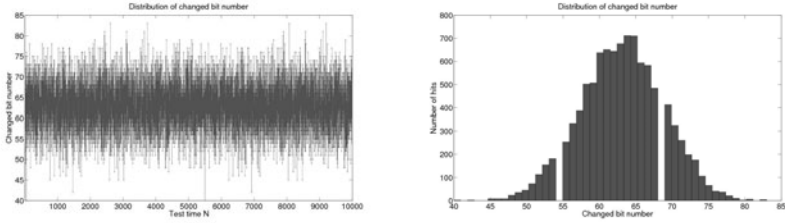


Fig. 2. Hamming distance to evaluate collision resistance

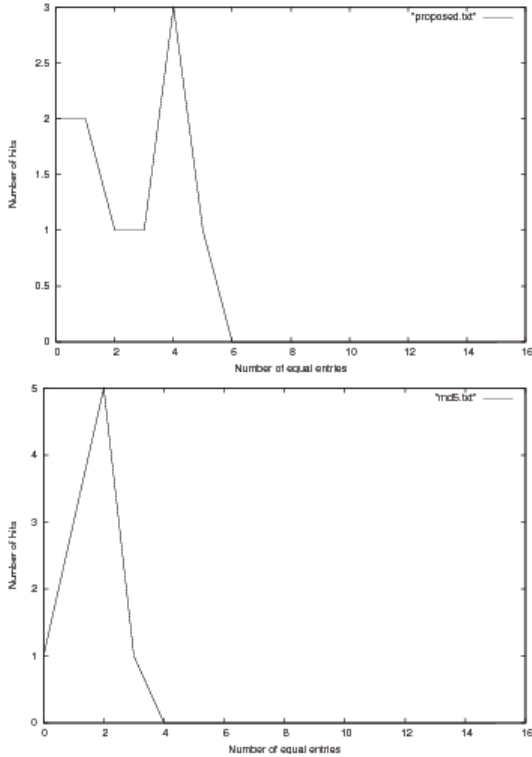


Fig. 3. Distribution of the number of ASCII characters with the same value at the same location of proposed function and MD5

4.3 Sensitivity of Control Parameter

Let $h(P)$ be the hash value obtained by taking the control parameter as 3.87 and $h(P')$ be the hash value obtained by changing the control parameter λ from 3.87 to $3.87 + 10^{-7}$. Then $d_H(h(P), h(P'))$ denotes the number of bits changed in output with a change in the control parameter. The d_{avg} over 2048 trials is 63.32. When the first 96-bits are considered d_{avg} is 47.52 which is very close to

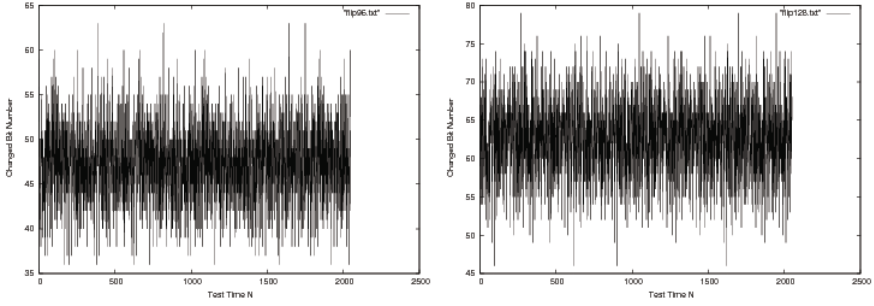


Fig. 4. Hamming distance of first 96-bits of the hash value and 128-bit hash value to evaluate collision resistance as the control parameter λ is varied

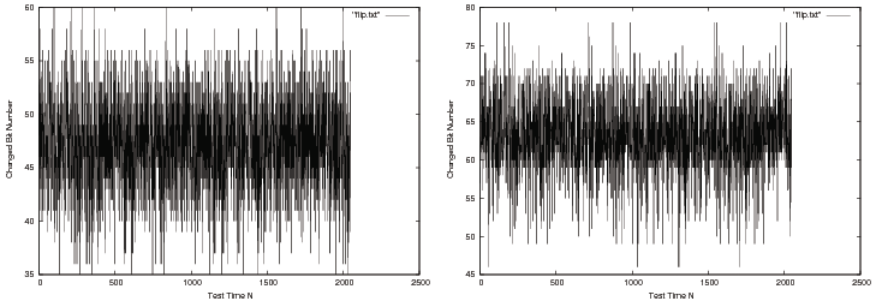


Fig. 5. Hamming distance of first 96-bits of the hash value and 128-bit hash value to evaluate collision resistance as the initial value x_0 is varied

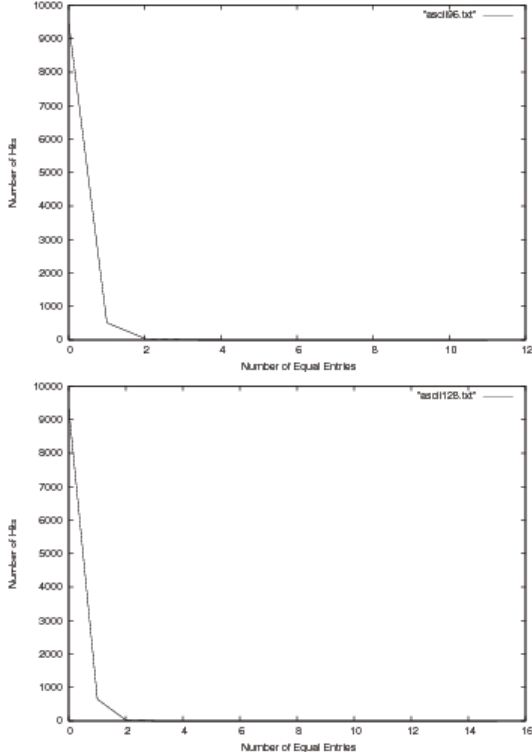
the ideal value of 50% probability. Figure 4 shows the hamming distance of first 96-bits of the hash value and 128-bit hash value. The space in which the tests are conducted is large enough to indicate that the values obtained by d_{avg} etc lie close to the true values of the distribution.

4.4 Sensitivity of Initial Value

Let $h(P)$ be the hash value obtained by taking the initial value x_0 as 0.232323 and $h(P')$ be the hash value obtained by changing the initial value x_0 from 0.232323 to $0.232323 + 10^{-7}$. Then $d_H(h(P), h(P'))$ denotes the number of bits changed in output with a change in x_0 . The d_{avg} over 2048 trials is 63.13. When the first 96-bits of 128-bit hash value is considered d_{avg} is 47.84 which is very close to the ideal value of 50% probability. Figure 5 shows the hamming distance of first 96-bits of the hash value and 128-bit hash value. The space in which the tests are conducted is large enough to indicate that the values obtained by d_{avg} etc lie close to the true values of the distribution.

Table 4. Absolute difference of two hash values

Maximum	Minimum	Mean
2153	447	1296.60

**Fig. 6.** Distribution of the number of ASCII characters with the same value at the same location in the of first 96-bits of the hash value and 128-bit hash value

4.5 Collision Analysis

ASCII character of the original and the new hash value are compared on the collision test performed for 10,000 times. The distribution of the number of ASCII characters with the same value at the same location in the hash value is shown in Fig 6. The maximum, mean, and minimum values are listed in Table 4.

The Figure 6 confirms that the proposed function is performing well with the maximum number of equal characters in hash values obtained being only 3 and the collision probability being quite low.

5 Conclusions

In this paper we show that using two chaotic maps judiciously achieves a keyed hash function which gives an output of 128-bits, in which the first 96-bits of hash value can be used as an ICV. The function is analysed and shown to possess collision, preimage and second preimage resistance. The keyed hash function is tested by varying the two secret keys infinitesimally and shown to achieve 50% variation in the output message digest and hence is strong against any collisions.

References

1. Buchmann, J.A.: Introduction to Cryptography, pp. 247–248. Springer, Heidelberg (2001)
2. Stinson, D.R.: Cryptography: Theory and Practice. CRC Press, Boca Raton (1995)
3. Stallings, W.: Cryptography and Network Security, Principles and Practice, pp. 493–500. Pearson, London (2004)
4. Wang, X., Yu, H.: How to break MD5 and other hash functions. In: Cramer, R. (ed.) EUROCRYPT 2005. LNCS, vol. 3494, pp. 19–35. Springer, Heidelberg (2005)
5. Wang, X., Yin, Y.L., Yu, H.: Finding collisions in the full SHA-1. In: Shoup, V. (ed.) CRYPTO 2005. LNCS, vol. 3621, pp. 17–36. Springer, Heidelberg (2005)
6. Baptista, M.S.: Chaos in cryptography. Phys. Lett. A 240, 50–54 (1998)
7. Taylor, R.: An integrity check value algorithm for stream ciphers. In: Stinson, D.R. (ed.) CRYPTO 1993. LNCS, vol. 773, pp. 40–48. Springer, Heidelberg (1994)
8. Jhansi Rani, P., Sambasiva Rao, M., Durga Bhavani, S.: Design of secure chaotic hash function based on logistic and tent Maps. In: Accepted for Publication in the Proceedings of CNSA-2011. Springer, Heidelberg (2011)
9. Delfs, H., Knebl, H.: Introduction of Cryptography. Springer, Heidelberg (2002)
10. Jiashu, Zhang, Wang, X., Zhang, W.: Chaotic keyed hash function based on feedforward-feedback nonlinear digital filter. Physics Letters A 362, 439–448 (2007)
11. Xiao, D., Liao, X., Wang, Y.: Parallel keyed hash function construction based on chaotic neural network. Neurocomputing 72, 2288–2296 (2009)
12. Xiao, D., Liao, X.F., Deng, S.J.: One-way hash function construction based on the chaotic map with changeable parameter. Chaos, Solitons & Fractals 24, 65–71 (2005)
13. Yi, X.: Hash function based on chaotic tent maps. IEEE Transactions on Circuits and Systems-II 52, 354–357 (2005)

An In-Depth Analysis of the Epitome of Online Stealth: Keyloggers; and Their Countermeasures

Kalpa Vishnani, Alwyn Roshan Pais, and Radhesh Mohandas

Dept. of Computer Science & Engg,
National Institute of Technology Karnataka, Surathkal, Srinivasnagar,
Mangalore - 575025, India
{kalpavishnani,alwyn.pais,radhesh}@gmail.com

Abstract. Malware came into existence ever since the inception of the computers itself and its spread has been gaining momentum as a result of persistent success and evolution of the Internet. Cyber world has been noticing a shift in the goals of malware writers, which would only become more insidious with time. Currently the matter of great concern for Internet users is that of online stealth. In this paper we discuss in detail about the epitome of online stealth, the keyloggers; present an analysis of few well known anti-keyloggers; list a set of counter-measures for the users based on our analysis; and also present our approach for client side authentication to reduce the attack surface available to the hackers.

Keywords: keylogger, password stealing, rootkits, malware.

1 Introduction

Impelled by the proliferation of high speed connections and the global coverage, Internet has become a powerful means for knowledge sharing as well as commercialization. The escalating dependence on the Internet, however, also makes it an obvious target for the miscreants to spread computer viruses and other types of malware.

Evidently, the world of cyber-crime has seen a tremendous rise in the number of money-oriented malware such as keyloggers which are made with the intention of stealing wealth by obtaining the victims' bank credentials, compromising their privacy and also for industrial espionage.

As the name suggests, keyloggers are used to record keystrokes on the target system, which could be then sent remotely to someone other than the computer's user. Keystroke loggers may steal data such as PIN codes and account numbers for e-payment systems, passwords to online gaming accounts, email addresses, user names, email passwords etc.

In this paper we discuss software keyloggers in detail in order to understand their severity. Hardware keyloggers have not been addressed in this paper since they require physical installation on the user's machine and hence cannot spread through Internet.

The remainder of this paper is organised as follows. In Section 2 we list some advanced keylogger characteristics. In Section 3, we discuss the methodology behind keyloggers specifically addressing their stealth characteristic which lets them remain undetected on the victim's machine. In Section 4, we call the attention of the readers to the alarming escalation of keylogger-oriented malware by presenting some statistics. In Section 5, we present the anti-keylogger analysis. Here we also present two of our proof of concept keyloggers. The two samples have been built using two different approaches, user mode hooks and kernel mode hooks respectively. We tested these keyloggers against a battery of anti-keyloggers. In Section 6, we discuss the mitigation techniques for users. In Section 7, we discuss the multi-factor authentication techniques and their effectiveness in combating password stealing malware. In Section 8, we present our client side authentication solution for countering the same. Finally, Section 9 concludes.

2 Advanced Keylogger Characteristics

Although the main purpose of keyloggers is to monitor a user's keyboard actions, they now have capabilities that extend beyond that function. They can track virtually anything running on a computer. Here is only a partial list of some of the information keyloggers sense, record, and transmit:

- Keystrokes
- Site Monitoring
- Chat Monitoring
- Application / Program Tracking
- Printing Activity Recording
- Invisible Mode
- System Logon/Logoff
- Clipboard Monitoring
- File/folder Monitoring Screenshots Recording
- E-mail Reporting and Alerting
- 'Find Keyword' monitoring
- Hot Key and Password Protection
- Some keyloggers, known as "screen scrapers," enable the visual surveillance of a target computer by taking periodic snapshots of the screen. The captured images can then be used to gather valuable information about the user.

3 Keylogger Methodology: Stealth and Persistence

Keyloggers are marked by two important characteristics. One is to log the keystrokes and the other to remain hidden in the target system in order to persistently and undetectably steal sensitive information of the victim. Monitoring methods for software keyloggers are operating system specific. Windows operating systems (WOS) contain an event mechanism. When a user presses a key in the WOS, the keyboard driver of the operating system translates a keystroke into a Windows message called WM_KEYDOWN. This message is pushed into the system message

queue. The WOS in turn puts this message into the message queue of the thread of the application related to the active window on the screen. The thread polling this queue sends the message to the window procedure of the active window. This kind of messaging mechanism works for other events like mouse messages [1]. There are three main methods for developing keylogger systems:

1. The Keyboard State Table method
2. The Windows Keyboard Hook method
3. The Kernel-Based Keyboard Filter Driver method

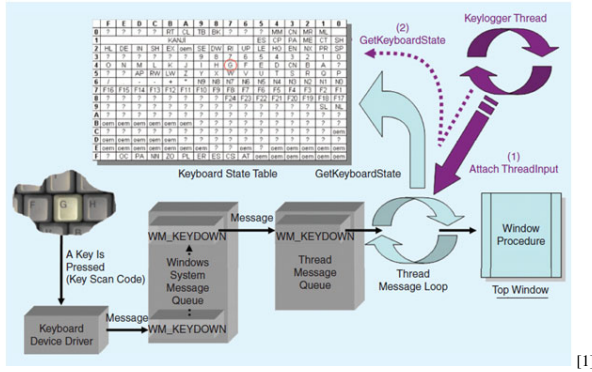


Fig. 1. Keyboard State Table method

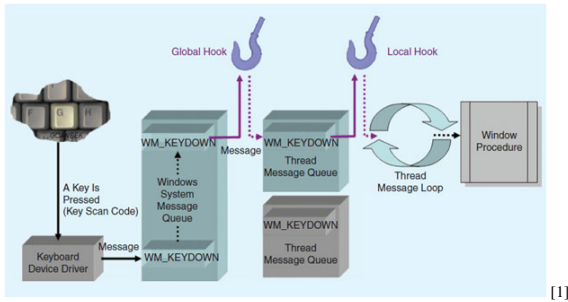


Fig. 2. Windows Keyboard Hook method

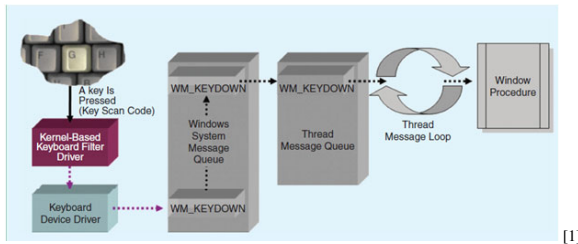


Fig. 3. Kernel-Based Keyboard Filter Driver method

Stealth

More challenging for the keyloggers, is the goal of remaining hidden. Most of the advanced keyloggers use the rootkit technology to accomplish this task. Rootkits can be difficult to detect, especially when they operate in the kernel. This is because a kernel rootkit can alter functions used by all software, including those needed by security software. A rootkit can prevent detection or anti-malware software from running or working properly by hiding files and concealing running processes from the computer's operating system [2]. Different keyloggers achieve different levels of stealth depending upon the rootkit mechanism they use.

4 Keylogger Malware Statistics

Keystroke logging, phishing and social engineering are currently the main techniques employed to carry out cyber fraud. We performed a simple search of a virus encyclopaedia [3] for keylogger malware examples. It displayed an overwhelming 67406 examples in the category! Let us now take a look at two famous keylogger malware examples and the damage they have caused hitherto.

1. Zeus

Also known as Zbot, PRG, Wsnpoem, Gorhax and Kneber, it is a Trojan horse that steals banking information by keystroke logging. Zeus is spread mainly through drive-by downloads and phishing schemes. Zeus' infection rate is higher than that of any other financial Trojan. In 2010, Zeus was the culprit behind one of the largest e-crime busts in UK history, causing a loss of \$31Million to the bank customers. [4]

2. Backdoor: W32/Haxdoor

This is a family of malicious remote administration programs (RAT) that are commonly used by attackers to steal sensitive information - passwords, account details, etc - from infected systems by their keylogging capability. Later variants in the Haxdoor family also engage in attacks against online payment systems and banking portals. Some Haxdoor variants also include rootkit capabilities and are able to hide their presence and actions on the computer system.

A theft of over \$1million from client accounts at the major Scandinavian bank Nordea in August 2006 was a consequence of Haxdoor infection on client systems [5].

5 Anti-keylogger Analysis

We built two proof of concept keyloggers. One is a simple keylogger implemented by hooking into the Windows API function, SetWindowsHookEx, and monitoring the event WH_KEYBOARD_LL. We will refer to it as *Sample A*. This keylogger spreads through infected flashdrive and sends the recorded keystrokes to the remote attacker via e-mail. However it is visible in the task manager. Although, there exist several ways to hide a program from the task manager/process explorer as well as registry

entries, we chose to keep it simple. Generally an average internet user does not identify such suspicious files and processes if they are given names resembling system processes.

Another keylogger was implemented as a kernel mode rootkit (a device driver) which remains hidden from the task manager and all such other applications e.g. Process Explorer. We will refer to this as *Sample B*. We implemented this sample by hooking the keyboard interrupt.

We submitted both our keylogger samples to www.virustotal.com [6]. There, our samples were scanned against a total of 43 well known anti-viruses/anti-malware products and we got the following results:

- Only 9 could detect Sample A as malicious.
- Only 10 could detect Sample B as malicious.

In general, there exist two types of anti-keyloggers. One category detects keyloggers and the other merely prevents them from logging keystrokes. We saw that anti-keyloggers that fall in the detection category cannot be relied upon completely as many could not detect our proof of concept keylogger samples. However we tested few anti-keyloggers that prevent the keyloggers from logging keystrokes and found them to be very effective. We tested the following products in this category:

1. KeyScrambler Pro
2. Anti-Keylogger Shield
3. Guarded ID

We tested these anti-keyloggers against the following keylogging software:

1. Refog Personal Monitor
2. All in One Keylogger
3. PC Pandora
4. Keyboard Spectator
5. Quick keylogger
6. Blazing Tools PKFree
7. Our Proof of concept Keylogger

Table 1. Anti-keylogger Analysis

	Refog Personal Monitor	All in One Keylogger	PC Pandora	Keyboard Spectator	Quick Keylogger	Blazing Tools PKFree	Our Sample A	Our Sample B
KeyScramblerPro	✓	✓	✓	✓	✓	✓	✓	✓
Anti-Keylogger Shield	✓	✓	✓	✓	✓	✓	✓	✓
Guarded ID	✓	✓	✓	✓	✓	✓	✓	✓

We found that all these anti-keyloggers effectively prevented the keystrokes from being logged by the above mentioned keyloggers. However, it is important to note that most of these keyloggers log many other things than just keystrokes such as screenshots, clipboard activity, websites visited and so on. These activities cannot be

prevented by the above mentioned anti-keyloggers. Yet these applications certainly mitigate the risk of text based information stealing which is far more critical than other sort of information.

6 Mitigation Techniques Users

From our discussion in the previous sections, it can be inferred that it is very easy to fool the average computer users into stealing their sensitive information in no time. And now that there exist very sophisticated, undetectable, rootkit based keyloggers, the situation needs immediate attention in order to contain and control the huge losses being incurred due to them. The following are some techniques, which would lower this risk to a great extent.

1. Use anti-keylogger software, especially those that prevent keylogging activity. Following are the few examples we tested.

a. KeyScramber Pro

KeyScrambler Personal is a free plug-in for Web browser that protects everything you type from keyloggers[7]. It defeats keyloggers by encrypting the keystrokes at the keyboard driver level, deep within the operating system. When the encrypted keystrokes reach the browser, KeyScrambler decrypts them so you see exactly the keys you've typed. Keyloggers end up logging only the encrypted keys, which are completely indecipherable. Unlike anti-virus and anti-spyware programs that depend on recognition to remove keyloggers that they know about, KeyScrambler protects from both known and unknown keyloggers.

b. Anti-Keylogger Shield

Anti Keylogger Shield is a powerful, easy to use anti-spy software tool that prohibits operation of any keylogger, either known or unknown, professional or custom made[8]. Once installed, Anti Keylogger Shield silently runs in the System Tray, blocks the system mechanisms that are exploited by keyloggers, and protects user privacy immediately and constantly.

c. GuardedID

GuardedID also takes a proactive approach to stopping malicious keylogging programs by encrypting every keystroke at the point of typing the keys, and rerouting those encrypted keystrokes directly to the Internet Explorer browser through its own unique path [9]. GuardedID bypasses the typical places keyloggers normally reside, thereby helping to eliminate the vulnerability to keylogging attacks.

2. A good anti-malware and a firewall in place would further strengthen the protection. The following is a list of decently effective anti-Malware/ Anti-spyware software:

- Malwarebytes' Anti-malware
 - Spyware Terminator
 - Microsoft Security Essentials
 - Panda Cloud Anti-virus free Edition
 - Spybot Search and Destroy
3. Virtual Keyboard should be used in order to fill out sensitive information such as PIN, passwords, etc. Most of the online banks provide virtual keyboards on their websites. There are also many free virtual keyboard software available.
 4. No sensitive information must be given out while surfing through a cyber cafe. It should be done through one's personal computer only.
 5. In order to protect against the malicious password stealing add-ons and toolbars, browsing should be done in safe mode. This mode disables all the add-ons and toolbars.

7 Multifactor Authentication to Thwart Password Stealing [10]

We saw that anti-keyloggers in the prevention category show 100 percent containment results in our analysis. However, continuous and rapid evolution of malware prompts us to take the anti-keylogger defences to a higher level. Multifactor authentication schemes provide very strong protection to a user even when his credentials have been compromised by a keystroke logger. Multi Factor Authentication is based on two or more factors. These methods often provide greater security. However, they may also impose different levels of inconvenience to the end user.

The success of these methods depends upon the degree of their acceptance by the users. Here we discuss some of these methods which could contain the password stealing attacks to a great extent.

7.1 Scratch List

Scratch or grid cards contain randomly generated numbers and letters arranged in a row and column format. They are cheap, durable and easy to carry and thus very popular. This type of authentication requires no training and, if the card is lost, replacement is relatively easy and inexpensive.

Used in a multi factor authentication process, the user first enters his or her user name and password in the established manner. Assuming the information is entered correctly, the user will then be asked to input, as a second authentication factor, the characters contained in a randomly chosen cell in the grid. The user will respond by typing in the data contained in the grid cell element that corresponds to the challenge coordinates.

7.2 Time Based

Timebased password systems, such as the RSA SecurID, are another type of onetime passwords. In such a system, the password changes every minute or so, by an algorithm known to the system and the user's authentication device. The algorithm to

create the password is based on a function of a key and the current time. The key is preset at the fabrication of the device. Further, the token has an internal clock. A timer creates a new password in a predefined time range (usually every 60 seconds), using the current time and the key as its input. The function itself is usually proprietary, but is basically a form of the following:

$$\text{password} = \text{encrypt}(\text{TIME with KEY})$$

One challenge of this system is the synchronization of the time.

7.3 Challenge Response Based

In comparison to the time based system, here the time is replaced by a challenge, and the SmartCard/Token is optionally protected by a PIN. Instead of using the time as input for creating the password, a server generated challenge is used:

$$\text{password} = \text{encrypt}(\text{CHALLENGE with KEY})$$

The above mentioned techniques would be highly effective against the password stealing attacks. However, we observed that they come with few significant disadvantages. They are discussed as follows.

- 1) They require a well planned infrastructure requiring enormous investments on the part of the web application developers.
- 2) It raises the customers' expenses in terms of purchasing and maintaining the smart card devices/ tokens etc.
- 3) Such schemes generally face high resistance from the users as they are more used to typing a username and a password.
- 4) The adoption of these schemes is limited to financial organisations such as banks for the purpose of online banking. For most other web applications facilitating email services, social networking, for instance, the expense of the schemes exceeds the value of the assets, rendering them useless.

However it is important to note that these schemes are worth all the investment in case of financial organisations such as banks since security is the topmost priority in their case!

8 Our Solution

Our solution is a more cost effective version of Challenge Response method discussed above. As per this technique, the user at the time of creating his user account with the website would be given a set of symbols/images for his personal identification. At the time of authentication, the user will not only be asked to enter his user-id and password, but also be presented with a large set of symbols on the screen.

Now the user will be asked if he could find any of those symbols on the screen that were allotted to him. If yes then he would have to type in their relative positions on the screen. And if no, then he will simply have to type in a zero. This process could be repeated again with different set of symbols displayed on the screen to further strengthen the security. We built a prototype website implementing this solution. The following steps demonstrate the scheme in more detail.

- Our server initiates a download activity for the client (on a secure channel) once he registers with our website.
- The client receives a total of 4-6 images in a zipped folder that were fetched randomly from our database. This set of images is now stored along with the user's credentials in a database on the server side.
- At the time of login, the user is first asked to enter his user-id and password.
- Once the correct user-id, password combination is entered, he is presented with 20 random images.
- These displayed images may or may not contain the images possessed by the user. We are randomly fetching these images from a database. However, a well planned algorithm would better suit the needs of scalability i.e. when we have a large number of images in our database.
- In case one or more user's images are present in the displayed list, he enters their relative positions. Else he enters a zero.
- This procedure is repeated once again irrespective of whether the correct positions were entered or not in the previous iteration. This is done to minimize the possibility of guessing the images by an attacker.
- Only after the correct position of images is entered both the times, the user is authenticated.

8.1 Advantages

- 1) The method can successfully thwart keylogger malware attacks because even though a keylogger may record the user's passwords, it would not be able to guess the symbols. Moreover the mechanism would become more reliable if the user is periodically asked to change his password along with the set of symbols.
- 2) Implementation overhead is considerably low.
- 3) The method is not complicated due to which it will hardly face any resistance from the users, if at all.
- 4) Most of the anti-keyloggers fail to detect rootkit based malware. With this mechanism in place even a rootkit based keylogger could be effectively countered.

8.2 Disadvantage

The scheme could fail if the keylogger malware on the victim's machine has screen capture facility. However, since all the symbols that the user was allotted need not be displayed at a time in the challenge, the attacker may be able to obtain only a partial list of symbols at a time. So if the user is periodically made to change his set of symbols, say in a period of 3 months, the failure rate of this technique would be minimal.

9 Conclusion

- At present, keyloggers – together with phishing and social engineering methods – are the most commonly used techniques employed to carry out cyber fraud.
- Statistics prove that there has been a tremendous rise in the number of malicious programs that have keylogging functionality.

- Malware sophistication is constantly evolving. There has been an increase in the use of rootkit technology in keylogging software, empowering them to evade detection and persistently gather sensitive information.
- Dedicated protection by using a combination of tools can mitigate the risk posed by this category of malware.
- Our approach aims at countering the password stealing attacks by imposing a kind of site key challenge, such that even if a keylogger records the password of the user, the attacker is still incapable of hacking into the user's account.

References

1. Keyloggers, Increasing Threat to Computer Security. IEEE Technology And Society Magazine, Fall (2009)
2. Ries, C.: Inside Windows Rootkits. VigilantMinds Inc. (May 2006)
3. Trend Micro Threat Encyclopedia | Latest information on malware, spam, malicious URLs, vulnerabilities, <http://about-threats.trendmicro.com/threatencyclopedia.aspx?language=us&tab=malware>
4. Hackers infected thousands of PCs with Zeus trojan to steal millions, <http://news.techworld.com/security/3241594/police-arrest-gang-behind-20-million-online-bank-fraud/>
5. Keyloggers: How they work and how to detect them - Securelist, http://www.securelist.com/en/analysis/204791931/Keyloggers_How_they_work_and_how_to_detect_them_Part_1
6. VirusTotal - Free Online Virus, Malware and URL Scanner, <http://www.virustotal.com>
7. QFX Software - Anti-Keylogging Software, <http://www.qfxsoftware.com/>
8. Anti Keylogger Shield, http://amictools.com/v-anti_keylogger_shield.html
9. GuardedID® – Next Generation Security, Anti Keylogger to Protect against Keylogger attacks, Identity Theft, and Clickjacking, <http://www.guardedid.com/demo.aspx>
10. Meyer, R.: Secure Authentication on the internet. SANS Institute (2007)

Cancelable Biometrics for Better Security and Privacy in Biometric Systems

Sanjay Ganesh Kanade*, Dijana Petrovska-Delacrétaz, and Bernadette Dorizzi

Institut TELECOM: TELECOM SudParis, Département Electronique et Physique,
9 Rue Charles Fourier, 91011, Evry, France

{Sanjay.Kanade,Dijana.Petrovska,Bernadette.Dorizzi}@it-sudparis.eu

Abstract. We present a simple scheme with three main features: (1) it induces revocability in biometrics based user authentication systems, (2) protects the biometric information, and (3) improves the verification performance. The user's biometric feature vector is shuffled with a user specific shuffling key to transform into a revocable template. Comparison between different templates is carried out in the transformed domain. If the template is compromised, it can be replaced with another template obtained by changing the shuffling key. This scheme makes cross-matching between databases impossible by issuing different templates for different applications. The performance evaluation of the system is carried out on two modalities: iris and face using publicly available databases. This scheme significantly improves the verification performance of the underlying biometric system, e.g., it reduces the Equal Error Rate (EER) from 1.67% to 0.23% on the NIST-ICE iris database. The EER on the NIST-FRGCv2 face database reduces from 8.10% to zero.

Keywords: Revocability, Cancelable biometrics, Iris, Face, Security and privacy.

1 Introduction

Biometrics is defined as automatic recognition of persons based on their physical or behavioral characteristics (such as fingerprint, iris, face, etc.). Since the biometric characteristics are implicitly associated with a user, they provide a strong proof of his identity. In the existing biometric systems that we denote as 'classical biometric systems' (shown in Fig. 1), the information needed for further comparisons, denoted as biometric reference or template, is stored in a database. However, because of the permanent association of biometric characteristics with the user, these templates remains substantially similar across databases if the modality and the biometric algorithm are the same, e.g., for minutiae based fingerprint systems, minutiae sets extracted from the same fingerprint in different

* The first author was supported by the French "Agence Nationale de la Recherche (ANR)" project BIOTYFUL, (ANR-06-TCOM-018).

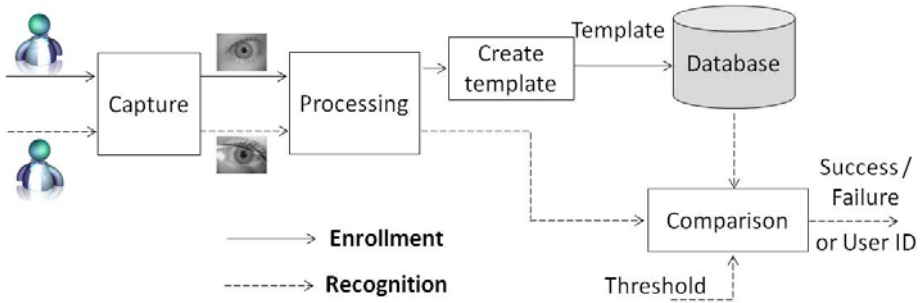


Fig. 1. Basic idea of a biometric based person recognition system. In verification mode, the result of the comparison is either success or failure. In identification mode, the result of comparison is the User ID.

systems are similar. If such template is compromised, it is not possible to replace it with a new one because the biometric characteristics (from which this information is extracted) are permanently associated with their owners. In other words, it is not possible to revoke or cancel a template. This phenomenon is called as lack of revocability.

The permanent association of biometric data with the user leads to another problem. Since the templates in all the systems based on the same biometric characteristic and using same biometric algorithms are similar, a compromised template from one biometric database can be used to access information from another system. This can be referred to as cross-matching between databases and is a threat to privacy. Moreover, in some cases, the stored information can be used to create a dummy representation of the biometric trait which can be used to access the system [2,6,25,7]. For example, a dummy finger can be constructed from a fingerprint image.

Because of these reasons, the property of cancelability or revocability is becoming a necessity. In order to induce revocability into biometric systems, cryptographic techniques are a good candidate. Many systems that induce these characteristics are proposed in literature. A summary of such systems is presented in Section 2. In this paper, we present a simple shuffling scheme to create cancelable templates from biometric data. This system was first proposed in our earlier paper on crypto-biometric key regeneration [12]. This scheme involves two factors: biometrics and a shuffling key. Because of this additional parameter, the proposed scheme significantly improves the verification performance of the baseline biometric system.

In general, an authentication system should possess following characteristics:

1. **Identity verification and non-repudiation:** The system should be able to confirm the identity of the user with high degree of confidence. It also indicates that the system should resist repudiation attempts carried out by the users. Involvement of biometrics helps achieve this property.

2. **Revocability:** If the stored user template is compromised, it should be possible to cancel that template and reissue a new one. Additionally, the newly issued template should not match with the previously compromised template. Thus revocability does not mean just to cancel the old template and issue a new one; it also means that, the authentication rights of the old authenticator are revoked. The system should be able to reject a person if he provides the authenticator linked with the old template. Note that, biometrics alone cannot provide this property because biometric characteristics cannot be changed while systems using passwords and tokens have excellent revocability.
3. **Template diversity:** It should be possible to issue different templates for different applications related to the same user. These templates should not match with each other and should make cross-matching impossible. Password and token based systems are good at that, though practically, password diversity can be argued. Biometrics, by itself, cannot have template diversity.
4. **Privacy protection:** These systems should protect the privacy of biometric data, privacy of information protected by the system, and user identity privacy.

The system presented in this paper satisfies all these desired characteristics. In this scheme, the biometric features are combined with a user specific random key to obtain a revocable template. The scheme improves the biometric system performance in an ideal case when the user specific keys are kept secret. If the keys for all the users are stolen, the system is as secure as the underlying biometric system.

The remainder of the paper is organized as follows: recent developments in the field of revocable biometrics based authentication are given in section 2. In section 3, the algorithm to obtain revocable iris template is described. The experimental setup for the performance evaluation, including baseline biometric systems, databases and experimental protocols are explained in section 4 and results are reported in section 5 along with experimental security analysis. Finally, the conclusions and perspectives are given in section 6.

2 Cancelable Biometrics: Related Work

There are many solutions found in literature which aim at inducing cancelability/revocability in biometric systems. These systems apply some sort of (one-way) transformation on the biometric data. Some of these transformation methods include, Cartesian, polar and functional transformations of Ratha et al. [23,24], BioHashing of Jin et al. [10], cancelable filters of Savvides et al. [26], improved BioHashing of Lumini and Nanni [17], Revocable biotokens of Boulton et al. [4], and transformations proposed by Maiorana et al. [18].

The drawback of many of these cancelable biometric systems is that their performance degrades compared to the baseline biometric system. In some cases,

the performance improves, however, the improvement is because of the additional parameter (such as password, PIN, key, token, etc.). Such systems should be analyzed for their verification performance in the stolen key (also called as stolen token) scenario. Such analysis is not reported in most of these works. For BioHashing based systems, the performance in the stolen key scenario degrades compared to the baseline biometric system.

Biometric-based cryptographic key (re)generation systems [11, 27, 9, 28, 5, 19, 12, 14, 13] can provide cancelable templates based on biometrics, but their main aim is to obtain a cryptographic key. Hence we will not discuss these schemes here.

Our previous work on cryptographic key regeneration [12] incorporates the shuffling scheme, presented in this paper, to have better separation between genuine and impostor users. But in that work, we do not analysis the scheme from a revocable biometrics point of view which we present in this paper. The biometric data shuffling scheme is described in the following section.

3 A Biometric Data Shuffling Scheme to Create Cancelable Biometric Templates

The shuffling scheme described in this section can work with any biometric modality provided the biometric features are represented as an ordered set. In this scheme, a randomly generated shuffling key is used to shuffle the biometric data. The shuffled biometric data represents the cancelable template. It is not feasible to recover the original biometric data from this cancelable template. This scheme can be considered analogous to classical symmetric encryption technique because, as in encryption, a key is used to protect the biometric data. But contrary to classical encryption, the user discrimination properties of biometric data are retained by the transformed data, and hence, comparison between two such transformed biometric data can be carried out in the transformed domain. The shuffling technique is explained in details in the next subsection.

3.1 The Shuffling Technique

The shuffling scheme that we introduce requires a binary shuffling key \mathbf{K}_{sh} of length L_{sh} . Since this key is a long bit-string, it is stored on a secure token or it can be obtained using a password. The biometric feature vector is divided into L_{sh} blocks each of which has the same length. To start the shuffling, these L_{sh} blocks of the feature vector are aligned with the L_{sh} bits of the shuffling key \mathbf{K}_{sh} . In the next step, two distinct parts containing biometric features are created: the first part comprises all the blocks corresponding to the positions where the shuffling key bit value is one. All the remaining blocks are taken into the second part. These two parts are concatenated to form the shuffled biometric feature vector which is treated as a revocable template. Figure 2 shows a schematic diagram of this shuffling scheme.

The original and shuffled feature vectors have one-to-one correspondence. A block from the original vector is placed at a different position in the shuffled

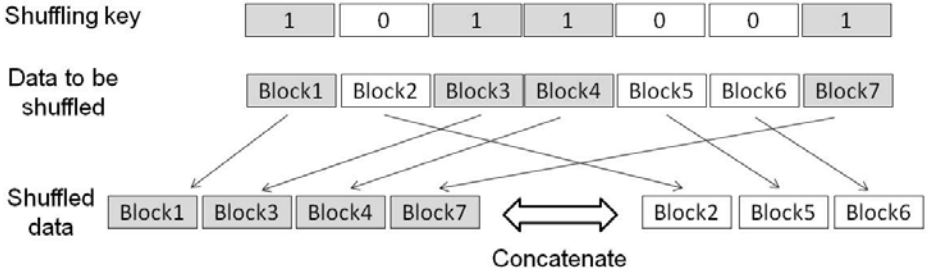


Fig. 2. The shuffling scheme

vector. Thus, only the alignment of the feature blocks is changed by the scheme with no change in the actual values of the features. The length of the biometric feature vector does not change because of the shuffling. Hence, the matching algorithms used for calculating the similarity (or dis-similarity) score between two biometric feature vectors are still applicable for the shuffled data.

Note that the effectiveness of this scheme is because it changes the alignment of the feature vectors. If the feature vectors do not require any particular order (e.g., fingerprint minutiae sets), this system is ineffective. This system can work only if the biometric data is in form of an ordered set.

3.2 Advantages of Using the Proposed Shuffling Scheme

The proposed shuffling scheme has the following advantages:

1. **Revocability:** The shuffled feature vector, which is treated as a cancelable template, is a result of combination of an intrinsic identifier (i.e., a biometric characteristic) and an assigned identifier (the shuffling key). Therefore, in case of compromise, it can be canceled and a new template can be generated by changing the shuffling key \mathbf{K}_{sh} (the assigned credential).
2. **Performance improvement:** Another advantage of using the shuffling scheme is that it improves the verification performance. The shuffling process changes the alignment of the feature vector blocks according to the shuffling key. When two biometric feature vectors are shuffled using the same shuffling key, the absolute positions of the feature vector blocks change but this change occurs in the same way for both of the biometric feature vectors. Hence, the Hamming distance (in case of binary vectors) between them does not change. On the other hand, if they are shuffled using two different keys, the result is randomization of the feature vectors and the Hamming distance increases. In fact, the shuffling process acts like a randomizer and moves the average Hamming distance for such cases close to 0.5.

A unique shuffling key is assigned to each subject during enrollment and he has to provide that same key during every subsequent verification. This means, in ideal case, that the genuine users always provide the correct shuffling key and hence, the Hamming distance for genuine comparisons remain

unchanged. On the contrary, in case of random impostor attempts where an impostor tries to get verified with his own credentials, he provides his biometric data along with his shuffling key (or a random shuffling key) to match against other users. The feature vectors for such impostor comparisons are shuffled with two different shuffling keys and the result is that the Hamming distances increase. This effect can be seen in Fig. 3. The separation between the genuine and impostor Hamming distance distributions shows the ability of the system to distinguish genuine users from impostors. As can be seen from Fig. 3, shuffling increases the separation between the two distributions. In this way, the shuffling scheme improves the verification performance of the system.

3. **Template diversity:** With the help of the shuffling technique, different templates can be issued for different applications by using different shuffling keys with the same biometric data. This particularly helps to avoid cross-database matching. In order to make the template-diversity effective, it is suggested that the shuffling key should be generated randomly and protected by a password.
4. **Protection against stolen biometric data:** If a feature vector is shuffled using two different shuffling keys, the resulting shuffled vectors appear to be originating from two different subjects. They can be seen as comparing two random sequences and hence they do not match. Therefore, if a stolen biometric data of a legitimate person is used by an impostor to get verified, the system can still resist such attack due to the use of shuffling key.
5. **Biometric data protection:** It is not computationally feasible to recover the original biometric feature vector from the shuffled data without the proper shuffling key. However, as in classical encryption, the security depends on the secrecy of the shuffling key.

These effects can be better understood from the experimental results and analysis presented in the next section.

4 Experimental Setup

The cancelable biometric system is based upon an underlying baseline biometric system. Therefore, for fair comparison, first the biometric verification performance of the baseline biometric system is reported followed by the performance of the proposed cancelable system.

The proposed cancelable biometric system is evaluated on two biometric modalities: iris and face. For iris, the Open Source Iris Recognition System (OSIRIS) described in [22] and available online at [1] is used to extract binary iris code features from the iris images. The CBS database [22] (BiosecureV1 OKI device subset) is used for development in order to find out the optimum length of the shuffling key. The system is then evaluated on the NIST-ICE database [21]. As described in the ICE evaluation, we carried out two experiments: ICE-Exp1 involving comparisons between right-eye images whereas ICE-Exp2 involving

left-eye images. In total, 12,214 genuine and 1,002,386 impostor comparisons are carried out in ICE-Exp1, whereas in ICE-exp2, 14,653 genuine, and 1,151,975 impostor comparisons are performed.

For face, a Gabor filter based approach is applied to extract features from the face image [16]. The face image is first geometrically normalized using the CSU Face Recognition System [3], and then processed using log-Gabor filters having four scales and eight orientations using the MATLAB source code available at [15]. Magnitude of the filtered output is calculated, downsampled, and concatenated to form a 3,200-element feature vector. The values in this vector are then binarized to obtain a 3,200-bit string called face code. The binarization process used is fairly simple. The median of the values in a feature vector is taken as a threshold. The elements having higher value than the threshold are made one while the remaining are made zeros.

The development and evaluation data sets for face experiments are from a subset of the NIST-FRGCv2 database [20]. This subset is composed of 250 subjects each of which has 12 images. Data from the first 125 subjects are used for development and the remaining 125 subjects are used for evaluation. For each subject, there are eight images captured in controlled lighting conditions while four images in uncontrolled conditions.

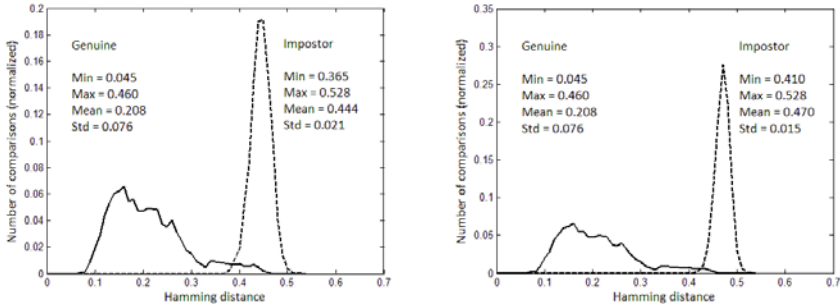
Two separate experiments are carried out during development as well as evaluation: FRGC-Exp1* – where the enrollment as well as test images are captured under controlled conditions, and FRGC-Exp4* – in which the enrollment images are from controlled conditions while the test images are from uncontrolled conditions. For the FRGC-Exp1*, 3,500 genuine and 496,000 impostor comparisons are carried out while for FRGC-exp4*, 4,000 genuine and 496,000 impostor comparisons are performed.

5 Experimental Results and Security Analysis

5.1 Results and Security Analysis on Iris Modality

Results on Iris Modality. The genuine and impostor Hamming distance distributions for the CBS-BiosecureV1 data set before and after shuffling are shown in Fig. 3. As described in Section 3.2, the shuffling process increases the impostor Hamming distances while the genuine Hamming distances remain unchanged. This can be seen from the Fig. 3. In this figure, the mean of the impostor Hamming distance distribution of the baseline system shifts from 0.44 to 0.47 when the shuffling scheme is applied. Note that, the genuine Hamming distance remains unchanged. This reduces the overlap between the genuine and impostor distribution curves which improves the user discrimination capacity of the system thereby increasing the verification accuracy.

The better separation between genuine and impostor Hamming distance distribution curves improves the verification performance of the system. The verification performance in terms of Equal Error Rate (EER) on the development database (CBS database) is reported in Table 1.



(a) Baseline iris biometric system

(b) Proposed cancelable biometric system (baseline iris system with shuffling)

Fig. 3. Normalized Hamming distance distributions for genuine and impostor comparisons on the CBS-BioSecureV1 [22] development data set

Table 1. Verification results of the baseline biometric system (which is based on the OSIRISv1) and the proposed cancelable system on iris modality; in terms of EER in %. Values in bracket indicate the error margins for 90% confidence intervals.

Experiment	CBS-BiosecureV1	ICE-Exp1	ICE-Exp2
Baseline	2.63[±0.34]	1.71[±0.11]	1.80[±0.10]
Proposed cancelable	0.93[±0.20]	0.23[±0.04]	0.37[±0.05]
OSIRISv1 [22]	2.83[±0.35]	1.52[±0.12]	1.71[±0.12]
Stolen biometric scenario	1.50[±0.26]	0.27[±0.08]	0.44[±0.09]
Stolen key scenario	2.63[±0.34]	1.71[±0.11]	1.80[±0.10]

The system is then evaluated on the evaluation (NIST-ICE) database. These results are also reported in Table 1. As noted before, we carried out separate experiments according to the common protocol for ICE evaluation for right (ICE-Exp1) and left (ICE-Exp2) iris comparisons. A clear improvement in performance can be seen by comparing the EER of the baseline system with the proposed cancelable system. For example, in case of ICE-Exp1, the EER for the baseline system is 1.71% which reduces to 0.23% when the cancelable scheme is applied. Similarly, for the ICE-Exp2, the EER reduces from 1.80% to 0.37% because of the shuffling scheme. For the sake of comparison, the EER values reported in the documentation of the OSIRISv1 on these two data sets are also reported in this table.

Security Analysis on Iris Modality. The cancelable biometric system proposed in this chapter has two factors: biometrics and a shuffling key. In order to test the robustness of the system, we carried out the performance evaluation in two extreme hypothetical impostor scenarios: (i) stolen biometric and (ii) stolen key.

¹ The baseline iris system is based on OSIRISv1; the difference is that the matching module is re-implemented to cope with the iris rotations.

In the stolen biometric scenario, we consider a hypothetical extreme situation when the biometric information for all the users is stolen. Here, an impostor will try to provide the stolen biometric data along with a wrong shuffling key. In this situation, the EER increases compared to that of the cancelable system with both factors secret. But, it is still less than the EER of the baseline biometric system. For example, as shown in Table 1, for the ICE-Exp1, the EER of the cancelable system is 0.23% when both the factors are secret. Considering that the iris image is stolen for all the users, the EER increases to 0.27% which is still less than the EER for baseline system (1.71%). Thus, use of the shuffling scheme prevents the impostors from being successfully verified using stolen biometric data.

In the stolen key scenario, we consider another extreme situation when the shuffling keys of all the users are compromised. As in the stolen biometric scenario, the EER increases compared to that of the cancelable system having both parameters secret. But, the EER is equal to the EER of the baseline biometric system meaning that the system in this stolen key scenario is still as good as the baseline biometric system (see Table 1). In fact, the proposed shuffling scheme is such that, it increases the Hamming distance between two iris codes if and only if they are shuffled with different keys. If the same key is used to shuffle two codes, the Hamming distance remains intact. Thus in the stolen key scenario, the Hamming distance distribution is exactly the same as that for the baseline system, and hence, yields the same result as that of the baseline biometric system. This is a distinct advantage of our system over other cancelable systems found in literature. For most of the cancelable systems found in literature, the performance degrades if the keys (or the cancelable parameters used) are compromised. Only the Farooq et al. [8] system is shown to have the performance equal to the baseline biometric system in the stolen key scenario.

Detection Error Tradeoff (DET) curves for the proposed cancelable system along with the security threats are shown in Fig. 4 for the iris modality. These curves show the performance on the evaluation database – the NIST-ICE database – for the ICE-Exp1 experiment. The DET curves for the baseline system and that for the stolen key scenario overlap with each other which indicates that the performance of the system in stolen key scenario is same as the baseline system.

The stolen biometric scenario also proves the template diversity property. It shows that, if the biometric feature vector is shuffled with two different keys, the two shuffled codes appear to be random and do not match. In order to prove our point, we carried out an additional test. We shuffled one iris code with 100,001 randomly generated shuffling keys. The first shuffled iris code is compared with the remaining 100,000 shuffled iris codes. The distribution of Hamming distances obtained from these comparisons is shown in Fig. 5. This distribution is also close to the random impostor distribution which validates our claim of template diversity.

In case of compromise, the cancelable template can be revoked. In order to revoke the template, the user is asked to re-enroll into the system. The fresh

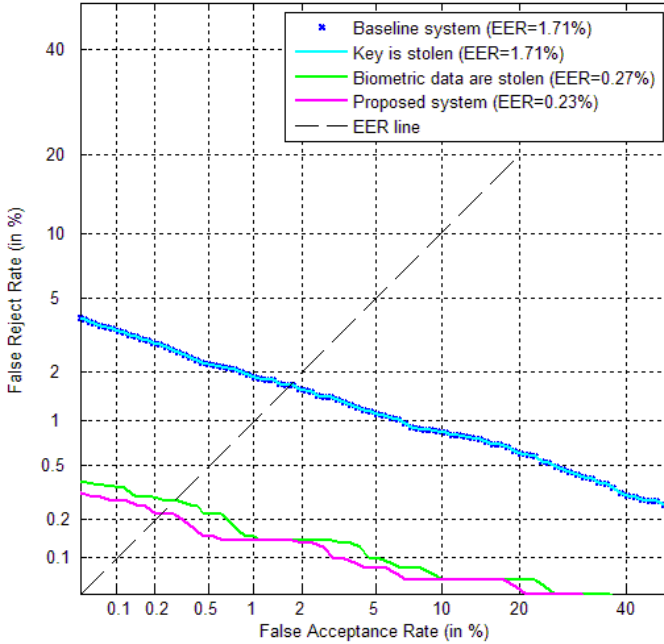


Fig. 4. DET curves for the proposed system performance along with the possible security threats for iris modality on the NIST-ICE database(evaluation data set) [21]; ICE-Exp1

biometric data is shuffled with a newly generated random shuffling key. Since this shuffling key is different than the one used earlier in enrollment, the old template and the newly issued template cannot match with each other. If an attacker obtains an iris code of the user from previously compromised template or from another biometric system, that iris code cannot be used by the impostor to get verified because the new shuffling key resists such attacks.

5.2 Results and Security Analysis on Face Modality

Results on Face Modality. The Hamming distance distribution curves for genuine and impostor comparisons before and after shuffling on the development data sets are shown in Fig. 6. The curves for both, FRGC-Exp1* and FRGC-Exp4*, experiments are shown.

As was observed in case of iris, the impostor Hamming distances increase because of the shuffling process. Note that the genuine Hamming distances remain unchanged. A clear separation between genuine and impostor Hamming distance distributions is observed for both the experiments. This complete separation results in zero EER. The results of the proposed cancelable system for the FRGC-Exp1* and FRGC-Exp4* on the development data sets are reported in Table 2.

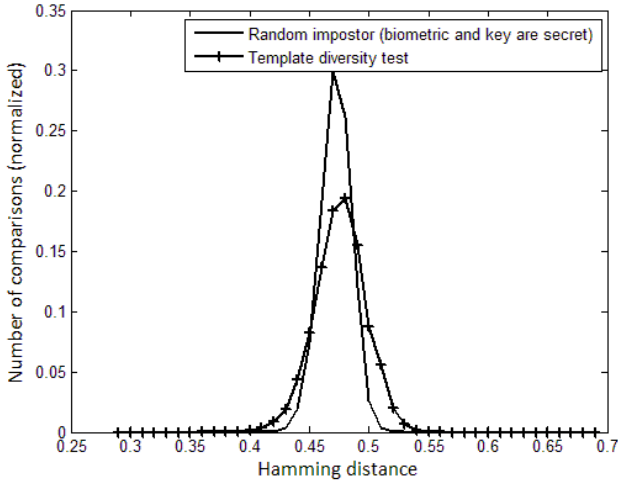
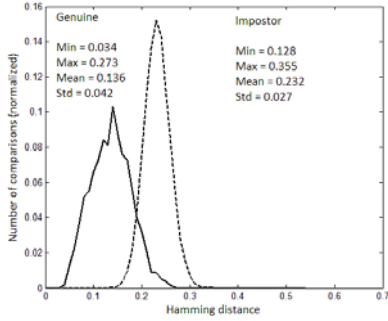


Fig. 5. Impostor Hamming distance distributions for the proposed system along with the Hamming distance distributions for the template diversity test on iris modality on the NIST-ICE database [21] (ICE-Exp1)

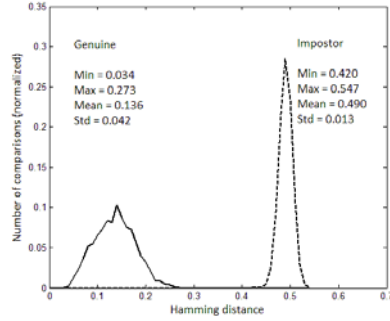
Table 2. Verification results of the proposed cancelable system on face modality along with the security analysis in terms of EER in %. Values in bracket indicate the error margins for 90% confidence intervals.

Test	Development set		Evaluation set	
	FRGC-Exp1*	FRGC-Exp4*	FRGC-Exp1*	FRGC-Exp4*
Baseline	8.10[±0.41]	35.90[±0.68]	7.65[±0.40]	35.00[±0.68]
Proposed cancelable	0	0	0	0
Stolen biometric	0	0	0	0
Stolen key	8.10[±0.41]	35.90[±0.68]	7.65[±0.40]	35.00[±0.68]

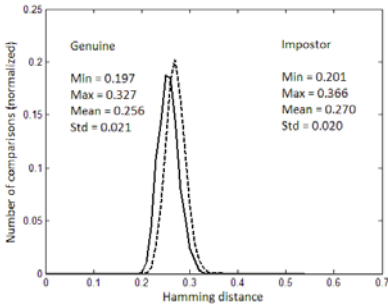
Note that, the improvement in performance is because of the increase in impostor Hamming distances. The shuffling scheme works as a randomization process which shifts the mean of the impostor Hamming distance distribution close to 0.5. Therefore, if the mean of the original (un-shuffled) impostor Hamming distance distribution is small, the improvement in performance will be more prominent. This can be visualized by comparing the improvements for iris and face modalities. For example, on the development data set CBS-BiosecureV1 for iris, as shown in Fig. 3, the average impostor Hamming distance for iris is 0.44, which after shuffling, increases to 0.47. Similarly, for face, on the development data set Exp1 (Fig. 6), the average impostor Hamming distance is 0.23, which moves to 0.49 after shuffling. Thus, the increase in the separation between



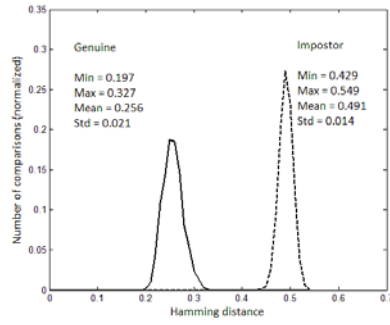
(a) Baseline face biometric system (FRGC-Exp1*)



(b) Baseline face system with shuffling (FRGC-Exp1*)



(c) Baseline face biometric system (FRGC-Exp4*)



(d) Baseline face system with shuffling (FRGC-Exp4*)

Fig. 6. Normalized Hamming distance distributions for genuine and impostor comparisons on the NIST-FRGCv2 development data set for FRGC-Exp1* and FRGC-Exp4*

genuine and impostor Hamming distance curves is more in case of face than for iris. Therefore, the improvement in performance is higher in case of face than in case of iris.

The proposed cancelable system is then evaluated on the evaluation data sets. As it is seen for the experiments on development sets, a clear separation is obtained on the evaluation sets also. The outcome of this separation is zero EER as reported in Table 2.

Security Analysis on Face Modality. The experimental security analysis of the proposed system carried out for the iris modality is also performed for the face modality. The two scenarios: (i) stolen biometric scenario and (ii) stolen key scenario, are followed. During these tests, it is observed that the proposed cancelable system behaves in a similar way as it did on iris. The performance in

case of the stolen biometric case remains unchanged. In the stolen key scenario, the performance is exactly the same as that of the baseline biometric system. The results for these tests in terms of EER are reported in Table 2.

6 Conclusions and Perspectives

Classical biometric systems lack the important properties of revocability and template diversity because the biometric traits are permanently associated with the user. Cancelable biometric systems overcome these drawbacks of classical biometric systems. The shuffling scheme proposed in this paper employs a randomly generated shuffling key to randomize the biometric feature codes. The shuffled feature vectors act as cancelable templates. The system can issue different templates for different applications using the same biometric which preserves privacy. If the stored template is compromised, it can be canceled and a new template can be issued by changing the shuffling key. Such use of shuffling key prevents an attacker from getting verified by providing the compromised template or stolen biometric data. One distinct advantage of this system is that the performance of the baseline system increases by more than 80% due to shuffling. And even if one of the two secret factors, the biometric data and the shuffling key, is compromised, the EER of the system in such scenario still remains less than or equal to that of the baseline biometric system.

The drawback of this shuffling scheme is that it is not noninvertible. Practically, it works as a classical symmetric encryption where data can be encrypted by a key and the encrypted data can be decrypted by providing the same key. If an attacker succeeds to obtain the shuffling key, he can de-shuffle the cancelable template to obtain the reference biometric data. However, when such compromise is detected, the system can revoke the old template and issue a new one and the earlier attack becomes irrelevant.

A limitation of this shuffling scheme in its current form is that it can only be applied to biometric systems when the templates are in form of an ordered set. It cannot be applied to unordered sets such as a set of fingerprint minutiae.

The proposed shuffling scheme is very effective and therefore can be used as a means to induce revocability in other key regeneration systems.

References

1. http://svnnext.it-sudparis.eu/svnview2-eph/ref_syst/
2. Adler, A.: Sample Images Can be Independently Restored from Face Recognition Templates. In: Canadian Conference on Electrical and Computer Engineering, CCECE (2003)
3. Beveridge, J.R., Bolme, D., Raper, B.A., Teixeira, M.: The CSU Face Identification Evaluation System. *Machine Vision and Applications* 16(2), 128–138 (2005)
4. Boulton, T.E., Scheirer, W.J., Woodworth, R.: Revocable fingerprint biotokens: Accuracy and security analysis. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (June 2007)

5. Bringer, J., Chabanne, H., Cohen, G., Kindarji, B., Zémor, G.: Optimal Iris Fuzzy Sketches. In: IEEE Conference on Biometrics: Theory, Applications and Systems (2007)
6. Cappelli, R., Lumini, A., Maio, D., Maltoni, D.: Can Fingerprints be Reconstructed from ISO Templates? In: 9th International Conference on Control, Automation, Robotics and Vision, ICARCV (2006)
7. Cappelli, R., Maio, D., Lumini, A., Maltoni, D.: Fingerprint Image Reconstruction from Standard Templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(9), 1489–1503 (2007)
8. Farooq, F., Bolle, R.M., Jea, T.Y., Ratha, N.: Anonymous and Revocable Fingerprint Recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2007)
9. Hao, F., Anderson, R., Daugman, J.: Combining crypto with biometrics effectively. *IEEE Transactions on Computers* 55(9), 1081–1088 (2006)
10. Jin, A.T.B., Ngo, D., Ling, C., Goh, A.: Biohashing: two factor authentication featuring fingerprint data and tokenised random number. *Pattern Recognition* 37(11), 2245–2255 (2004)
11. Juels, A., Wattenberg, M.: A fuzzy commitment scheme. In: Proceedings of the Sixth ACM Conference on Computer and Communication Security (CCCS), pp. 28–36 (1999)
12. Kanade, S., Camara, D., Krichen, E., Petrovska-Delacrétaz, D., Dorizzi, B.: Three Factor Scheme for Biometric-Based Cryptographic Key Regeneration Using Iris. In: The 6th Biometrics Symposium (BSYM) (September 2008)
13. Kanade, S., Petrovska-Delacrétaz, D., Dorizzi, B.: Generating and Sharing Biometrics Based Session Keys for Secure Cryptographic Applications. In: IEEE International Conference on Biometrics: Theory, Applications, and Systems, BTAS (2010)
14. Kanade, S., Petrovska-Delacrétaz, D., Dorizzi, B.: Obtaining Cryptographic Keys Using Feature Level Fusion of Iris and Face Biometrics for Secure User Authentication. In: IEEE CVPR Workshop on Biometrics (June 2010)
15. Kovese, P.: Matlab and octave functions for computer vision and image processing (2005), <http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/>
16. Lades, M., Vorbrüggen, J.C., Buhmann, J., Lange, J., von der Malsburg, C., Wüertz, R.P., Konen, W.: Distortion Invariant Object Recognition in the Dynamic Link Architecture. *IEEE Transactions on Computers* 42(3), 300–311 (1993)
17. Lumini, A., Nanni, L.: An improved biohashing for human authentication. *Pattern Recognition* 40(3), 1057–1065 (2007)
18. Maiorana, E., Campisi, P., Ortega-Garcia, J., Neri, A.: Cancelable Biometrics for HMM-based Signature Recognition. In: IEEE Conference on Biometrics: Theory, Applications and Systems, BTAS (2008)
19. Nandakumar, K., Jain, A.K., Pankanti, S.: Fingerprint-based fuzzy vault: Implementation and performance. *IEEE Transactions of Information Forensics and Security* 2(4), 744–757 (2007)
20. National Institute of Science and Technology (NIST): Face Recognition Grand Challenge (2005), <http://www.frvt.org/FRGC/>
21. National Institute of Science and Technology (NIST): Iris Challenge Evaluation (2005), <http://iris.nist.gov/ice>

22. Petrovska-Delacrétaz, D., Chollet, G., Dorizzi, B. (eds.): Guide to Biometric Reference Systems and Performance Evaluation. Springer, Heidelberg (2009)
23. Ratha, N.K., Connell, J.H., Bolle, R.M.: Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal* 40(3), 614–634 (2001)
24. Ratha, N.K., Chikkerur, S., Connell, J.H., Bolle, R.M.: Generating cancelable fingerprint templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(4), 561–572 (2007)
25. Ross, A., Shah, J., Jain, A.K.: From template to image: Reconstructing fingerprints from minutiae points. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(4), 544–560 (2007)
26. Savvides, M., Kumar, B.V., Khosla, P.: Cancelable biometric filters for face recognition. In: Proceedings of the 17th International Conference on Pattern Recognition (ICPR 2004), vol. 3, pp. 922–925 (August 2004)
27. Soutar, C., Roberge, D., Stoianov, A., Gilroy, R., Kumar, B.V.: Biometric encryption. In: *ICSA Guide to Cryptography*. McGraw-Hill, New York (1999)
28. Uludag, U., Jain, A.: Securing fingerprint template: Fuzzy vault with helper data. In: Proc. of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, pp. 163–170 (June 2006)

Advanced Clustering Based Intrusion Detection (ACID) Algorithm

Samarjeet Borah, Debaditya Chakravorty,
Chandan Chawhan, and Aritra Saha

Department of Computer Science & Engineering
Sikkim Manipal Institute of Technology,
Majitar, East Sikkim-737136
{samarjeetborah,debaditya.chakravorty,chandanchawhan,
aritra.s484}@gmail.com

Abstract. Computer security or network security has become one of the biggest issues now-a-days. Intrusion Detection process detects malicious attacks which generally includes theft of information or data. Traditional IDS (Intrusion Detection System) detects only those attacks which are known to them. But they rarely detect unknown intrusions. Clustering based method may be helpful in detecting unknown attack patterns. In this paper an attempt has been made to propose a new intrusion detection method based on clustering. The algorithm is experimented with KDD99 dataset and is found to produce satisfactory results.

Keywords: Security, Network, Intrusion Detection, Clustering, IDS, Training Data Set, Knowledge Base.

1 Introduction

1.1 Intrusion

Intrusion [1] is defined as “the act of wrongfully entering upon, seizing, or taking possession of the property of another”. They are serious threats in a network environment. They attempt to destabilize the network and attempt to gain unauthorized access to the network.

1.2 Intrusion Detection and Its Importance

Intrusion Detection is the process of detecting the malicious attacks or threats to any computer or network. The basic task of Intrusion Detection is to audit the log data of a computer which includes network as well as host based data.

Intrusion detection process helps to make the network more secure for data transmission. An Intrusion Detection System (IDS) helps to determine the attackers and ensures secure transmission of data. It also facilitates to stabilize and increase the lifetime of the network.

2 Clustering

2.1 Introduction

The process of grouping a set of physical or abstract objects into classes of similar objects is called clustering [2]. A cluster comprises of objects that are similar to objects within the same cluster and are different from those present in other clusters. By clustering, one can identify crowded and sparse regions in a data set and therefore recognize patterns and interesting co-relations among data attributes. So, it can be used in applications involving pattern matching and data analysis [7]. These techniques are always found helpful in detecting unknown patterns.

2.2 Clustering in Intrusion Detection

Clustering is used to group data instances of the training set together into clusters using a simple distance-based metric. Once the data is clustered, certain techniques can be used to classify these clusters either as anomaly or normal instances [3][4][6]. After the clustering process is over and the clusters have been labeled accordingly, the system is ready to accept the network data instances and compare them with the already formed clusters and thus classify them as possible threats or safe to pass instances, thus detecting possible intrusions.

3 Advanced Clustering Based Intrusion Detection (ACID) Algorithm

3.1 Introduction

ACID is not a signature based algorithm but rather uses Misuse Detection [4][9] and Anomaly Detection modules [4][9]. During the clustering and classification of training data, the classification process is supervised using the labeled KDD data set [5][8][9] and during Misuse Detection, the algorithm efficiently classifies the different attack types without being supervised. Thus, it can possibly identify unknown attack types at later stages.

3.2 Analysis of the Problem

Connection packets in a communication network comprise of various fields which give these packets their unique signatures. The signatures of intrusion data are different from those of normal data; thus, they can be distinguished by comparing their signatures.

3.3 Solution Strategy

The problem can be solved in three phases:

- i. Perform clustering of the training data set so that data of similar types fall within same clusters.

- ii. Classify clusters based on their population and the knowledge base available (10% labeled KDD data set).
- iii. Perform clustering operation on the test data set (from the network) and classify their types according to the classification of the clusters (from the above step) they fall into.

3.4 Various Modules or Functions Used

In ACID Algorithm, seven major modules are used, which are as follows:

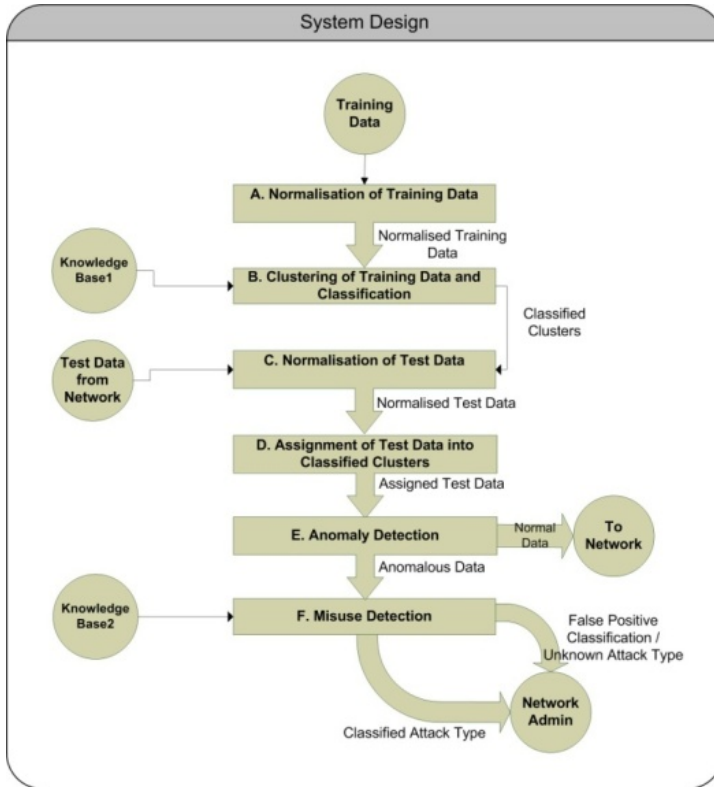


Fig. 1. System Design of ACID

Normalization of Training Data: It is found that the various connections in the training data come from different arbitrary state spaces. Again, they vary in the values of their features widely. Therefore, it is required to normalize [8] all these data instances into one standard form, so that these can be used in the intrusion detection process. It is performed by using the following equations:

$$\sum_{i=1}^N \text{avg_vector}[j] = \left(\sum \text{instance}_i[j] \right) / N \quad (i)$$

$$\sum_{i=1}^N \text{std_vector}[j] = \left(\left(\sum_{i=1}^N \text{instance}_i[j] \right) - \text{avg_vector}[j] \right)^2 / (N - 1) \quad (\text{ii})$$

$$\text{new_instance}[j] = (\text{instance}[j] - \text{avg_vector}[j]) / (\text{std_vector}[j]) \quad (\text{iii})$$

where, N is the number of data instances in the training data set.

In this way an instance can be transformed from its own space to the standard space, using the above mentioned formulae which are based on standard statistical information derived from the training set.

At first, data instances are read from the training data set and the *avg_vector*, *std_vector* and *new_instance* are calculated for the same using equations (i), (ii) and (iii). Thus *new_instance* represents the normalised data instances.

Clustering of Training Data and Classification: The normalized training data instances are grouped into clusters, using Y-means algorithm [5]. It aims at clustering without sacrificing its simplicity. After the clustering phase, the clusters formed are classified based on the population of the clusters and the knowledge base available (KDD99). A cluster is marked as normal if the majority of instances in the cluster are normal else the cluster is marked as an anomaly.

First, normalized training data obtained from stage A are taken into consideration. Then, the number of clusters (C) is initialized between 1 and C . Cluster centroids are randomly chosen from the training data set. Distance (D) between each of the instances and all the existing centroids is calculated based on the Euclidean distance. In a 2-dimensional plane, if $\mathbf{p} = (\mathbf{p}_1, \mathbf{p}_2)$ and $\mathbf{q} = (\mathbf{q}_1, \mathbf{q}_2)$ then the distance is given by,

$$D(\mathbf{p}, \mathbf{q}) = ((\mathbf{p}_1 - \mathbf{q}_1)^2 + (\mathbf{p}_2 - \mathbf{q}_2)^2)^{1/2}$$

Now, each instance from the training data set is assigned to its closest cluster centroid, in case the instance lies within the centroid's confidence area i.e.

$$\text{dis_mat}[i][j] \leq 2.32 * \text{std_v}[j]$$

where, i is the number of the cluster and j is the number of the instance and $\text{dis_mat}[i][j]$ stores the value of the distance of each instance from each cluster.

The assignment of a data instance to a cluster is represented in $\text{ass_mat}[\text{no_of_clusters}][\text{no_of_instances}]$, i.e. if instance j is assigned to cluster i , then $\text{ass_mat}[i][j] = 1$ and 0 for all other values of i . After this, the features of the $\text{ass_mat}[][]$ are copied to another matrix and the cluster centroids are recalculated by taking the mean of the newly assigned instances and already existing instances in the cluster. If all the instances in the data set have not been assigned to some existing cluster or the other, then the maximum distance of an instance from a cluster is calculated and this instance is made the cluster centroid of a new cluster. The number of existing clusters is updated accordingly and $\text{ass_mat}[\text{no_of_clusters}][j]$ is assigned the value of one where j is the instance number of the furthest instance. The above process is iterated until all instances have been assigned to some cluster. This marks the end of the clustering process. Next, the classification of these clusters is done based on the majority of the instances inside a particular cluster i.e. if number of

normal instances is in majority inside a cluster, then that cluster is marked as normal otherwise it is marked as an anomalous cluster. These classifications will be used later to aid in the classification of test data.

Normalization of Test Data: Normalization is also performed on the test data using the same equations (i), (ii) and (iii). Data instances are taken from the test dataset and *avg_vector*, *std_vector* & *new_instance* are calculated for the same. Thus normalized test data instances are obtained.

Assignment of Test Data into classified clusters: The *Distance(D)* of each normalized test instance from the classified cluster centroids is calculated using Euclidean distance again. The instance is then assigned to the cluster it is closest to.

Anomaly Detection: The test instances assigned to classified clusters in the previous stage are given the classification of the clusters they are assigned to, which is either normal or anomalous. The normal data instances are treated as safe. The anomalous instances are however passed on to misuse detection model.

Misuse Detection: In this stage, the anomalous instances received from Anomaly Detection stage are classified with the different attack types and also the class of attack is determined. A knowledge base contains the standard features of all the different attack types and their classes. The features of the anomalous instances (from the previous stage) are compared against these standard features of attacks. If a match is found then this anomalous instance is classified as a genuine threat and is listed with an attack type and class same as that of the standard feature. However, if a proper match is not found then this instance rings a false positive alarm which can be taken up by the network administrator for further investigation. These instances might be normal or in some cases, hitherto unknown types of attacks whose signatures are not available.

Performance Analyzer: The various performance measure parameters of the algorithm are *Detection Rate (DR)*, *False Positive Rate (FP)* and *Accuracy* [5][8][9]. The equations are given below:

$$\text{Detection Rate (DR)} = \frac{\text{Number of detected attacks}}{\text{Number of attacks}} \times 100\% \quad (\text{iv})$$

$$\begin{aligned} \text{False Positive Rate (FP)} \\ = \frac{\text{Number of misclassified connections}}{\text{Number of normal connections}} \times 100\% \end{aligned} \quad (\text{v})$$

$$\text{Accuracy} = \left(\frac{\text{Number of correct classified connections}}{\text{Number of connections}} \right) \times 100\% \quad (\text{vi})$$

4 Results and Discussion

The algorithm is implemented using C programming language in Windows environment. The dataset used in experiment is 10% labeled KDD data set. In this

experiment, the algorithm achieves a Detection Rate of about 95%, the False Positive Rate is very low (less than 0.5%) and the Accuracy achieved is about 92.75%. It is found that the accuracy is adversely affected when the training sample is not substantial. Therefore, the training samples must be of considerable size to ensure the accuracy of classification.

The number of clusters to be formed is not hard-coded but instead is decided automatically depending on the number of training data instances present. The cluster width is also not hard-coded but is adjusted to accommodate newly formed clusters. An initial number of seed clusters are however provided to start the iteration. As the iterations proceed, the number of clusters formed and the cluster width are automatically adjusted.

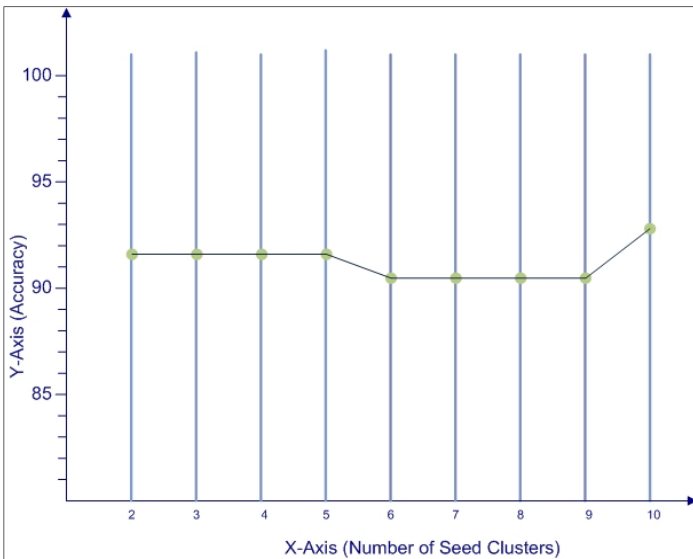


Fig. 2. Number of Seed Clusters v/s Accuracy

Table 1. Number of Seed Clusters v/s Accuracy

Number of Seed Clusters	2	3	4	5	6	7	8	9	10
Accuracy	91.75	91.75	91.75	91.75	90	90	90	90	93

It is seen that the performance measure remains constant when the number of seed clusters is between 2 and 5. There is a decrease in the Performance Measures for number of seed clusters between the ranges of 6 to 9. Finally, there is an increase in the performance measures when the number of seed clusters is 10, which is even more than the cluster ranges between 2 to 5. So, the initial number of seed clusters is taken to be 10.

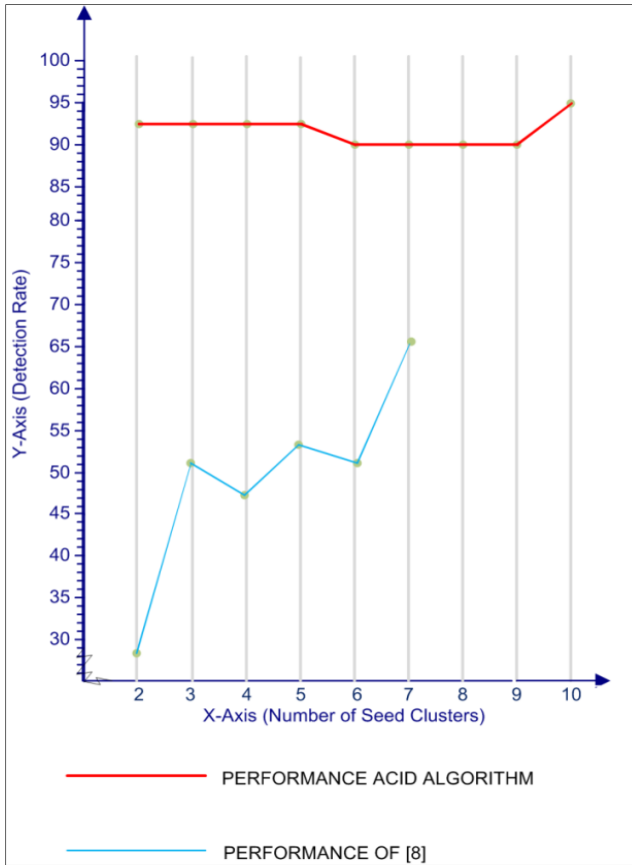


Fig. 3. Number of Seed Clusters v/s Detection Rate

Table 2. Comparing Detection Rate of ACID algorithm with [8]

Number of Seed Clusters	2	3	4	5	6	7	8	9	10
Detection Rate ACID	92.5	92.5	92.5	92.5	90	90	90	90	95
Detection Rate [8]	28.5	51.3	47.2	53.3	50.9	65.7	-	-	-

It is evident that detection rate of ACID is superior to that of [8]. This is because in ACID, the cluster width is not hard coded instead it keeps on changing as the algorithm proceeds automatically towards creating the optimum number of clusters whereas in [8], the cluster width is hard coded. ACID is consistent as the performance measure remains constant for groups of adjacent values of seed clusters. It does not deviate erratically with every change in number of seed clusters. This kind of consistency is a desirable property.

5 Conclusions

The proposed ACID Algorithm performs clustering in two phases: first on the training data and then on the test data. During clustering of the training data new clusters are formed based on the difference between the features of the data instances. The number of clusters to be formed is decided automatically. Also the algorithm avoids cluster degeneracy. The clustering of the test data is unsupervised; the data instances are assigned to the already formed clusters during the training phase. The test data is assigned to the cluster it is closest to.

The Algorithm consists of an anomaly detection model and a misuse detection model. Thus it filters a large number of records, using the anomaly detection model, and performs a second detection with the misuse detection model, when the data instance is determined to be abnormal. Therefore, it efficiently detects intrusion, and avoids the resource waste. Finally, it integrates the outputs of the anomaly detection and misuse detection models with a decision making model. This determines the presence of an intrusion, and classifies the type of attack. The output of the decision making model can then be reported to an administrator for follow-up work.

The Algorithm has a major advantage of being able to perform better than most of the existing signature based algorithms. This algorithm can detect new types of attacks using the anomaly detection model. However, the exact classification of the attack would not be possible due to the absence of its signature in the training data.

References

1. Sabahi, F., Movaghar, A.: Intrusion Detection: A Survey. In: The Proceedings of 3rd International Conference on Systems and Networks Communications, ICSNC 2008. IEEE, Los Alamitos (2008)
2. Berkhin, P.: Survey of clustering data mining techniques. Technical report, Accrue Software, San Jose, CA (2002)
3. Terry Brugger, S.: Data Mining Methods for Network Intrusion Detection. University of California, Davis (2004)
4. Prerau, M.J., Eskin, E.: Unsupervised anomaly detection using an optimized K-nearest neighbors algorithm. Master's thesis, <http://www.music.columbia.edu/~mike/publications/thesis.pdf>
5. Guan, Y., Ghorbani, A., Belacel, N.: Y-means: A Clustering Method for Intrusion Detection. In: Proceedings of Canadian Conference on Electrical and Computer Engineering, Montreal, Quebec, Canada, May 4-7 (2003)
6. Bloedorn, E., Christiansen, A.D., Hill, W., Skorupka, C., Talbot, L.M., Tivel, J.: Data mining for network intrusion detection: How to get started (August 2001), <http://citeseer.nj.nec.com/523955.html>
7. Lee, W., Stolfo, S.J.: Data Mining Approaches for Intrusion Detection. In: Proceedings of the 1998 USENIX Security Symposium (1998)
8. Portnoy, L., Eskin, E., Stolfo, S.: Intrusion Detection with Unlabeled Data Using Clustering. In: Proceedings of the ACM CSS Workshop on Data Mining Applied to Security (DMSA- 2001), Philadelphia, PA, November 5-8 (2001)

9. Yan, K.Q., Wang, S.C., Liu, C.W.: A Hybrid Intrusion Detection System of Cluster-based Wireless Sensor Networks. In: Proceedings of the International MultiConference of Engineers and Computer Scientists 2009, IMECS 2009, Hong Kong, March 18-20, vol. I (2009)
10. Zhong, S., Khoshgoftaar, T.M., Seliya, N.: Clustering-based network intrusion detection (2007)
11. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification (2001)
12. Zanero, S., Savaresi, S.M.: Unsupervised learning techniques for a intrusion detection system (2004)
13. Lazarevic, A., Ertöz, L., Kumar, V., Ozgur, A., Srivastava, J.: A Comparative study of anomaly detection schemes in network intrusion detection (2003)
14. Zhang, R., et al.: BIRCH: An efficient data clustering method for very large databases (1996)
15. Lee, S.: Data mining approaches for intrusion detection (1998)

Measuring the Deployment Hiccups of DNSSEC

Vasilis Pappas and Angelos D. Keromytis

Computer Science Dept.
Columbia University
New York, NY
{vpappas,angelos}@cs.columbia.edu

Abstract. On May 5, 2010 the last step of the DNSSEC deployment on the 13 root servers was completed. DNSSEC is a set of security extensions on the traditional DNS protocol, that aim in preventing attacks based on the authenticity and integrity of the messages. Although the transition was completed without major faults, it is not clear whether problems of smaller scale occurred. In this paper we try to quantify the effects of that transition, using as many vantage points as possible. In order to achieve that, we deployed a distributed DNS monitoring infrastructure over the PlanetLab and gathered periodic DNS lookups, performed from each of the roughly 300 nodes, during the DNSSEC deployment on the last root name server. In addition, in order to broaden our view, we also collected data using the Tor anonymity network. After analyzing all the gathered data, we observed that around 4% of the monitored networks had an interesting DNS query failure pattern, which, to the best of our knowledge, was due to the transition.

1 Introduction

The Domain Name System is one of the core components of the Internet, used by virtually all the network applications. It provides name to IP (network layer) address translation in a form of a distributed database. Although DNS proved to be robust and highly scalable through everyday use, later discovered vulnerabilities [5,4] opened the doors to attacks. DNS Security Extensions (DNSSEC) [3,9,2,1] were proposed in order to address these issues. These security extensions incorporated cryptographic signatures along with each DNS message sent by the servers. Replying with signed messages gives the ability to recipients to verify the authenticity of the source and also, in most of the cases, to verify the integrity of the message itself. This, not only protects against current attacks, but it could also prevent future ones.

Designing even an extension to such a vital and highly critical component of the Internet infrastructure, is not an easy task. The first RFC document about the security extensions was written in 1997, updated in 1999 and 2001 and took its final form in 2005. Even worse, the deployment proved to be harder. Although DNSSEC was first implemented in BIND9 [1] in 1999, the deployment stalled until

¹ <http://www.isc.org/software/bind>

2001, when a problem in the design of key handling was found. The operational problems caused by that, rendered the deployment almost impossible at that time. The first Top Level Domain to deploy the security extension was .SE (Sweden) in 2005².

However, more recent studies during the deployment of DNSSEC [14,12] revealed adaptation challenges that were not anticipated in the design. Middle boxes – such as firewalls, NATs, etc. – proved to be troublesome. The main reason behind that is the increase of the DNS messages size [10], in order to facilitate the security extensions’ data. In this paper, we specifically study these availability problems, using data from as many vantage points as possible, during the deployment of DNSSEC on the last of the root name servers. Our results show that the rate of DNS query failures in about 4% of the monitored networks, follows an interesting pattern that we believe is most probably caused by the problems described above. To the best of knowledge, this is the most global study on the availability of DNSSEC.

2 Background

In this section, we firstly briefly describe how DNS works and then, we move one step further to also describe how the Security Extensions (DNSSEC) work in practice.

2.1 Domain Name System (DNS)

In summary, the Domain Name System (DNS) can be viewed as a large distributed database of host name to IP address pairs. Although that simplification helps in order to understand the service provided to the end users/clients, it does not give any insights about the internals of the system. DNS is comprised mainly from two components: resolvers and name servers. Name servers store information about their “zone”. This information includes name to IP address pairs, as well as delegations – that is, pairs of subdomains and name servers, lower in the hierarchy. Resolvers, on the other hand, are used to query the name servers in order to find the IP address corresponding to a host name (and vice versa).

For example, suppose that an application (web browser, e-mail client, etc.) wants to find the IP address for a specific host name, e.g. `www.example.com`. The first step for the application would be to send that query to a stub resolver on its host. If the result is not previously cached locally, the stub resolver will forward the query to a configured local resolver, usually located within the user’s network (ISP, corporate network, etc.). Similarly, this resolver will also try its local cache and if that fails, it will start a recursive search. Firstly, it will query one of the root name servers in order to find the name server responsible for the top level domain, `com` in our case. Then, it will repeat the same procedure

² The source of the historical data on DNSSEC is:

<http://www.nlnetlabs.nl/projects/dnssec/history.html>

recursively, querying the `com` name server for the `example.com` name and so forth. If the resolver manages to find the corresponding pair, it will return that to the local resolver and it will be then forwarded to the application. If there is no such pair or an another error occurred, an appropriate error will be returned.

It should be mentioned here though, that even the above definition by example is still abstract. In reality, there is more complexity added by replicas and caches throughout the system, but the main functionality remains the same.

2.2 DNS Security Extensions (DNSSEC)

As mention before, the Domain Name System was found to be vulnerable to a number of attacks (cache poisoning, etc.). In order to shield it against them and maybe even prevent future attack techniques, DNS Security Extensions were proposed. Basically, DNSSEC extends DNS through a hierarchal cryptographic signing of the DNS messages. Cryptographic signatures provide both authentication and integrity for the transmitted message. When a node receives a signed message, it can firstly establish the source and also verify the integrity of that message.

DNSSEC implementation added four new resource record types, namely: resource record signature (RRSIG), DNS public key (DNSKEY), delegation signer (DS) and next secure (NS). In order for a zone to enable the security extensions, it needs to generate a public/private key pair. Then, the records of each response should be signed using the private key and the RRSIG and DNSKEY records should contain the signature and the public key respectively. When a resolver receives a signed response from a DNSSEC capable server, it firstly verifies the DNSKEY, starting from the root servers and continuing recursively using the DNSKEY and DS records found in the messages. If the DNSKEY is trusted, the resolver can proceed with verifying the message itself using both the DNSKEY and RRSIG fields.

3 Methodology

The main burden of accurately examining the effect of the DNSSEC transition is the fact that it needs to be done on the end networks. Thus, the more available end points (in distinct networks), the more global the view will be and more accurate the results. For the purposes of our study, we chose to use two distributed systems, in order to maximize the coverage. These were PlanetLab [6] and Tor [8]. Although these systems have different goals – distributed testbed versus anonymity – they have one thing in common. Both of them are comprised by a (relatively large) set of distributed nodes. The size of these systems rely on the voluntariness of users to deploy more nodes and as both of them are closely related to academia, most of their nodes are hosted on academic institutions. At this point we have to mention that we do understand that having most of the end points in academia networks (and not so many in ISPs or corporate networks) could pose a limitation to our methodology. But, we believe that the

coverage we have is the best possible, taking into account the readily available and freely usable distributed systems on the Internet. In sections 3.1 and 3.2 we describe in more details PlanetLab and Tor respectively and how we utilized each system to gather data.

3.1 PlanetLab

PlanetLab is a distributed testbed for networking and distributed systems research. It was established in 2003 by the Princeton University and as of May, 2010 it consists of 1,086 nodes at 506 sites³. To run an experiment on PlanetLab, initially, the user chooses a set of nodes and creates a “slice” containing them. Then, she is given remote access to each of the nodes in the slice, isolated by any other users.

For the purposes of our study, we created a slice consisting of 308 unique per site and alive (functional) nodes. On each of them, we periodically executed `dig`, which is a tool bundled with BIND to query DNS servers, for a list of host names, every 15 minutes. The host names we used are the top 100 from the Alexa web site (<http://www.alexa.com/topsites>), as of May, 4. In order to avoid some of the caching effects on the DNS resolver, each time we were resolving a host name, we additionally resolved another one, uniquely constructed. More precisely, we were prepending the current timestamp in seconds from the Epoch time to each host name. For example,

```
google.com -> 1273109411.google.com
```

Note that distinguishing between different kinds of resolve failures is possible. More specifically, the status field of each response is set to `NXDOMAIN`, if the host name does not exist and `SERVFAIL` if another error occurred during the resolving.

3.2 Tor

Tor is the second generation *Onion Routing* system. It provides low-latency anonymous communication between two parties, by routing traffic through multiple nodes and applying multiple layers of encryption. As an example, suppose that a user wants to anonymously browse a website using Tor. The steps she has to follow in order to achieve that would be:

1. Create a *circuit* by choosing a number of nodes (3 by default) and exchange session keys using Diffie-Hellman key exchange protocol – which provides *perfect forward secrecy*. During the creation of the circuit, the client only communicates with the first node in the circuit (*entry node*). The final node, called *exit node*, is the one who will actually communicate with the target web server.
2. Transmit data. After the circuit is set up and the connection between the client and the web server is established, the user is able to browse the website anonymously.

³ Source: <http://www.planet-lab.org/>

We have to mention here that not all nodes support acting as *exit* ones. There are nodes that can be used only in the beginning or in the middle of circuits and never connect to hosts outside the Tor network. This behavior is based on the configuration of each node's *exit policy*. Recall that for our purposes we are only interested in utilizing hosts in as many end networks as possible. So, we are only interested in nodes that have less restrictive exit policy and support communicating with other target hosts, outside the Tor network.

Currently, the Tor network consists of 1536 nodes, of which 594 support acting as *exit nodes*⁴. In order to take full advantage of the Tor's exit nodes, we tried to periodically send a HTTP request to each of the 100 hosts used in [3.1](#) using a different exit node each time. This was repeated hourly and the result of whether the connection succeeded or not was stored. In order to implement that, we used the very convenient TorFlow software package (<http://fscked.org/projects/torflow>). Note though that we did not use any uniquely constructed host names in this case, as we would not be able to distinguish between failed resolves and non existent domain names.

Although using the Tor network we broaden our view with more end points in distinct networks, the level of useful information that we were able to collect was significantly lower. The interface of Tor to the client applications is through a SOCKS [\[11\]](#) proxy, so, the only information that we could get is whether a connection succeeded or not at the exit node.

4 Data Collection

In order to have sufficient data to examine the effects of the last step of the DNSSEC transition we monitored the DNS behavior, using the above methods, for roughly a 5 day period.

After that time period, each PlanetLab node had invoked `dig` for roughly 100,000 times. Thus, creating an equal number of files containing the output of each DNS query. In total, there were about 30 million files occupying 60 giga bytes of storage when stored as plain text files and 8.1 giga bytes when compressed using `gzip`.

On the other hand, the data collected using the Tor network were stored to a single file. Unfortunately, due to a bug in the script we used to scan the exit nodes, each iteration over them was immediately terminated after a single circuit had failed. More precisely, the total number of circuits created during the measuring period was 4,745 instead of about 9,600 – we roughly lost half of the data. Although we did lose a big portion of valuable data due to this bug, the rest of them was nonetheless important enough. Out of the 594 Tor nodes that can be *exit nodes* only 184 of them had the flags *Exit*, *Fast*, *Stable*, *Running* and *Valid* set to true at the same time and these were the ones we were able to effectively use.

⁴ Source: <http://torstatus.kgprog.com/>

Table 1. TLD distribution for both Tor and PlanetLab nodes

Top Level Domains	# of nodes	Top Level Domains	# of nodes
edu	98	uk ch	9
net	50	ca	8
de	49	tw se es	7
com	34	ru	6
NA	26	pt kr br	5
pl	15	sg nz il gr cn au	4
org jp it	14	nl hu hk fi cz be at	3
fr	12	tr ro no eu dk ar	2
ve uy to th su si re jo info eg cy cx			1

Table 2. Categorization of the nodes based on their failure patterns

Category	Nodes	Description
<i>Zero</i>	222	No failures
<i>Few</i>	108	A few failures
<i>Interesting</i>	15	Fair amount of failures, interesting patterns
<i>Constant</i>	7	A lot of failures, constant rate though

Table 1 summarizes the distribution of the nodes’ Top Level Domains, both for Tor and PlanetLab systems. As previously mentioned, most of the nodes are deployed in academic networks, for example the .edu TLD.

5 Results

After we collected the datasets described in the previous section, we analyzed them in order to extract useful insights related to the DNSSEC transition. In this section, we will thoroughly discuss our main findings from the analysis.

5.1 Node Categorization

The question we tried to answer here was: “*Do we see any DNS query failures that could probably be due to the transition?*” We first analyzed the failures in the dig logs and gathered the timestamps in which each node experienced a DNS lookup failure. Based on that information we divided the PlanetLab nodes in the following four categories (summarized in Table 2):

- **Zero.** The nodes that had no DNS query failures during the monitoring period fall into this category. As expected, the majority of the nodes are in this category.
- **Few.** Some of the nodes sporadically encountered a small number of failures. These were most of the times less than 10, with just a few cases that had up to almost a hundred. Still, we consider these failures not caused by the DNSSEC transition since they were few in number and occurred randomly.

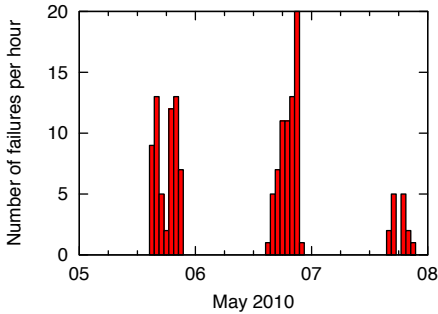


Fig. 1. Interesting failure pattern 1

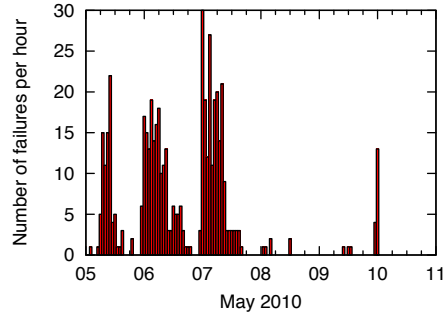


Fig. 2. Interesting failure pattern 2

- **Interesting.** This category contains the nodes that had a fair amount of failures and most importantly, their failure rates were forming an interesting pattern. In average, these nodes failed for a few hundreds of queries.
- **Constant.** Finally, there were some nodes that were constantly failing for a number of queries. Clearly that could not be due to the DNSSEC transition as this failure rate was constant during the whole monitoring period. Probably, that should be the result of a misconfiguration or blacklisting – the top 100 hostnames we used contained some adult oriented sites too.

5.2 Failure Patterns

In this section we examine a few specific failure patterns as observed on individual PlanetLab nodes. More precisely, we closely examine three failure patterns of what we called *Interesting* in the previous section and one example of a *Constant* failure pattern. Although, we are more concerned to the first ones, as the failures in these cases were probably be caused by the transition, we include the last example (constant failure rate) for completeness.

Figures 1, 2 and 3 show the failure patterns for three nodes that we categorized as *Interesting*.

In the first case (Figure 1) we see that the failures started on the 5th of May in the afternoon. Same time as the transition of the last root server to DNSSEC was performed. This is a strong evidence that that could be their cause. It is also interesting that we see a similar failure pattern on the next day (6th of May) that could be due to caching effects. One common Time-to-Live (TTL) value for the host name to IP address pairs is 86,400 seconds, which equals to one day. The same was observed on the third day too, but with lower magnitude. Finally, the most interesting thing is that this node had no more failures for the next two days, until 10th of May, when we stopped the monitoring. Our explanation for that is that there could be indeed an error that was fixed by the network administrators in the first couple of days.

The next one, Figure 2, follows almost the same pattern as the previous one. We see that the failures start on the day of the transition and they have a diurnal

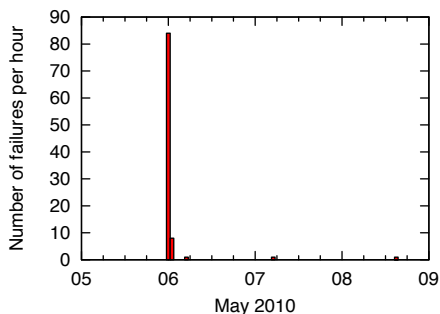


Fig. 3. Interesting failure pattern 3

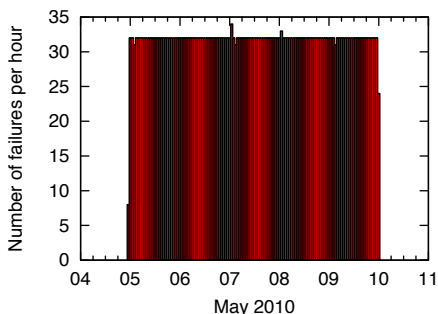


Fig. 4. Constant failure rate

pattern. Most likely again due to the fact that the TTL set by the name servers is one day. The interesting thing here though is that although the problem seemed to have been fixed on the fourth day (9th of May), we do see more failures on the fifth day. That is hard to explain by this figure alone, but, correlating more diverse data from the same network could reveal more insights.

The third *Interesting* node we chose to discuss follows a different failure pattern than the previous two (Figure 3). In this case, (i) the failures do not start on the 5th of May and (ii) there is only a tiny diurnal pattern. This is again a both interesting and hard to explain situation without having more details. We speculate that it could have been indeed caused by the the DNSSEC transition and the failures were reported later either due to a wrong configured clock on the node or due to an uncommon caching policy on the local resolver.

We should mention here that out of the fifteen *Interesting* cases, most of them follow the pattern of the first two. Which is what we believe was caused by the transition. But, we decided to show this case too, because of its uniqueness.

Finally, Figure 4 shows a case of a node that had a constant DNS query failure rate (i.e., example of a *Constant* node). As shown in the figure, this particular node fails to resolve about 32 host names per hour. Recall from Section 3.1 that each node was periodically trying to resolve 100 host names every fifteen minutes. In addition, each of these queries was followed by another query to a uniquely constructed hostname, under the same domain. So, each domain name was queried eight times per hour. Dividing the failure rate with the number of queries for each individual domain per hour, results in the number four. This constant failure rate of 32 queries per hour could mean that there were just four domain names that always failed to resolve. After investigating the log files of that specific case, we confirmed that that was the case. This node was always failing to resolve four domains, of which one of them was an adult oriented site, another was a webmail service and the other two, social networking websites. In conclusion, this confirms our first intuition about the nodes that had constant failure rate, which was blacklisting.

As previously stated in Sections 3.2 and 4, the information we gathered from the Tor network was not that comprehensive. In addition to that, we also lost a big portion of the data due to a programming bug in our scripts. Being not inclusive enough the dataset, we were not able to draw any conclusions from it alone. But, we did use this data in order to help us verify some moot cases.

6 Related Work

To the best of our knowledge, the most relevant work to our is Osterweil’s et al. study on the deployment of DNSSEC [14]. The authors gathered historical data over 2 years – since the initial deployment of the security extensions. After analyzing their dataset, they derived three basic metrics to quantify the effectiveness of the DNSSEC’s deployment, namely *Availability*, *Verifiability* and *Validity*. In addition, their results exposed previously undocumented open issues during the deployment, such as lack of Availability due to middle boxes (NATs, firewalls, etc.), poor Verifiability as a result of not having an external key authentication and finally they showed that cryptographic verification is not necessarily equivalent to validation.

Although in terms of their work we only looked at *Availability*, our dataset was gathered from hundreds of distinct networks whereas they used five locations. We believe that using such a geographically diverse monitoring infrastructure gave us a more global view of the problem. And most importantly, our results do confirm what was anticipated by the authors about the Availability metric, in a small scale though.

The implementation of the system used in the above study, namely *SecSpider*, is described in detail in [13]. Some early observations on the data gathered by SecSpider are reported in [12]. One of their observations, which was the need to reduce signatures lifetime, is addressed in [15].

Finally, Curtmola et al. compared the performance of PK-DNSSEC versus SK-DNSSEC, arguing that a hybrid approach would leverage the benefits from both worlds [7].

7 Conclusion

DNSSEC’s goal is to shield the domain name system to any attacks based on the authenticity and integrity of the messages. This is achieved by the incorporation of cryptographic signatures along with each message sent by the name servers. Although the deployment of the security extensions started after years of designing and testing, studies showed that problems related to availability could still be experienced in some networks.

Our study revealed that in a relative small number of the networks we monitored, availability issues were evident. We consider these issues to have been most probably caused by the DNSSEC deployment on the last root name server.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Grant No. CNS-09-14312. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

1. Arends, R., Austein, R., Larson, M., Massey, D., Rose, S.: Protocol modifications for the dns security extensions. RFC 4035 (March 2005)
2. Arends, R., Austein, R., Larson, M., Massey, D., Rose, S.: Resource records for the dns security extensions. RFC 4034 (March 2005)
3. Ateniese, G., Mangard, S.: A new approach to dns security (dnssec). In: CCS 2001: Proceedings of the 8th ACM Conference on Computer and Communications Security, pp. 86–95. ACM, New York (2001)
4. Atkins, D., Austein, R.: Threat analysis of the domain name system (dns). RFC 3833 (August 2004)
5. Bellovin, S.M.: Using the domain name system for system break-ins. In: SSYM 1995: Proceedings of the 5th Conference on USENIX UNIX Security Symposium, p. 18. USENIX Association, Berkeley (1995)
6. Chun, B., Culler, D., Roscoe, T., Bavier, A., Peterson, L., Wawrzoniak, M., Bowman, M.: Planetlab: an overlay testbed for broad-coverage services. SIGCOMM Comput. Commun. Rev. 33(3), 3–12 (2003)
7. Curtmola, R., Del Sorbo, A., Ateniese, G.: On the performance and analysis of DNS security extensions. In: Desmedt, Y., Wang, H., Mu, Y., Li, Y. (eds.) CANS 2005. LNCS, vol. 3810, pp. 288–303. Springer, Heidelberg (2005)
8. Dingledine, R., Mathewson, N., Syverson, P.: Tor: the second-generation onion router. In: SSYM 2004: Proceedings of the 13th Conference on USENIX Security Symposium, p. 21. USENIX Association, Berkeley (2004)
9. Friedlander, A., Mankin, A., Maughan, W.D., Crocker, S.D.: Dnssec: a protocol toward securing the internet infrastructure. Commun. ACM 50(6), 44–50 (2007)
10. Gudmundsson, O.: Dnssec and ipv6 a6 aware server/resolver message size requirements (2001), <ftp://ftp.rfc-editor.org/in-notes/rfc3226.txt>
11. Leech, M., Ganis, M., Lee, Y., Kuris, R., Koblas, D., Jones, L.: Socks protocol version 5. RFC 1928 (March 1996)
12. Osterweil, E., Massey, D., Zhang, L.: Observations from the dnssec deployment. In: NPSEC 2007: Proceedings of the 2007 3rd IEEE Workshop on Secure Network Protocols, pp. 1–6. IEEE Computer Society, Washington, DC, USA (2007)
13. Osterweil, E., Massey, D., Zhang, L.: Deploying and monitoring dns security (dnssec). In: ACSAC 2009: Proc. of the 2009 Annual Computer Security Applications Conference, pp. 429–438. IEEE Computer Society, Washington, DC, USA (2009)
14. Osterweil, E., Ryan, M., Massey, D., Zhang, L.: Quantifying the operational status of the dnssec deployment. In: IMC 2008: Proceedings of the 8th ACM SIGCOMM Conference on Internet Measurement, pp. 231–242. ACM, New York (2008)
15. Yan, H., Osterweil, E., Hajdu, J., Acres, J., Massey, D.: Limiting replay vulnerabilities in dnssec. In: 4th Workshop on Secure Network Protocols, 2008, NPSec 2008, pp. 3–8, 19–19 (2008)

Self-organizing MAC Protocol Switching for Performance Adaptation in Wireless Sensor Networks

Fan Yu and Subir Biswas

Electrical and Computer Engineering
Michigan State University, East Lansing, MI 48824, USA

Abstract. This paper presents a distributed MAC protocol switching mechanism for maximizing network throughput in the presence of traffic and topology heterogeneity. The key idea behind dynamic MAC switching is for each node to use its local topology and traffic density information to decide the most suitable MAC protocol that can maximize the MAC layer throughput in the neighborhood. A formal MAC switching rule is developed using analytical formulation of the MAC throughput available in the literature. NS2 based simulation experiments demonstrate that with the proposed MAC switching strategy, nodes in a mesh network are able to achieve maximum MAC throughput by adaptively choosing the appropriate MAC protocol in the presence of heterogeneity in terms of data rate and node population.

Keywords: MAC, Self Organization, Protocol Switching.

1 Introduction

1.1 Background

Wireless sensor networks are motivated by applications such as intrusion detection, battle-field surveillance, and environment and health monitoring. Although there have been significant recent innovations at different protocol layers, maximizing network throughput at the MAC layer in a multi-mission network is still a key design challenge. The traffic heterogeneity in a multi-mission network can manifest through: a) spatial and temporal variance in data rates [1] across different parts of a network supporting different applications, and b) temporal variance in number of active transmitters at a given time and at a given part of the network. The problem we address in this paper is how to develop a self organizing MAC layer that can maximize network throughput in the presence of the above traffic and network heterogeneity.

1.2 Related Work

The existing MAC layer self-organization can be divided into two broad categories, namely, intra-MAC and inter-MAC approaches. The intra-MAC approaches include contention-based [2-4] and schedule-based [5-7] protocols. The intra-MAC self-organization is achieved by dynamically adjusting the state machine and parameters

of a specific MAC protocol. CSMA/CA [2] and TDMA [5-7] are two examples of such intra-MAC approaches. CSMA/CA has the advantages of simplicity, flexibility and robustness, and it does not need any infrastructure support or clock synchronization. The most notable advantage of CSMA/CA is that a node can access all available wireless bandwidth in its neighborhood in an as-needed manner. However, the medium access collisions and the corresponding unbounded access delay are a concern for CSMA/CA and all other protocols in that category, mainly because of their underlying random access. TDMA protocols, on the other hand, allocate fixed, collision-free, and guaranteed bandwidth for all nodes in the network. While guaranteeing maximum delay bounds, the primary disadvantage of TDMA is that it does not perform well under dynamic bandwidth requirements. In other words, TDMA performs well under highly symmetrical load, but performs poorly under asymmetric load. Under the latter situation, CSMA based protocols outperform TDMA.

The authors in [8] propose an inter-MAC self-organization called Funneling-MAC. This is designed to address the traffic funneling effect near sink nodes caused by gradual aggregation of multipoint-to-point traffic. With Funneling-MAC, nodes near a sink (i.e. those with heavy traffic) are said to belong within an *intensity area* and they run a TDMA protocol which is managed by the sink node. Nodes outside the *intensity area* (i.e. those with relatively lighter traffic) run CSMA without any coordination from the sink. The Funneling-MAC [8] provides beacon based protocol syntaxes that the sink uses for dynamically deciding the boundary of the *intensity area*. Through a dynamic depth-tuning algorithm, the network throughput is maximized and the packet loss rate is minimized at the sink point. Although it provides a novel way of accomplishing inter-MAC (i.e. between CSMA and TDMA) self-organization, one notable limitation of Funneling-MAC is that it is suitable only for multipoint-to-point applications and not for peer-to-peer traffic, which is often required by sensor applications with distributed data fusion requirements.

1.3 Proposed Dynamic MAC Protocol Switching

In the proposed approach in this paper the operating MAC protocol within a node is dynamically switched between CSMA and TDMA based on the instantaneous traffic and topological property of the neighborhood of the node. The mechanism is designed to be general so that unlike Funneling-MAC [8], it can be applied for handling both multipoint-to-point and peer-to-peer data traffic. Each node monitors its neighborhood traffic and topology conditions, determines the appropriate individual MAC protocol to run, and switches its protocol as needed.

The contributions of the paper are as follows. First, it proposes a self organizing MAC protocol switching paradigm to address the problem of how to achieve the maximum throughput in a network with traffic heterogeneity. Second, it introduces syntax extensions to the protocols CSMA/CA and TDMA so that they can co-exist in immediately neighboring nodes. Third, it develops dynamic protocol switching rules based on an analytical model. Finally, a detailed simulation model is developed for experimentally validating the concept of dynamic MAC protocol switching in wireless networks.

2 Protocol Adaptations

The key idea behind dynamic protocol switching is that depending on the specific traffic pattern in its neighborhood, a node can choose to run one of multiple MAC protocols. This capability in a network can lead to situations in which a node can have multiple neighbors that are running different protocols. This will require the node in question to be able to communicate using multiple different MAC protocols simultaneously in order to maintain end-to-end network connectivity. This MAC coexistence can be particularly challenging due to the fact that a sensor node needs to accomplish this over a single wireless interface. In this paper, coexistence of two MAC protocols, namely a random access CSMA/CA and a schedule based TDMA, are studied. CSMA/CA and TDMA are chosen because they serve as good representations of contention based and schedule based protocols respectively.

2.1 Adaptations of CSMA/CA for TDMA Coexistence

The objective here is to adapt the CSMA/CA protocol syntaxes such that a node running CSMA/CA can also communicate with a node running TDMA. The basic RTS-CTS-DATA cycle in CSMA/CA transmissions is maintained. The Acknowledgement (ACK) is not utilized, (meaning there is no link layer), and the regular exponential back-off are still applied. Adaptations of the CSMA/CA syntaxes are made such that no access collisions are imposed on the TDMA nodes in a CSMA/CA node's neighborhood.

Adapted Send Logic: A CSMA/CA sender is able to estimate the duration of a successful transmission in terms of RTS, CTS, the data message, and the various inter frame spacing. If it knows the TDMA slots of all its neighbors, then the sender can determine if the following data transmission will overlap with the neighbors' TDMA slots. If there is at least one anticipated overlapping, the sender simply defers its transmission till the end of all its neighbors' TDMA slots. Otherwise, the data transmission could carry on as usual.

```

    /* Adapted CSMA/CA Logic */

    Sender Side:
        Estimate the duration of a successful transmission;
        if ( any overlapping with TDMA neighbors' slots)
            defer transmission till end of neighbors' TDMA slot;
        else // no overlapping within knowledge of sender side
            start RTS transmission;

    Receiver Side:
        Upon receiving RTS, check the duration of the intended
        message transmission;
        if ( any overlapping with TDMA neighbors' slots)
            send NTS message back to the sender;
        else // no overlapping within knowledge of receiver side
            start CTS transmission;
  
```

Fig. 1. Pseudo code for adapted CSMA/CA logic

Adapted Receive Logic: Upon receiving an RTS message, from its NAV field, the intended receiver is able to know the expected duration of the following data message. If that expected duration overlaps with the TDMA slots of the receiver's TDMA neighbors, the receiver will send a Not-to-Send (NTS) message, containing the duration of the overlapping period, back to the sender. Upon receiving the NTS message, the sender will then decide as to when to restart the current data transmission in order to avoid the slots of all TDMA neighbors of the sender as well as the receiver. The logic of this adapted CSMA/CA is presented in the pseudo code in Fig. 1.

2.2 Adaptation of TDMA for CSMA/CA Coexistence

The same slot and frame structure as in regular TDMA is maintained. Meaning, each TDMA node is allocated a slot in each frame, which is periodic. As an adaptation, an RTS-CTS syntax is added within the TDMA slots so that before sending a data message, a node needs to perform the RTS/CTS handshake with the receiver node. This is done irrespective of whether the receiver is currently using CSMA/CA or TDMA.

```

                /* Adapted TDMA Logic */
    Sender Side:
        if ( in sender's own TDMA slot)
            start RTS transmission and wait for CTS;
    Receiver Side:
        if ( receiving RTS)
            send CTS message and wait for data message;
```

Fig. 2. Pseudo code for adapted TDMA logic

The addition of this RTS-CTS within TDMA allows the node to send data to both CSMA/CA and TDMA nodes. Note that a TDMA sender node does not need to sense the channel before it sends the CTS since all nodes in its neighborhood (CSMA/CA or TDMA) respect the TDMA slot allocation and the channel is guaranteed to be free after the sender's slot has started. The RTS-CTS mechanism, in fact, is used to make sure that the receiver node is in a state to receive. The pseudo code for the adapted TDMA logic is shown in Fig. 2.

3 MAC Protocol Switching Logic

3.1 Influencing Switching Parameters

Consider a network of n nodes and each node's MAC layer is modeled as an M/M/1 queue. If λ and μ represent the data rate and the MAC layer service rate at each

node, the unsaturated network-wide throughput can be written as $n \cdot \lambda$, and the saturated network-wide throughput can be expressed as $n \cdot \mu$. The network-wide throughput of a fully connected n -node TDMA network can be written as [9]:

$$S_{TDMA} = f(\lambda, n, F), \quad (1)$$

where F is the TDMA frame duration which is fixed for a given network. Similar throughput for a CSMA/CA network can be written as [10-15]:

$$S_{CSMA/CA} = g(\lambda, n, W_{\min}, m), \quad (2)$$

W_{\min} is the minimum contention window size for CSMA/CA, and m is the maximum back-off stage.

Therefore, the throughput of a network with n_{TDMA} and $n_{CSMA/CA}$ number of nodes running the protocols TDMA and CSMA/CA respectively is:

$$S = S_{TDMA} + S_{CSMA/CA} = f(\lambda, n_{TDMA}, F) + g(\lambda, n_{CSMA/CA}, W_{\min}, m) \quad (3)$$

For a given network, the quantities F , m and W_{\min} are all constants. Thus, the throughput is decided based only on load λ and the node counts in Eqn. 3.

Implications of different data rates (λ): The maximum MAC layer service rate of a node running TDMA is bounded by the TDMA frame duration, and can be reported as a fixed quantity $\mu_{MAX-TDMA}$. Meaning, as long as the data rate λ at a node is less than or equal to $\mu_{MAX-TDMA}$, the MAC layer throughput is same as the load λ .

For a node running CSMA/CA, however, there is no such absolute bound for the achievable throughput. The per-node throughput here depends on the amount of MAC layer collisions contributed by other CSMA/CA nodes in the neighborhood. In other words, the MAC throughput of a CSMA/CA node is less than or equal to the MAC load λ ; the exact value depends on the degree of collisions.

Implications of the number of active nodes (n): If $S_{\sin gle-TDMA}$ represents the MAC throughput of a single node running TDMA MAC, then the aggregated throughput of a network cluster of n active nodes running TDMA can be expressed as $n \cdot S_{\sin gle-TDMA}$. For CSMA/CA, however, the aggregated throughput of the cluster is less than $n \cdot S_{\sin gle-CSMA/CA}$, since with n nodes running CSMA/CA in a cluster reduces the effective MAC throughput (due to contentions and collisions) of each node to be less than $S_{\sin gle-CSMA/CA}$. Such throughput reduction does not apply for TDMA because of the lack of bandwidth contention and collisions.

The above analysis indicates that in order to maximize the aggregated MAC layer throughput of a network, the MAC protocols for each node should be individually chosen (CSMA/CA or TDMA) based on its data load λ and the number of active nodes n in its immediate neighborhood.

```

/* MAC Protocol Switching Logic decisions with  $\lambda$  and  $n$  */
if (  $\lambda \leq \mu_{MAX-TDMA}$  )
    run TDMA MAC;
else if (  $\lambda \leq \mu_{CSMA/CA}$  and  $\mu_{CSMA/CA} > \mu_{MAX-TDMA}$  )
    //  $\lambda > \mu_{MAX-TDMA}$ 
    // and for moderate  $n$ ,  $\mu_{CSMA/CA} > \mu_{MAX-TDMA}$ 
    run CSMA/CA MAC;
else if (  $\mu_{CSMA/CA} \leq \mu_{MAX-TDMA}$  )
    run TDMA MAC; // large  $n$  reduces  $\mu_{CSMA/CA}$ 

```

Fig. 3. Logic for protocol switching decision with λ and n

3.2 Switching Criteria

According to the analysis presented in Section 3.1, when for a node the condition $\lambda \leq \mu_{MAX-TDMA}$ is true, the node should choose TDMA as the MAC protocol for the best throughput. That is irrespective of the number of active neighbors n . When $\lambda > \mu_{MAX-TDMA}$, the quantity n needs to be considered. If the effective MAC service rate $\mu_{CSMA/CA}$ at the current n for the node is smaller than the constant $\mu_{MAX-TDMA}$, then also TDMA should be chosen. Finally, if for the current n , the quantity $\mu_{CSMA/CA}$ is larger than $\mu_{MAX-TDMA}$ then CSMA/CA should be chosen for maximizing the MAC throughput. The protocol switching criteria is summarized in the pseudo code in Fig. 3.

In the above decision process, the rate λ can be measured locally, and the quantity $\mu_{MAX-TDMA}$ is dimensioned based on the preset frame duration. The only unknown quantity is $\mu_{CSMA/CA}$ which needs to be computed as a function of all system parameters, especially considering the number of active neighbors n . It can be computed from the following analysis [10-15] which involves solving a set of equations as follows.

$$\frac{1}{b_{0,0}} = \begin{cases} \frac{2pW[1-(2p)^m](1-p) + p(1-2p)(1-p^m)}{2(1-2p)(1-p)} + \frac{(2-p_b)(W+1)}{2} + \frac{1-p_b}{p_a}, m \leq m' \\ \frac{2pW[1-(2p)^m](1-p) + p(1-2p)(1-p^m)}{2(1-2p)(1-p)} + \frac{(2-p_b)(W+1)}{2} + \frac{1-p_b}{p_a} \\ + \frac{p^{m+1}(W_{max}+1)(1-p^{m-m'})}{2(1-p)}, m > m' \end{cases} \quad (4)$$

$$\tau = \sum_{i=0}^m b_{i,0} = \frac{1-p^{m+1}}{1-p} b_{0,0} \quad (5)$$

$$p = 1 - (1 - \tau)^{N-1} \quad (6)$$

$$1/\mu = \frac{1-p-p(2p)^m}{1-2p} \frac{W}{2} \sigma + t_x + t_c \frac{p}{1-p} + \rho(N-1)(t_x + t_c \frac{p}{1-p}) \quad (7)$$

$$p_a = 1 - e^{-\lambda\sigma} \quad (8)$$

$$p_b = \rho = \lambda/\mu \quad (9)$$

$$t_x = RTS + SIFS + CTS + SIFS + Header + DATA + DIFS \quad (10)$$

$$t_c = RTS + DISF \quad (11)$$

where m is the maximum number of retransmissions, m' is the maximum number of back-off stages, σ is the slot duration of CSMA/CA, W is the minimum congestion window size, N is the number of contenders, λ is the data rate and μ is the MAC service rate. The quantity $\mu_{CSMA/CA}$ can be obtained from Eqn. 9 after the above system of equation (Eqns. 4 through 11) is solved [10-15].

4 Experimental Evaluation

The proposed protocol switching logic has been implemented in a 100-node mesh network within the NS2 simulator [16]. A wireless data rate of 2 Mbps has been chosen with fixed packet duration of 2 ms, which is also the TDMA slot duration. Unless stated otherwise, we have fixed the TDMA frame size to 20 slots for all experiments. The numbers of active nodes and data generation rate have been varied. Before running protocol switching logic, we run a distributed TDMA protocol ISOMAC [17] to assign TDMA slots to the network nodes. The other baseline parameters for the simulation are shown in Table 1.

Table 1. Baseline experimental parameters in simulation

TDMA MAC Related	
Frame Size (number of slots)	20
TDMA Slot Duration	2ms
CSMA/CA MAC Related	
Minimum Congestion Window Size	32
Maximum Number of Back-off Stage	5
CSMA/CA Slot size	20us
Maximum Number of Retransmission	7
SIFS Duration	10 us
DIFS Duration	30 us
Protocol Switching Related	
Evaluation Time	W=3
Percentage of Decision Count	D=55%

4.1 Switching in Response to Changes in Data Rate

As presented in Section 3.1, since the data rate λ is a key criterion for protocol switching, its variation in a network neighborhood is expected to trigger such switching for maintaining high network throughput. In practice, the data rate variation in wireless sensor network can be caused by several factors including start or termination of applications and changes in application data rates. In the simulated 100-node mesh network, at any given time each node sends packets to one of its 1-hop neighbors at a given rate. Such rates are changed from 100 packets/second to 20 packets/second at time 21.0 second, and then changed back again at time 31.0 second. The corresponding network dynamics in terms of MAC protocol switching and its impacts on the network throughput are shown in Fig. 4.

In addition to the data rate graph (the solid line) for each node, Fig. 4 reports three other lines representing network-wide throughput for three distinct operating situations. The first one demonstrates the network throughput when all nodes run CSMA/CA. The second one shows what happens when all nodes run TDMA. Finally, the one marked switching, demonstrates the network throughput in the presence of adaptive MAC protocol switching following the logic as described in Section 3.

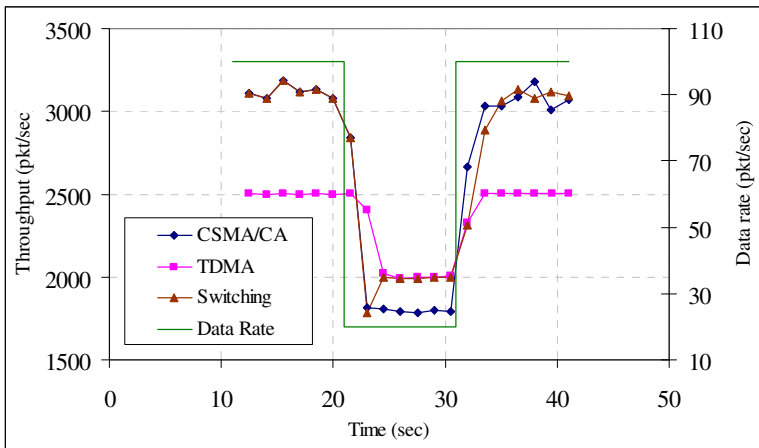


Fig. 4. Impacts of protocol switching with varying data rate

As can be seen in Fig. 4, initially when the data rate is high, the CSMA/CA mode of operation provides better effective throughput than the TDMA mode since the TDMA throughput is bounded by the frame size. Note, however, that the CSMA/CA mode throughput is lower than the maximum possible throughput, indicating certain amount of throughput loss due to the CSMA/CA contention and collisions. These contention and collisions for CSMA/CA persist even when the data rate is lowered (at time 20 sec.). At lower rates, however, TDMA can sustain the offered load since it is able to provide completely collision free MAC access to the channel. As a result, as evident from Fig. 4, the TDMA mode of operation provides better network-wide throughput at lower loading conditions.

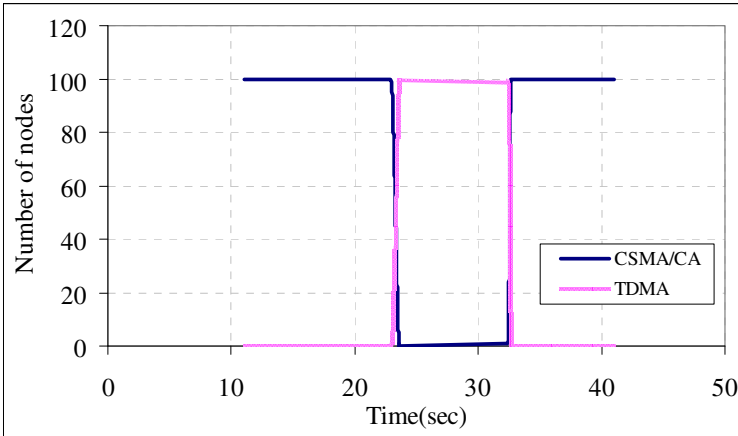


Fig. 5. MAC protocol dynamics with varying data rate

The throughput line representing protocol switching in Fig. 4 demonstrates that via exercising such MAC layer switching, the network is able to track the best possible throughput (those of CSMA/CA at high load, and TDMA at low load) as a response to the change in loading conditions. This partially validates the switching logic presented in Section 3.

The dynamic behavior of MAC switching is also shown in Fig. 5. As the data rate changes from high to low, the number of nodes running CSMA/CA reduces and the number of nodes running TDMA rises. Note that there is a lag between when the data rate changes and when the nodes actually switch their protocols. It was experimentally observed that this delay is caused due to the rate measurement latency at the MAC layer.

4.2 Switching in Response to Number of Active Nodes

As presented in Section 3.1, the number of active nodes in a neighborhood is the second key parameter that influences the quantity $\mu_{CSMA/CA}$, thus affecting the decisions for protocol switching. The variation of the number of active nodes in a network neighborhood is expected to trigger protocol switching for maintaining high network throughput.

An experiment was conducted to study the impacts of variable number of active nodes on the performance of protocol switching. The throughput results in Fig. 6 correspond to an initial active node-count of 25 which is increased to 100 at time 21 second, and then changed back to 25 at time 31 second. When there are only 25 active nodes, they are randomly selected out of all 100 network nodes. All active nodes send data to all 1-hop neighbors at the rate of 35 packets/second.

Fig. 6 demonstrates that with lower number of active nodes (i.e. till time 21sec and then after 31 second) CSMA/CA performs better because of its larger MAC service rate in the absence of collisions as explained in Section 3.1. When the number of contender nodes increases, the contention brings down the service rate of CSMA/CA

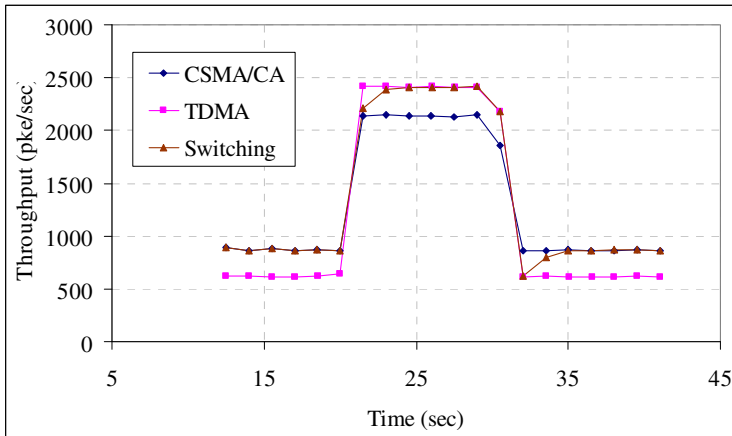


Fig. 6. Protocol switching with varying active node count

compared to that during the TDMA mode of operation. This explains why CSMA/CA outperforms TDMA in low active node-count situations, but not when it is high (i.e. between 21 sec to 31 sec).

The throughput line representing protocol switching in Fig. 6 demonstrates that by exercising MAC layer protocol switching, the network is able to track the best possible throughput (those of CSMA/CA at low active node-count, and TDMA at high active node-count). This completes the validation of the switching logic presented in Section 3.

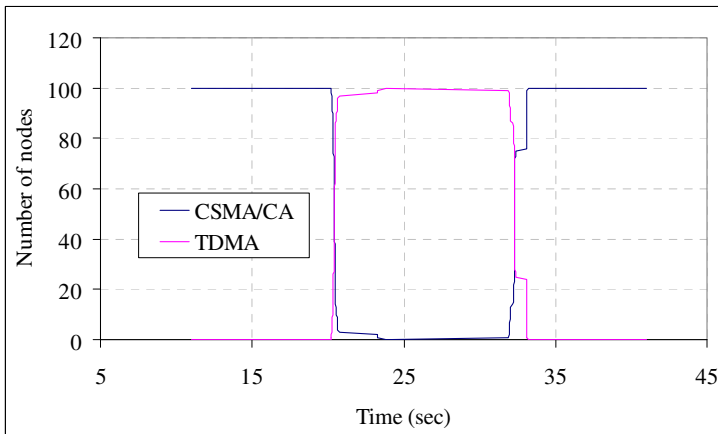


Fig. 7. Protocol dynamics with varying active node count

The dynamics of the number of nodes executing CSMA/CA and TDMA are depicted in Fig. 7. As in Fig. 5, the latency in switching transient is caused due to the rate measurement latency at the MAC layer.

4.3 Switching with Spatially Clustered Traffic

All experimental results so far describe protocol switching in a network with homogeneous traffic, thus causing all the active nodes to run the same protocol and to switch at the same time. In this Section we describe experiments with spatial traffic heterogeneity, which can cause different network nodes to run different MAC protocols at the same time, and to not necessarily switch in a synchronous manner.

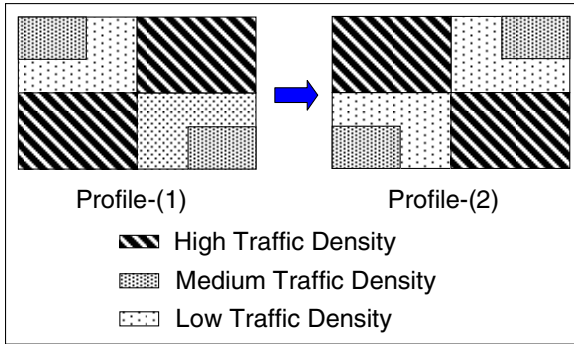


Fig. 8. Network with spatial traffic heterogeneity

Fig. 8 shows a rectangular sensor field equipped with a mesh network topology consisting of 400 nodes. Two different heterogeneous traffic profiles, each consisting of low, medium, and high density traffic regions, are used. In high density region, each node sends to all its 1-hop neighbors with a data rate of 3 packets/second per neighbor. Each node in the medium density region sends only to one of its 1-hop neighbors at a data rate of 140 packets/second. A node in the low traffic density region also sends only to one of its 1-hop neighbors at a rate of 12.5 packets/second. During a 60sec long experiment, the network traffic profile has been changed once from Profile-(1) to Profile-(2) at time approximately 31.0 second.

A node in the high traffic density region is expected to run TDMA to achieve higher throughput, because the MAC service rate of CSMA/CA is less than that of TDMA at high traffic level. CSMA/CA is expected in a medium traffic region, and TDMA is expected again in a low density region.

Fig. 9 reports the impacts of time-varying traffic profile on the throughput dynamics with and without MAC protocol switching. After the network is initiated at around 12 sec, observe that in the all-TDMA scenario, the network throughput settles down at around 18sec. Then it remains steady till 31sec which is when the traffic profile switches from Profile-(1) to Profile-(2). As a result of this, the throughput goes through a transient state before it settles back down roughly at the same value. Very similar trends can be also observed for the all-CSMA/CA scenario, except that with all-CSMA in this case the achievable network throughput happens to be higher than that of the all-TDMA case.

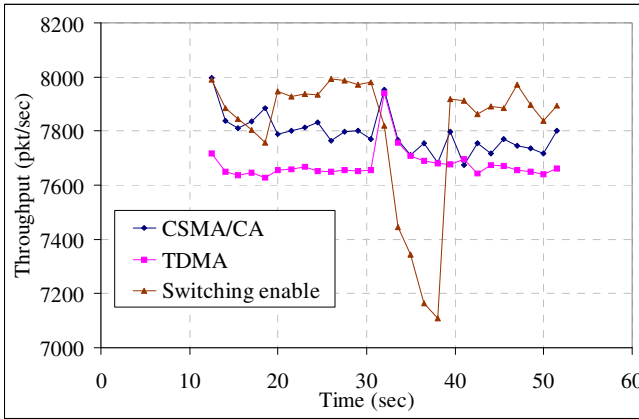


Fig. 9. Throughput dynamics with traffic heterogeneity

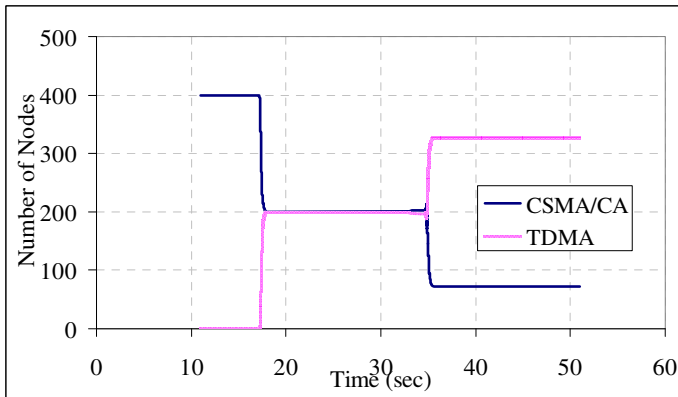


Fig. 10. Protocol dynamics with traffic heterogeneity

Observe that with protocol switching, the achieved throughput is higher than the two without switching cases during both before and after the traffic profile changes. Note that there is a significant transient drop for the with-switching case during the profile change. This drop in performance is caused because certain packets experience increased contention and collisions during the profile change. Such contentions and collisions lead to MAC layer drops and subsequent loss of throughput. After the transient situation is over, the protocol switching mechanism can track the new traffic profile distribution and able to select appropriate MAC protocols at the network nodes to bring the throughput back higher than the all-TDMA and all-CSMA/CA scenarios.

Two observations can be made from Fig. 10 which reports the number of nodes running TDMA and CSMA/CA during the entire experiment. First, nodes change their protocols in a period of few seconds after the traffic profile actually changes. This is due to the traffic rate measurement latency at the MAC layer. Second, the distribution of numbers of TDMA and CSMA/CA nodes during Profile-(1) and

Profile-(2) are different. During when Profile-(1) is executed, the number of TDMA nodes and the number of CSMA/CA nodes are equal. However, during when Profile-(2) is executed, there are 328 nodes are running TDMA and only 72 nodes are running CSMA/CA. It was experimentally found that this variation is a result of the irregularity in network connectivity. Due to this topological irregularity, even nodes in the same traffic density region experience different number of contender nodes n , which impacts the protocol switching decision.

5 Conclusions and Ongoing Work

In this paper we have presented a distributed MAC protocol switching mechanism for maximizing MAC layer throughput in the presence of traffic and topology heterogeneity. The key idea behind dynamic MAC switching is for each node to use its local topology and traffic density information to decide the most suitable MAC protocol that can maximize the MAC layer throughput in the neighborhood. A formal MAC switching rule has been developed using analytical formulation of the MAC throughput available in the literature. NS2 based simulation experiments demonstrate that with the proposed MAC switching strategy, nodes in a mesh network are able to achieve maximum MAC throughput by adaptively choosing the appropriate MAC protocol in the presence of heterogeneity in terms of data rate and node population. Ongoing work on this topic includes generalizing the proposed mechanisms for a broader set of MAC protocols beyond CSMA/CA and TDMA.

References

- [1] Balakrishnan, H.: Opportunities and challenges in high-rate wireless sensor networking. In: 29th Annual IEEE International Conference on Local Computer Networks (November 2004)
- [2] Kleinrock, L., Tobagi, F.: Carrier Sense Multiple Access for Packet Switched Radio Channels. In: International Conference on Communications, Minneapolis, Minnesota, pp. 21B-1–21B-7 (June 1974)
- [3] Ye, W., Heidemann, J., Estrin, D.: An Energy-Efficient MAC Protocol for Wireless Sensor Networks. In: INFOCOM, pp. 1567–1576 (June 2002)
- [4] Dam, T., Langendoen, K.: An Adaptive Energy-Efficient MAC Protocol for Wireless Sensor Networks. In: 1st ACM International Conference on Embedded Networked Sensor Systems, pp. 171–180 (November 2003)
- [5] Chen, Z., Khokhar, A.: Self organization and Energy Efficient TDMA MAC Protocol by Wake up for Wireless Sensor Networks. In: 1st Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks 2004 (IEEE SECON 2004), pp. 335–341 (October 2004)
- [6] Rajendran, V., Obraczka, K., Garcia-Luna-Aceves, J.J.: Energy-Efficient, Collision-Free Medium Access Control for Wireless Sensor Networks. In: International Conference on Embedded Networked Sensor Systems, pp. 181–192 (November 2003)
- [7] Rhee, I., Warrier, A., Min, J., Xu, L.: DRAND: Distributed Randomized TDMA Scheduling for Wireless Ad-hoc Networks. In: The 7th ACM International Symposium on Mobile Ad Hoc Networking and Computing, Florence, Italy, May 22-25 (2006)

- [8] Ahn, G.-S., Miluzzo, E., Campbell, A.T., Hong, S.G., Cuomo, F.: Funneling-MAC: A Localized, Sink-Oriented MAC For Boosting Fidelity in Sensor Networks. In: Fourth ACM Conference on Embedded Networked Sensor Systems (SenSys 2006), Boulder, Colorado, USA (November 2006)
- [9] Funneling-MAC Technical Report,
<http://www.cs.dartmouth.edu/~sensorlab/funneling-mac/TAPTR-2006-08-003.pdf>
- [10] Sudarev, J.V., White, L.B., Perreau, S.: Performance Analysis of 802.11 CSMA/CA for Infrastructure Networks under Finite Load Conditions. In: The 14th IEEE Workshop on Local and Metropolitan Area Networks, 2005, LANMAN 2005 (September 2005)
- [11] Chung, M.Y., Jung, M.-H., Lee, T.-J., Lee, Y.: Performance Analysis of HomePlug 1.0 MAC With CSMA/CA. *IEEE Journal on Selected Areas in Communications* 24, 1411–1420 (2006)
- [12] Lee, W., Wang, C., Sohrawy, K.: On Use of Traditional M/G/1 Model for IEEE 802.11 DCF in Unsaturated Traffic Conditions. In: IEEE Wireless Communications and Networking Conference (WCNC 2006), Las Vegas, Nevada, pp. 1933–1937 (April 2006)
- [13] Barowski, Y., Biaz, S.: The Performance Analysis of IEEE 802.11 Under Unsaturated Traffic Conditions (August 2004),
<ftp://ftp.eng.auburn.edu/pub/techreports/csse/04/CSSE04-08.pdf>
- [14] Bianchi, G.: Performance Analysis of the IEEE 802.11 Distributed Coordination Function. *IEEE Journal on Selected Areas in Communications* 18, 535–547 (2000)
- [15] Tickoo, O., Sikdar, B.: A queuing model for finite load IEEE 802.11 random access MAC. In: IEEE International Conference of Communication (ICC), pp. 175–179 (2004)
- [16] The network simulator: NS-2, <http://www.isi.edu/nsnam/ns>
- [17] Yu, F., Wu, T., Biswas, S.: Towards In-Band Self-Organization in Energy-Efficient MAC Protocols for Sensor Networks. *IEEE Transaction of Mobile Computing* (accepted) (2007)

DFDNM: A Distributed Fault Detection and Node Management Scheme for Wireless Sensor Network^{*}

Indrajit Banerjee¹, Prasenjit Chanak³,
Biplab Kumar Sikdar², and Hafizur Rahaman¹

¹ Department of Information Technology

² Department of Computer Science

³ Purabi Das School of Information Technology

Bengal Engineering and Science University, Shibpur, Howrah, India

{ibanerjee, rahaman_h}@it.becs.ac.in,

prasenjit.chanak@gmail.com,

bipla@cs.becs.ac.in,

Abstract. This paper presents a distributed fault detection and node management scheme with the help of cellular automata in wireless sensor networks. The node failure is detected in distributed way and the cellular automata are used for network management. The linear cellular automata run on every sensor nodes and each sensor node set own state with the help of CA. The faulty are detected by neighbour node's temporal and spatial correlation of sensing information and heart beat message passing by the cluster head. The proposed distributed fault detection and node management scheme (DFDNM) can detects and manages the faulty nodes in an energy efficient manner. The simulation result shows how DFDNM perform well comparison to others faults detection algorithms.

Keywords: Wireless sensor network (WSN), fault detection (FD), cellular automata (CA), base station (BS).

1 Introduction

The wireless sensor network (WSN) is an ad-hoc network that encompasses of small inexpensive low power device, deployed in large number at remote geographical region, office building or in industrial plants [1]. A WSN is used for environment monitoring which includes air, soil and water, habitat monitoring, military surveillance, inventory tracking, condition base maintenance etc [2].

Due to low cost, battery powered and possible deployment at harsh and hostile environment, the sensors are prone to failure. Faulty sensor nodes may cause wrong data sensing, erroneous data processing and incorrect data communications.

^{*} This work is partially supported by the Pramoda Lodh Laboratory, PDSIT, BESU.

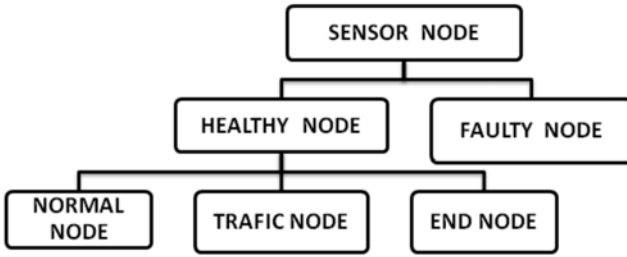


Fig. 1. Classification sensor nodes with respect to fault

The main components of a sensor node are a microcontroller, transceiver circuits, memory, power source and one or more sensors. The microcontroller is mainly responsible for data processing and management of other components of sensor nodes. The functionality of both transmitter and receiver are combined into a single device known as transceivers. The most relevant kind of memory is on-chip memory of a microcontroller. The power is stored in batteries, both rechargeable and non-rechargeable, are the main source of power supply for sensor nodes. The sensor in a node is a hardware device that is responsible for measuring physical data of the parameter to be monitored. The WSN node faults are due to failure of any one of its hardware components' as discussed above. The node status in WSN'S can be divided into two types: healthy and faulty (shown in Fig 1). A node is defined as faulty if its battery power reaches the threshold limit or if its microcontroller does not work or if the transmitter circuit of the transceiver is not working properly. In a node if the transmitter circuit is not healthy, even if all other hardware components are in good shape, the node is declared as a faulty node. The faulty node must be replaced by a new node; otherwise, its responsibility has to be shared by other available healthy nodes. The healthy node may again be categorized into three groups: traffic node, normal node, and end node. In a healthy node where the transceiver is operational but the sensor device is malfunctioning, then we may use this node as a traffic node. A traffic node can act as a router in multi-hop wireless data communication. The normal healthy node where all components of a sensor node are in good shape, may be used for any type of job in WSN. In the end node, the receiver circuit of the transceiver is malfunctioning, hence it can sense the parameter of the monitoring field and is able to transmit to the base station via other nodes. However, an end node cannot receive data from its neighbour node; therefore, it cannot be used as a router in WSN.

The proposed techniques have different fault detection techniques. On the basis of types of fault detected in the node, they are classified into different categories as discussed above. Then we have proposed a cellular automata-based faulty network management scheme. The CAs are installed in all the nodes which work on the basis of local information. The cluster head, after gathering the fault information, takes a decision about its member node's responsibility and sends them instructions by passing the appropriate CA rules.

Table 1. Truth table for rule 90, 70

Present State	111	110	101	100	011	010	001	000
70	0	1	0	0	0	1	1	0
90	0	1	0	1	1	0	1	0

This paper is organized as follows: Section 2 introduces an overview of cellular automata. Section 3 describes the mathematical model. Section 4 describes fault detection and recovery technique and algorithms. Section 5 reports the simulation results of our algorithm. Section 6 concludes the paper.

2 Cellular Automata

In this section we briefly describe the cellular automata theory relevant to the work [3]. A cellular automata (CA) is a discrete time decentralized system. Cellular automata consists of a number of cells, cell change their states according to their neighbour cells information with respect to time. In 3-neighbour CA cells change their states according to equation number 1.

$$S_i^{t+1} = f_i(S_{i-1}^t, S_i^t, S_{i+1}^t) \quad (1)$$

The S_{i-1}^t is the left neighbour cells information at time t, S_i^t is the own state information of the cells at time t and S_{i+1}^t is the right neighbour cell information at time t. f_i defines the next state logic for the cell i. The table 1 represent truth table of f_i for two such rules 70 and 90.

An n-cell CA is represented by the rule vector $R = \langle R_1, R_2, \dots, R_i, R_{i+1}, \dots, R_n \rangle$, where R_i is the rule selected for the i^{th} cell. The state of an n-cell CA at time t is

$$S^t = (S_1^t, S_2^t, \dots, S_i^t, \dots, S_n^t) \quad (2)$$

If all the cells of a CA obey the same rule, the CA is said to be a uniform CA, otherwise it is a hybrid CA. If any CA is defined left most cell and right most cell are 0 state then this CA is called Null Boundary CA. A CA is said to be a Periodic Boundary CA if the extreme cells are adjacent to each other.

3 Mathematical Analysis

3.1 Energy Requirement for Wireless Communication

Communication energy consumption is happened because of transmission or receiving data from other nodes. The energy loss by each node for single message transmission is represented by the linear equation [4].

$$E_t = (\phi_t + \phi_d r^n) d_i \quad (3)$$

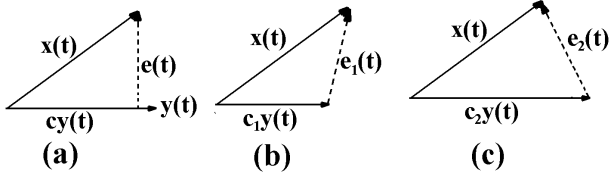


Fig. 2. Ideal condition of data sensing by the sensors when all neighbouring nodes are sensing same data. Fig.2 (b) and Fig. 2(c) shows the sensor circuit fault condition, when $e_1(t)$ and $e_2(t)$ are greater than the threshold value.

Where, ϕ_t is the energy loss per bit by the transmitter electronics circuit, and ϕ_d is the dissipated energy in the transmit op-amp. Transmission range is r^n . The parameter n is power index for the channel path loss of the antenna. d_i message size which is transmitted by each sensor nodes.

Receiving energy loss of each node is represented by the following formula [4]

$$E_r = \phi_r d_i \quad (4)$$

Where, ϕ_r is energy per bit which is consumed by the receiver's electronics circuit. d_i message size which is received by each sensor node.

3.2 Fault Detection

In sensor network sensor nodes are distributed in a particular area. The neighbouring nodes are closely deployed in the sensing region. Therefore the sensed information of the neighbouring node varies very small. In DFDNM sensor nodes are sensing information at the same rate from monitoring environment. If the neighbour node sensing information is $x(t)$ and current node sensing information is $y(t)$. Then difference between two sensed information is very small because they are closely deployed (Fig.2).

Theorem 1. If information sensing by the sensor node is $y(t)$ and neighbour node sensing information is $x(t)$, are two vector products for a given time (t_1, t_2) and the $cy(t)$ is $x(t)$ along $y(t)$ then information difference vector $e(t)$, is represented by

$$e(t) = \begin{cases} x(t) - cy(t) & t_1 \leq t \leq t_2 \\ 0 & \text{otherwise} \end{cases}$$

If different vector $e(t)$ less than threshold value then the sensing information by the sensor circuit is good. Otherwise sensor circuit of the node is faulty.

Proof: Let the component of $x(t)$ along $y(t)$ is $cy(t)$. Geometrically the component of $x(t)$ along $y(t)$ is the projection of $x(t)$ on $y(t)$, and is obtained by drawing a perpendicular from the tip of $x(t)$ on the vector $y(t)$.

$$x(t) = cy(t) + e$$

Where $x(t)$ is represented in terms of $y(t)$ plus another vector, called the information different vector. If we approximate $x(t)$ by $cy(t)$

$$x(t) \cong cy(t)$$

The difference in this approximation is the vector

$$e = x(t) - cy(t)$$

The parameter c is chosen to minimize the length of the different vector. For convenience we define the dot product of two vectors $x(t)$ and $y(t)$ as

$$x(t).y(t) = |x(t)|.|y(t)|\cos\theta$$

Where θ is the angle between vector $x(t)$ and $y(t)$. Using this definition, we can express $|y(t)|$. The length of a vector $y(t)$, is

$$|y(t)|^2 = y(t).y(t)$$

Now, the length of the component of $x(t)$ along $y(t)$ is $|x(t)|\cos\theta$, but it is also $cy(t)$. Therefore,

$cy(t) = |x(t)|\cos\theta$ Multiplying both side by $|y(t)|$ yields
 $c|y(t)|^2 = |x(t)|.|y(t)|\cos\theta = x(t).y(t)$ and

$$c = \frac{x(t) \cdot y(t)}{y(t) \cdot y(t)} = \frac{1}{|y(t)|^2} x(t) \cdot y(t)$$

When $x(t)$ and $y(t)$ are perpendicular, or orthogonal, then $x(t)$ has a zero component along $y(t)$; consequently, $c=0$. Therefore defined $x(t)$ and $y(t)$ to be orthogonal if the dot product of the two vectors is zero, that is , if

$$x(t).y(t) = 0$$

The sensing information of nodes are equal then the difference vector $e(t)$ is zero. This is the ideal condition because in this case no environmental briars are present i.e. each and every node collects same information.

4 Proposed DFDNM Approach

This section describes our proposed distributed fault detection and node management scheme, referred to as the DFDNM. In this section first we briefly describe different fault detection techniques of nodes; second we are discussing about the utilization of cellular automata for faulty network management in DFDNM. After this we describe DFDNM algorithm.

The DFDNM technique detects five types of node's hardware failures such as battery failure, microcontroller failure, sensor failure, transmitter circuit failures and receiver circuit failure (Table 2). According to the type of failure the sensor nodes are used in different purpose such as *normal node*, *traffic node* and *end node* The node itself can detect the battery/power failure by periodical check up of its energy level. If the battery energy level of sensor node is less than

Table 2. Node component and corresponding fault

Type of fault	Detected by	Method of detection
Battery	Self	When battery energy level goes below threshold value
Microcontroller	Cluster Head	By diagnosing heart beat message
Sensor	Self	By neighbours node data comparison
Transmitter circuit	Cluster Head	By diagnosing the heart beat message
Receiver circuit	Self	By diagnosis received data

Table 3. Node fault information

Bit representation	Node Condition		
	Sensor circuit	Battery/Microcontroller/Transmitter Circuit	Receiver Circuit
111	Healthy	Healthy	Healthy
110	Healthy	Healthy	Faulty
101	Healthy	Faulty	Healthy
100	Healthy	Faulty	Faulty
011	Faulty	Healthy	Healthy
010	Faulty	Healthy	Faulty
001	Faulty	Faulty	Healthy
000	Faulty	Faulty	Faulty

threshold value then battery fault occurs. Therefore, it will announce itself as faulty node by sending message to its neighbour. In DFDNM technique each node periodically sends a *heartbeat message* to its neighbour. However, if this *heartbeat message* is not received for certain period of time then the cluster head can declare the corresponding node as *faulty node* because of its microcontroller or transmitter circuit failure. In DFDNM each node is responsible for detecting its sensor fault. The sensor failure is detected by each node with the help of neighbour nodes data comparison. The node collects data from their nearest neighbour nodes and compares it with own sensing data. In theorem 1 we have calculated the data comparison results $e(t)$, if that crosses the threshold value then sensor circuit of sensor node is declared as faulty. If any node finds its sensor fault it will declare itself as a *traffic node* In DFDNM each node can detect its receiver circuit failure by simply diagnosing its received data. A node can declare itself as *end node* if its receiver circuit is faulty.

In the DFDNM every sensor node, cluster heads and base station runs linear CA which stores the neighbour node failure information as shown in Fig.3. The node failure information is represented by three bit that is stored in three different linear CA in each node of WSN. The i^{th} bit of first CA represents sensor

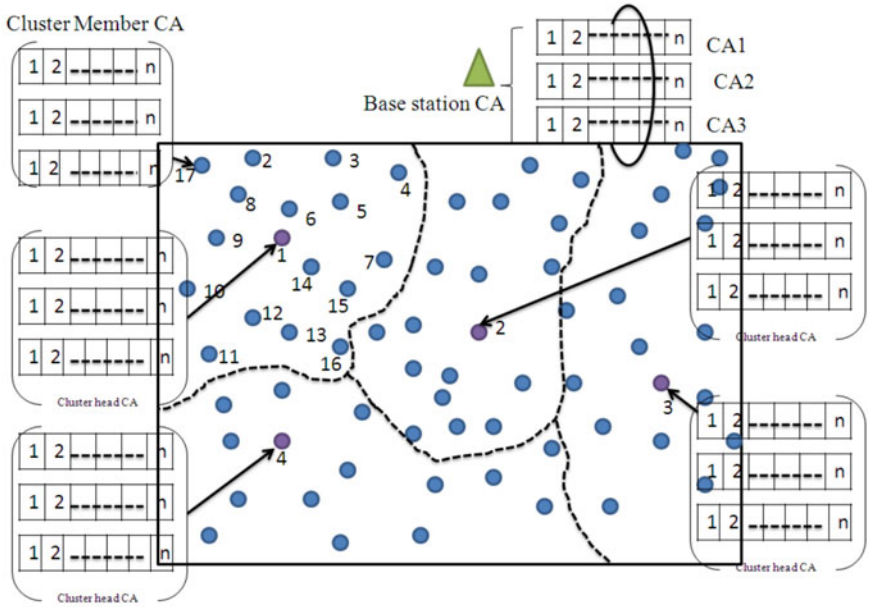


Fig. 3. CA structure in each node

Table 4. Classification of nodes depending on CA state

1	1	1	1	0	0	0	0
1	1	0	0	1	1	0	0
1	0	1	0	1	0	1	0
Normal Node	End Node	Faulty Node	Faulty Node	Traffic Node	Faulty Node	Faulty Node	Faulty Node

circuit condition, the i^{th} bit of middle CA represent the battery, microcontroller and transmitter circuit condition and i^{th} bit of last CA represents the receiver circuit condition (Table 3).

The cluster head collects different failure information of its member nodes which is then feed into its three linear CA. The cluster head is now determining the condition of its member node by analysing the CA as per Table 4. Therefore the nodes responsibility is determined depending on the bit status of the CA.

The three neighbours CA changes its state according to three neighbour's information as equation number 1. We have classified the three neighbours' CA rules into two groups according to their state change dependency (i) Partially dependent CA rules (ii) Fully dependent CA rules (Table 5).

Definition 1. Partially dependent CA, the next state of a cell is determined depending on its two neighbour cells state information only, not three neighbour cells state information.

Table 5. Classification of CA

CA Class		CA rules
Partially dependent CA	Partially dependent left CA	12,15,48,51,192,204,240
	partially dependent right CA	17,118,136,153,232,187
Fully dependent CA	8,50,59,49,60,62,61,64,128,129,130,200...	

Table 6. Next state of nodes depending upon the CA State

CA rule	111	110	101	100	011	010	001	000
192	1	1	0	0	0	0	0	0
136	1	0	0	0	1	0	0	0
200	1	1	0	0	1	0	0	0
128	1	0	0	0	0	0	0	0
64	0	1	0	0	0	0	0	0
8	0	0	0	0	1	0	0	0

In CA, if left most cell content δ_L information, middle cell content δ_M information and right most cell content δ_R information then partially dependent CA (PDCA) rules depends on any two information between δ_L , δ_M and δ_R .

Example 1. The CA following the rules $R_i < 12, 15, 17, \dots, 192, \dots >$ in table 5 is partially dependent CA that changes its state depending on left most cell information (δ_L) and middle cells information (δ_R). The right cells information (δ_R) does not affect the state change of the CA cell.

Definition 2. Fully dependent CA, the next state of cell change depends on its three neighbour cells state information.

Example 2. In fully dependent CA, the state change is depends on the three neighbour's information. The CA follows the rules $R_i = < 128, 64, 8, \dots, >$, the next state change is depending on its three cell state information. In these types of rules their no state change will occur depending only on its two cell's information. Partially dependent CA can farther be classified into two groups (i) Partially dependent left CA (ii) Partially dependent right CA.

Definition 3. Partially dependent left CA, the next state of the cell depends on the left and middle cells state only.

In Partially dependent left CA, right neighbour cell state does not affect the next state of the CA. In rules $R_L = < 12, 15, 48, 192, \dots >$ the next state of cell 'I' is depending on the left most cells condition (δ_L) and on middle cell condition (δ_M). In this rule right most cell state (δ_R) 0 or 1, does not affects the next state 'I' of the CA.

Definition 4. Partially dependent right CA, the next state of the cell depends on the right and middle cell state.

In Partially dependent right CA, left neighbour cell state information is not affected the next state of the CA. In rules $R_R = \langle 17, 118, \dots, 136 \dots \rangle$ the next state 'I' of CA depends on right (δ_R) and middle cells state information (δ_R). In this rules left cell condition (δ_L) 0 or 1 does not affect the next 'I' state of the CA.

Theorem 2. The partially dependent CA rules are sub set of classical dependant CA rules.

Proof: Let parent set of rule is $R = \langle R_1, R_2 \dots R_n \rangle$. The child set of CA rule is $R_i = \langle R_1, R_2, \dots \rangle$ sub set of R. The classic CA rule vector $R = \langle R_1, R_2 \dots R_n \rangle$ and PDCA rule vector $R_i = \langle R_1, R_2, R_3, \dots, R_n \rangle$ is a sub set of rule vector R. The fully dependent CA rule vector $R_k = \langle R_1, R_2 \dots R_n \rangle$ is also a subset of rule vector R. The fully dependent rule vector $R_k = R - R_i$. The partially dependent left CAs rule vector $R_L = \langle R_1, R_2 \dots R_n \rangle$ is a sub set of rule vector R_i . The partially dependent right CAs rule vector $R_R = \langle R_1, R_2 \dots R_n \rangle$ is also a subset of rule vector R_i . Therefore, rule vector $R = R_i \cup R_k$ and rule vector $R_i = R_L \cup R_R$.

DFDNM Algorithm

```

WHILE network is active DO
  FOR each node DO
    IF node is active THEN
      Check microcontroller fault
      Check battery fault
      Check transmitter circuit fault
      CALL neighbour information verification
      Check receiver circuit
      Decrement node energy by active value
    ELSE
      CALL Time Verification Procedure
    END IF
  END FOR
  Sensor nodes run their own CA.
END WHILE

```

The DFDNM used partially dependant left CA rules for *end* sensor node or *normal* node management, whereas the partially dependent right CA rules are used for *traffic* nodes or *normal* nodes management. The fully dependent CA is used for only *normal* node, *traffic* node, *end* nodes management, shown in table 7. The cluster head in DFDNM collects all fault information from cluster member nodes. According to fault information, cluster head will determine the

responsibility of the member nodes. Then the cluster head will inform its member node the status of the other member node by sending appropriate rule to each other. For next few round until the fault status of the cluster member node is unchanged, the individual nodes will understand the status of its neighbour from the CA itself.

The detail for the implementation of proposed DFDNM algorithm is described next.

Neighbour Information Verification

```

FOR each neighbour nodes DO
    Send SINGLE message
END FOR
Calculate nodes information Difference from Theory 1
IF it is less the threshold THEN
    Nodes is Active non faulty
ELSE
    Nodes is faulty 0 state
END IF

```

Timer Verification Procedure

```

IF node energy is zero THEN
    Declare the node is dead
ELSE
    IF time is zero THEN
        Consult CA to renew the state
        Reset the timer
    ELSE
        Decrement the timer
    END IF
END IF

```

5 Experimental Results

This section reports the simulation results received while running DFDNM scheme. In order to evaluate the performance of our proposed scheme, described in Section V, three traditional metrics have been considered:

1 Time elapsed: This is the average time taken to determine the fault in the network.

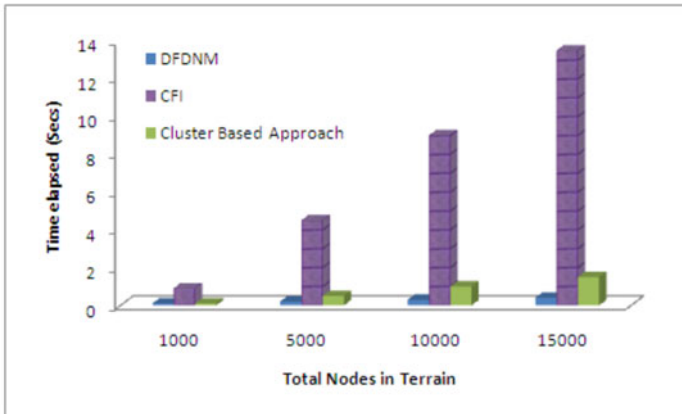
2 Energy loss rate: The energy loss by the network with respect to time.

Table 7. CA rule for different node function

CA Rule Class	Function of Node
Fully dependent CA	Normal node
	End node
	Traffic node
Partially dependent left CA	End/Normal node
Partially dependent right CA	Traffic/Normal node

Table 8. Simulation parameters

Sensor Deployment Area	128128
Number of node	16384
Data Packet Size	800bit
Initial Energy	0.5J
ϕ_t energy loss by transmitter electronics circuit	50 nJ/bit
ϕ_d dissipated energy by transmit op-amp	10 pJ/bit/ m^2

**Fig. 4.** Time Elapsed for Fault Detection

3 Number of healthy node: The number of healthy node signifies the total number of nodes that are in working condition. The normal node, end node and traffic node are considered to be as healthy node.

In this simulation we have considered total 16384 numbers of distributed sensor nodes. The values adopted for simulation are shown in table 8.

In Fig.4 we have compared the time required for fault detection between proposed DFDNM technique, crash faults identification (CFI) [5] and cluster base approach to fault detection and recovery [6]. It can be observed that with increase of network size the DFDNM can detect fault more quickly than other two approaches.

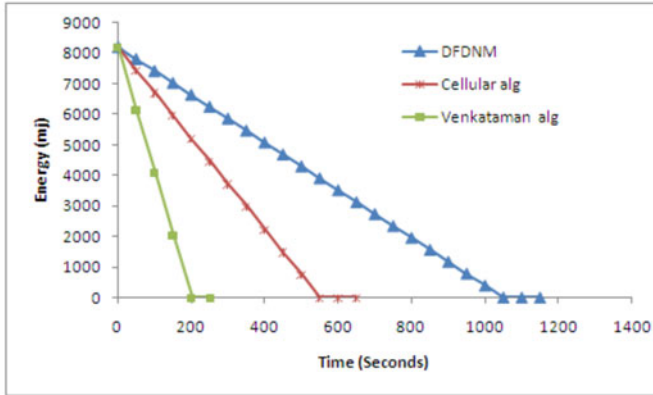


Fig. 5. Number of active nodes in WSN

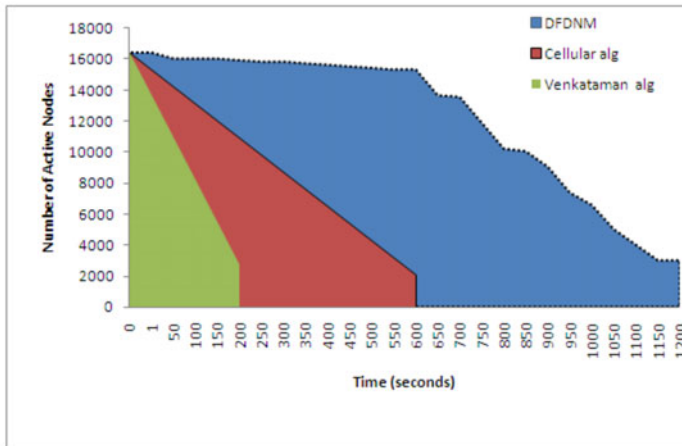


Fig. 6. Number of active nodes in WSN

The rate of energy loss in WSN is given in Fig. 5. The diagram depicts that in DFDNM the rate of energy loss is better than other existing techniques. The proper utilization of network energy in DFDNM technique have been increased the network life time up to 75% compare to cellular algorithm [7] and 90% comparison to Venkataraman algorithm [8].

The DFDNM technique improves the faulty node utilization compare to other existing techniques. The Fig. 6 compares the number of active nodes in DFDNM, cellular algorithm [7] and venkataman algorithm[8]. In DFDNM technique

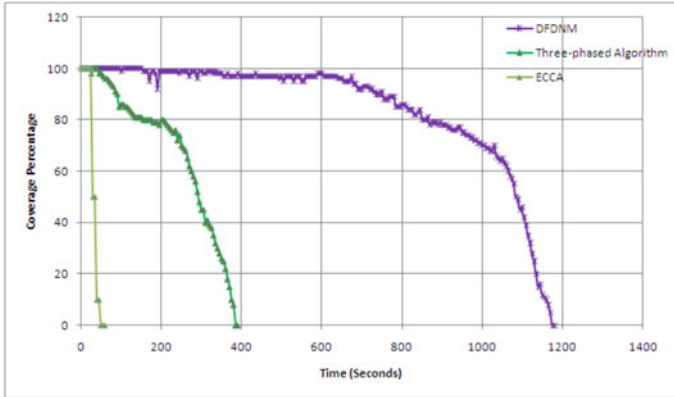


Fig. 7. The network coverage in WSN

number of healthy nodes is very high compare to other fault detection techniques because in this technique we have classified the node depending upon their failure type. Therefore, the node utilization is better in proposed DFDNM scheme.

Fig. 7. Shows that networks coverage comparison result between DFDNM, three phase algorithm, ECCA. The network coverage of the DFDNM technique which is 67.5% better then three phase algorithm and 90.65% better then ECCA algorithm.

6 Conclusions

In this paper we propose fault detection and cellular automata based faulty node management scheme in wireless sensor networks. We have proposed different fault detection techniques for different hardware components failure of the node. The nodes are classified into normal, traffic and end node depending on their hardware failure conditions and therefore, we have reused a faulty node for WSN. The cluster head after collection of member nodes' fault condition manage the network with the help of cellular automata. The CA helps us to manage the network with minimum data transmission and thereby we are saving a considerable amount of energy of the nodes.

References

1. Venkataraman, G., Emmanuel, S., Thambipillai, S.: A cluster Base Approach to fault Detection and Recovery in Wireless Sensor Networks. In: IEEE 2007 International Conference, ISWCS (2007)
2. Asim, M., Mokhtar, H., Merabti, M.: A cellular approach to fault detection and recovery in wireless sensor networks. In: IEEE 2009 Third International Conference on Sensor Technologies and Application (2009)

3. Banerjee, I., Das, S., Rahaman, H., Sikdar, B.K.: An Energy Efficient Monitoring of Ad-Hoc Sensor Network with Cellular Automata. In: IEEE International Conference on System Man and Cybernetics, Taiwan, October 8-11 (2006)
4. Banerjee, I., Chanak, P., Sikdar, B.K., Rahaman, H.: Energy Efficient Routin. In: Wireless Sensor Network. In: IEEE International Conference, IIT, Kharagpur, India (2011)
5. Chessa, S., Santi, P.: Crash Faults Identification in Wireless Sensor Network. *Computer Comm.* 25(14), 1273–1282 (2002)
6. Venkataraman, G., Emmanuel, S., Thambipillai, S.: A Cluster-Based Approach to fault Detection and Recovery in Wireless Sensor Networks. In: IEEE ISWCS (2007)
7. Asim, M., Mokhtar, H., Merabti, M.: A cellular approach to fault detection and recovery in wireless sensor networks. In: IEEE Third International Conference on Sensor Technologies and Applications (2009)
8. Banerjee, I., Chanak, P., Sikdar, B.K., Rahaman, H.: EERIH: Energy Efficient Routing via Information Highway in Sensor Network. In: IEEE International Conference on Emerging Trends in Electrical and Computer Technology, Kanyakumari, India, March 23-24 (2011)
9. Venkataraman, G., Emmanuel, S., Thambipillai, S.: Energy-efficient cluster-base scheme for failure management in sensor networks. *IET Commun.* 2(4), 528–537 (2008)

An Optimized Reduced Energy Consumption (OREC) Algorithm for Routing in Wireless Sensor Networks

Joydeep Banerjee¹, Swarup Kumar Mitra²,
Pradipta Ghosh¹, and Mrinal Kanti Naskar¹

¹ Advanced Digital and Embedded Systems Laboratory, Department of ETCE,
Jadavpur University, Kolkata

² Department of ECE, M.C.K.V.I.E, LiLuah, Howrah
jogs.1989@rediffmail.com,
{swarup.subha, iampradiptaghosh}@gmail.com,
mrinalnaskar@yahoo.co.in

Abstract. Wireless Sensor Network (WSN) is constrained by limited available energy for data communication. Utilization of energy and enhancing network lifetime is a challenging task for the next generation wireless sensor network. In this paper we have designed an algorithm that performs the routing with low energy consumption. We have experimentally optimized the number of clusters that should be formed for escalating the lifetime of the network with provisions made to include equal number of nodes in each cluster. Our analysis and simulation results show that with realistic radio model we have achieved better load balance than several existing protocols, like LBEERA, HDS, SHORT, PEGASIS, LEACH and BINARY. A suitable node deployment strategy was adopted for ensuring wireless connectivity between each node. Moreover we have made simulations in NS-2 which supports our propositions to a greater extend.

Keywords: Wireless Sensor Network, Network Lifetime, Realistic Radio Model, node deployment, NS-2.

1 Introduction

Wireless sensor networks (WSNs) consist of sensor nodes that are randomly deployed in a large area, collecting salient information from the sensor field. These sensor nodes are equipped with limited energy resources and hence, the energy consuming operations such as data collection, transmission and reception, must be kept at a minimum [1]. Also, it is often infeasible to replace or recharge the sensors nodes deployed in inaccessible terrains. The sensor networks are required to transmit gathered data to a base station (BS) or sink, often distantly located from the sensor field. Network lifetime thus becomes an important metric for efficiency of a sensor network. In case of WSNs, the definition of network lifetime is application specific [2]. It may be taken as the time from inception to the time when the network becomes nonfunctional. A network may become non-functional when a single node dies or when a particular percentage of nodes perishes depending on requirement. However,

it is universally acknowledged that equal energy dissipation for equalizing the residual energy of the nodes is one of the keys for prolonging the lifetime of the network [2]. Node density is another performance parameter which is determined by ratio of number of nodes by coverage area of the sensing field. Each sensor node is provided with transmit power control and omni-directional antenna and therefore can vary the area of its coverage [3]. It has been established in [4] that communication requires significant amount of energy as compared to computations. In this paper, we consider a wireless sensor network where the base station is fixed and located far off from the sensed area. Furthermore all the nodes are static, homogeneous, energy constrained and capable of communicating to the BS. The remainder of the paper is organized as Section 2 describes the related work and the radio propagation path loss, Section 3 deals with Proposed Method, Section 4 contributes about simulation results and finally conclusion and future works in Section 5.

2 Related Work

Hierarchical or cluster-based routing, originally proposed in wire line networks, are well-known techniques with special advantages related to scalability and efficient communication. This means that creation of clusters and assigning special tasks to cluster heads can greatly contribute to overall system scalability, lifetime, and energy efficiency. Heinzelman *et al.* in [5] developed a cluster-based routing scheme called Low-Energy Adaptive by connecting the last node on the chain to its closest unvisited neighbor. PEGASIS [2] greatly reduces the total communication distance and achieves a very good energy and lifetime performance for different network sizes and topologies. CDMA capable and non-CDMA-capable sensor nodes, the chain-based BINARY [4] and 3-Level Hierarchy schemes were proposed respectively in [4] to achieve better “energy \times delay” performance than PEGASIS. In [6], a cluster-based Load Balance and Energy Efficient Routing Algorithm (LBEERA) are presented. LBEERA divides the whole network into several equal clusters and every cluster works as in an optimized PEGASIS routing. A tree-structure routing scheme called Shortest HOP Routing Tree (SHORT) [3] offers a great improvement in terms of “Energy \times Delay” [4] with a good performance for network lifetime. An amalgamation of clustering and shortest hop pairing of the sensor nodes is called hybrid data gathering scheme HDS [7]. We have included a realistic power model [8] for a more realistic and efficient power consumptions. But the first order radio model, in power calculation, is also used for the simulation of the algorithm.

3 Proposed Method

The proposed method is segregated into three parts. We begin with describing the adopted node distribution protocol, and then the realistic power consumption model that is considered for energy calculations is elaborated and at the last Optimized Reduced Energy Consumption (OREC) algorithm for routing is explained.

3.1 The Node Deployment Protocol

For an efficient and economic approach and for ensuring the connectivity between each node for data gathering one must optimize the deployment of sensors. This is the part of deployment protocol. For achieving this we divide the field in n squares of edge length $'a/\sqrt{n}'$ for the deployment of $'n'$ sensor motes in a square field of edge length $'a'$. This is shown in Figure 1. The nodes are deployed within each such sub squares on a randomly occupying any position in that. For explanation we deployed two motes in one sub square and it can be seen that the sensing region of those nodes are overlapping at the lowest possible power level [8]. Thus there is no need to place two sensors within such close proximity or in more generalized way in same such square block. But if it is so it would be more power saving to switch one of the sensors off while the other does not get exhausted in terms of power. Now it can be also seen in Figure 1 that by following this protocol each sensor has eight sensors surrounding its sensing region. Now two particular sensors communicate at the lowest power level settings and hence the message transmit cost will also be low and hence enhances the lifetime of the network.

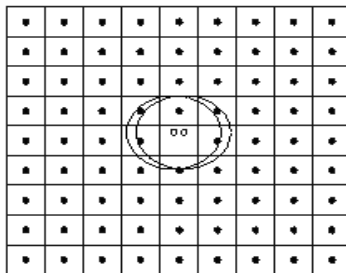


Fig. 1. Representation of sensor deployment protocol to be adopted for enhancement of lifetime in wireless sensor network

3.2 The Realistic Power Model

The realistic power model as described in [8] gives a realistic power consumption scenario. According to this model a sensor node works in 32 identifiable power levels with different power output for each level. This power level gets adjusted according to the distance of the node to which it wants to communicate. The power levels of a CC2420 [9] (we have used the data sheet of this chip for our calculations) transceiver is shown in table 1.

Out of this 32 power levels [8] has justified that only power levels 3, 7, 11, 19, 23, 27 and 31 are used. The distance range in which a particular power level operates and the energy consumed for sending data packets in that power level is shown in table 2. We have used the data of table 2 for calculating the energy consumptions and this gives a realistic power consumption model which can be used for simulation within involving any hardware.

Table 1. The table shows the various Power output for the discrete power levels available

Power Level	Power Output	Power Level	Power Output	Power Level	Power Output
0	-37.90	11	-10.00	22	-3.520
1	-33.00	12	-9.120	23	-3.000
2	-28.70	13	-8.330	24	-2.470
3	-25.00	14	-7.630	25	-1.950
4	-21.80	15	-7.000	26	-1.450
5	-19.20	16	-6.440	27	-1.000
6	-16.90	17	-5.940	28	-0.610
7	-15.00	18	-5.470	29	-0.310
8	-13.40	19	-5.000	30	-0.091
9	-12.10	20	-4.520	31	0
10	-11.00	21	-4.030	-	-

Table 2. The table shows the various Power output for the discrete power levels available

Power level (k)	P _{out} [dBm]	Distance (in meters)	I _x (mA)	P _{TX} (mW)	E _{tx/bit} [μJ]
3	-25.00	d<8.7m	17.04	15.15	0.0606
7	-15.00	8.7m<d<21.86	15.78	17.47	0.0699
11	-10.00	21.86<d<34.61	14.63	19.62	0.0785
19	-5.00	34.61<d<54.97	12.27	22.08	0.0883
23	-3.00	54.97<d<62.22	10.91	26.33	0.1050
27	-1.00	62.22<d<71.67	9.71	28.40	0.1136
31	0	74.47<d<81	8.42	30.67	0.1227

3.3 Optimized Reduced Energy Consumption (OREC) Algorithm

For increasing the energy efficiency of any algorithm which employs cluster or chain formation it is necessary that equal number of nodes is included in each cluster. For this purpose we have divided the field in vertical sections with each section having equal number of sensor nodes. This may lead to unequal area distribution of each vertical section but ensures that clusters of equal size are formed. The figure 2 and 3 shows random distribution of 40 sensor nodes in a square field of length 40 meter. The number of clusters formed is taken to be 5. So that indicates ideally each cluster would include 8 nodes. But in figure 2 where the field was divided into 5 equal areas the cluster size is very uneven with a minimum of 3 nodes to maximum of 18 nodes. This reduces energy efficiency to a great extent. Though this is eradicated to a great

extend when the node deployment protocol described above is used but for any other distribution where node density is not uniform throughout the network this problem entails. So we have divided the field as shown in figure 3 as given in the following algorithm.

Input: n = number of nodes, x = array of length n containing x co-ordinate of each node (data available to the Base Station), c = number of clusters to be formed
Output: z = two dimensional array of length c containing the node ids of each node that is included in a particular cluster

1. sort the array x and store it in an array say b and set k and $p=1$
2. For i starting from 1 to n
 - 2.1. For j starting from 1 to n
 - 2.1.1. If $b(i)$ is equal to $x(j)$
 - 2.1.1.1. $z(k, p)=j$;
 - 2.1.1.2. increment p by 1
 - 2.1.1.3. If p is equal to n/c
 - 2.1.1.3.1. set p equal to 1 and increment k by 1
 - 2.1.1.4. end If statement of 2.2.3
 - 2.1.2. end If statement of 2.2
 - 2.2. End For loop of 2.1
3. End For loop of 2
4. End program

By employing the above algorithm we can effectively form clusters of equal number of nodes as shown in figure 3.

The number of clusters to be formed is also optimized. In figure 4, 5 and 6 we have implemented OREC algorithm by varying the number of partitions or clusters formed for 40 nodes with initial energy of 250 mJ in a square field of length 40 m with base station at a co-ordinate of (20, 100). We see that for each case of number of round

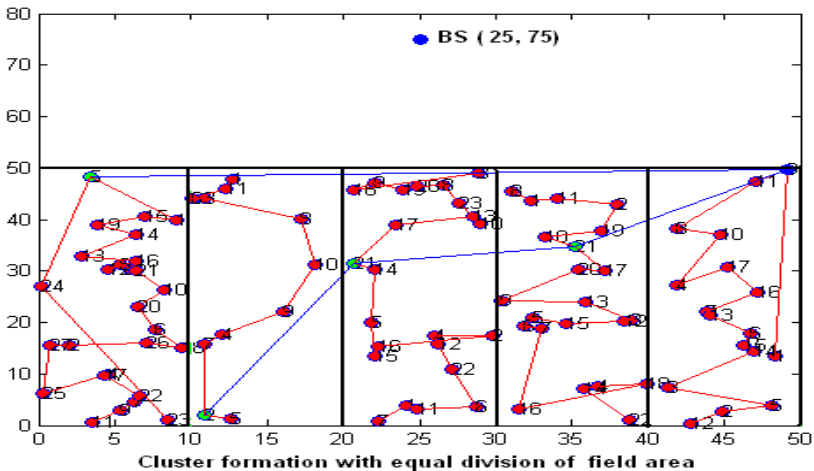


Fig. 2. The Clusters formed with equal division of the field area

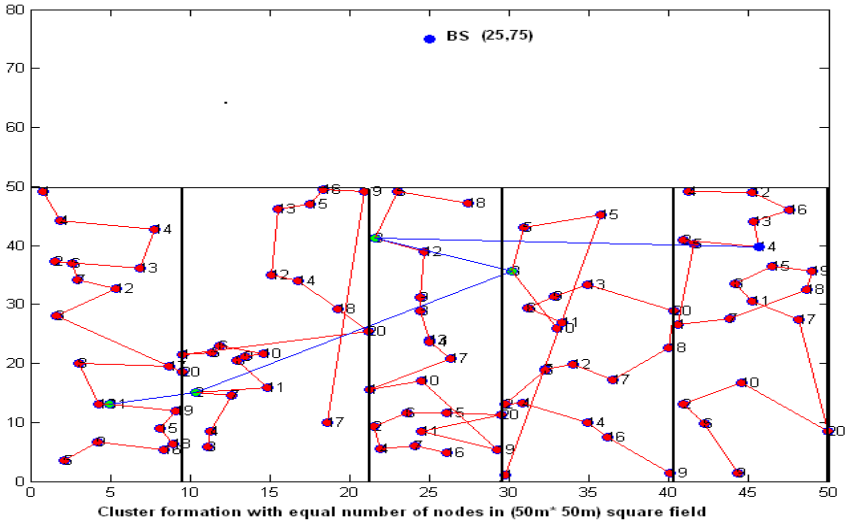


Fig. 3. Clusters formed with equal division of number of nodes

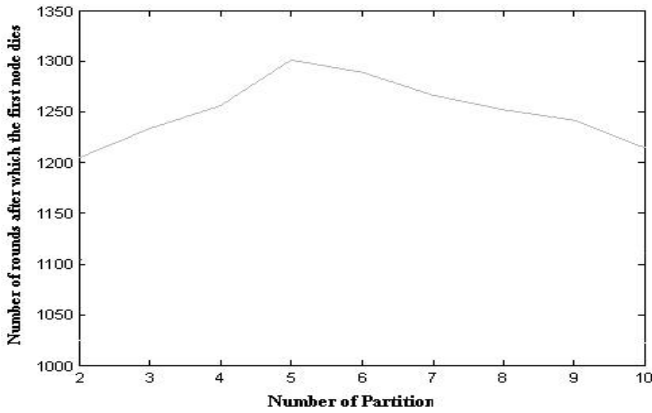


Fig. 4. The plot of the number of rounds after which the first node dies with varying number of partitions or clusters

when first node dies, 10% nodes die and 50% of the nodes die is greatest for number of partitions when it is equal to 5. So it is justified that by setting the number of clusters to be equal to 5 we get the greatest energy efficiency and hence it increases the lifetime of the network.

After setting the number of partitions to be 5 and implementing equal node distribution in the cluster or chain we discuss the cluster or chain formation, the leader selection and super leader selection of OREC. The chain is formed by highly optimizing the PEGASIS algorithm. For any cluster the chain starts with the farthest node from the Base Station (BS). It then includes the node which is closest to it. The

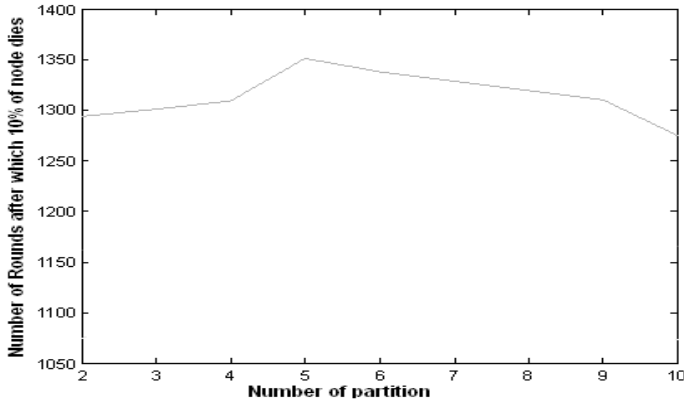


Fig. 5. The plot of the number of rounds after which the 10% of nodes die with varying number of partitions or clusters

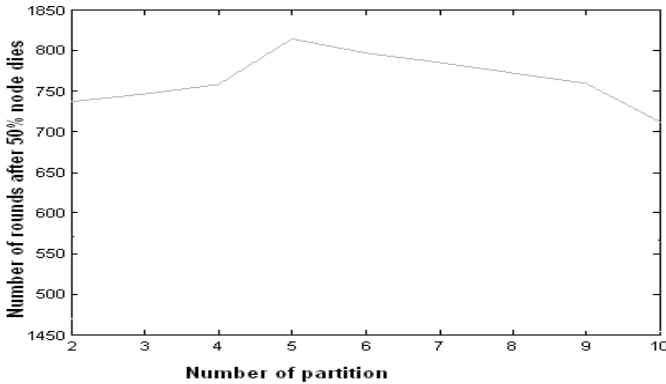


Fig. 6. The plot of the number of rounds after which the 50% of nodes die with varying number of partitions or clusters

included node then again performs this algorithm but it also includes the node which is second closest to it. This concept reduces the exclusion of certain nodes, which may have been connected to a node with much lower distance, which is in turn connected with a node at much greater distance than its closest node. The node in any cluster for which its remaining energy per square of its distance from the base station is maximum is selected as the leader of that cluster. Finally a super leader is selected which has the maximum value of remaining energy per square of its distance from the base station among the leaders. In any round the leaders gather the data from its own cluster and sends it to the super leader where as the super leader performs the same function but instead sends the total collected data to the BS. The chain formed by the leaders follows the same algorithm as in the cluster or chain formation. The leader selection is done after every 10 rounds for reducing the delay in the BS in calculations regarding the leader and super leader selections. The OREC algorithm can be described in the algorithmic form as given below.

4 Simulation Results

We have made simulation of all the algorithms and noted the first node death and half of the node death condition in the network employing both the realistic power model and the first order radio model. We have implemented the results in NS 2.33 in Linux (FEDORA 12 version). The number of nodes was initially taken to be 50 and spread across a square field of side 50 meter. The initial energy of each node was set to be 500 mJ and the packet length of each communication was set to 2000 bits. The location of the base station was set at (25,125) for all the cases, the co-ordinate were calculated by treating one of the corner of the square field to be at origin with two of its sides forming the perpendicular axes. In table 3 we have included the above mentioned results for both the radio models. From there we see that OREC outplays HDS by 18.32% and 3.34% in half of node dies criteria incorporating first order radio model and discrete radio model respectively. The figure 7, 8, 9 and 10 represents the first node dies condition by first order radio model, first node dies condition by realistic radio model, half of the nodes dies condition by first order radio model and half of node dies condition by first order radio model respectively for all condition remaining same except the length of the field which has been altered for varying the node density of the network against which all the above mentioned algorithms were plotted for all the conditions as stated. From the figure we see that OREC always exhibit a greater lifetime of the network for all condition. This justifies the suitability of it in terms of network lifetime and load balancing. So it qualifies for better routing in a network of wireless sensor.

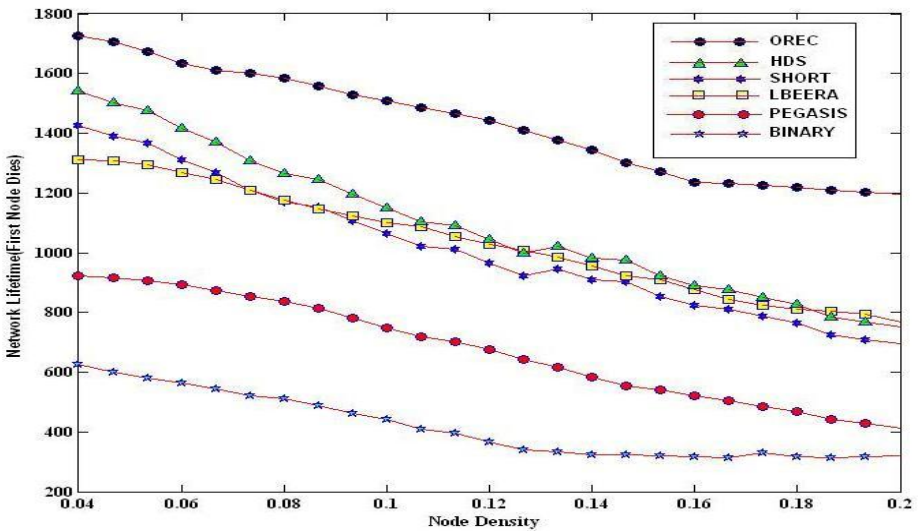


Fig. 7. Network lifetime (First Node Dies) versus Node Density for First order radio Model

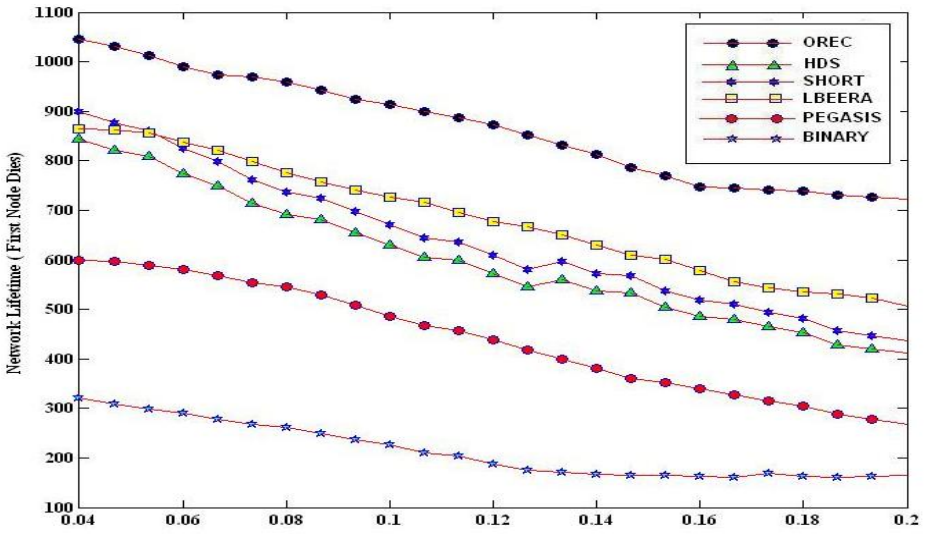


Fig. 8. Network lifetime (First Node Dies) verses Node Density for realistic radio model

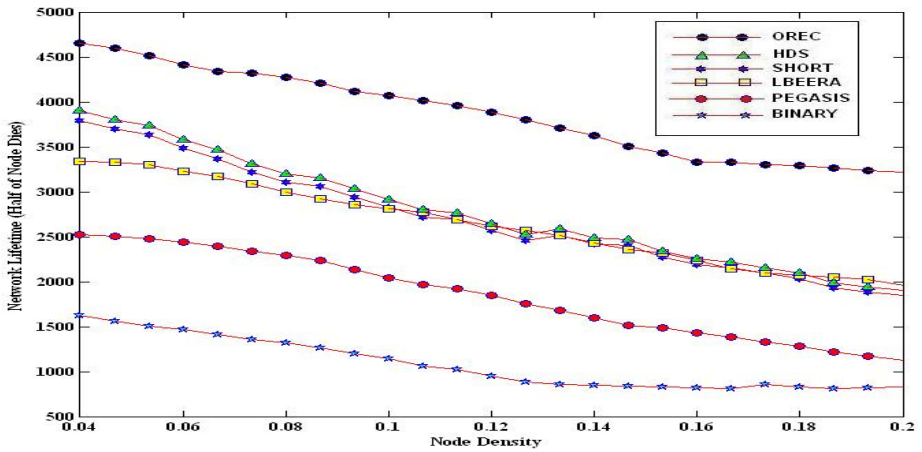


Fig. 9. Network lifetime (Half of node dies) verses Node Density for First order radio Model

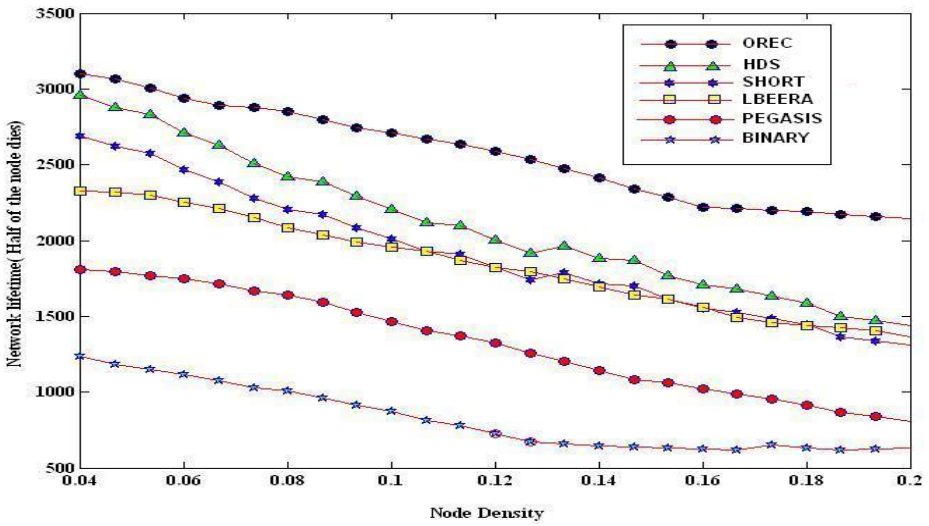


Fig. 10. Network lifetime (Half of node dies) verses Node Density for realistic radio Model

Table 3. The First node dies and half of node dies result for all the algorithms incorporating both the radio models with the specification as detailed in Section 5

Radio Model		Data Gathering Scheme Network Lifetime (No of Rounds) for (50* 50) square field with 100sensors					
		OREC	HDS	SHORT	LBEERA	PEGASIS	BINARY
First Order Radio Model	FND	1730	1550	1410	1300	940	620
	HND	4650	3930	3850	3350	2500	1600
Realistic Radio model	FND	1050	845	900	855	600	320
	HND	3090	2990	2700	2350	1800	1245

5 Conclusion and Future Work

We can interpret from the results that our method is much efficient in terms of lifetime of the network. The average lifetime of the sensor networks implemented using our method is much more than the other existing algorithm. The effectiveness is already presented in terms of simulation results. The result shows that the average time after which the first node dies is highest for our method. So obviously the overall lifetime of the network is highest among all the algorithms. So we can certainly propose our method as one of the best alternatives in the field of routing in sensor networks. Our further researches will be focused on improving this technique further. Also we will try to incorporate our own routing technique in spite of using the existent ones.

References

1. Clare, Pottie, Agre: Self-Organizing Distributed Sensor Networks. In: SPIE Conference on Unattended Ground Sensor Technologies and Applications, pp. 229–237 (1999)
2. Lindsey, S., Raghavendra, C.S.: PEGASIS: Power Efficient Gathering in Sensor Information Systems. In: Proceedings of IEEE ICC 2001, pp. 1125–1130 (2001)
3. Yang, Y., Wu, H.H., Chen, H.H.: SHORT: Shortest Hop Routing Tree for Wireless Sensor Networks. In: IEEE ICC 2006 Proceedings (2006)
4. Lindsey, S., Raghavendra, C.S., Sivalingam, K.: Data Gathering in Sensor Networks using energy*delay metric. In: Proceedings of the 15th International Parallel and Distributed Processing Symposium, pp. 188–200 (2001)
5. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: Energy- Efficient Communication Protocol for Wireless Microsensor Networks. In: IEEE Proceedings of the Hawaii International Conference on System Sciences (2000)
6. Yu1, Y., Wei, G.: Energy Aware Routing Algorithm Based on Layered Chain in Wireless Sensor Network, 1-4244-1312-5/07/\$25.00 © 2007. IEEE, Los Alamitos (2007)
7. Chakraborty, A., Mitra, S.K., Naskar, M.K.: An Efficient Hybrid Data-Gathering Scheme in Wireless Sensor Networks. In: Janowski, T., Mohanty, H. (eds.) ICDCIT 2010. LNCS, vol. 5966, pp. 98–103. Springer, Heidelberg (2010)
8. Mitra, S.K., Banerjee, J., Chakraborty, A., Naskar, M.K.: Data Gathering in Wireless Sensor Network using Realistic Power Control. In: Proceedings ICCCS 2011 (2011) (published by ACM)
9. C.T.Inc.,
http://www.xbow.com/Products/Product_pdf_files/Wireless_pdf/MICAZ_Datasheet.pdf
10. Mallinson, M., Drane, P., Hussain, S.: Discrete radio power level consumption model in wireless sensor networks. In: Second International Workshop on Information Fusion and Dissemination in Wireless Sensor Networks, Sensor Fusion (2007)

Improving Energy Efficiency of Underwater Acoustic Sensor Networks Using Transmission Power Control: A Cross-Layer Approach

Sumi A. Samad¹, S.K. Shenoy², and G. Santhosh Kumar¹

¹ Cochin University of Science and Technology, Kochi

² Naval Physical and Oceanographic Laboratory, DRDO, Kochi

sumi.a.samad@gmail.com

Abstract. Underwater Acoustic Sensor Networks (UW-ASNs) have found a wide range of applications from ocean monitoring to military surveillance. The underwater environment is energy constrained and hence it is very important to improve the life expectancy of sensor nodes. In this paper we propose a new MAC protocol (RMAC-PC) which uses transmission power control to enhance the energy efficiency. The protocol is developed as an extension to the RMAC protocol which schedules the transmissions of control and data packets depending on the latency calculations. Here, we utilize a cross-layer interaction between the MAC and physical layers to compute the optimum transmit power based on inter-nodal distance.

Keywords: Energy efficiency, Underwater sensor networks, MAC protocols, Cross-layer design, Acoustic communication.

1 Introduction

Underwater Acoustic Sensor Network (UW-ASN) refers to a set of independent sensor nodes (static and/or mobile) deployed over a given ocean area. The nodes sample some parameters of the environment and relay the data wirelessly with each other or to a central station for collaborative monitoring tasks. The radio waves used in terrestrial sensor networks have very high attenuation in conductive sea water. Therefore, the typical physical layer technology in underwater networks is acoustic communication [1][2].

The sensor nodes are powered by batteries. Since solar energy cannot be exploited underwater, it is difficult to recharge batteries in the nodes. Alternate methods such as generating energy from ocean currents are yet to be exploited. Since recharging or replacing the batteries is extremely difficult, it is very important to save energy expended in every function executed by the nodes of UW-ASNs. Here we present an extension to the existing RMAC protocol (RMAC-PC), for improving the energy efficiency of UW-ASNs by making the data transmission between the nodes more efficient.

The rest of the paper is organized as follows. In section 2 we discuss the energy efficiency of the existing MAC protocols for UW-ASNs. Section 3 illustrates how

transmission power control improves the energy efficiency and explains the proposed protocol with transmission power control. Section 4 analyses the performance through simulations, by comparing RMAC with and without power control. Section 5 concludes the paper and summarizes the results.

2 Related Work

Collisions of data transmissions are the major source of energy waste. It causes the transmitted packet to be corrupted and to be discarded. Further, retransmissions increase energy consumption. The contention free protocols such as *TDMA*, *FDMA* and *CDMA* avoid collisions. But these protocols are not suitable for UW-ASNs for various reasons as brought out in [1]. The Contention based protocols can be random access methods, carrier sense access methods or collision avoidance with handshaking access methods. The random access methods like *ALOHA*, which do not try to prevent packet collision are not suitable for UW-ASNs because the number of retransmissions increases due to collisions and thus increases the energy consumption of sensor nodes. Carrier Sense Multiple Access (*CSMA*) [3] and its variations have been widely used in terrestrial networks to prevent collisions between multiple nodes transmitting at the same time. Since the propagation delay in underwater is high, *CSMA* protocols are not very efficient and cannot be used in UW-ASNs.

Slotted FAMA [4] protocol for UW-ASN does not take care of the energy waste due to idle listening and overhearing. It also suffers from control packet overhead. This protocol also requires clock synchronization between nodes which is difficult to achieve in UW-ASNs. In *Su-MAC* (underwater S-MAC) and *Tu-MAC* (underwater T-MAC) [5] which use sleep/wakeup cycles for energy efficiency, sender will not send data immediately up on receiving CTS from the intended receiver, it waits until the CTS propagates through the whole transmission range of the receiver. The resultant high propagation delay leads to energy wastage in the overhearing and idle state. *UWAN-MAC* [6] avoids energy wastage due to handshaking over heads and idle listening by making use of sleep/wakeup schedules and does not require globally synchronized schedules as in *Su-MAC* and *Tu-MAC*. In [7] Wu et al. describe a power control mechanism to save energy as an enhancement over *UWAN-MAC*. *RMAC* [8] is a reservation based protocol which utilizes sleep/wake up schedules for energy efficiency. It schedules the transmissions of control and data packets to avoid data packet collision.

3 Transmission Power Control to Enhance Energy Efficiency

The MAC protocols employing sleep/wakeup schedules are more energy efficient than other protocols in the UW-ASN context. To further improve the energy efficiency of these protocols, transmission power control techniques can be utilized, particularly as the power consumed for transmission is much higher than reception in UW-ASNs. We propose a new MAC protocol with transmission power control technique on top of the basic RMAC protocol using MAC-PHY cross-layer interaction.

3.1 Overview of RMAC

RMAC [8] operates in three phases. (i) Latency detection (ii) Period announcement (iii) Periodic operation. In latency detection phase each node calculates the latency to its neighbors. In the second phase each node randomly selects its own listen/sleep schedules and broadcasts. The first two phases are for synchronization of the nodes. No centralized scheduling and synchronization are required in RMAC. The data packets are transmitted in the third phase.

In RMAC, nodes communicate through the exchange of REV/ACK-REV/DATA/ACK-DATA messages. When a node has data to send, it first sends a REV (reservation message) to reserve a time slot at the receiver. The receiver notifies all its neighbors by sending ACK-REV about the reserved time slot by ACK-REVs. The nodes receiving ACK-REV other than the sender keep silent in their corresponding time slots.

The first part of the listen window is reserved exclusively for ACK-REV packets. This reserved part is called R-Window, which is the maximum possible duration of a control packet. If a node receives an ACK-REV in its R-Window, then this node knows the duration of the subsequent data transmission and keeps silent during that period. If there is an ACK-REV collision, it will back off. All nodes have to schedule the transmission of control and data packets. Only ACK-REV packet can propagate to any node in its R-Window, all other control packets such as REV and ACK-DATA are scheduled to arrive at the target in its listen window and data packets are scheduled to arrive at the receiver in the reserved time slot.

3.2 RMAC with Power Control (RMAC-PC)

In RMAC, each node has knowledge of latencies to all its neighbors after the latency detection phase. The basic idea is to utilize this information and use the optimum power for transmission depending on inter-nodal distance. The inter-nodal distance can be directly calculated from latency at the MAC layer and it is passed to the PHY layer. This interaction between the MAC and PHY layer makes it possible to optimize the transmission power based on inter-nodal distance rather than a uniform maximum power for all transmissions.

Here, the traditional layered structure is preserved; while each layer is informed about the conditions of other layers for possible optimizations at the adjacent layer based on knowledge of this information. This direct communication between layers is one of the cross layer design proposals and is reported in the survey by Srivastava and Motani [11]. In this design we have used the protocol header HDR_RMAC to allow flow of information between layers. This proposal is appealing where just a few cross-layer information exchanges are to be implemented in systems that were originally designed in conformance with layered architectures. In that case, one can conceivably “punch” a few holes in the protocol stack while still keeping it tractable [11].

The optimum power based on the inter-nodal distance from the MAC layer is computed at the PHY layer. To compute the optimum power we have used the following passive sonar equation [9].

$$DT \geq SL - TL - NL + DI \quad (1)$$

Where DT is the Detection Threshold, SL is the source level, NL is the ambient noise level and DI is the directivity index.

All the quantities in Equation (1) are in $dB \text{ re } \mu Pa$, where the reference value of $1 \mu Pa$ amounts to $0.67 \times 10^{-22} \text{ Watts/cm}^2$.

$$\text{The source level, } SL = 10 \log \frac{I_1}{I_r} \quad (2)$$

I_1 = Intensity of source at standard range

I_r = Reference Intensity

If the omni-directional power output is P watts, then the source intensity at 1 meter is $P/12.6 \text{ W/m}^2$ and SL becomes

$$\begin{aligned} SL &= 10 \log \left(\frac{P/12.6}{0.67 \times 10^{-18}} \right) \\ &= 10 \log P + 170.8 \text{ dB} \end{aligned} \quad (3)$$

The Spreading loss is given by $PL = k \cdot 10 \log l$.

Where, k is the Spreading factor and l is the range. $k=1$ for cylindrical, and $k=2$ for spherical spreading.

TL is the transmission loss which is the sum of spreading loss and absorption loss.

$$10 \log A(l, f) = k \cdot 10 \log l + l \cdot 10 \log \alpha(f) \quad (4)$$

$\alpha(f)$ is the attenuation coefficient in dB/km for a given frequency f and is given by Thorp's approximation:

$$10 \log \alpha(f) = \begin{cases} 0.11 \frac{f^2}{1+f^2} + 44 \frac{f^2}{4100+f} + 2.75 \cdot 10^{-4} f^2 + 0.003, & f \geq 0.4 \\ 0.002 + 0.11 \frac{f}{1+f} + 0.011f, & f < 0.4 \end{cases} \quad (5)$$

NL is the ambient noise level which is the sum of noise due to turbulence, shipping, wind and thermal. The following empirical relations give the power spectral density of the noise components in dB re $1 \mu Pa$ per Hz, where f is in KHz.

$$10 \log N_t(f) = 17 - 30 \log f \quad (6)$$

$$10 \log N_s(f) = 40 + 20(s - 0.5) + 26 \log f - 60 \log(f + 0.03)$$

$$10 \log N_w(f) = 50 + 7.5w^{1/2} + 20 \log f - 40 \log(f + 0.4)$$

$$10 \log N_{th}(f) = -15 + 20 \log f,$$

where N_t is the noise due to turbulence, N_s is the noise due to shipping, N_w is the noise due to wind, and N_{th} represents the thermal noise. The variable s indicates shipping activity having values in the range (0,1) and w is the wind speed in m/s.

The overall noise power spectral density for a given frequency f is then:

$$N(f) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f). \quad (7)$$

For omni-directional antenna the Directivity Index (DI) is considered to be zero. For directional antennas DI refers to antenna gain. DT is the detection threshold which is the Minimum Discernable Signal at the receiver for a given probability of detection and probability of false alarm.

For the given transmission frequency and range, the PHY layer first compute the transmission loss TL using equation (4). If NL, DI and DT are known (these are fixed for a given transmitter/receiver configuration and environment), SL can be directly calculated from equation(1). To achieve this source level, the omni-directional power output P is computed by equation (3). Then, the transmission power required to produce this output power P is computed by taking the power amplifier efficiency and projector efficiency into account.

4 Simulation and Results

Although the transmit power requirement based on the distance is straight forward, it is important to study the effect of transmission power control on the specific MAC protocol. To measure the energy consumption in the new protocol, we have used Aqua-sim[12], an NS2[13] based simulator for underwater acoustic networks. It can simulate the attenuation and propagation of acoustic signals. It can also simulate packet collisions in underwater sensor networks. Aqua-Sim adopts the signal attenuation model given in equation (4) with the absorption loss calculated using the Thorp's approximation by equation (5). The propagation speed of sound adopted in Aqua-Sim is 1500m/s. The channel model used in Aqua-Sim takes care of fast-varying multipath propagation.

Table 1. Simulation Parameters

Spreading factor	2 (spherical spreading)
Ambient noise	90 dB
Detection threshold	10 dB
Transmission frequency	25 KHz
Number of transmit power	5
Max transmit power	40 W
Receive power	0.08 W
Idle power	0.01 W
Power amplifier efficiency	30%, Class AB Type
Projector efficiency	50%
Data rate	10 kbps
Simulation time	1000 s

The network scenario shown in Figure 1 is used in the simulation. The simulation scenario consists of six ordinary nodes and one sink node. The ordinary nodes transmit data to the sink node at 0.03 pkts/sec and the packet length is 480 bits.

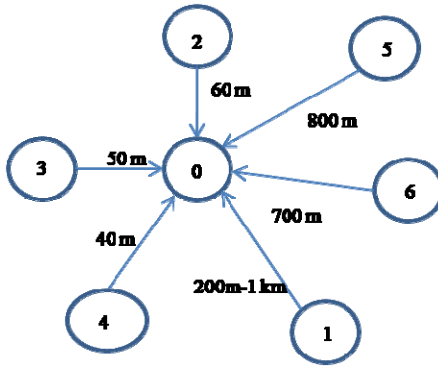


Fig. 1. Simulation Scenario

The distance between the node1 and sink node is varied between 200m to 1000m. Since the transmission power grows exponentially as the distance increases, using the optimum power based on distance can significantly reduce the power consumption.

The performance of the network is computed in terms of energy per bit consumption. The energy/bit consumption computed in dB is plotted (Figure 2) for RMAC and RMAC-PC with five power levels between 1W and 40W.

The modem’s capability to handle more number of discrete power levels can further improve the energy efficiency.

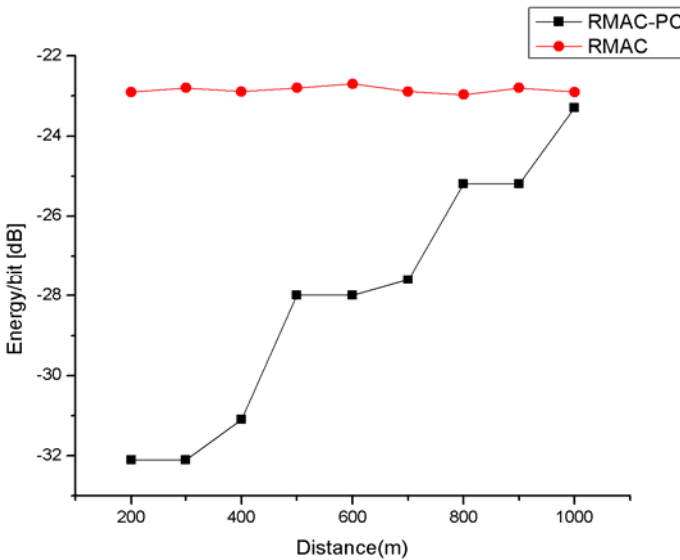


Fig. 2. Energy/bit [dB] vs. Distance (m)

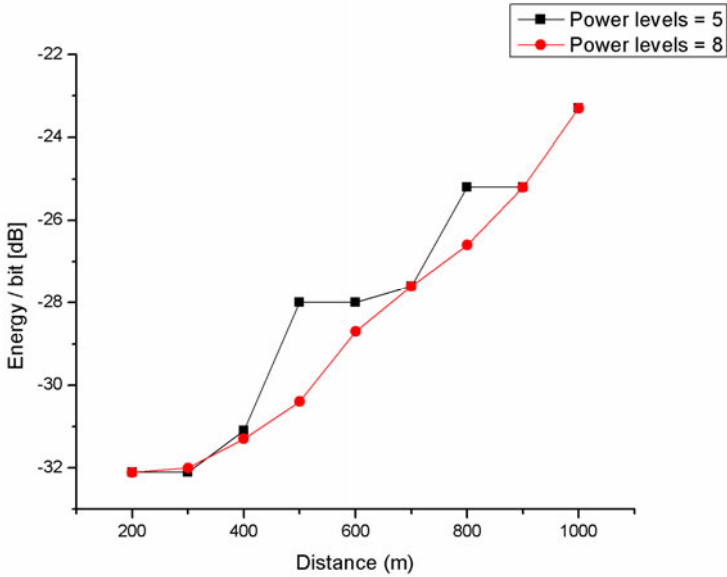


Fig. 3. The energy/bit consumption of RMAC-PC for different power levels

The energy/bit consumption of the network measured with different number of discrete power levels is depicted in Figure 3. It has been found that as the number of power levels increases the energy/bit consumption decreases, but the modem design will become complex.

The total energy consumption of the network over the simulation time (1000 Sec) with varying traffic rate (0.02 pkts/sec to 0.1 pkts/sec with 60 bytes/pkt is plotted in Figure 4.

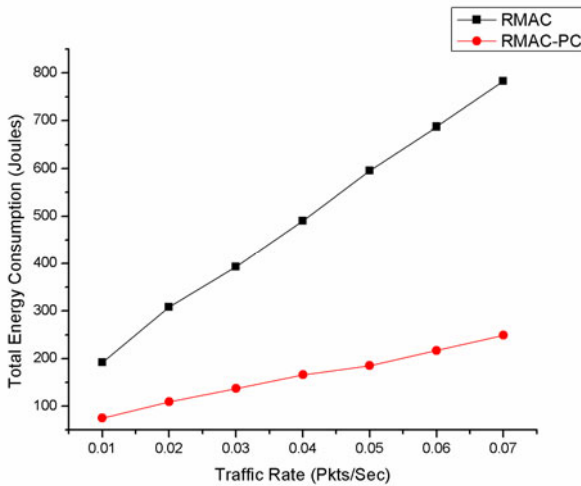


Fig. 4. Total Energy Consumption of the network vs. Traffic Rate

It is clear from the simulation results that RMAC-PC is much more energy efficient than RMAC and at the same time, RMAC and RMAC-PC have the same delivery ratio (number of packets received/ total number of packets sent) as given in (Figure 5).

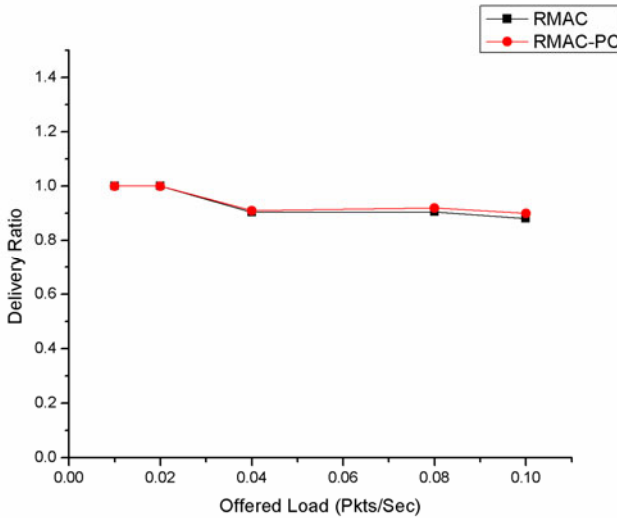


Fig. 5. Delivery Ratio vs. Offered Load

5 Conclusion

Energy efficiency is a critical design issue in underwater acoustic sensor networks. In this paper, we have proposed an extension of the RMAC protocol with transmission power control (RMAC-PC), to improve the energy efficiency of RMAC protocol using MAC-PHY cross-layer optimization. Simulation results show that RMAC-PC outperforms the existing protocol in energy efficiency, while preserving all the advantages of RMAC.

The RMAC-PC is well suited for network with static nodes. As part of our future work, we intend to integrate node mobility into the RMAC-PC protocol.

References

1. Pompilli, D., Rutgers, Akyildiz, I.F.: Overview of Networking Protocols for Underwater Wireless Communications. *IEEE Communications Magazine* (January 2009)
2. Akyildiz, F., Pompilli, D., Melodia, T.: Underwater Acoustic Sensor Networks: Research challenges. *Ad Hoc Net.* 3(3), 257–279 (2005)
3. Kleinrock, L., Tobagi, F.A.: Packet switching in radio channels: part I carrier sense multiple-access modes and their throughput-delay characteristics. *IEEE Trans. on Commun.* COM-23, 1400–1416 (1975)
4. Molins, M., Stojanovic, M.: Slotted FAMA: a MAC protocol for underwater acoustic networks. In: *Proceedings of the IEEE OCEANS 2006 Asia Conference* (2006)

5. Xie, P., Cui, J.-H.: Exploring Random Access and Handshaking Techniques in Large-Scale Underwater Wireless Acoustic Sensor Networks. In: MTS/IEEE Oceans 2006, Boston, MA (September 2006)
6. Park, M.K., Rodoplu, V.: UWAN-MAC: An Energy-Efficient MAC Protocol for Underwater Acoustic Wireless Sensor Networks. IEEE Journal Of Ocean Engineering 32(3) (July 2007)
7. Wu, S.-L., Tseng, Y.-C., Sheu, J.-P.: Intelligent Medium Access for Mobile Ad hoc networks with Busy Tones and Power Control. IEEE Journal on Selected Areas in Communications 18(9) (September 2000)
8. Xie, P., Cui, J.-H.: R-MAC: An Energy-Efficient MAC Protocol for Underwater Sensor Networks. In: Proceedings of International Conference on Wireless Algorithms, Systems, and Applications (WASA 2007), Chicago, Illinois, USA, August 1-3 (2007)
9. Waite, A.D.: Sonar for Practising Engineers, 3rd edn. (2002)
10. Urick, R.J.: Principles of Underwater Sound. McGraw-Hill, New York (1983)
11. Srivastava, V., Motani, M.: Cross-Layer Design: A Survey and the Road Ahead. IEEE Communications Magazine (December 2005)
12. Xie, P., Zhou, Z., Peng, Z., Yan, H., et al.: Aqua-Sim: An NS-2 Based Simulator for Underwater Sensor Networks. In: Underwater Sensor Networks Lab, University of Connecticut. OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges (2009)
13. The Network Simulator – NS2, <http://www.isi.edu/nsnam/ns>

A Collaborative, Secure and Energy Efficient Intrusion Detection Method for Homogeneous WSN

T. Mohamed Mubarak¹, Syed Abdul Sattar², Appa Rao³, and M. Sajitha⁴

¹ Department of CSE, Royal college of Engineering and Technology, Trichur, Kerala
mubarak.fma@gmail.com

² Royal Institute of Technology and Science, Professor, Dean,
Department of Computer Science
Hyderabad, Andra Pradesh State, India
syed49in@yahoo.com

³ GITAM Institute of Technology, GITAM University, HOD,
Department of Computer Science,
Visakhapatnam, Andra Pradesh State, India
apparao_999@yahoo.com

⁴ MES College of Engineering, Department of Computer Science,
Kuttippuram, Kerala State, India
saji_moidu@yahoo.com

Abstract. Intrusion detection in WSN has to be done with the trusted nodes in order to secure the process. This paper deals with selection of a set of trusted nodes and does the intrusion detection only with this set of nodes. If the detection is carried out by other types of nodes, we propose a multi-detection model in order to secure the process. Here we also analyze the probability of intrusion detection with these set of nodes. Our proposed selection algorithm also helps in energy efficiency as the information is routed only through these set of nodes. This method is more secure as the detection process is carried out by sensor nodes which are trusted. It is found that our analytical results validate the simulation

Keywords: Intrusion detection, Trust, node density, sensing range, Wireless Sensor Network (WSN).

1 Introduction

A wireless sensor network (WSN) consists of spatially distributed autonomous sensors to cooperatively monitor physical or environmental conditions. For e.g. the wireless sensor network is mainly used in military applications such as in borders for finding out the infiltrations. It is also used in industrial process monitoring and control, machine health monitoring, environment and habitat monitoring, healthcare applications, home automation and traffic control [1].

The Sensor nodes are tiny and limited in power. Sensor types vary according to the application of WSN. Whatever be the application, the resources such as power, memory and band width are limited. Moreover, most of the sensors nodes are throw

away in nature. Therefore it is vital to consider energy efficiency so as to maximize the life time of the WSN. Great efforts have been devoted to minimizing the energy consumption and extending the lifetime of the network. One common way is to put some sensor nodes in sleep mode to save energy and wake them up under some strategies.

Work towards maximizing the life time of WSN has been studied in many research works. Some of them lead to the need of heterogeneous WSN deployment. Lee et al. [2] analyze heterogeneous deployments both mathematically and through simulations in different deployment environments and network operation models. In [3], Hu et al. investigate some fundamental questions for hybrid deployment of sensor network, and propose a cost model and integer linear programming problem formulation for minimizing energy usage and maximizing lifetime in a hybrid sensor network.

Intrusion detection plays an important role in the area of computer security, in particular network security, so an attempt to apply the idea in WSNs makes a lot of sense. However, there are currently only a few studies in this area. Da Silva et al. [4] and Onat and Miri [5] propose similar IDS systems, where certain monitor nodes in the network are responsible for monitoring their neighbors, looking for intruders. They listen to messages in their radio range and store in a buffer specific message fields that might be useful to an IDS system running within a sensor node.

The rest of this paper is organized as follows. There are five sections. First section includes the related works. The papers which we referred to start this work are mentioned in this section. Following this contribution and network deployment sections are there, which specifies our idea to intrusion detection. Network deployment specifies the way we deployed the sensors. Intrusion detection in homogeneous wsn includes the algorithm and probability analysis. The simulation results are specified in simulation and verification section. Finally, the paper is concluded in the last section.

2 Related Works

There exist several tools for security in networks and IDSs are important tools. Many solutions have been proposed in traditional networks but it cannot be applied directly to WSN because the resources of sensor nodes are restricted. Ad-hoc and WSNs security has been studied in a number of proposals. Zhang and Lee [6] are among the first to study the problem of intrusion detection in wireless Ad-hoc networks. They proposed architecture for a distributed and cooperative intrusion detection system for Ad-hoc networks; their scheme was based on statistical anomaly detection techniques. But the scheme need much time, data and traffic to detect intrusion. In WSNs, the nodes can not afford the cost.

Detecting a moving intruder is a crucial application in wireless sensor networks, thus, attracting considerable research attention in the literature. Intrusion detection is defined as the first contact time when the intruder hits the sensing range of a sensor belonging to the large sensor cluster. Liu et al. [7] have explored the effects of sensor mobility on sensing coverage and detection capability in a mobile WSN.

Wang et al. [8] have provided a unifying approach in relating the intrusion detection probability with respect to the intrusion distance and the network parameters

(i.e., node density, sensing range and transmission range), under single-sensing detection and multiple-sensing detection models, in both homogeneous and heterogeneous WSNs. A straight line or linear motion intrusion path is assumed for an intruder. An intruder can attack the network following a curved path or even a random walk in order to improve its attacking probability. Yun Wang, Yoon Kah Leow, and Jun Yin [9] have provided an approach where the intruder takes a curved path. They propose a novel Sine-curve mobility model to explore the effects of different intrusion paths on the intrusion detection probability using single-sensing and K-sensing detections in a given wireless sensor network.

Yang Xiao [10] have provided the performance of the randomized scheduling algorithm via both analysis and simulation in terms of intrusion coverage intensity when an intrusion object occupies an area. They also study the impact of the size of intrusion object on the sensor network's configuration. Xi Peng et al [11] proposed a security management model for self organizing wireless sensor networks based on intrusion detection. It can prevent most of attacks. Then an analysis of each layer of networks in security model is discussed and the security management measures in the data link layer and network layer are described in detail especially.

Typically a wireless sensor network uses cryptography to secure itself against unauthorized external nodes gaining entry into the network. Brutch and Ko classify intrusion detection systems (IDS) into two categories: *host-based* and *network-based*. They further classify intrusion detection schemes into those that are signature based, anomaly based, and specification based [12]. In GTMS [21], the authors propose a group based trust management scheme for WSN and it employs clustering. It reduces the cost of trust calculation. In [13], the authors describe a trust management scheme for Sensor network security and it also provides a secure routing based on PLUS.

3 Contributions

In this paper, we have developed an algorithm which helps for more secure intrusion detection process in WSN. Also, we have analyzed the probability of intrusion detection while applying this algorithm.

In short, our work does two things

1. Develops and applies an algorithm for secure and energy efficient Intrusion detection.
2. Analyze the intrusion detection probability.

It finds out all the trusted nodes in the given network. And finds out a set of selected trusted nodes which covers the entire area using our algorithm [14]. And enable the intrusion detection modules for these set of nodes. Here the sink node will communicate only with the trusted nodes so that communication is set to minimum. An untrusted node on finding out an intruder passes the information to the nearest trusted node. The trusted node collects the information from all the sensors in that area and finds the probability of detection if more than threshold (Three nodes in the nearby area) passes the same information to the sink. This will help the detection process more reliable.

4 Network Deployment

The sensors are uniformly and independently deployed in a squared area. A sensor can only sense the intruder within its sensing coverage area that is a disk with radius r_s centered at the sensor. Consider figure 1, here the intruder is coming from the boundary and the distance moved by the intruder is D , the intruder is detected only when there is any sensor in the area moved by the intruder. The intruder can move through a straight line or through a curved path. The intruder follows the straight path only when it knows the exact destination. Otherwise it follows curved path. In this paper we are considering only straight path. Figure 1 show the case when the intruder enters from the boundary. Here the area moved by the intruder

$$S=2*D*r_s+ \pi r_s^2/2.$$

If the intruder is entering the WSN area from a random point, i.e, the intruder is dropped from the air, then the area moved by the intruder is also shown in figure 1. This area is given by

$$S=2*D*r_s+ \pi r_s^2.$$

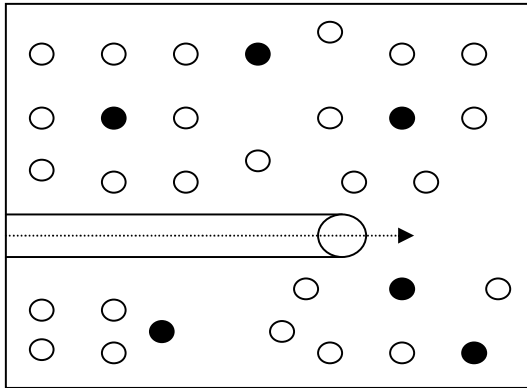


Fig. 1. The area moved by the intruder

5 Algorithm and Probability Analysis

This algorithm is executed at sink node and after finding out the nodes for intrusion detection, the sink passes this information to the nodes by one message. The algorithm first selects the trusted nodes from the WSN area. The trusted nodes can be selected by using different methods. Here in this paper we are taking group based trust management scheme for finding out the trusted nodes. After finding the trusted nodes, a small set of nodes are selected by using the algorithm given below from the set of trusted nodes. Here we are assuming that sink will communicate only with trusted nodes for more security. This technique can be enhanced by selecting another set of nodes after a predefined amount of time. This approach will select different set of nodes each time which will result in energy efficiency.

Algorithm

Step 1: Find out trusted nodes in WSN.
Step 2: Select a set of trusted nodes which covers the entire deployment area based on the algorithm shown below.
Step 3: The selected trusted node collects the information from all the neighboring sensors in that area.
Step 4: The selected trusted node finds the probability of detection. And if it is more than threshold, it passes the information to the sink
Step 5: sink node will communicate only with trusted nodes

Algorithm for selecting Nodes for intrusion detection:

Let U be the set of trusted nodes in WSN.
 Let N (i) is the set of neighbors of node i.
 repeat
 If N(i)≠∅
 Find i with min N(i)
 Put i in Stack;
 I={a/distance between a and i is<tr/2}
 if N(i)>1
 U=U-[i U I];
 else
 U=U-I;
 Until U is Null

5.1 Single Sensing Detection Model

As we explained before, the intruder is detected only when it enters the sensing range of sensor nodes. When the intruder enters the area through the boundary and the boundary is covered by the sensors, then the intruder will be detected as soon as it enters the WSN area. Otherwise it has to move a certain distance D before detected by any of the sensors.

Theorem 1

The probability $P(D)$ of detecting an intruder in a homogeneous WSN can be given by

a. $P(D = 0) = 1 - e^{-n_1}$

b. $p(D \leq \eta) = 1 - e^{-n_2}$

Where n_1 and n_2 are the number of sensors participating in intrusion detection area $\pi[r_s^2/2]$ and $2\eta r_s + (1/2)\pi r_s^2$ respectively.

Proof: According to Poisson distribution

$$p(m, A) = \frac{(A\lambda)^m e^{-A\lambda}}{m!}$$

Where m is the number of sensors and A is the area. Here λ is the probability of sensors which is doing the intrusion detection. In this paper only $n1$ sensors are doing intrusion detection at a time. So the $\lambda = n1/A$. If there is no sensors in the area A , then the probability will be equal to $P(0, A) = e^{-A\lambda}$. Based on complement of the probability, the probability that there is at least one sensor in the area A and the intruder is detected by any of these sensors $1 - P(0, A) = 1 - e^{-n1}$. Hence the theorem part (a) is proved. Similarly we can prove part b.

5.2 Multi Sensing Detection Model

Here multiple sensors have to detect an intruder at the same time. Three sensors are considered. The intruder is within the sensing range of three sensors.

Theorem 2

Let $P_m(D=0)$ be the probability that an intruder is detected in a WSN with node density and sensing range r_s in multisensing detection model. Then probability is given by

$$P_m(D=0) = 1 - \sum_{i=0}^{m-1} e^{-n1}, \text{ where } n1 \text{ is the number of selected sensors within the area } A = \pi r_s^2 / 2.$$

Proof: This theorem can be proved just like above theorems. Here the area is only one half circles with radius r_s . $P(i, A)$ gives the probability of detecting the intruder with i sensors. $\sum_{i=0}^{m-1} P(i, A)$ Gives the sum of the probabilities of detecting the intruder with less than m sensors. So the complement will give the multi sensing probability.

6 Simulation and Verification

In order to get the result we are varying the parameters such as sensing range, transmission range, number of sensors and so on. The sensors are uniformly distributed in a two dimensional space of 1000×1000 meters. The sensing range is varied from 0 to 50 meters and maximal allowable intrusion distance is 50 meters. The graph shows the detection probability. It is found that the detection probability remains same as in the case of analytical results thus proving the correctness of the analytical model. The fig 2 show Single-Sensing detection. It is evident that the single sensing detection probability is higher than that of multi sensing- detection probability. This is because the multi-sensing detection imposes a stricter requirement on detecting the intruder (e.g., at least 3 sensors are required).

Here the graph is obtained by changing the sensing range from 0 to 40. The each point in the graph is a result of 100 simulations. i.e. to get each point we need to execute our simulation and find out the probability from the result of this 100 executions. Here we can see that single sensing is possible at lower ranges also. But

for multi sensing it will take a little time to get the result. Because it needs more than one sensor (here, in this simulation 3 sensor information) information to detect the intruder.

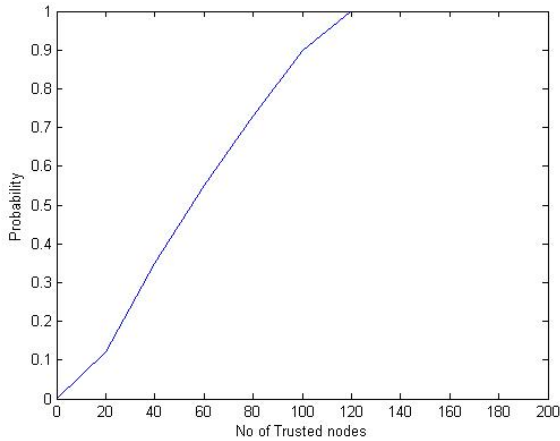


Fig. 2. Probability analysis

Fig. 3 demonstrates the average number of nodes selected by using this algorithm specified above. The density of trusted nodes is varied to check how many nodes are activating its IDS module. The sensing range and transmission range are set to 40. The numbers of trusted nodes are varied in each execution and find out how it will affect the selection process.

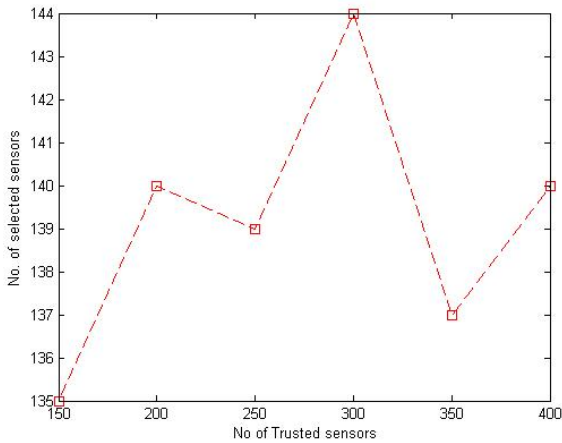


Fig. 3. Selection of nodes

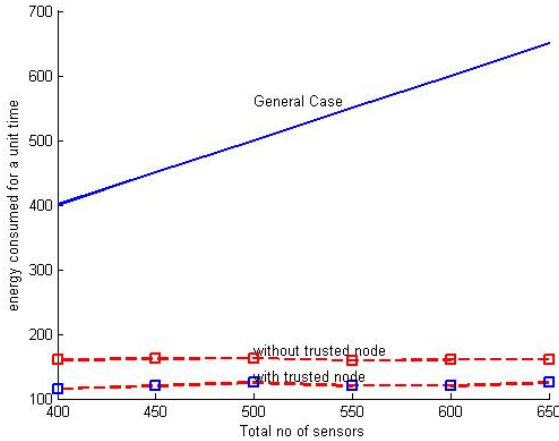


Fig. 4. Energy efficiency analysis

The energy used by this algorithm is analyzed in the figure 4 given below. Here we compared this approach with two cases. In the first case, was all sensors are participating in intrusion detection [15] and in second case, we are selecting the sensors without using trust node concept [14]. We assumed that the energy used by one node for a unit time is one unit. The graph clearly shows the energy efficiency.

7 Conclusions

The information that an intruder is detected with untrusted nodes is risky as it may cause unwanted traffic, energy lose and activation of preventive measures in the WSN. Therefore we propose a solution which allows the trusted nodes to do the intrusion detection and others, if detects the intruder, make sure that the information is genuine. This will help the WSN to save energy, avoid unwanted traffic, and will provide more security to the network. The correctness of the analytical model is proved by simulation.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A Survey on Sensor Networks. *IEEE Communication Magazine* 40(8), 102–114 (2002)
2. Lee, J.J., Krishnamachari, B., Kuo, C.C.J.: Impact of Heterogeneous Deployment on Lifetime Sensing Coverage in Sensor Networks, *IEEE SECON* (2004)
3. Hu, W., Chou, C.T., Jha, S., Bulusu, N.: Deploying Long-Lived and Cost-effective Hybrid Sensor Networks. *Elsevier Ad-Hoc Networks* 4(6), 749–767 (2006)
4. da Silva, A.P., Martins, M., Rocha, B., Loureiro, A., Ruiz, L., Wong, H.C.: Decentralized intrusion detection in wireless sensor networks. In: *Proceedings of the 1st ACM International Workshop on Quality of Service & Security in Wireless and Mobile Networks*

5. Onat, I., Miri, A.: An intrusion detection system for wireless sensor networks. In: Proceeding of the IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, Montreal, Canada, vol. 3, pp. 253–259 (August 2005)
6. Zhang, Y., Lee, W.: Intrusion Detection in Wireless Ad-Hoc Networks. In: Proc. ACM MobiCom, pp. 275–283 (2000)
7. Liu, B., Brass, P., Dousse, O., Nain, P., Towsley, D.: Mobility improves coverage of sensor networks. In: Proceedings of the 6th ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc (2005)
8. Wang, Y., Wang, X., Xie, B., Wang, D., Agrawal, D.P.: Intrusion detection in homogeneous and heterogeneous wireless sensor networks. *IEEE Transactions on Mobile Computing* 7(6), 698–711 (2008)
9. Wang, Y., Leow, Y.K., Yin, J.: Is Straight-line Path Always the Best for Intrusion Detection in Wireless Sensor Networks. In: 15th International Conference on Parallel and Distributed Systems (2009)
10. Xiao, Y., Chen, H., Zhang, Y., Du, X., Sun, B., Wu, K.: Intrusion Objects with Shapes under Randomized Scheduling Algorithm. In: The 28th International Conference on Distributed Sensor Networks, in Computing Systems Workshops (2008)
11. Peng, X., Wu, Z., Xiao, D., Yu, Y.: Study on Security Management Architecture for Sensor Network based on Intrusion Detection. In: 2009 International Conference on Networks Security, Wireless Communications and Trusted Computing (2009)
12. Brutch, P., Ko, C.: Challenges in intrusion detection for wireless ad-hoc networks. In: 2003 Symposium on Applications and the Internet Workshops (SAINT 2003 Workshops), 2003 (Q2SWinet 2005), pp. 16–23. ACM Press, New York (2005)
13. Shaikh, J., Lee, H., Lee, S., Song, Y.-J.: Group Based Trust Management Scheme for Clustered WSN. *IEE Transaction on Parallel and Distributed Systems* 20, 1698–1712 (2009)
14. Mubarak, M., Sattar, S.A., Rao, A., Sajitha, M.: Energy efficient intrusion detection in three dimensional wireless sensor networks. In: 2010 International Conference on Computational Intelligence and Computing Research (December 2010)
15. Mubarak, M., Sattar, S.A., Rao, A., Sajitha, M.: Intrusion detection: A probability Model for 3D Heterogeneous WSN. *International Journal of Computer Applications* 6(12) (September 2010)

An Efficient and Hybrid Key Management Scheme for Three Tier Wireless Sensor Networks Using LU Matrix

Manivannan Doraipandian, Ezhilarasie Rajapackiyam,
P. Neelamegam, and Anuj Kumar Rai

School of Computing, SASTRA University, Thanjavur- India
dmv@cse.sastra.edu, crezhil@yahoo.com,
neelkeer@eie.sastra.edu, anujlucky007@gmail.com

Abstract. Security plays a vital role in Wireless Sensor Networks and attention on key part is essential to encrypt the information processing among the sensor nodes in the network. Due to sensor node constraint, Key Management plays an important role. Existing Key Management Schemes (KMS) for Cluster-based architecture either supports Group communication or Node-to-Node communication. The proposed hybrid KMS concentrates on both Group communication and Node-to-Node communication using LU matrix and to enhance the strength of the security between cluster head and base station the ElGamal Public key encryption techniques is used. The main feature of this proposed protocol is 100% Node-to-Node connectivity and perfect resilience is achieved when Sensor node/Cluster Head is compromised. The scheme and its detailed performance analysis are discussed in this paper.

Keywords: LU matrix, Wireless Sensor Network, Group communication, Node-to-Node connectivity.

1 Introduction

Sensor network is a collection of sensor nodes generally referred as motes and each node has different capability in terms of memory, processing, and transmission range. There are many applications based on wireless sensor network such as Health monitoring, Industrial Automation, Military application, Area Monitoring etc., where sensor network plays a very important role. Since sensor nodes are deployed in unfriendly area, security becomes a major issue in WSN. Before any exchange of data, encryption key must be known. Key Management plays a significant role in Security. Many Key Management schemes are available, but due to sensor node constraints they are not effective.

Various Keying mechanisms: Key Pre-distribution schemes, Self-enforced schemes and arbitrated schemes. But Key Pre-distribution keying mechanism is accepted widely because of motes constraints. For communication either Group keying for group communication or Pair-wise keying for Node-to-Node communication are used. Communication Patterns are selected based on the Architecture of Sensor Network and the types of applications.

In proposed three-tier architecture, in between base station and cluster head, public key cryptography is used and between cluster head and sensor node, Symmetric key cryptography is used. While designing Key Management protocol, the metrics to be evaluated are:

Key Connectivity: Each node communicates with every other node in the clustered region.

Resilience: When node gets compromised, how secure the remaining communication links, i.e., resistance against node capture

Scalability: Capability to support when large number of nodes or cluster is added to the network.

Efficiency: In terms of storage, communication and computation.

2 Related Works

KMS are mostly based on key pre-distribution schemes i.e., keys will be distributed to Sensor nodes before deployment. A Key Pre-distribution scheme is widely classified into two classes. 1. Random key Pre-distribution 2. Deterministic Approach.

Many Random Key Pre-distribution schemes are available. The E-G scheme [1] is the basic scheme, in which large key pool (2^{20}) is generated P. From the key pool P, randomly draw m keys and assign to Sensor nodes. Then each sensor node broadcasts the Key ID's of m key to find a common key and make use of that key as shared secret key. If any pair of nodes does not share any key, it will communicate through secure links. Q- Composite scheme [2] is the improvement of E-G scheme in which instead of one shared key between nodes, q (>1) keys are needed for Communication. Advantage of this scheme when compared to E-G scheme is, resilience will be more.

Matrix Based scheme is proposed by Blom [3], to establish a pairwise key between any pair of nodes. Blom's scheme have two matrices one public $(\lambda + 1) * n$ matrix G and another $(\lambda + 1) * (\lambda + 1)$ symmetric matrix D. These two matrices are used to calculate $n * (\lambda + 1)$ matrix, $A = (D.G)^T$. Matrix A should be secret and in each sensor node one row from matrix A will be stored. Since G matrix is based on Vandermonde matrix, no need to broadcast the entire column, just by broadcasting the column id / seed, the sensor node is capable to compute the matrix. Though it provides perfect connectivity, scalability and resilience will be an issue. If λ nodes get captured, the entire network will get captured. To improve the Blom's scheme, Du et al. [4] proposed a pairwise key pre-distribution schemes. In that instead of one key space, multiple key spaces are used.

To increase the resilience, Liu and Ling [5] proposed a Polynomial pool based key pre-distribution scheme, which incorporated the idea of Blundo's scheme [10] and Random Key Pre-distribution [1]. In all the above schemes, when numbers of compromising nodes reach to a certain threshold λ , the entire network will get capture.

Deterministic approaches use combinatorial designs i.e., grouping of certain elements based on certain properties. For Group communication EBS (Exclusion Basis System) scheme [11] has been used. SHELL [6] incorporated EBS. Though

number of Re-keying message is less, collision of nodes ($k < m$) will reveal all the administrative keys. Tree based approach is also used for group communication. If one node gets compromised, the cluster key gets revealed. So in the case of group communication, the cluster key should be changed periodically. For Key refreshing, i.e., updating a new key periodically, energy consumption will be more. So network lifetime will get reduced.

Achieving all metrics in a single key management protocol is difficult. For group communication since changes takes place periodically, number of Re-keying message and scalability can be evaluated. In Node-to-Node communication, metrics to be evaluated are Resilience and Key connectivity. To increase the network lifetime, node-to-node connectivity i.e., pair wise schemes can be used.

2.1 Our Contributions

In proposed scheme, LU matrix is introduced for Group communication and Node-to-Node communication whereas PKI is used for communication between Cluster head and Base Station. The important assumption in proposed scheme is that the base station will never get compromised.

In Proposed scheme:

1. Node-to-Node Communication within Cluster is achieved by using LU matrix. Any pair exchanges their row with each other to get secret key. ElGamal Public key encryption scheme is used for secure communication between Cluster head and Base Station.
2. Divide the cluster into sensor groups for group communication. Communication between group and cluster head can be established using LU matrix.
3. Group Key Manager (GKM) node is introduced to reduce the overhead of cluster head for refreshing the Group key.
4. Perfect Resilience and Full network Connectivity is achieved.
5. Performs well in terms of Storage and Scalability.

2.2 Preliminaries

2.2.1 LU Matrix

Lower Triangular matrix is formed from randomly generated key pool. Upper Triangular matrix can be constructed based on assumption that product of L and U will yield symmetric matrix. For example consider 3*3 matrixes

$$\begin{bmatrix} l_{11} & 0 & 0 \\ l_{21} & l_{22} & 0 \\ l_{31} & l_{32} & l_{33} \end{bmatrix} \cdot \begin{bmatrix} u_{11} & u_{12} & u_{13} \\ 0 & u_{22} & u_{23} \\ 0 & 0 & u_{33} \end{bmatrix} = \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}$$

From above the value of each K element can be calculated

$$\begin{aligned} l_{11} \cdot u_{11} &= k_{11} ; l_{11} \cdot u_{12} = k_{12} ; l_{11} \cdot u_{13} = k_{13} ; l_{21} \cdot u_{11} = k_{21} ; l_{31} \cdot u_{11} = k_{31} \\ l_{21} \cdot u_{12} + l_{22} \cdot u_{22} &= k_{22} \quad l_{21} \cdot u_{13} + l_{22} \cdot u_{23} = k_{23} \quad l_{31} \cdot u_{12} + l_{32} \cdot u_{22} = k_{32} \\ l_{31} \cdot u_{13} + l_{32} \cdot u_{23} + l_{33} \cdot u_{33} &= k_{33} \end{aligned}$$

Since K is symmetric, K_{ij} will be equal to K_{ji}

$$\begin{aligned} k_{12} &= k_{21} = l_{11}.u_{12} = l_{21}.u_{11} \Rightarrow u_{12} = (l_{21}.u_{11}) / l_{11} \\ k_{13} &= k_{31} = l_{11}.u_{13} = l_{31}.u_{11} \Rightarrow u_{13} = (l_{31}.u_{11}) / l_{11} \\ k_{23} &= k_{32} = l_{21}.u_{13} + l_{22}.u_{23} \\ &= l_{31}.u_{12} + l_{32}.u_{22} \\ \Rightarrow u_{23} &= (l_{31}.u_{12} + l_{32}.u_{22} - l_{21}.u_{13}) / l_{22} \end{aligned}$$

By using these equations, randomly set the values for u_{11} , u_{22} & u_{33} based on that calculate the values for remaining unknown variables of U matrix so that their product would be symmetric.

2.2.2 ElGamal Scheme

Between Cluster head and base station, ElGamal PKI scheme is used because communication between cluster head and base station will have to be more secure than sensor node communication. Communication between base station and cluster heads are infrequent; but the level of secrecy is more and Communications between cluster heads and nodes and among nodes is frequent but the level of secrecy is normal than the former. In regular PKI scheme there is a possibility of Man-in-the-Middle sAttack due to its broadcasting nature. In this proposed scheme ElGamal Public key is not broadcasted, so there is no possibility of Man-in-the-Middle attack.

2.3 Notations Used

m :	number of Cluster heads.	L_{nl,C_i} :	a row from nL_i matrix assigned to cluster Head C_i
n :	number of sensor node in each Cluster.	U_{gUC_i} :	a column from nU_i matrix assigned to cluster Head C_i
nodeId :	unique Id for each node in one Cluster.	L_{gl,C_i} :	a row from gL_i matrix assigned to Cluster Head C_i
groupId :	unique id for each group in Cluster.	U_{gUC_i} :	a column from gU_i matrix assigned to cluster Head C_i
N_{ij} :	j^{th} node in i^{th} Cluster.	$GKM_{i,j}$:	j^{th} key group manager of i^{th} Cluster
nL_i :	lower triangular matrix for generating Symmetric matrix for i^{th} Cluster.	$K_{i,jk}$:	key between $N_{i,j}$ and $N_{i,k}$ in Cluster i calculated by node k
gL_i :	lower triangular matrix for generating Symmetric Matrix.	$K_{i,GKM}$:	Key between GKM and node j
nU_i :	upper triangular matrix for generating Symmetric Matrix for i^{th} Cluster.	$K_{gU,j}$:	Key Between group j and cluster head C_i
gU_i :	upper triangular matrix for generating Symmetric Matrix	P_{gUC_i} :	Public key of cluster head C_i
C_i :	Cluster head of i^{th} Cluster.	P_{nb} :	Public key of base station.
$L_{gl,i,j}$:	a row in gL_i matrix in Cluster i for group j	P_{rci} :	Private key of Cluster head.
$U_{gU,i,j}$:	a Column in gU_i matrix for group j in cluster i	P_{rb} :	Private key of base station.
$L_{nl,i,j}$:	a row in nL_i matrix in Cluster i assign to sensor node j	P_{gKMe_j} :	Public key of j^{th} Group Key Manager of i^{th} Cluster.
$U_{nt,i,j}$:	Column in nU_i matrix in cluster i assign to sensor node j	P_{rGKMe_j} :	Private key of j^{th} Group Key Manager of i^{th} Cluster.

3 WSN Architecture

The proposed architecture comprises of Sensor nodes, Base Station, Cluster Head, and Group Key Manager. Since the architecture is a cluster based, base station communicates with all Cluster heads. Inside the Cluster there will be sensor groups consisting of some number of nodes. Sensor groups are formed for group communication, in which sending a single message encrypted with group key rather than sending individual message to each and every node and mainly used for commanding the nodes. Transfer of data from sensor nodes to Cluster head is through shared secret key. For refreshing the group key, Group Key Manager is used which updates the key. There will be more than one GKM in Cluster, like GKM_1 take care of Group 1, 2, 3 and GKM_2 takes care of Group 4, 5, 6 likewise.

Base Station: Base station is assumed to be highly secure with high processing speed and memory. The entire nL_i , nU_i , gL_i and gU_i matrix (for $i=1$ to m) of all the clusters, one public and one private key for communication between the Cluster Head and Base Station and public key of all the Cluster Heads are saved in the Base Station.

Cluster Head: It is assumed to have more resources than sensor node in terms of transmission range, memory, and processing speed. Each cluster head saves one public, and one private key for communication with Base Station. The public key of base station and public keys of Group Key managers of its cluster are saved in the cluster head which also saves one row from nL_i and one column from nU_i matrix for node to node connectivity, another one row of gL_i and one column of gU_i matrix for Group communication between cluster Head and group members. It is assumed that Cluster head is able to reach all the nodes within a cluster.

Group Key Manager: Group Key Manager is used for maintaining the group and refreshing the key, so that security in-group communication is maintained. GKM stores the information of which node belongs to which group and which nodes come under its maintenance. There will be more than one GKM in a cluster. GKM of cluster i stores one row of nL_i and one column of nU_i matrix so that it can unicast message to the node for refreshing key. One public, one private key and public key of cluster head are saved for communication with cluster Head. In the proposed three-tier architecture, within a cluster there will be sensor group as shown in the figure Fig 1.

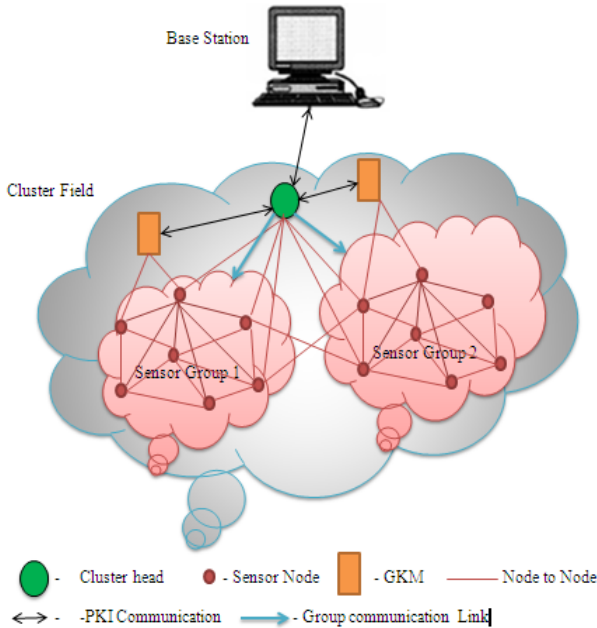


Fig. 1. WSN Architecture

Sensor Node: Each sensor node has limited memory, less processing speed, small transmission range that can be vulnerable also. Each sensor node in Cluster i save one row from nL_i and one column from nU_i matrix which is used for node to node communication, and one row of gL_i and one column of gU_i matrix which is for group communication. Each node in the group has same row of gL_i and column of gU_i matrix, so that they all can have the same key with the cluster Head.

4 Proposed Scheme

Phase I: Key Pre-distribution

Step 1: The base station first randomly generates a large pool of numbers (2^{20}) and randomly picks number to construct nL_i matrix and gL_i matrix for all m clusters.

Step 2: Calculate nU_i upper triangular matrix for each nL_i lower triangular matrix such that their product is symmetric matrix.

Step 3: A large number of public and private key pair is generated using Diffie-Hellman Protocol. Since this is pre-deployed scheme, all keys will be stored in the nodes before deployment, so possibility of Man-in-the-Middle attack is avoided

Step 4: For assigning sensor node in i^{th} cluster randomly select one row and column from nL_i and nU_i matrix and store it in sensor nodes.

Step 5: For assigning sensor node to group j in cluster i store U_{gU_i-j} and L_{gL_i-j} from gU_i and gL_i respectively which is common to all the nodes in the group j .

Step 6: Cluster head C_i , of i^{th} cluster is assigned a row and column from nL_i and nU_i respectively, for communicating with all the nodes within the cluster. For communicating with the Base station, a pair of private key P_{rci} , public key P_{uc_i} is selected randomly and saved in the cluster head. Public key P_{uB} of Base Station and public Key of all the GKM are stored in the cluster head

Step 7: In Group Key Manager GKM_{i-j} one row and one column of nL_i , nU_i is randomly selected, a pair of public key $P_{uGKM_{i-j}}$ and private $P_{rGKM_{i-j}}$ from the pool is randomly selected and both are saved. Public key of cluster head P_{uc_i} is also loaded. Group key manager is also loaded with the table, which contains the nodeId and its groupId to which it belongs.

Phase II: Pair wise Key Establishment

A. In-between sensor nodes

After key deployment each sensor node needs to establish pairwise key with its neighbors and cluster head. To establish a pair wise key between N_{i-j} and N_{i-k} of cluster i , following steps are required:

Step 1: Node $N_{i,j}$ sends its row $L_{nL_{i,j}}$ to node $N_{i,k}$

Step 2: Receiving $L_{nL_{i,j}}$ from $N_{i,j}$, $N_{i,k}$ multiplies its column $U_{nU_{i,k}}$ and gets $K_{i,jk}$ and replies with $(L_{nL_{i,k}}, F(K_{i,jk}))$ and sends it to node $N_{i,j}$, where $F(K_{i,jk})$ is the hash value of key .

Step 3: Node $N_{i,j}$ receives $L_{nL_{i,k}}$ from $N_{i,k}$, multiply it with its column $U_{nU_{i,j}}$ and calculate $K_{i,kj}$ and computes the hash value of it and checks whether $F(K_{i,jk})=F(K_{i,kj})$

Step 4: If hash value of both are same then node $N_{i,j}$ sends (ok, $F(K_{i,kj})$) to $N_{i,k}$ otherwise an error message (err, $N_{i,k}$) is broadcasted. The error message will authenticate by the receiver using TESLA [12].

If the number of error message cross a threshold for $N_{i,k}$ then messages from $N_{i,k}$ is discarded by other member of cluster i

B In-Between Cluster Head and Base Station

Encrypt and decrypt the messages shared between Cluster head and Base Station using pre-deployed public and private key.

Phase III: Group Key in the Cluster

Step1: Cluster head C_i sends an encrypted message $P_{uB}(\text{Group}, \text{groupId})$ to the Base station

Step2: Base station decrypt the message by its private key P_{rB} . Base station sends the row $(P_{uCi}(L_{gLi_j}))$ of group j to C_i encrypted by public key P_{uCi} of cluster head

Step 3: Cluster head decrypt the message by its private key P_{rCi} and gets the row L_{gLi_j} and it multiplies it with $U_{gU_{ci}}$ gets group Key K_{gi_j} and then broadcast group message in format $(\text{Group}||\text{groupId}||L_{gLi_j}||K_{gi_j}(\text{Message}))$

Step 4: Upon receiving the message every node checks whether the group id in the message is same as its groupId then it multiplies received Cluster's group row L_{gLi_j} with $U_{gU_{i,j}}$ and gets the key K_{gi_j} and use this key to decrypt the group message.

Step 5: If the groupId is not same as nodes groupId message is discarded.

Phase IV: Key Refreshing by Group Key Manager

Step 1: Base Station sends a encrypted message $P_{uCi}(\text{KeyRef}||\text{groupId}||L_{gLi}||U_{gUi})$ to the Cluster head C_i for refreshing L_{gLi} & U_{gUi} of Group members having groupId.

Step 2: Cluster Head decrypts the message using P_{rCi} .

Step 3: Cluster head encrypts the message with public key of Group key Manager $P_{uGKMi_j}(\text{KeyRef}\parallel\text{groupId}\parallel L_{gLi}\parallel U_{gUi})$ and sends it to the Group Key Manager.

Step 4: Group Key Manager Decrypt the message With P_{rGKMi_j} and then it sends the new group row and column received to all nodes having groupId using Unicast message to all the node in the group.

It is the responsibility of the Group Key Manager not to send new row and column to the compromised sensor node, so that new Group key is not revealed to the compromised node. The Group Key manager maintains the detail of compromised nodes that will be updated by the cluster Head.

5 Security Analysis

Assumption: Base station never get compromised and all information are stored on the Base station i.e., nL_i , nU_i , gL_i and gU_i matrix for $0 < i \leq m$. Security analysis shows better result than existing methodology when Node, Cluster head and Group Key Manager are compromised.

A. When Sensor Node gets compromised

One row and one column of nL_i , nU_i matrix will get revealed. This will not affect the communication between the remaining nodes within a cluster, because each link will use different key for Node-to-Node connectivity. In that case group key will get revealed, which won't affect the rest of the sensor groups, only the group in which sensor node is present will get affected, that can also can be solved by GKM. GKM will send new row and column to all the nodes in the group except the compromised node, by unicasting the message using individual key K_{j_GKM} between GKM and node.

B. When Cluster Head gets compromised

One row of nL_i , one column of nU_i , public and private key of Cluster head C_i will get revealed when C_i gets compromised, but it will not affect other communication between the nodes in a cluster. Since Public Key infrastructure communication is used between base station and cluster head, it won't affect the remaining links between base station and cluster head. When compared with scheme [9], since LU matrix is used so there is a possibility of having same Keys between base station and Non-compromised cluster head. This is the added advantage of the proposed scheme.

C. When Group Key Manager gets compromised

None of the communication links will get affected. But it can be able to change the group key. But this also can be overcome by following the procedure to check whether GKM has changed group key of nodes or not:

1. Cluster head sends nonce encrypted by using previous group key and broadcast the message
2. Sensor nodes within sensor groups will decrypt the message and again gets encrypted by individual key thereby sends that message to Cluster head.
3. If the Cluster head receives the same message, just replace the GKM as it is compromised; else replace the GKM and change Group row and column of each group.

For the security purpose, the key should be changed periodically, which can be done by refreshing the row and column of nodes using previous key.

6 Performance Analysis

A. Node-to-Node Connectivity

Compared with E-G scheme, proposed scheme has 100% node-to-node connectivity, since each node shares its row elements to get shared secret key of any other node. In E-G scheme, probability of sharing same key between two nodes, having Key pool P, and no of keys stored in each node is m, is given by,

$$Probability = 1 - \frac{(P-m)!}{P!(P-2m)!}$$

In Fig.2 for P=1000, if m=90 the probability of key sharing will be 0.99, if m=40 it will be 0.811 but in proposed scheme connectivity does not depend on the number of keys stored in node and the probability of connectivity will be 1.

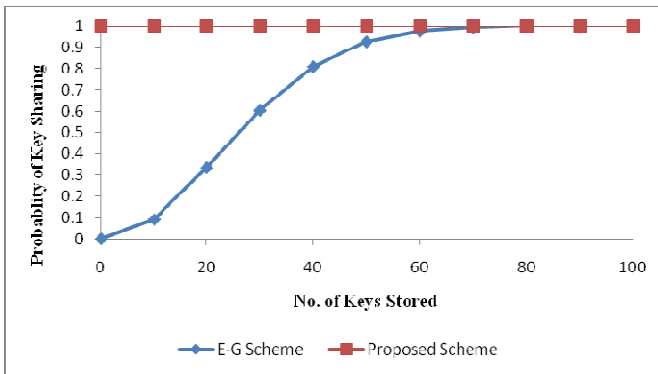


Fig. 2. Key Connectivity

B. Scalability

When the number of nodes increases, proposed scheme provides good scalability than other protocols without compromising the security. When nodes are added in the scheme using LU matrix, either size of matrix should be changed or randomly pick

row and column and assign to nodes. In the scheme given in [8] when new cluster is introduced, size of the matrix should be changed so computation and communication cost is increased. In proposed protocol, all the clusters are having different matrix for node-to-node connectivity and group communication, and different private and public key for communicating with the Base station. So we can introduce a new cluster with new set of nL_i , nU_i , gL_i , gU_i matrix, and public and private key, which reduce computation and communication cost, without compromising the security.

C. Resilience

Perfect resiliency is achieved. In scheme [8] when number of cluster head compromising increases, possibility of compromising communication links will be increases. While randomly picking up keys, possibility of assigning same key will get increases because there will be single LU matrix for cluster head and base station communication. But in proposed scheme since PKI is incorporated, compromising cluster head will not affect the remaining communication links between other cluster head and base station as shown in the Fig.3.

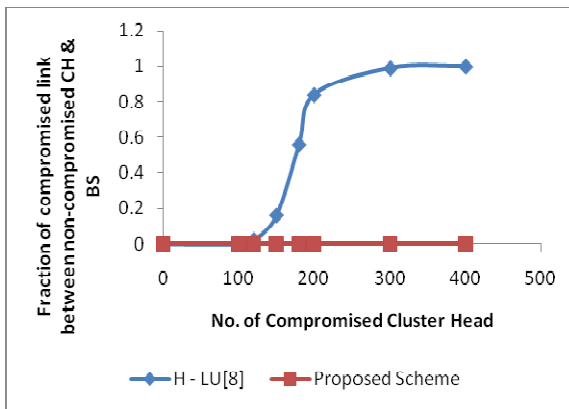


Fig. 3. Cluster Head Compromised

When a node is compromised it will not affect the communication between the non-compromised nodes up to certain threshold.

Conclusion

An efficient and Hybrid KMS for WSN is discussed. LU matrix will play a major role to achieve 100% Node-to-Node connectivity. In this architecture, secured ElGamal Public Key Cryptosystem is used between Cluster Head and Base Station and also between Cluster Head to GKM in order to improve the security. Re-keying is done using GKM. It is also good in Resilience and Scalability. Security analysis and Performance analysis show that this scheme is the most suitable one in real time active Wireless Sensor Networks applications.

References

- [1] Eschenauer, L., Gligor, V.D.: A key management scheme for distributed sensor networks. In: Proc. of the 9th ACM Conference on Computer and Communication Security (2002)
- [2] Chan, H., Perrig, A., Song, D.: Random Key Predistribution Schemes for Sensor Networks. In: Proc. of the 2003 IEEE Symposium on Security and Privacy, pp. 197–213, May 11–14 (2003)
- [3] Blom, R.: An optimal class of symmetric key generation systems. In: Beth, T., Cot, N., Ingemarsson, I. (eds.) EUROCRYPT 1984. LNCS, vol. 209, pp. 335–338. Springer, Heidelberg (1985)
- [4] Du, W., Deng, J., Han, Y.S., Varshney, P.K.: A pairwise key pre-distribution scheme for wireless sensor network. In: Proc. of the 10th ACM Conference on Computer and Communications (SecurityCCS 2003), pp. 42–51 (2003)
- [5] Liu, D., Ning, P.: Establishing pairwise keys in distributed sensor networks. In: Proc. of the 10th ACM Conference on Computer and Communications Security (CCS 2003), pp. 52–61 (2003)
- [6] Younis, M.F., Ghumman, K., Eltoweissy, M.: Location-Aware Combinatorial Key Management Scheme for Clustered Sensor Networks. *IEEE Transactions on Parallel and Distributed Systems* 17(8), 865–882 (2006)
- [7] Stallings, W.: *Cryptography and Network Security: Principles and Practices*, 4th edn. Prentice Hall, Englewood Cliffs (2005)
- [8] Wen, M., Zheng, Y., Li, H., Chen, K.: A hierarchical composition of LU matrix-based key distribution scheme for sensor networks. In: Washio, T., Zhou, Z.-H., Huang, J.Z., Hu, X., Li, J., Xie, C., He, J., Zou, D., Li, K.-C., Freire, M.M. (eds.) PAKDD 2007. LNCS (LNAI), vol. 4819, pp. 608–620. Springer, Heidelberg (2007)
- [9] Dai, H., Xu, H.: Key Predistribution Approach in Wireless Sensor Networks Using LU Matrix. *IEEE Sensors Journal* (August 2010)
- [10] Blundo, C., DeSantis, A., Herzberg, A., et al.: Perfectly-secure key distribution for dynamic conference. *Inf. Comput.* 146(1), 1–23 (1998)
- [11] Eltoweissy, M., Heydari, H., Morales, L., Sadborough, H.: Combinatorial Optimization of Key Management in Group Communications. *J. Network and Systems Management* 12(1), 33–50 (2004)
- [12] Perrig, A., Canetti, R., Tygar, J.D., Song, D.: The Tesla Broadcast Authentication Protocol. *RSA CryptoBytes* 5 (2002)

Grey System Theory-Based Energy Map Construction for Wireless Sensor Networks

Vivek Katiyar, Narottam Chand, and Surender Soni

Department of Computer Science and Engineering
National Institute of Technology Hamirpur
Hamirpur (H.P.), India
{vivek.kat,nar.chand,surender.soni}@gmail.com

Abstract. Energy is one of the most important resources in wireless sensor networks (WSN). Due to unattended nature of WSNs, it should be used smartly and efficiently to maximize lifetime. A map representing the residual energy of sensor nodes in the sensor field can be constructed, which is called as energy map. Depletion of energy in sensor nodes can be modeled as time-series. The grey models are considered to be the best tool for time-series prediction. In this paper, we propose a grey system theory-based prediction approach to construct the energy map for WSN. Simulation results show that our proposed approach outperforms various prediction based approaches for energy map construction.

Keywords: energy map, grey system theory, prediction, sensor nodes, statistical approach, WSN.

1 Introduction

An emerging area of research in the field of wireless communication and networking that attracted the attention of researches throughout the globe is wireless sensor network (WSN). WSN consists of large number of low cost sensor nodes deployed randomly in sensor field. The sensor nodes communicate wirelessly in ad-hoc manner to form a sensor field. Such networks have gained attention in application areas like industrial process monitoring and control [1, 2], machine health monitoring [3], environment and habitat monitoring, healthcare applications, home automation, and traffic control [1, 4]. A good survey of advances in WSNs and future research directions has been presented in [5].

In WSNs, the sensor devices are very constrained in terms of battery power. Sensor nodes in WSNs have non-rechargeable batteries. At the same time, it is not easy to replace batteries because WSNs are deployed generally in inhospitable environments like forests, sea and battlefields. The only way to make the WSN alive for longer time is to use the battery power efficiently. Power optimization must be taken into account at each layer of network model including physical and application layer. Since a large fraction of the energy of a sensor node is consumed in data transmission, so most of the energy efficient protocols are designed at network layer.

To optimize the energy consumption in sensor nodes, Y.J. Zhao et al. [6] have designed a residual energy scan for whole sensor network to monitor the energy

consumption in every part of the sensor network. R.A.F. Mini et al. extended their work of energy scan construction and named it as energy map [7, 8].

An energy map is a scan of available remaining energy at sensor nodes in WSNs. Energy maps can also be useful in increasing lifetime of sensor network by adaptive clustering, energy centric routings, data aggregation, etc. With the help of energy map, we can determine if any part of the sensor network is about to fail in near future due to depleted energy. In [6], authors described the aggregation based approach for energy map construction. In this approach a composite scan is created by combining all local scans by sensor nodes. Another approach for Energy map construction used in [7, 8] is based on prediction. In prediction based approach a sensor node can predict its energy consumption based on its past history. Based on that prediction, energy map can be constructed. In the mechanism proposed by Mini et al. [7, 8] every sensor node has not to send energy information to the monitoring node, it can just send its available energy and parameters of energy consumption model. In this way authors minimized the cost of energy map construction. C. Song et al. [9] proposed an energy map construction approach based on non-linear manifold learning algorithm. We should devise some efficient methods for energy map construction so that utility of the energy information compensate the amount of energy spent in this process.

Energy map is a very important phenomenon in the context of WSN because the energy information is usually very crucial to develop a good sensor network protocol stack such as clustering algorithms and routing protocols. In [10] authors presented a clustering algorithm that takes clustering decision based on energy information in the sensor network. The algorithm also dynamically selects the clusterheads based on residual energy information. A routing algorithm named Energy Centric Routing (ECR) has been proposed by N. Jamal et al. [11]. This algorithm takes routing decision based on residual energy in the network. The protocol proposed in [12] is an example of a routing protocol that could take advantage of the energy map. In this work author described the trajectory based forwarding protocol that is a new forwarding algorithm suitable for routing packets along a predefined curve. Many other applications of energy map can be found in data gathering [13] and data dissemination [14]. Some other possible applications are reconfiguration algorithms, query processing and data fusion that could take advantage of the energy map. In fact, we cannot think of an application or an algorithm that is not benefited with the use of energy map.

In this paper, we propose a grey system theory-based energy map construction algorithm for WSN. If a sensor node can predict its future energy consumption based on past history, the clusterhead will be able to prolong the lifetime of sensor node by distributing load. But the effectiveness of this concept depends on how accurate the prediction is. Depletion of energy in sensor nodes can be modeled as time-series. The grey models are considered to be one of the best tools for time-series prediction [15]. The strength of grey models over conventional statistical models is that it requires only a limited amount of data to estimate the behavior of unknown systems [16].

Rest of the paper is organized as follows. Section 2 describes the related work. Energy map and prediction based protocol to construct the energy map is described in Section 3. Section 4 presents the fundamentals of grey system theory and grey system theory based prediction algorithm. Simulation results are presented in Section 5. Lastly the paper is concluded in Section 6.

2 Related Work

In this section we describe the popular and recent approaches for energy map construction such as aggregation based, prediction based, statistical based, contour map based, etc. Residual energy scan has been computed by many authors with different names as eScan [6], Energy map [7, 8, 9], Continuous Residual Energy Monitoring (CREM) [17] and Iso-Map [18].

Zhao et al. [6] first tried to design a monitoring scheme for sensor networks in the same manner as SCAN [19] that provides a multicast based continuous monitoring infrastructure. This scheme is based on residual energy scan for monitoring sensor network health. Any network failure due to energy depletion can be avoided. An aggregation based approach is used for energy map construction.

First each node calculates its local scan with residual energy and its location. Sensor node reports to the monitoring node only when its energy level drops significantly since last time it reported its scan. All local scans are aggregated with the help of aggregation tree.

The pioneering work of Zhao et al. [6] can be thought of as revolution in the field of energy efficient WSNs, however some problem exists. Sometimes there may be so much message exchange for energy scan construction that will not justify the energy saved because there was no clear cut assumption of hierarchical structure. Moreover there is no topology control mechanism. Nodes near to the base station will consume more energy. To overcome these shortcomings and reduce the cost of collecting residual energy scan of sensor network, the authors proposed a monitoring tool called *digest* which is an aggregate of some network properties [20].

Another improvement in eScan approach has been proposed by H. Song et al. [17] that follows a hierarchical approach for continuously collecting residual energy scans to construct the final scan at base station. Continuous Residual Energy Monitoring (CREM) divides whole sensor network into clusters, represented by clusterheads. A backbone is constructed with the help of clusterhead nodes. One of the major advantages of CREM over previous approaches is topology maintenance. Topology maintenance is very much important because clusterhead nodes have to transmit and receive continuously for energy scan construction and hence will deplete its energy resource at a much higher rate.

Aggregation base approaches generally suffer from heavy transmission traffic and sometimes a large computational overhead on each sensor node. M. Li et al. proposed an energy efficient contour mapping scheme named Iso-Map [18]. Iso-Map constructs a contour map that reduces the network traffic and computation overhead by selecting some nodes intelligently to report energy data. These nodes are called as *isoline* nodes.

A scheme that abstracts the energy concentration in the sensor field into prescribed levels like a topographic map has been proposed by Al-Karaki et al. [11]. The scheme provides an aggregated view of the residual energy levels of different regions in the sensor field instead of detailed information of residual energy at individual sensor nodes. The scheme is named as ECscale.

In [8], an approach to form energy map by predicting the energy consumption of sensor node is proposed. It predicts the energy consumption based on past history of sensor nodes. According to the model, if any sensor node efficiently predicts its

energy consumption in future, then a significant amount of message flow to collect the energy information will be reduced. At the same time, effectiveness of the approach depends upon the accuracy of the prediction model.

To predict the energy dissipation, probabilistic model based on Markov chain is used. Sensor nodes with M operation modes are modeled as Markov chain with M states. In this model, there is a fixed probability of sensor nodes transiting from one state to other in next time step. The n -step transition probability can be defined as $P_{ij}^{(n)} = \sum_{k=1}^M P_{ik}^{(r)} P_{kj}^{(n-r)}$ for $0 < r < n$, where i is the current state and j is the next state. With the help of transition probability $P_{ij}^{(n)}$, each sensor node can find out the expected amount of energy spent in next T times.

$$E^T(i) = \sum_{s=1}^M \left(\sum_{t=1}^T P_{is}^{(t)} \right) \times E_s$$

Using this $E^T(i)$, each node can calculate the energy dissipation rate ΔE . Now, each sensor node sends remaining energy and ΔE to monitoring node. Monitoring node collects energy information from all nodes and constructs a final energy map of entire sensor network. Simulation results show that this approach saves a significant amount of energy in energy map construction in comparison with naive approach.

A statistical model to forecast the available energy is defined in [7]. Energy map can be constructed with forecasted values. The energy drop of a sensor node can be represented as time series [21]. In this work, author uses the AREMA model [22] to predict the future values of time series.

In another work of Mini et al. [23], an approach for energy map construction under finite energy budget is discussed. Finite energy budget means that each node will spend a fixed amount of energy for energy map construction. Authors assume number of packets as metric for energy budget. This approach of energy map construction is based on probability.

The grey models are very efficient tool for time series prediction for all engineering applications. These can also be useful in various fields of the wireless sensor network. One of such recent application is grey model based data aggregation (GMDA) [25] for WSN. In this work authors designed a prediction based data aggregation technique to reduce redundant data transmission. Authors have used double queue mechanism to synchronize the predicted data series at the sensor node and the sink node. In this way this prediction technique avoids commutative errors of continuous predictions. In this approach sensor node energy is saved by avoiding the unnecessary data transmission. The sensor node sends data to sink node only when the prediction error is more than a pre-configured value.

3 Prediction Based Energy Map

The information about amount of available energy at every sensor in sensor network can be very useful in designing better energy efficient protocols. A very basic approach to collect the energy information from each sensor node is that every sensor node periodically sends its energy information to the sink or any monitoring node. But such approaches may require a lot of energy for energy map construction and the energy information would not compensate the amount of energy spent in this process.

In this paper, we propose a prediction based approach for energy map construction. This approach is based on prediction based cooperation between the sensor node and the sink node. Both sensor node and sink node will use the same prediction model and same energy data for prediction. Sensor nodes need not to send the energy information to the sink every time when there is an energy drop. Sink node can predict the energy drop in each sensor and an energy map can be constructed based on the predicted value of energy in each sensor node. Thus a significant amount of energy can be saved if sensor node and sink node efficiently predict the amount of energy sensor node will spend in future.

3.1 Energy Map

Just like a weather map or air traffic radar images, a scan of sensor network can describe the geographical distribution of network resources or activity of a sensor field. We can also draw an energy scan for a sensor network in which we group the sensor nodes according to their residual energy. This energy scan is also known as energy map. An energy map for a typical WSN is shown in Fig. 1. In other words, an energy map is a scan of available remaining energy at sensor nodes in WSNs. Energy map can help in detecting the network failure due to depleted energy in next rounds. In this way, the energy map can be useful in deployment of additional nodes in the regions where the energy of the sensors is likely to deplete soon.

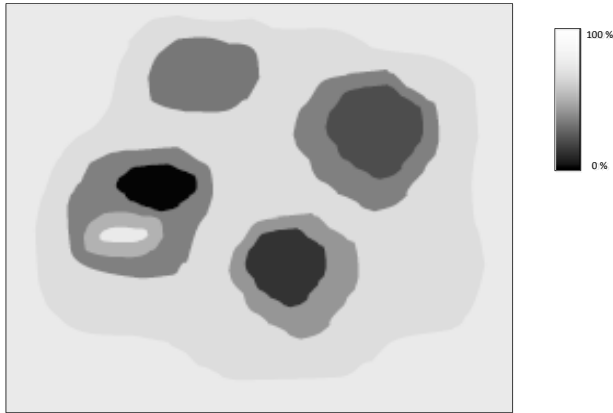


Fig. 1. Energy Map for a typical Wireless Sensor Networks

A node is assigned to collect the scans from all parts of the sensor network. This node is known as monitoring node. Selection of monitoring node depends upon the application. Generally speaking the selection of monitoring node is done from the region having comparatively more residual energy. Nodes near the monitoring node probably spend more energy because they are used more frequently to relay packets to the monitoring node. Energy map construction methods should be energy efficient so that utility of the energy information compensates the amount of energy spent in this process.

3.2 The Protocol

Assumptions

- (1) Sensor node can send energy information only once in a period.
- (2) Both the sensor node and sink node will use the same prediction model.
- (3) Sink node have sufficient computing power, energy and storage.
- (4) There is reliable data transmission between sensor node and sink node.

The sink node broadcasts its maximum acceptable prediction error \mathcal{E} , to all sensor nodes. Two data queues, prediction energy queue, $PEQ_{sensor,i}$ at each sensor and corresponding $PEQ_{sink,i}$ at sink node for each sensor i is constructed. The length of the $PEQ_{sensor,i}$ and $PEQ_{sink,i}$ are equal and specified by prediction algorithm. Initially $PEQ_{sink,i} = PEQ_{sensor,i}$, i.e. both $PEQ_{sink,i}$ and $PEQ_{sensor,i}$ store the same predicted values of energy of sensor node i at sink node and sensor node respectively. Actual energy value (AEV) at any period t is compared with $PEQ_{sensor,i}[t]$, the t^{th} item in sensor prediction queue. If the difference between the $PEQ_{sensor,i}[t]$ and AEV is greater than \mathcal{E} , the $PEQ_{sensor,i}[t]$ is replaced with AEV and the value of $PEQ_{sensor,i}[t]$ is sent to the sink. Sink node will update the t^{th} item in its $PEQ_{sink,i}$ queue. Now the prediction model uses the updated energy information for further predictions.

4 Grey System Theory-Based Energy Map

Grey system theory is an interdisciplinary scientific area that was first introduced in early 1980s by Deng [16]. Since then, the theory has become quite popular with its ability to deal with the systems that have partially unknown parameters. White system is the system in which the information is assured and the data is whole. Black system is the system in which the information is not assured and the data is little. A system is grey system in which some information is known and some is not. It has been used well in social facets since the establishment of grey theory.

Grey models are powerful tools in predicting the future values of a time series based only on a set of the most recent data depending on the window size of the predictor. In general, $GM(n,m)$ denotes a grey model, where n is the order of the difference equation and m is the number of variables. $GM(1,1)$ type of grey model is a time series forecasting model that is most widely used in the literature, pronounced as “Grey Model First Order One Variable”.

4.1 The $GM(1,1)$ Model

$GM(1,1)$ is used to capture long-term trends from recent behavior of sensor node. Before making prediction, few historical data values should be stored to construct the initial data sequence for $GM(1,1)$ model. It is denoted as $X^{(0)}$.

$$X^{(0)} = (x^{(0)}(1), x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(t)). \quad \text{for } t \geq 4 \quad (1)$$

Where $x^{(0)}(j)$, $j=1, 2, \dots, t$, represents energy information. Model $GM(1,1)$ uses the energy information of most recent t periods. We apply the Accumulation Generation Operator (AGO) to $X^{(0)}$ to obtain the following series $X^{(1)}$,

$$X^{(1)} = (x^{(1)}(1), x^{(1)}(2), x^{(1)}(3), \dots, x^{(1)}(t)) \quad \text{for } t \geq 4 \quad (2)$$

where $x^{(1)}(k) = \sum_{i=1}^k x^{(0)}(k)$ and $k=1, 2, 3 \dots t$.

The generated mean sequence $Z^{(1)}$, of $X^{(1)}$ is defined as

$$Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), z^{(1)}(3), \dots, z^{(1)}(t)) \quad (3)$$

where $z^{(1)}(k)$ is the mean value of adjacent data, i.e.

$$z^{(1)}(k) = \frac{1}{2}(x^{(1)}(k) + x^{(1)}(k - 1)) \text{ and } k=2, 3, \dots, t.$$

Therefore $GM(1,1)$ model can be established as follows:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = b \quad (4)$$

a and b are sequence of parameters and can be found as follows

$$[a, b]^T = (B^T B)^{-1} B^T A \quad (5)$$

$$A = \begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \dots \\ x^{(0)}(t) \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} -z^{(1)}(2) & 1 \\ -z^{(1)}(3) & 1 \\ \dots & \dots \\ -z^{(1)}(t) & 1 \end{bmatrix}$$

According to Eq. (4),

$$x_p^{(1)}(k + 1) = e^{-ak} \left(x^{(1)}(1) - \frac{b}{a} \right) + \frac{b}{a} \quad (6)$$

Therefore the predicted data $x_p^{(1)}(k + 1)$ of the data sequence $X^{(0)}$ can be computed as follows

$$x_p^{(0)}(k + 1) = x_p^{(1)}(k + 1) - x_p^{(1)}(k) \quad (7)$$

Now the prediction error can be computed as follows

$$\Delta(t + 1) = |x_p^{(0)}(t + 1) - x^{(0)}(t + 1)|$$

If $\Delta(t + 1) < \epsilon$, the sensor node need not to transmit $x^{(0)}(t + 1)$ to the sink node.

Otherwise it must send $x^{(0)}(t + 1)$ to sink node. The sink node will also execute the same prediction algorithm for same energy data. It is very important to synchronize $PEQ_{\text{sink},i}$ and $PEQ_{\text{sensor},i}$.

5 Simulation Results

In this section we present the simulation results of naive approach and prediction based approach for energy map construction that is based on grey system theory. For this work we have assumed the sensor network to be static and homogeneous and replacement of batteries is unfeasible and impossible. To analyze the performance of proposed system, we have implemented a grey system theory based energy map construction scheme on ns-2. Compared to other prediction based approaches, grey system theory has been found light weight.

5.1 Optimal Sequence Length

Generally, if longer sequence of data is used for prediction, it may produce more accurate prediction. But, at the same time it will consume more storage at sensor node and computational complexity. As shown in Fig. 2, the optimal value of sequence length is 3, at which prediction accuracy is maximum. We have chosen randomly five sensor nodes, each having 30 continuous values of their energies. We performed simulation for $t=0$ to 9 for all five sensor nodes and plotted average value of prediction accuracy at each value of t .

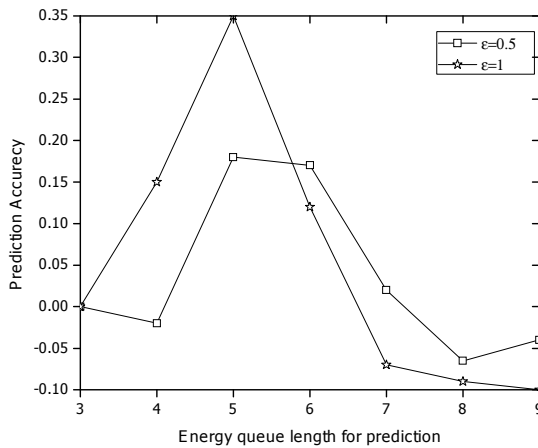


Fig. 2. Optimal queue length

5.2 Predicted Value and Correct Value of Energy

We simulated our approach for 100 sensor nodes having 1J of initial energy for 1000s. Fig. 3 shows that by using grey system theory based prediction we have to send only three packets (167, 515, 789) when the prediction error is less than maximum acceptable error, \mathcal{E} , while naive approach will send eight packets (120, 173, 399, 502, 639, 781, 875, 985). This shows a significant reduction in number of packets sent for energy map construction, hence saving in energy.

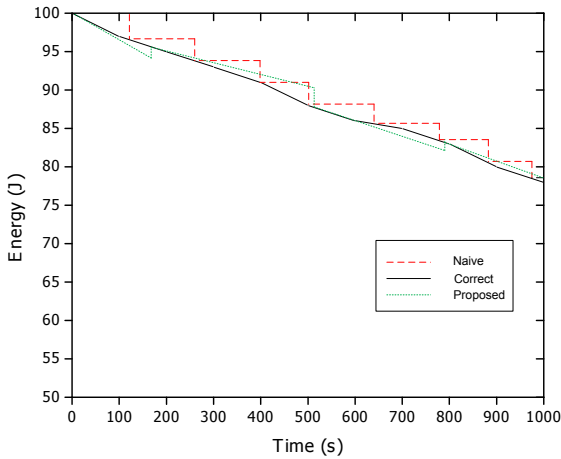


Fig. 3. Energy packet transmission for energy map construction

6 Conclusion

Energy maps can be very useful in lifetime maximization of wireless sensor networks. In this paper, we have proposed a prediction based approach of energy map construction. The algorithm makes the prediction based on grey system theory. Grey system theory is a widely used prediction method for various engineering streams. In the proposed approach the sink node collects the energy information predicted by sensor node if the prediction error is greater than the maximum acceptable prediction error. Simulation results prove the effectiveness of the proposed approach. It reduces up to 30% message exchange for energy map construction than naive approach.

References

1. Kay, R., Mattern, F.: The Design Space of Wireless Sensor Networks. *IEEE Wireless Communications* 11(6), 54–61 (2004)
2. Haenselmann, T.: *Sensornetworks*. GFDL Wireless Sensor Network textbook, http://pi4.informatik.uni-mannheim.de/~haensel/sn_book (retrieved August 29, 2006)
3. Tiwari, A., Ballal, P., Lewis, F.L.: Energy-efficient wireless sensor network design and implementation for condition-based maintenance. *ACM Transactions on Sensor Networks (TOSN)* 3(1) (2007)
4. Hadim, S., Mohamed, N.: Middleware: middleware challenges and approaches for wireless sensor networks. *IEEE Distributed Systems Online* 7(3), 1 (2006)
5. Katiyar, V., Chand, N., Chauhan, N.: Recent Advances and future trends in Wireless Sensor Networks. *International Journal of Applied Engineering Research* 2(1), 43–55 (2010)
6. Zhao, Y.J., Govindan, R., Estrin, D.: Residual Energy Scans for Monitoring Wireless Sensor Networks. In: *IEEE Wireless Communications and Networking Conference*, pp. 356–362 (2002)

7. Mini, A.F., Antonio, L.A.F., Nath, B.: The distinctive design characteristic of a wireless sensor network: the energy map. *Computer Communications* 27, 935–945 (2004)
8. Mini, R.A.F., Machado, M.V., Loureiro, A.A.F., Nath, B.: Prediction-based Energy map for Wireless Sensor Networks. *Ad Hoc Net. J.* 3, 235–253 (2005)
9. Song, C., Guizani, M.: Energy map: Mining Wireless Sensor Network Data. In: *International Conference on Communications, 2006, ICC 2006*, vol. 8, pp. 3525–3529. IEEE, Los Alamitos (2006)
10. Song, C., Guizani, M., Sharif, H.: Adaptive clustering in wireless sensor networks by mining sensor energy data. *Computer Communications* 30, 2968–2975 (2007)
11. Al-Karaki, J.N., Ghada, Al-Mashaqbeh, A.: Energy-centric routing in wireless sensor networks. *Microprocessors and Microsystems* 31, 252–262 (2007)
12. Niculescu, D., Nath, B.: Trajectory-based forwarding and its applications. In: *Rutgers University Technical Report DCS-TR-488*, pp. 1–18 (2002)
13. Goussevskaia, O., Machado, M.V., Mini, R.A.F., Loureiro, A.A.F., Mateus, G.R., Nogueira, J.M.: Data Dissemination Based on the Energy map. *Topics in Ad-hoc Networking, IEEE Communications Magazine*, 134–143 (2005)
14. Rhazi, A.E.L., Pierre, S.: A Data Collection Algorithm Using Energy maps in Sensor Networks. In: *Third IEEE International Conference on Wireless and Mobile Computing, Networking and Communications, WiMob 2007* (2007)
15. Kayacan, E., Ulutas, B., Kaynak, O.: Grey system theory-based models in time series prediction. *Expert Systems with Applications* 37, 1784–1789 (2010)
16. Deng, J.L.: Introduction to grey system theory. *The Journal of Grey System* 1(1), 1–24 (1989)
17. Han, S., Chan, E.: Continuous Residual Energy Monitoring in Wireless Sensor Networks. In: *Cao, J., Yang, L.T., Guo, M., Lau, F. (eds.) ISPA 2004. LNCS*, vol. 3358, pp. 169–177. Springer, Heidelberg (2004)
18. Li, M., Liu, Y.: Iso-Map: Energy-Efficient Contour Mapping in Wireless Sensor Networks. *IEEE Transactions on Knowledge and Data Engineering* 22(5), 699–710 (2010)
19. Reddy, A., Estrin, D., Govindan, R.: Large Scale Fault Isolation. *IEEE Journal of Selected Areas in Communication, Special Issue on Network Management*, 733–743 (2000)
20. Zhao, J., Govindan, R., Estrin, D.: Computing aggregates for monitoring wireless sensor networks. *Technical Report 02-773, USC* (September 2003)
21. Brockwell, P.J., Davis, R.A.: *Introduction to Time Series and Forecasting*, 2nd edn. Springer, New York (2002)
22. Box, G.E.P., Jenkins, G.M.: *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco (1976)
23. Mini, R.A.F., Loureiro, A.A.F., Nath, B.: Energy map Construction for Wireless Sensor Network under a Finite Energy Budget. In: *MSWiM 2004*, pp. 165–169 (2004)
24. Tenenbaum, J.B., Silva, V.D., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. *Science* 290, 2319–2323 (2000)
25. Wei, G., Linga, Y., Guoa, B., Xiaob, B., Vasilakos, A.V.: Prediction-based data aggregation in wireless sensor networks: Combining grey model and Kalman Filter. *Computer Communication* 34(6), 793–802 (2011)

An Entropic Approach to Data Aggregation with Divergence Measure Based Clustering in Sensor Network

Adwitiya Sinha and D.K. Lobiya

Jawaharlal Nehru University, Mehrauli Road,
New Delhi, India
{mailto:adwitiya,lobiyal}@gmail.com

Abstract. In wireless sensor network, data fusion is considered an essential part for preserving energy. Periodic data sampling leads to enormous collection of raw facts, the transmission of which would rapidly deplete the sensor power. In this paper, we have performed data aggregation on the basis of entropy of the sources. The entropy is computed from the local and global probability models. The models provide assistance in extracting high precision data from the sensor nodes. Further, we have proposed an energy efficient method for clustering the sensor nodes. Initially the sensors sensing same category of data are placed within a distinct cluster. The remaining unclustered sensors estimate their divergence with respect to the clustered neighbors and ultimately join the least-divergent cluster. The performance of our proposed methods is evaluated using ns-2 simulator in terms of entropy, aggregation cycles and energy utilization. The simulation results confirm the validity and efficiency of our approach.

Keywords: Wireless sensor network, node clustering, Kullback-Leibler directed divergence measure, Jeffrey's symmetric divergence measure, data aggregation, entropy, fuzzy-entropy, local and global probability measure.

1 Introduction

The energy consumption in wireless sensor network (WSN) has created enormous awareness among the researchers for increasing the network lifetime. The sensor network is considered to have prospective results in terms of dynamism and diversity in everyday applications. A WSN is formed with small electronic devices possessing self-configuring capability that are either randomly deployed or manually positioned in huge bulk [1]. Moreover, the sensor network exhibits less mobility than the mobile adhoc networks. The sensor network is basically designed to monitor hostile environments, as a result of which it becomes extremely difficult to replace or recharge the sensor batteries. Such constraints demand energy efficient resolutions to enhance the performance of sensor network. The proliferation of sensor network has created the urge of exploring novel ideas for data aggregation, thereby lessening the transmission cost. However, the aggregation schemes would require efficient clustering protocols to well-implement it's functioning. Therefore, we have proposed an aggregation and clustering protocol for the prolonged existence of sensor network.

The remainder of this paper is organized as follows. A brief survey of previous research carried out in the related field is included in section 2. Our proposed clustering technique based on divergence measure is provided in section 3. In section 4 the proposed fuzzy-entropy based aggregation scheme has been elaborated. Section 5 shows the performance evaluation of our proposed method. Finally, the paper is concluded in section 6 along with directions for further scope.

2 Related Work

Significant amount of research has been carried out in achieving the goal of increasing the network lifespan through clustering and aggregation. Several methods of data aggregation [2] depend on the topology of the sensor network. For instance, a tree-based data aggregation protocol [3] constructs a simple topology based on a parent and child association. However, large transmission delays and poor rate of aggregation makes it unsuitable for the dynamic applications. Further, we have centralized aggregation protocol [4], in which aggregation is done only at the sink (data processing center). As a result, such protocols lead to heavy workload and unnecessary packet drops. There are other clustering schemes based on static [5-7] and dynamic cluster aggregation [8-10]. In case of static environment, the clusters are formed in the initial stage and the aggregation is carried out by the cluster heads. The clusters once formed remain unchanged throughout the network lifespan. This procedure is suitable for area monitoring (recording earthquake, temperature, humidity, etc.), but not supported over wide range of applications, like- forest fire supervision, wildlife monitoring, target tracking, etc. Therefore most of the research awareness can be found in dynamic cluster aggregation schemes, where clusters are formed dynamically and updated on sensing environmental parameters followed by aggregation at the cluster head. The clusters formed in this case, are also known as *adaptive clusters*.

In this paper, we propose a dynamic cluster aggregation algorithm to perform aggregation of data at two levels: sensor node and cluster head. With the use of entropy and information theory, we attempt to reduce the transmission and processing cost with increase in the relevance of aggregated data.

For the evaluation of the performance of our proposed strategy, we make a comparative analysis with two well-known clustering protocols: Hybrid Energy-Efficient Distributed Clustering (HEED) [19] and an inference clustering protocol based of Belief Propagation (BP) [20]. HEED is a distributed clustering approach that operates in energy efficient manner and helps in prolonging network lifetime. It is scalable over large network sizes and performs load balancing within clusters. However, frequent computation of communications cost and broadcasting among neighbors degrades its performance. As a strong counterpart, BP clustering method offers energy effective solutions based on belief calculations with potential functions. Though BP performs better than HEED in terms of clustering the network and packet delivery performance, but long-length messages induce larger overheads in message passing. This makes transmission cost higher in case of BP. Previous simulations have shown a marginal difference in network lifetimes contributed by these protocols.

3 Proposed Divergence Measure Based Clustering Technique

Clustering is the process of assigning a set of sensor nodes, with similar attributes, to a specified group or *cluster*. There is a wide hierarchy of clustering algorithms available for different types of sensor network applications.

In our research, we have proposed a new energy efficient clustering algorithm that operates in two phases: preliminary and final clustering phase. In preliminary phase, sensor nodes sensing same category of data are placed in a distinct cluster. In final phase, the remaining unclustered sensors estimate their divergence with respect to the clustered neighbors and ultimately join the least-divergent cluster.

3.1 Preliminary Clustering Phase

The formation of preliminary clusters is purely distributed and is based on the sensed data. The proposed clustering method is independent of predetermination of number of clusters, geographic positioning and distance measures.

We have used a window function $\phi(\cdot)$ to normalize the sensed data so as to scale the value within the range $[0..1]$. Let us assume, a and b to be the minimum and maximum value of the environmental parameter to be monitored and $x_{avg}(t)$ be the average of the set of data sensed for the time interval t . The window function can be defined as follows:

$$\phi\left(\frac{x_{avg}(t)}{b-a}\right) = \begin{cases} 1 & \left(\frac{x_{avg}(t)}{b-a}\right) \in [0,0.2[\\ 2 & \left(\frac{x_{avg}(t)}{b-a}\right) \in [0.2,0.4[\\ 3 & \left(\frac{x_{avg}(t)}{b-a}\right) \in [0.4,0.6[\\ 4 & \left(\frac{x_{avg}(t)}{b-a}\right) \in [0.6,0.8[\\ 5 & \left(\frac{x_{avg}(t)}{b-a}\right) \in [0.8,1.0[\\ 0 & otherwise \end{cases} \quad (1)$$

The sensors use the window function to map the data into one of the formats [11]. All the nodes that sense the same format in 1-hop distance groups together to form a *preliminary cluster*.

In the initial phase, the node with maximum energy within the preliminary cluster is appointed as the cluster head. It maintains a *duration timer* to keep track of the period for which it remained cluster head. Once appointed the node functions as cluster head till its duration timer expires. On the expiration of the timer, the role of cluster head rotates to other probable nodes whose residual energy qualifies above a minimum predefined energy threshold (Ω_E). The head rotation performs load balancing within the clusters. The cluster head assigns a unique cluster *id* to all the cluster members and also keeps account of the cluster probability.

Though the idea of preliminary stage of cluster formation is simple to implement but due to some situations (boundary value or out-of-bound data sensing) few nodes in the network might still remain unclustered. This problem is solved by our *final clustering phase*.

3.2 Final Clustering Phase

The final clustering phase ensures that all the nodes in the sensor network get clustered. The process begins with an unclustered node discovering one or more clustered neighbour in its direct hop. The node then obtains the array of probabilities of the sensed data from its neighbours that are distinctly clustered. However, if there is no clustered neighbour in 1-hop vicinity, then the node will wait till it discovers one. Since most of the nodes would be clustered in the first phase (in which spatial property of the node is exploited, i.e. node sensing same category data are put together in similar groups), fewer nodes would confront such isolation.

The sensor nodes maintain the following information in its database, in order to calculate the divergence measure required for final clustering.

$$\Delta_n^s = \left\{ P^s = (p_1^s, p_2^s, p_3^s, \dots, p_n^s), p_i^s \geq 0, \sum_{i=1}^n p_i^s = 1 \right\} \quad (2)$$

where p_i^s is the probability of i^{th} data format from the sensor s and the probability sequence is denoted by P^s .

Derivation of divergence measure. We know that the entropy of the source can be given by the *Shannon's entropy* $H(P)$:

$$H(P) = - \sum_{i=1}^n p_i \ln p_i \quad (3)$$

where $p_i \in P^S$ and P is *Host Probability Model (LPM)* of host sensor node. Moreover, the *inaccuracy* in data is given by:

$$H(P \parallel T) = - \sum_{i=1}^n p_i \ln t_i \quad (4)$$

where $t_i \in T^S$ and T is *Remote Probability Model (LPM)* of remote sensor node. On subtracting equation (4) from (3), we get *Kullback-Leibler directed divergence measure* [12]:

$$\begin{aligned} D(P \parallel T) &= H(P \parallel T) - H(P) = - \sum p_i \ln t_i + \sum p_i \ln p_i \\ &= \sum p_i \ln p_i - \sum p_i \ln t_i = \sum p_i \ln \left(\frac{p_i}{t_i} \right) \end{aligned} \quad (5)$$

However, the directed divergence $D(P \parallel T)$ is not symmetric divergence measure. This means $D(P \parallel T) \neq D(T \parallel P)$, which would lead to error prone results. Therefore,

we consider the symmetric version of *Kullback-Leibler*, known as *Jeffrey's (J) divergence measure* [13], by deriving it from equation (5):

$$J(P \parallel T) = D(P \parallel T) + D(T \parallel P) = \sum p_i \ln \left(\frac{p_i}{t_i} \right) + \sum t_i \ln \left(\frac{t_i}{p_i} \right)$$

$$J(P \parallel T) = \sum p_i \ln \left(\frac{p_i}{t_i} \right) - \sum t_i \ln \left(\frac{p_i}{t_i} \right) = \sum (p_i - t_i) \ln \left(\frac{p_i}{t_i} \right) \quad (6)$$

Application of divergence measure. Every unclustered sensor node makes use of the *J-divergence measure* derived in equation (6) to calculate the divergence between itself and every other clustered (neighboring) sensor nodes.

$$\left. \begin{array}{l} J(T^1 \parallel P^s) \\ J(T^2 \parallel P^s) \\ \vdots \\ J(T^z \parallel P^s) \end{array} \right\} = \min J(T^x \parallel P^s) \quad (7)$$

where $J(T^i \parallel P^s)$ denote the *J-divergence measure* between the i^{th} clustered node and s^{th} sensor node to be clustered. The unclustered sensor s will join the node x , with which its divergence is the least as compared to other clustered nodes. This process of clustering recursively continues till all nodes in the network are clustered.

4 Proposed Data Fusion Algorithm Using Fuzzy-Entropy

In the proposed work, we apply the data fusion approach for monitoring the variation in the temperature. However, generalization can be done to other environmental parameters, for instance- pressure, humidity, etc.

Fuzzification of input data. We consider five data formats for recognizing the range of sensed data. The formats are represented by two well-known fuzzy membership functions - *Sigmoidal* and *Generalized-bell membership function* [14-15]. The *Sigmoidal membership function* is given by $y = \text{sigmf}(x, [a \ c])$

$$f(x; a, c) = \left[\frac{1}{1 + e^{-a(x-c)}} \right] \quad (8)$$

The *Generalized-bell membership function* is given by $y = \text{gbellmf}(x, [a \ b \ c])$

$$f(x; a, b, c) = \left[\frac{1}{1 + \left| \frac{x-c}{a} \right|^{2b}} \right] \quad (9)$$

We have selected *Generalized-bell* membership function to model the moderate data formats: m_2 (cold temperature) m_3 (normal temperature) m_4 (hot temperature). Moreover, *Sigmoidal* membership function has been chosen to model extreme data formats: m_1 (very cold temperature) m_5 (very hot temperature). These functions are

known for best representing maximum variation and smoothness. The temperature is continuous parameter which requires functions that can well represent its characteristics. Therefore, both the membership functions are appropriate in this regard. The simulation parameters of the membership functions are provided in table 1 and 2 respectively.

Table 1. GBELLMF parameters table

Data Formats	a	b	c
m_2	2.86	2.43	12.60
m_3	4.10	3.86	24.36
m_4	2.86	2.43	35.60

Table 2. SIGMF parameters table

Data Formats	a	c
m_1	-1.57	6.857
m_5	+1.57	41.640

Sampling process. We assume that the sensors sense data for a time period of t seconds. Further, let $M = \{m_i \mid i = 5\}$. After t seconds, a sequence $\delta(t)$ of L messages is generated:

$$\delta(t) = m_{i1}, m_{i2}, m_{i3}, \dots, m_{iL} \tag{10}$$

The frequency f_i of message m_i is recorded.

Local probability measure. On the basis of the frequency of occurrence of each message (m_i) with respect to sensor s , local probability is computed as:

$$p_i^{(s)} = \frac{f_i}{\sum_{i=1}^n f_i} \tag{11}$$

such that $\sum_{i=1}^n p_i^{(s)} = 1$. This probability function has been designed to capture the maximum variation. Finally, the entropy is calculated locally at each sensor s as the following:

$$H_r^{(s)}(M) = \sum_{i=1}^n p_i^{(s)} \log_r \left(\frac{1}{p_i^{(s)}} \right) \tag{12}$$

All the sensors send ($sensor_id, H_r^{(s)}(M)$) to the cluster head. The cluster head then computes an entropy threshold and sends an *ack* (acknowledgement) to selected sensors that qualify the threshold. This results in *sensor compression*. On receiving *ack*, sensors compute the mathematical expectation of the data value:

$$d_{(exp)}^{(s)} = \sum_{i=1}^n p_i^{(s)} d_i \tag{13}$$

Finally, the sensors send ($sensor_id, d_{(exp)}^{(s)}$) to the cluster head.

Global probability measure. On receiving data and entropy from selected sources the cluster head computes a global probability p_q to capture the focused information (rather than capturing the variation in the information). Finally, the cluster head computes the expected value of the actual set of data received from the selected sensors:

$$d_{\text{expc}} = \sum_{q=1}^{|Q|} p_q d_i \tag{14}$$

The cluster head then sends $(\text{cluster_id}, d_{(\text{expc})})$ to the data processing node (sink). The performance of the aggregation further shows that our proposed approach is relevant as well as energy-efficient.

Network timeline diagram. The network timeline diagram shows the working slots for initial cycle of our proposed work (figure 1). The network initiates with the gathering of data by individual nodes, also known as *random sensing*. The next stage in the cycle is our proposed *preliminary clustering phase* (PCP), on the completion of which data aggregation and *final clustering phase* (FCP) are executed in parallel thereby performing better time utilization. The data aggregation or *data fusion phase* (DFP) is further classified in local and global phases respectively. The local DFP is carried out by individual nodes with the help of local probability measure and the global DFP is performed by the cluster heads using the global probability measure. Since, the aggregation process is involved only within the cluster, the FCP can continue in parallel (outside the clusters) without collision. This efficient utilization of time ultimately results in significant energy savings.

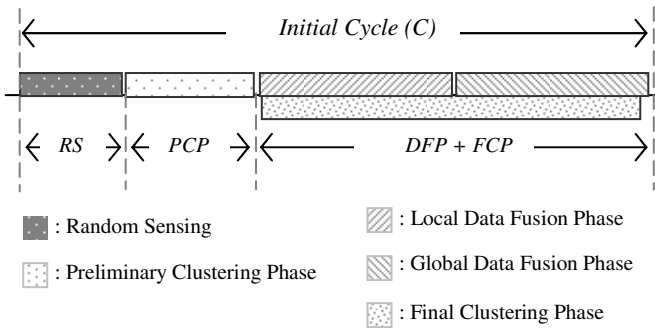


Fig. 1. Network timeline diagram for the initial cycle

5 Simulation and Performance Evaluations

The simulation of the proposed clustering and entropy-based aggregation is performed using network simulator [16-18]. Moreover, on the basis of the tracing data generated in ns-2 and other log files, the graphical evaluation is generated using Matlab. In our simulation, the sensor nodes are randomly deployed over a network of

dimension 1000 x 1000 square meters. The simulation parameters used for the experimentation are specified in table 3. We have used Gaussian bell and Sigma membership functions to monitor the fuzzy environmental parameter (temperature). The trend of the membership function, over the range of temperature being monitored, is highlighted in figure 2.

Our proposed clustering method uses divergence measure to discover clusters in the network. The clustered nodes keep track of the frequency of data formats sensed during the sampling period. The bar graph plotted in figure 3 displays the variation of local probability of different data formats with respect to a randomly chosen cluster in the sensor network.

Further, figure 4 represents the local entropy sent by the sensors to their cluster head in a cluster. It is evident from the graph that node with id 5 gives highest entropy owing to the least variation of the same node in figure 3. Therefore, the fact that, least divergence is the implication of maximum entropy is verified

Table 3. Simulation parameters used for performance evaluation

Parameter	Value
Network dimension	1000 x 1000 meters ²
Number of nodes	50 nodes
Sensor radius	115 meters
Simulation Time	150 seconds
Routing protocol	DSDV
Sampling time	5 seconds
Number of samples	16 samples
Number of data formats	5 formats
Initial energy	100 joules
Transmission power	2.500 watts
Reception power	2.119 watts
Data packet	24 bytes
Entropy packet	22 bytes
Ack packet	14 bytes

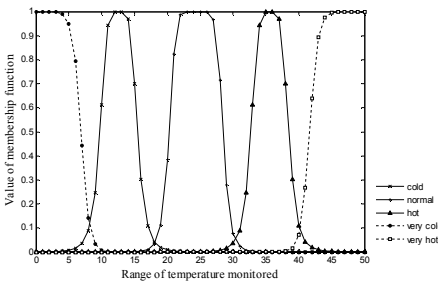


Fig. 2. The membership function for the monitored range of temperature

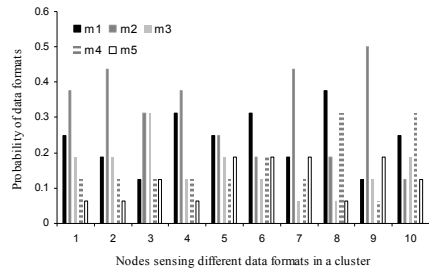


Fig. 3. Variation in the probability of data formats sensed by nodes in a cluster

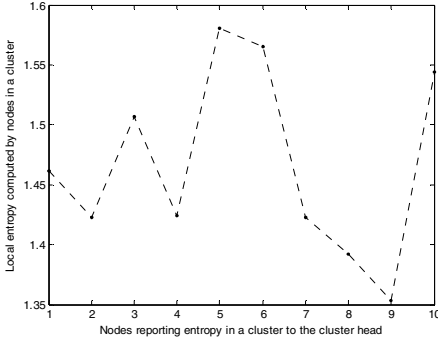


Fig. 4. The local entropy sent by sensor nodes to the cluster head in a particular cluster

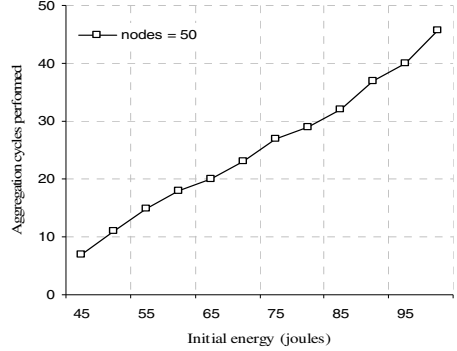


Fig. 5. The number of aggregation cycles for varying energy

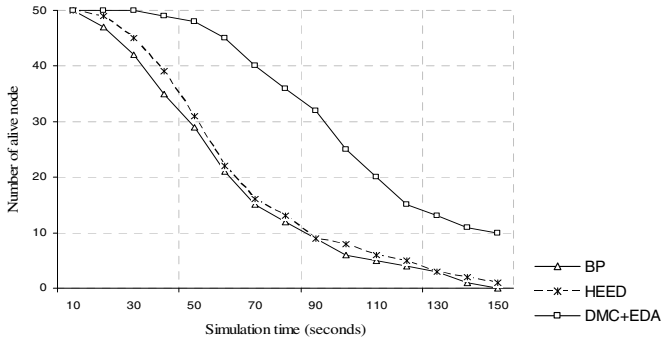


Fig. 6. Network lifetime in terms of alive nodes in the network

In figure 5, network lifetime is highlighted in terms of the number of aggregation cycles performed with a given amount of energy. The elevation in the trend apparently shows that the increase in number of aggregated samples is achieved at the minor cost of minimum packets transmission.

The graph presented in figure 6 shows the lifetime of the sensor network contributed by our proposed Divergence Measure based Clustering (DMC) + Entropy based Data Aggregation (EDA) along with HEED and BP [19-20]. The maximum message size of BP is 74 bytes which are forwarded frequently in the network for the purpose of updation of local belief by individual nodes. As a result of which we find degradation in the performance of BP. Though BP performs lesser re-clustering than HEED, but owing to the smaller size of the messages in HEED (29 bytes) it achieves marginally better results. However, our proposed DMC+EDA protocol presents best results comparatively. The packet size of our protocol is maximum 24 bytes (minimum size being 14 bytes) which reduces the transmission cost to a greater extent. Moreover, sending the entropy of nodes in the first phase of data aggregation followed by actual data (expected value) in second phase actually reduces the bulk of packets transfer, thereby increasing the network lifetime.

6 Conclusion

In this research, we have demonstrated that our proposed clustering protocol in wireless sensor network provides significant energy savings. The clustering process is purely distributed and is based on the sensed data, regardless of geographic positioning and distance measures. Furthermore, the simulations of our proposed methods have shown outperforming results. The entropy measurement facilitates the efficient selection of maximum information bearing nodes, which further makes more accurate aggregation at the cluster head. The simulation results have also clarified that our proposed data aggregation technique performs in energy efficient manner. Therefore, it can be concluded that entropy based fusion is relevant in terms of information, network lifetime as well as energy utilization.

Thus far, we have calculated the precision of sensor data on the basis of local probability model. This approach can be further extended to a global probability measure, to be performed at the cluster head. Furthermore, the working slots to aggregate data can be explicitly defined for the intermediate cycles, once the whole network is clustered. Moreover, the energy consumption in the network can also be analyzed for several aggregation cycles.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless Sensor Networks: A Survey. *J. of Computer Networks* 38, 393–422 (2002)
2. Chitnis, L., Dobra, A., Ranka, S.: Aggregation Methods for Large-Scale Sensor Networks. *ACM Transactions on Sensor Networks* 4(2), article 9, 1–36 (2008)
3. Castelluccia, C., Chan, A.C.-F., Mykletun, E., Tsudik, G.: Efficient and Provably Secure Aggregation of Encrypted Data in Wireless Sensor Networks. *ACM Transactions on Sensor Networks* 5(3), article 20, 1–36 (2009)
4. Xiong, N., Svensson, P.: Multi-Sensor Management for Information Fusion: Issues and Approaches. *Information Fusion* 3, 163–186 (2002)
5. Heinzelman, W., Chandrakasan, A., Balakrishnan, H.: An Application-Specific Protocol Architectures for Wireless Microsensor Networks. *IEEE Transactions on Wireless Communications* 1(4), 660–670 (2002)
6. Ghiasi, S., Srivastava, A., Yang, X.J., Sarrafzadeh, M.: Optimal Energy Aware Clustering in Sensor Networks. *Special Issue: Special Section on Sensor Network Technology and Sensor Data Management* 2, 258–269 (2004)
7. Srinivasan, S.M., Azadmanesh, A.: Data Aggregation in Static Adhoc Networks. In: *Third IEEE International Conference on Industrial and Information Systems, ICIIS*, pp. 1–6 (2008)
8. Wang, X., Li, J.: Precision Constraint Data Aggregation for Dynamic Cluster-Based Wireless Sensor Networks. In: *5th International Conference on Mobile Ad-hoc and Sensor Network (MSN 2009)*, pp. 172–179 (2009)
9. Zhao, F., Shin, J., Reich, J.: Information-Driven Dynamic Sensor Collaboration. *IEEE Signal Processing Magazine* 19, 61–72 (2002)
10. Commuri, S., Tadigotla, V.: Dynamic Data Aggregation in Wireless Sensor Networks. In: *IEEE 22nd International Symposium on Intelligent Control*, pp. 1–6 (2007)

11. Kong, L., Chen, Z., Yin, F.: Optimum Design of a Window Function Based on the Small-World Networks. In: IEEE International Conference on Granular Computing, p. 97 (2007)
12. Eguchi, S., Copus, J.: Interpreting Kullback–Leibler Divergence with the Neyman–Pearson Lemma. *Journal of Multivariate Analysis* 97, 2034–2040 (2006)
13. Chang, H., Yao, Y., Koschan, A., Abidi, B., Abidi, M.: Improving Face Recognition via Narrowband Spectral Range Selection Using Jeffrey Divergence. *IEEE Transactions on Information Forensics and Security* 4(1), 111–123 (2009)
14. Duch, W.: Uncertainty of Data, Fuzzy Membership Functions, and Multi-layer Perceptrons. *IEEE Transaction on Neural Networks* 20, 1–12 (2004)
15. Mandal, S.N., Choudhury, J.P., De, D., Chaudhuri, S.R.B.: Roll of Membership Functions in Fuzzy Logic for Prediction of Shoot Length of Mustard Plant Based on Residual Analysis. *World Academy of Science, Engineering and Technology* 38, 378–384 (2008)
16. Fall, K., Varadhan, K.: *The ns Manual*. The VINT Project (2009)
17. Altman, E., Jemenez, T.: *NS simulator for beginners* (2003)
18. Greis, M.: *Tutorial for Network Simulator* (2009)
19. Younis, O., Fahmy, S.: Distributed Clustering in Ad-hoc Sensor Networks: A Hybrid, Energy-Efficient Approach. *IEEE INFOCOM*, 1–12 (2004)
20. Anker, T., Bickson, D., Dolev, D., Hod, B.: Efficient clustering for improving network performance in wireless sensor networks. In: Verdone, R. (ed.) *EWSN 2008*. LNCS, vol. 4913, pp. 221–236. Springer, Heidelberg (2008)

Energy Efficient Routing Protocols for Wireless Sensor Networks Using Spatial Correlation Based Collaborative Medium Access Control

A. Rajeswari¹ and P.T. Kalaivaani²

¹ Department of ECE,
Coimbatore Institute of Technology,
Coimbatore -14

² Department of ECE,
Jansons Institute of Technology,
Coimbatore -59

rajeswari@cit.edu.in, kalaivani.p.t@gmail.com

Abstract. Wireless Sensor Networks (WSNs) are wireless networks consisting of a group of small, inexpensive nodes, which collect and disseminate significant data. Nodes in a wireless Sensor Networks have various energy and computational constraints due to their inexpensive, random method of deployment. Wireless sensor networks are utilized in a wide range of applications including battlefield surveillance, smart home environments, habitat exploration of animals and vehicle tracking. Energy efficiency is one critical issue for sensor networks because many sensors are equipped with batteries that have limited life time. In this paper, comparison of energy efficient routing protocols for Wireless Sensor Networks using spatial correlation based collaborative medium access control is proposed. A model for WSN is proposed using spatial correlation based collaborative medium access control (CC-MAC) with AODV and DSR and DSDV. Simulation results have been obtained by using Ns2. From the results, the performance of AODV is better compared to the other two protocols such as DSR and DSDV.

Keywords: Wireless Sensor Network (WSN), Spatial Correlation, Medium Access Control (MAC), Correlation based Collaborative Medium Access Control (CC-MAC), Adhoc On Demand Distance Vector Routing (AODV), Dynamic Source Routing (DSR), Destination Sequenced Distance Vector (DSDV).

1 Introduction

An emblematic sensor node in WSN consists of a wireless communication unit and a processing unit, a sensing unit, and a power supply unit. The energy source of a node is generally considered non-rechargeable. Therefore energy conservation is a major research challenge in WSN. For practical deployment, Wireless Sensor Networks must be energy efficient. Using Energy as a vital resource in WSNs, Several MAC protocols have already been proposed to get higher energy efficiency during long idle

period of the sensors[3] . Each sensor node in a WSN has three basic units, The sensing unit can sense various phenomena including light, temperature, sound and motion around its location. Process unit packetizes the sensed data and the transmission unit sends the packetized data to a base station, data receiving and data sending process are carried out by communication unit. Sensing unit consumes less energy than communication unit because most of the energy is consumed during data receiving and data sending which are carried out in communication unit. In Wireless Sensor Networks, an energy efficient medium access control protocol is required for obtaining higher energy efficiency in very difficult operating conditions, where node and link failures are common. The MAC protocol manages radio transmissions and receptions on a shared wireless medium. Therefore MAC has a very high effect on network performance and energy consumption.

2 Related Work

A brief literature survey is presented in the following section:

Due to the spatial correlation between the sensor nodes, it may not be necessary for every sensor nodes to transmit its data to sink node [1]. A draw back of the work is that only one type of phenomenon is sensed by the sensor nodes in a network [1]. A Spatial Correlation Based Medium Access Control protocol is developed to filter out redundant data. Therefore energy consumption of the network is reduced [2]. The specific requirement of wireless sensor real time MAC protocol is developed and the taxonomy for real-time MAC protocol in wireless sensor networks is proposed [3]. The concepts of MAC layer related sensor network properties and the reasons of energy waste are discussed [4]. An energy aware management of sensor networks that maximizes the lifetime of the sensors and reduces energy consumption by dynamically assigning the routes and arbitrate medium access [5]. Under minimum energy usage the spatio-temporal sampling rate of the network is determined [6].

3 Proposed Model

A wireless radio is the most energy consuming unit of a sensor node in WSN. It can operate in four different states. They are transmit, receive and idle and sleep. When the nodes are in active state, almost all the nodes consume the same energy. There are several ways to reduce the energy consumption in WSN. One way of reducing the energy consumption is, by using only a required set of nodes as active and thereby reducing redundant network traffic, decreasing packet forwarding delay, to help in avoiding packet collisions. The other way is to put few sensor nodes into sleep state and use only necessary node to be in active mode for sensing and communication.

In proposed work, the routing protocols such as AODV, DSR and DSDV are considered only to analyze the behavior of CC-MAC protocol also the proposed work considers spatial correlation which is exploited on the medium access control layer. A model for wireless sensor network has been designed and shown in Figure 1. Event source is represented as S. Total number of nodes available in an event area is N, Each node in the sensor field observes the noisy version of event information $X_i(n)$,

$S_i(n)$ is spatially correlated to event source S . Each node has to encode its observation for the purpose of reporting an event information to sink node. Sink is available at the other end which is responsible for decoding the information to get an estimated value of \hat{S} [1]. Network Simulator (Ns2) has been used for simulation of comparison of energy efficient routing protocols for wireless sensor network using Collaborative based medium access control. Parameters such as network area, initial energy, and transmitting power, receiving power, idle power, data rate and bandwidth values are tabulated in table I which have been used in the proposed model.

Table 1. Simulation Parameters

S.No	Parameters	Sizes
1.	Network Area	1500 X 1500
2.	Number of Nodes	50
3.	Packet Length	250 bytes
4.	Initial Energy	1000 joules
5.	Bandwidth	2 MHz
6.	Data Rate	1 Mbps
7.	TransmittingPower	1 mW
8.	Receiving Power	1 mW
9.	Idle Power	1 mW
10.	Sleep Power	0.001 mW

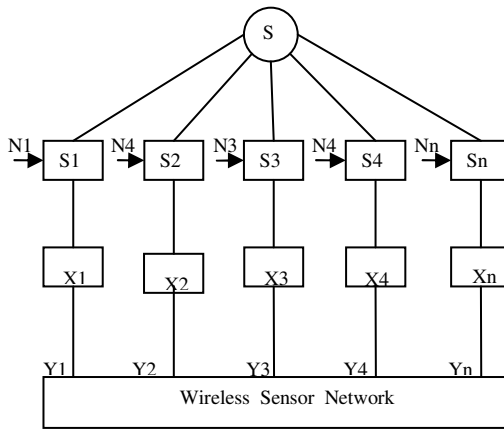


Fig. 1. Model for Wireless Sensor Networks – Encoder part

In Figure 1, The encoder part is labeled as E and the decoder part is labeled as D in Figure 2. Spatial Correlation based MAC protocol is developed (CC-MAC) to prevent redundant transmission from closely located neighbors and to regulate the medium access. Spatial correlation MAC protocol achieves high performance in terms energy.

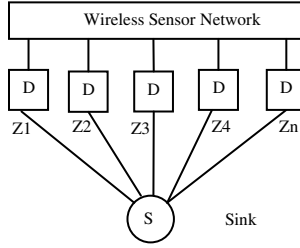


Fig. 2. Model for Wireless Sensor Networks – Decoder part

Total number of nodes available in an event area is N , where the event source is S . Each node in the sensor field observes the noisy version of event information $X_i(n)$, $S_i(n)$ is spatially correlated to event source S . Each node has to encode its observation for the purpose of reporting an event information to sink node. Sink is available at the other end which is responsible for decoding the information to get an estimated value of \hat{S} . At a time n , each observed sample is denoted as $X_i(n)$, equation is formed as,

$$X_i[n] = S_i[n] + N_i[n] \tag{1}$$

where the symbol i denotes the spatial location of the node n_i , i.e. (x_i, y_i) , $S_i(n)$ is the realization of the space-time process $s(t, x, y)$ at time $t = t_n$ and $(x, y) = (x_i, y_i)$ and $N_i(n)$ is the observation noise [1]. $\{N_i(n)\}$ is a sequence of i.i.d Gaussian random variables of zero mean and variance σ_N^2 . We assume that the noise each sensor node encounters is independent of each other, i.e., $N_i(n)$ and $N_j(n)$ are independent. Each observation $X_i(n)$ is then encoded into $Y_i(n)$ by the source-coding at the sensor node as,

$$Y_i[n] = f_i(X_i[n]) \tag{2}$$

The information is transferred to the sink through the network. The sink on the other hand decodes the received data to reconstruct an estimation \hat{S} of the source S .

$$\hat{S} = g(Y_1[n_1], \dots, Y_1[n_\tau]; \dots; Y_N[n_1], \dots, Y_N[n_\tau]) \tag{3}$$

Based on the data received from N nodes in the event area over a time period τ , the time difference between t_{n1} and $t_{n\tau}$ is expressed as,

$$\tau = t_{n\tau} - t_{n1} \tag{4}$$

4 Spatial Correlation in Wireless Sensor Networks

Spatial correlation between the sensor nodes helps to prevent redundant data during transmission. In wireless sensor network, when an event occurs in a sensor field, the nodes which are very nearer to that event area detect the event information and it is sensed by the neighborhood nodes. Every node transmits its own data to sink which is

highly correlated that results in redundant transmission. It is not requisite for all the nodes in the sensor field to send the data to the sink. Redundant transmission in a network is reduced by decreasing the selecting the subset of sensor nodes. The subset nodes efficiently transmits the data to sink. The filtration of redundant data and the maximization of network lifetime is increased by designing the efficient MAC protocol. The spatial correlation region is defined as the region in which all the sensor nodes send the readings which are similar in nature and therefore it is enough to send a single report to represent the correlation region [8].

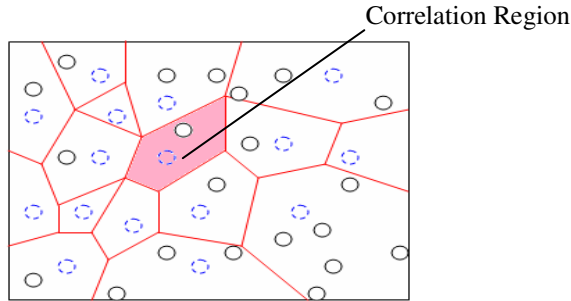


Fig. 3. Representation of Correlation Region

Fig.3. shows the representation of correlation region.

$$\text{corr} \{ S_i, S_j \} = \rho_{i,j} = K_v(d_{i,j}) = \frac{E[S_i S_j]}{\sigma_s^2} \quad (5)$$

where $d_{i,j} = \rho_{s_i} - \rho_{s_j}$ where ρ denotes the distance between nodes n_i and n_j located at coordinates S_i and S_j , respectively, and $k_v(\cdot)$ is the correlation model or covariance model. The covariance function is assumed to be non-negative and decrease monotonically with the distance. Covariance model is categorized into four types, they are,

- i) Spherical Model
- ii) Power Exponential Model
- iii) Rational Quadratic Model
- iv) Matern Model

In this paper Power exponential model has been considered [1] . The covariance function is given by

$$K_v^{PE}(d) = e^{(-d/\theta_1)^{\theta_2}} ; \theta_1 > 0, \theta_2 \in (0, 2] \quad (6)$$

From the above equation (6), K_v represents covariance function and PE represents power exponential model. For $\theta_2 = 1$, the model becomes exponential, while for $\theta_2 = 2$ squared exponential. The parameter θ_1 is important because it controls the relationship between the distance $d_{i,j}$ and $\rho_{i,j}$.

4.1 Distortion Constraint

Distortion is one of the reliability constraint. The distortion increases when the sensor nodes fail to report the event from within the defined correlation region. The correlation region is changed dynamically according to the observed reliability. At sink node, the distortion is given by,

$$D = E[d(S, \hat{S})] \tag{7}$$

where D is the Distortion value and S is the event and \hat{S} is the estimated value of S. Mean-squared error is used as the distortion metric

$$D(M) = E[(S - \hat{S}(M))^2] \tag{8}$$

D(M) shows the distortion achieved at the sink as a function of number of nodes M that send information to the sink and correlation coefficients $\rho_{(i,j)}$ and $\rho_{(s,i)}$ between nodes n_i and n_j , and the event source S and node n_i , respectively [1]. It is denoted as,

$$D(M) = \sigma_s^2 - \frac{\sigma_s^4}{M(\sigma_s^2 + \sigma_N^2)} \left(2 \sum_{i=1}^M \rho_{(s,i)} - 1 \right) + \frac{\sigma_s^6}{M^2(\sigma_s^2 + \sigma_N^2)^2} \sum_{i=1}^M \sum_{j \neq i}^M \rho(i,j) \tag{9}$$

4.2 Simulation Results for Distortion Constraint

The wireless sensor network is implemented in Ns2 with fifty nodes in random deployment. Representative nodes are selected randomly among 50 nodes and the distortion function is calculated according to the locations of these nodes. From the simulation, the correlation coefficient between the sensor nodes and the event and also the average distortion are calculated, the distribution of the distortion for each number of representative nodes is shown in figure 4 & 5. For the analysis, $\theta_2 = 1$ and $\theta_1 = \{10, 50, 100, 500, 1000, 5000, 10000\}$ have been considered in the covariance model for the covariance function.

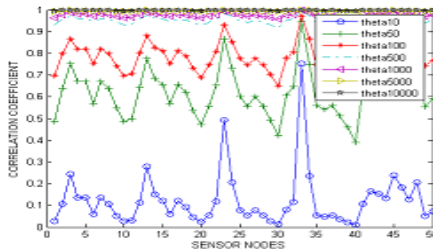


Fig. 4. Correlation coefficient between the sensor nodes and the event

Figure 4 shows the correlation coefficient between the sensor nodes and the event. Figure 5 shows the distortion according to changing the representative node number.

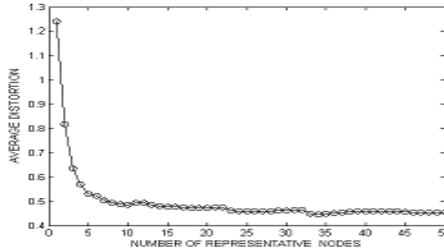


Fig. 5. Distortion according to changing representative node number

As shown in Figure 4 & 5, the achieved distortion stays relatively constant when the number of representative nodes is decreased to 15. This behavior is due to the highly redundant data sent by the sensor nodes that are close to each other. In addition, with increasing θ_1 , the observed event distortion decreases since close nodes become less correlated with increasing θ_1 .

5 Collaborative Based Medium Access control (CC-MAC)

When a specific source node, n_i , transmits its event record to the sink, all of its correlation neighbors have redundant information with respect to the distortion constraint, D_{\max} . This redundant information, if sent, increases the overall latency and contention within the correlation region and results in wasting of energy in WSN. The proposed work deals with collaborative based MAC protocol (CC-MAC). CCMAC aims to prevent the transmission of such redundant information and prioritize the forwarding of filtered data to the sink.

In WSN, the sensor nodes have the dual functionality of being both data originators and data routers. Hence, the medium access is performed for two reasons:

- Source function
- Router function

Source nodes with event information perform medium access in order to transmit their packets to the sink and the router function is used to forward the packets received from other nodes to the next destination in the multi-hop path to the sink [1].

CC- MAC protocol contains two components, Initiate MAC and Route MAC corresponding to the source and router functionalities. Initiate MAC (I-MAC), filters out the correlated records. Route MAC (R-MAC) ensures prioritization of route-thru packets. Both I-MAC and R-MAC use a CSMA/CA based medium access control with appropriate modifications. The information about correlation formation is embedded inside the RTS/CTS/DATA/ACK packets. Each node is informed about the

correlation information about a node using these packets. A node performs I-MAC when it has to transmit its sensor reading to the sink, while R-MAC is performed when a node receives a packet and tries to forward it to the next hop.

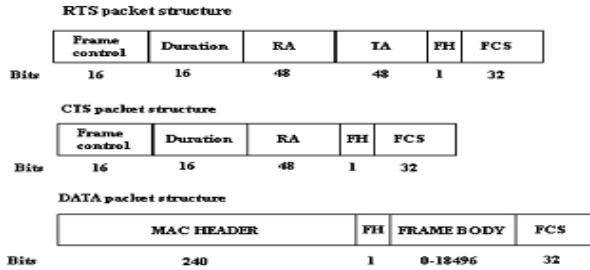


Fig. 6. Structure for RTS, CTS, Data Packet

5.1 Iterative Node Selection Algorithm

To select the representative node in a sensor field, the INS algorithm is used [1]. Vector Quantization (VQ) method is adapted to analyze correlated region in WSN field. It is named as voronoi region.

Pseudocode for the proposed model is

- Start Calculating $D(M)$ by setting $M=N$
- Check if $D(M) < D_{max}$
- Run VQ
 - (i) Generate Multiple Topologies with M nodes
 - (ii) Locations of Sensors nodes will be changed for each topology
 - (iii) Calculate $r(s,i), r(i,j)$ using q_1
 - (iv) Calculate $D(M)$
- Repeat the above steps while $D(M) < D_{max}$
- End the result by calculating the values of M^*

$$M^* = \operatorname{argmin} \{ D(M) < D_{max} \}$$

6 Routing Algorithms

Two types of routing protocols are Proactive protocols and reactive protocols [7]. Proactive Routing Protocols maintain fresh list of destinations and their routes by periodically distributing routing tables throughout the network. The main disadvantages are respective amount of data for maintenance and slow reaction on restructuring and failures. Reactive Protocols are also known as On-demand protocols. These type of protocol find a route on demand by flooding the network with route request packets. Three different routing protocols are considered to analyze the behavior of CC – MAC Protocol. They are AODV and DSR, DSDV. Dynamic Source Routing is shortly known as DSR and Adhoc On Demand Distance Vector Routing is known as AODV, DSDV is known as destination sequenced distance vector routing protocol. DSDV is a proactive table driven protocol, The basic improvements made include freedom from loops in routing tables, more dynamic and less convergence time [10]. DSDV is an elementary and moderately less complex protocol which is a suitable fit for less dense network i.e. targeted to function exquisitely on small node density. DSDV does work, essentially, by sharing routing information with neighboring nodes, which is stacked away by each node in the form of tables. DSR uses source routing. Timer activities are not directly involved in DSR also certain multiple route cache entries per destination are used. The main feature of DSR is source routing. i.e, the sender knows the complete hop-by-hop route to the destination. These routes are stored in a route cache. The data packets carry the source route in the packet header. Ad-hoc on demand vector is an on demand routing protocol which is used to find a route between the source and destination node as needed. AODV is preferable mostly because,

- Uses sequence numbers to avoid loops
- Quick adaptation to dynamic networks
- Unicasting and multicasting is possible
- Avoids counting to infinite problems

AODV is also appropriate for working in restrictive environments. It has the ability to intercommunicate with the endpoints which cannot be accessed directly.

7 Simulation Results

Energy efficient routing protocols for wireless sensor networks using spatial correlation based collaborative medium access control has been implemented using Ns2. The parameters such as end to end delay, packet delivery ratio and packet loss, energy consumption have been compared with AODV, DSR & DSDV

7.1 Comparison of End to End Delay with AODV, DSR, DSDV

End to End delay is defined as the ratio between sum of individual data packet delay to the total number of data packets delivered.

End to end delay = [(Sum of Individual data packet delay) / (Total number of data Packets delivered)]

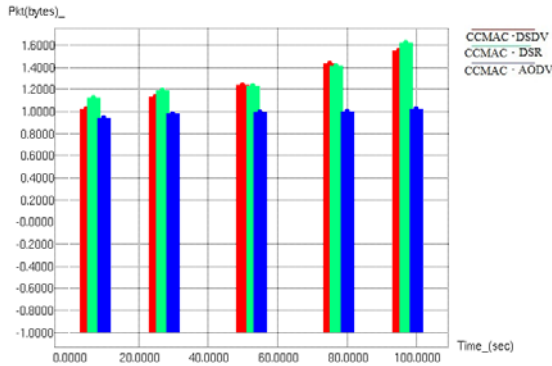


Fig. 7. End to End delay

Fig.7 shows the variation of end to end delay against time.

Table 2. End to End delay

S.No	Time in (sec)	Protocols	End to End Delay (ms)
1.	5	DSDV	1.02
2.	25		1.13
3.	50		1.24
4.	75		1.43
5.	95		1.55
1.	7	DSR	1.12
2.	27		1.19
3.	53		1.23
4.	77		1.41
5.	97		1.62
1.	10	AODV	0.94
2.	30		0.98
3.	55		0.99
4.	80		1.00
5.	100		1.02

It is observed from Figure 7 and Table 2 that as simulation time increases, the end to end delay with DSR and DSDV is higher than that of AODV. This is because DSDV is applicable for less dense network. DSR is designed for multihop wireless networks but AODV is a routing protocol which has ability to create a route to destination only on demand.

7.2 Comparison of Packet Delivery Ratio with AODV, DSR, DSDV

Packet Delivery ratio is the percentage of the ratio between total number of data packets successfully delivered to the total number of data packets sent.

$$Pdr = \frac{\text{(Total number of data packets successfully delivered)}}{\text{(Total number of data packets sent)}} * 100\%$$

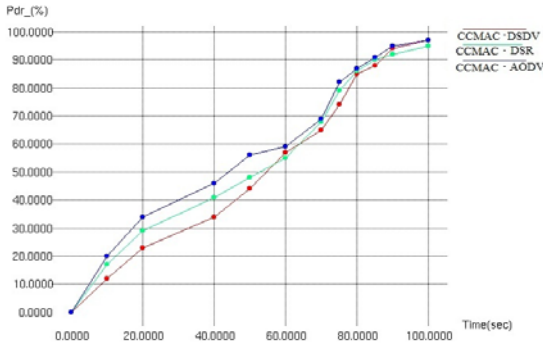


Fig. 8. Packet Delivery Ratio

Fig.8. shows the variation of packet delivery ratio against time

Table 3. Packet Delivery Ratio

S.No	Time in (Sec)	Packet Delivery Ratio (%)		
		DSDV	DSR	AODV
1.	0	0	0	0
2.	10	12	17	20
3.	20	23	21	34
4.	40	34	41	46
5.	50	44	48	56
6.	60	57	55	59
7.	70	65	68	69
8.	75	74	79	82
9.	80	85	86	87
10.	85	88	90	91
11.	90	94	92	95
12.	100	97	95	97

It is observed from Figure 8 and Table 3 that at simulation time of 10 sec, AODV has the highest (100%) of packet delivery ratio, DSR has 15% lesser packet delivery ratio compared to that of AODV and DSDV has 40% lesser packet delivery ratio compared to that of AODV. This is because DSR uses multiple route cache entries per destination and in DSDV, periodically advertises a node's interconnection topology with the other nodes. AODV uses traditional routing tables, one entry per destination.

7.3 Comparison of Packet Loss with AODV, DSR, DSDV

Packet Loss is obtained by subtracting the number of packets sent by the source to the number of packets received by sink.

$$\text{Packet Loss} = (\text{Number of packets sent by source}) - (\text{Number of packets received by sink})$$

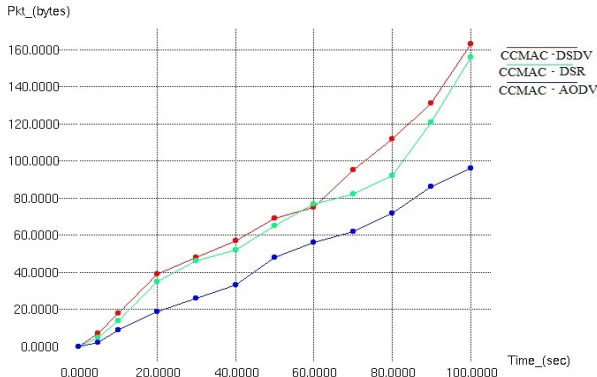


Fig. 9. Packet Loss

Fig. 9. shows the variation of packet loss against time

Table 4. Packet Loss

S.No	Time in (Sec)	Packet Loss (Bytes)		
		DSDV	DSR	AODV
1.	0	0	0	0
2.	5	7	5	2
3.	10	18	14	9
4.	20	39	35	19
5.	30	48	46	26
6.	40	57	52	33
7.	50	69	65	48
8.	60	75	77	56
9.	70	95	82	62
10.	80	112	92	72
11.	90	131	121	86
12.	100	163	156	96

It is observed from Fig 9 and Table 4 that at simulation time of 80 sec, AODV shows the lowest packet loss and DSR has a packet loss about 12% greater than that of AODV, DSDV has a packet loss about 15% greater than that of AODV . This is because AODV uses a broadcast route discovery algorithm and then Unicast route reply message. Also, AODV has the ability to provide Unicast, Multicast and broadcast communication.

7.4 Comparison of Energy Consumption with AODV, DSR, DSDV

Energy Consumption is defined as the ratio between sum of energy expended by each node to the total number of data packets delivered.

$$\text{Energy Consumption} = \left[\frac{\text{Sum of energy expended by each node}}{\text{Total number of data packets delivered}} \right]$$

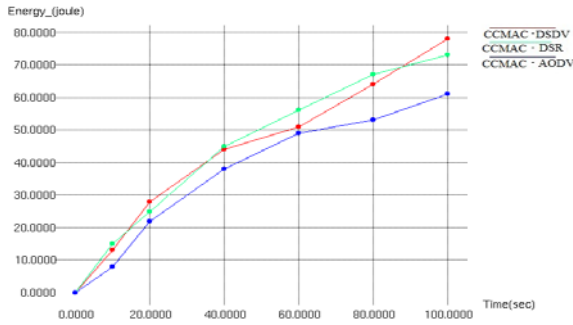


Fig. 10. Energy Consumption

Fig. 10. shows the variation of energy consumption against time

Table 5. Energy Consumption

S.No	Time in (Sec)	Energy Consumption in (joules)		
		DSDV	DSR	AODV
1.	0	0	0	0
2.	10	13	15	8
3.	20	28	25	22
4.	40	44	45	38
5.	60	51	56	49
6.	80	64	67	53
7.	100	78	73	61

It is observed from Fig 10 and Table 5 that at simulation time of 100 sec, AODV shows the lowest energy consumption. DSR has energy consumption about 12% greater than that of AODV and DSDV has an energy consumption of 13% greater than that of AODV. This is because DSR needs support from MAC layer to identify the link failure. In DSDV, an additional table is maintained by the mobile nodes that stores the data sent in the incremental routing information packets but in AODV, a route between the source and destination node is found only when they are needed.

From figure 7,8,9,10 and table 2,3,4,5, it is observed that the proposed energy efficient routing protocols for wireless sensor network using collaborative based medium access control with AODV performs better with higher packet delivery ratio, lesser end to end delay and lesser packet loss, lesser energy consumption than that of DSR and DSDV algorithms.

8 Conclusion and Future Work

Energy efficient routing protocols for wireless sensor network using collaborative based medium access control is proposed in this paper. The spatial correlation

between the sensor nodes is considered in the proposed algorithm. Using Ns2, wireless sensor networks are simulated and the proposed algorithm is implemented with spatial correlation based AODV, DSDV, DSR. The spatial resolution of nodes are controlled by deactivating the redundant nodes. The performance of CC – MAC is analyzed using Vector Quantization method. It yields better performance to achieve energy efficiency. Parameters such as, End to End delay and Packet drop rate and Packet delivery ratio and Energy Consumption are taken into account. Among the routing protocols AODV gives better results than DSR, DSDV. By reducing the redundant data from redundant nodes the spatial correlation based method proves that it is the suitable technique to attain energy efficient in WSN . In future work , field of grid and cluster head based algorithm will be detailed. Also, more number of routing algorithms will be considered to accomplish the energy efficient operation in a better way. In future, various network topology may be considered to proceed the work in different directions.

References

1. Mehmet, C., Vuran, I.F., Akyildiz: Spatial Correlation – Based Collaborative Medium Access Control in Wireless Sensor Networks. *IEEE / ACM Transaction on Networking* 14(2) (2006)
2. Zheng, G., Tang, S.: Spatial Correlation Based MAC Protocol for Event - Driven Wireless Sensor Networks. *Journal of Networks* 6(1) (2011)
3. Teng, Z., Kim, K.I.: A survey on Real-Time MAC protocols in Wireless Sensor Networks. *Journal of Communication and Networks* 2, 104–112 (2010)
4. Demirkol, I., Ersoy, C., Alagoz, F.: MAC Protocols for Wireless Sensor Networks: A survey. *IEEE Communication Magazine* (2006)
5. Khaled, A., Arisha, M.A., Youseef, M.F., Younis: Energy Aware TDMA-Based MAC for Sensor Networks. In: *Proceedings of the IEEE Integrated Management of Power Aware Communications, Computing and Networking, IMPACCT* (2002)
6. Vuran, M.C., Akan, O.B., Akyildiz, I.F.: Spatio-temporal correlation: theory and applications for wireless sensor networks. *Comput. Networks Journal* 45(3), 245–259 (2004)
7. Samir, R., Das, C.E., Perkins, E.M., Royer: Performance Comparison of Two On – demand Routing Protocols for Adhoc Networks. In: *Proceedings of the IEEE Conference on Computer Communications (INFOCOM)*, Tel Aviv, Israel (2000)
8. Ghalib A., Shah, Muslim Bozyigit.: Exploiting Energy aware Spatial Correlation in Wireless Sensor Networks. In: *2nd International Workshop on Software for Sensor Networks, Sensor Wave 2007* (2007)
9. Linde, Y., Buzo, A., Gray, R.M.: An algorithm for vector quantizer design. *IEEE Trans. Commun.* COM-28(1), 84–95 (1980)
10. Radhika, D., Joshi, P.P., Rege: Energy Aware Routing in Ad Hoc Networks. In: *6th WSEAS International Conference on Circuits, Systems, Electronics, Control & Signal Processing*, Cairo, Egypt (2007)
11. The Network Simulator—ns-2, <http://www.isi.edu/nsnam/ns/index.html>

Palmprint Authentication by Phase Congruency Features

Jyoti Malik¹, G. Sainarayanan², and Ratna Dahiya¹

¹ Electrical engineering Department, National Institute of Technology, Kurukshetra
jyoti_reck@yahoo.com

ratna_dahiya@yahoo.co.in

² New Horizon College of Engineering, Bangalore, India
sai.jgk@gmail.com

Abstract. The abstract should summarize the contents of the paper and should Palmprint recognition is an effective biometric authentication method to automatically identify a person's identity. In this paper, phase congruency method is proposed to extract features from a palm-print image for authentication. The phase congruency is one of the promising methods to analyze the image as it is invariant to image contrast and therefore can extract reliable features under varying illumination conditions. The hand image is pre-processed to get the desired Region of Interest (ROI) / palmprint. The palmprint features are extracted by phase congruency method and are stored in feature vector. Euclidean Distance similarity measurement method is used to compare the similarity/dissimilarity between two feature vectors. Experiments were developed on a database of 600 images from 100 individuals, with five image samples per individual for training and one image sample per individual for testing.

Keywords: Palmprint, Palmprint Authentication, Phase Congruency, Euclidean distance.

1 Introduction

Personal identification using biometric methods is becoming popular in today's automated world. Biometric authentication methods utilize automated techniques to authenticate a person's identity based on his/her behavioral/physiological characteristics [1-3]. The main aim in biometric authentication is to find a suitable and effective biometric feature which can find the similarity or dissimilarity between individuals. The stable lines in the palmprint make it one of the most reliable biometric features for authentication.

Palmprint is universal, unique, permanent, collectible, consistent, comparable, inimitable and tamper-resistant biometric method. Palmprint has several features like geometry features, line features, point features, texture features etc [4-16]. In this paper, line feature is analyzed for palmprint recognition. Line feature includes principal lines, wrinkles and ridges. There are several conventional methods that can extract the palmprint lines like Sobel operator, Prewitt operator, and Canny's technique etc. [17, 18]. These methods works better on images captured under controlled illumination conditions and are highly sensitive to texture lines etc. A more

promising phase congruency line detection method is proposed. Phase congruency method is less sensitive to image distortions and poor illumination conditions. The phase congruency method performs better than the real wavelet transform method for palmprint authentication [9].

The following section of the paper is organized as follows: Section 2 describes the basic of palmprint authentication system. In section 3 the phase congruency-based feature extraction technique is proposed. Section 4 presents feature matching by Euclidean Distance similarity measurement procedure. The experimental results are presented in Section 5, while section 6 concludes the paper.

2 Palmprint Authentication System

In this paper, the palmprint authentication system is divided in following two subsystems:

- (a) Pre- Authentication System
- (b) Authentication System

In Pre-authentication system, a database of Phase Congruency features is prepared. In addition, Reference threshold value is also identified and stored in database. These values will be used in Authentication system.

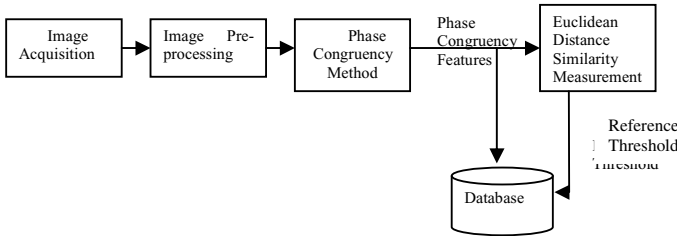


Fig. 1. Palmprint Pre-Authentication system

In Authentication system, the authenticity of a person is identified with the help of Reference threshold value stored in Pre-authentication system database.

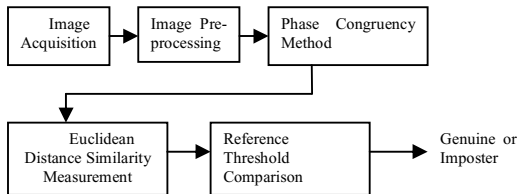


Fig. 2. Palmprint Authentication System

3 Feature Extraction Using Phase Congruency Features

The main aim of feature extraction is to get the desired features, here line feature from the palmprint image. Line feature includes principal lines, wrinkles and ridges. Line-feature extraction technique which extracts a set of phase congruency features from palmprint image is proposed. There exist several palmprint recognition systems that use line and gradient-based feature extraction methods like Sobel operators [17], Canny [18], line directional detectors calculates the points of high intensity gradients to extract the line features in different directions. All these palmprint recognition methods are based on intensity gradients and therefore got affected by the image contrast and brightness. Here, phase congruency model for line feature extraction is proposed that is invariant to changes in image brightness and contrast.

The phase congruency model [19] finds out the points in the palmprint where log-Gabor filter response over various scales (s) and orientations (o) is maximal in phase. The 2D phase congruency $PC_{2D}(x)$ is defined as:

$$PC_{2D}(x) = \frac{\sum_s \sum_o W_o(x) \lfloor A_{so}(x) \left(\cos(\phi_{so}(x) - \overline{\phi}_o(x)) - \left| \sin(\phi_{so}(x) - \overline{\phi}_o(x)) \right| \right) - T_o \rfloor}{\sum_s \sum_o A_{so}(x) + \epsilon} \quad (1)$$

where x is the pixel location in the spatial domain. $W_o(x)$ is weighing function of phase congruency by frequency spread at orientation o . $A_{so}(x)$ denotes the amplitude of the grey scale palmprint image. $\phi_{so}(x)$ denotes the phase response of palmprint image at scale s and orientation o of log-Gabor filter. $\overline{\phi}_o(x)$ represents the mean phase angle at orientation o . ϵ is small constant which prevents division by zero. $\lfloor \rfloor$ symbol denote that the enclosed quantity is equal to itself when its value is positive and zero otherwise. T_o is the estimated noise energy at orientation o .

The sample of Phase Congruency feature extracted from palmprint image is shown in Fig.3.

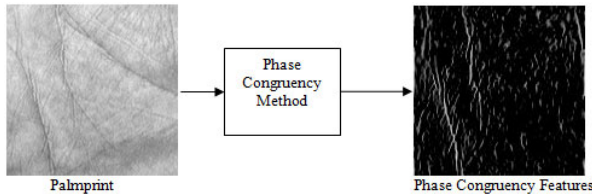


Fig. 3. Feature Extracted by Phase Congruency method on palmprint image

The Phase Congruency method is applied on the palmprint image and Phase Congruency Features extracted. The number of phase congruency features images depends on the number of orientations considered. Here the number of orientation considered is six. The phase congruency features (PCF) extracted and the binarized phase congruency features images are shown in Fig. 4.

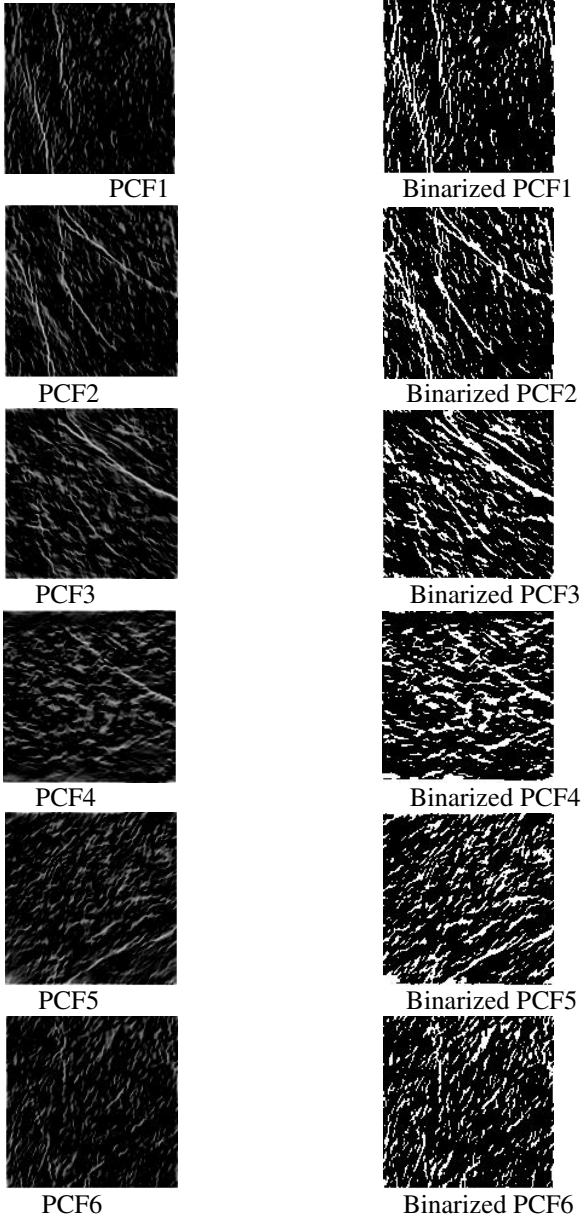


Fig. 4. Phase Congruency Features and Binarized Phase congruency features

Once a palm-print image is transformed by phase congruency model, the final feature vector FV is the energy at the considered value of orientation, constructed by computing the square and summation of the Phase Congruency image pixels (PCIP).

$$E_i = \sum_{x=1}^{rows} \sum_{y=1}^{columns} (PCF_i[x, y])^2 \tag{2}$$

$$FV = [E_1, E_2, \dots, E_N], N = 6 \tag{3}$$

4 Feature Matching

A matching algorithm describes the degree of similarity between two feature vectors. In a palm-print recognition system operating in verification mode the feature vector extracted from the given input palm-print image by the phase congruency method is compared (or matched) with the template associated with the claimed identity. The template is constructed during the enrolment stage. Here, Euclidean Distance similarity measure is used to produce the matching score.

4.1 Euclidean Distance Similarity Measurement

Euclidean distance involves computation of square root of the sum of the squares of the differences between two feature vectors as shown in (4).

$$ED = \sqrt{\sum (FV_1 - FV_2)^2} \tag{4}$$

In palmprint authentication, Euclidean distance is defined as the square root of the sum of the squares of the differences between two feature vectors, FV_{DB} , FV as shown in equation (5).

$$ED = \sqrt{\sum_{k=1}^m (FV_{DB,k} - FV_k)^2} \tag{5}$$

where FV_{DB} , FV are feature vectors from the database and the person came for authentication. ‘m’ is the number of orientation.

4.2 Reference Threshold

Euclidean Distance value “0” signifies both feature vectors are exactly same and a value approaching “0” signifies both feature vectors belongs to same hand. If the value of matching score (or Euclidean distance) defined by equation (5) is less than reference threshold value, then the feature vectors are considered to be from same hands (being genuine or authentic), otherwise different hands (being false or unauthentic).

The hand image samples are divided into two groups G1 and G2.

G₁ group

$$P_1 = [I_1, I_2, \dots, I_{(M-1)}], P_2 = [I_1, I_2, \dots, I_{(M-1)}], \dots \tag{6}$$

$$P_N = [I_1, I_2, \dots, I_{(M-1)}]$$

G₂ group

$$P_1 = [I_M], P_2 = [I_M], \dots, P_N = [I_M] \tag{7}$$

Where P_i denotes ith person in group G₁, G₂, I_j denotes the jth palm image in group G₁, G₂.

Table 1. Matching in group G₁ among person P₁

i \ j	1	2	3	M-1
1	X	ED ₁₂	ED ₁₃	ED _{1(M-1)}
2	ED ₂₁	X	ED ₂₃	ED _{2(M-1)}
⋮	⋮	⋮	⋮	⋮	⋮
⋮	⋮	⋮	⋮	⋮	⋮
M-1	ED _{(M-1)1}	ED _{(M-1)2}	ED _{(M-1)3}	X

In group G₁, each hand feature vector in P₁ is matched with all other (M-1) hands feature vector by Euclidean Distance similarity measurement method. The matching values are approaching “0” and are stored in threshold array.

$$TA_1 = \left[ED_{12}, ED_{13}, \dots, ED_{1M-1}, ED_{21}, ED_{23}, \dots, ED_{2M-1}, \dots, \dots, ED_{(M-1)1}, ED_{(M-1)2}, \dots, ED_{(M-1)(M-2)} \right] \tag{8}$$

Similarly, all N hand image samples matching results are stored in Threshold array (TA).

$$T_A = TA_1 + TA_2 + \dots + TA_N \tag{9}$$

The minimum and maximum of matching values are found out from the threshold array (TA₁, TA₂,.....TA_N) for each individual as shown in equation (10).

$$\left. \begin{aligned} T_{AiMIN} &= \min(T_{Ai}) \\ T_{AiMAX} &= \max(T_{Ai}) \end{aligned} \right\}_{i=1, \dots, N} \tag{10}$$

The accuracy of the system is identified by matching group G₂ samples with group G₁ samples using threshold values stored in threshold array. Finally, a threshold value is chosen where FAR and FRR is minimum, this value is called Reference threshold.

In real time authentication system, if a person’s hand is compared with the samples present in the database, the authenticity depends on the matching score. If matching score (Distance value) is less than reference threshold, the person is considered to be genuine (present in the database) otherwise imposter.

5 Experimental Results and Analysis

A database of 600 palmprint images from 100 palms with 6 samples for each palm is taken from PolyU palmprint database [22].

5.1 Palmprint Authentication System

The palmprint database is divided into two groups, first group (G_1) consists of 100 persons with each person having 5 palm sample images to train the system, and second group (G_2) contains 100 persons with each person having one palm image different from the first group images. Second group is used as testing sample. The image size is 284×384 pixels.

G_1 group

$$P_1 = [I_1, I_2, I_3, I_4, I_5], P_2 = [I_1, I_2, I_3, I_4, I_5], \dots, P_{100} = [I_1, I_2, I_3, I_4, I_5]$$

In G_1 group each hand P_i contains 5 sample image I_{1-5} .

G_2 group

$$P_1 = [I_6], P_2 = [I_6], \dots, P_{100} = [I_6]$$

In G_2 group each hand P_i contains only sample image I_6 .

Image is pre-processed to get the region of interest. Pre-processing includes image enhancement, image binarization, boundary extraction, cropping of palmprint/ROI. The ROI size is 64×64 pixels. Sample of ROI is shown in figure 5.

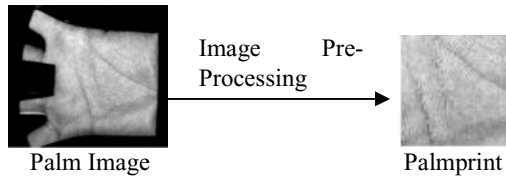


Fig. 5. Sample of ROI

Feature extraction is done by Phase Congruency method to get the phase congruency features from the palmprint image. The feature vector contains the energy of the phase congruency image pixels. Feature vector of all hand images samples is calculated and stored in database.

Euclidean Distance is used as a similarity measurement method for feature matching.

5.2 Reference Threshold Calculation

In group G_1 , each hand feature vector in P_1 is matched with all other 4 hands feature vector by Euclidean Distance similarity measurement method. Similarly, all 100 hand image samples 2000 matching values are stored in Threshold array (TA).

$$T_A = TA_1 + TA_2 + \dots + TA_{100}$$

$$\left. \begin{aligned} T_{AiMIN} &= \min(T_{Ai}) \\ T_{AiMAX} &= \max(T_{Ai}) \end{aligned} \right\}_{i=1, \dots, 100}$$

The maximum and minimum values are found out from threshold array.

The FAR and FRR are values are plotted with respect to threshold range values. From the graph the value of reference threshold is chosen where FAR and FRR are minimum. The authentication system accuracy is 98.1%. Plot of FAR and FRR is

shown in figure 6. The plot between accuracy and threshold is shown in Fig. 7. The values of FAR, FRR, Accuracy and Comparison time is tabulated in table II for scale 4 and orientation 6. The relation of number of orientation with comparison time is tabulated in table III and plotted in Fig. 8.

Table 2. Comparison of Reference Threshold, FAR and FRR with Accuracy

Reference Threshold	FAR	FRR	Accuracy
1.06E+02	4.76E-02	2.02E-03	98.1
1.95E+02	8.62E-02	1.82E-03	97.5
2.84E+02	1.56E-01	1.52E-03	95.6
3.73E+02	1.84E-01	1.42E-03	92.1
4.63E+02	2.05E-01	1.12E-03	90.7
5.52E+02	2.03E-01	1.12E-03	89.7

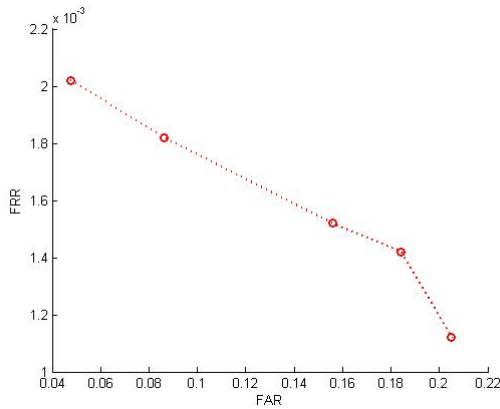


Fig. 6. FAR Vs FRR

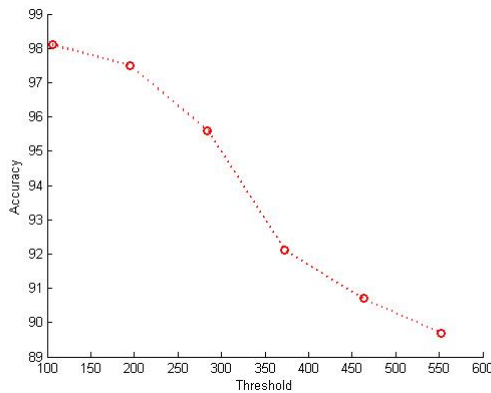
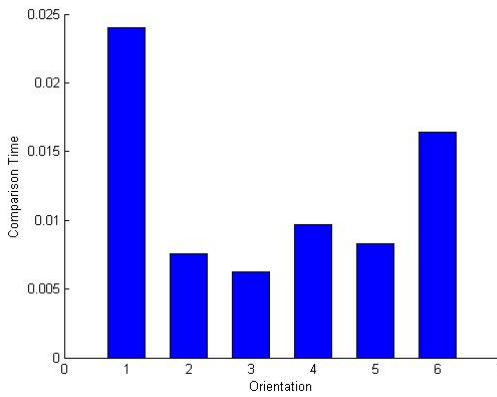


Fig. 7. Accuracy Vs Threshold

Table 3. Comparison between Time and Window Size

Number of orientation	Comparison Time
1	0.024
2	0.00755
3	0.00625
4	0.00969
5	0.00833
6	0.0164

**Fig. 8.** Comparison Time Vs Orientation

The comparison of different methods with respect to accuracy and FAR FRR is tabulated in table 4.

Table 4. Comparison between Different Methods

Method	FAR	FRR	Accuracy
Directional operator [8]	0.6	0.6	97.81
DLEF [9]	2.08	2.08	97.50
Sobel Code [23]	1.82	8.5	94.84
Proposed Approach	0.0476	0.00202	98.1

6 Conclusion

In this paper, we have presented a promising approach of phase congruency for the extraction of discriminative palm-print features at different orientations. The phase

information is obtained by using phase congruency method for detecting the edges of palmprint lines. The energy computed from the phase congruency image pixels is used for feature matching. The authentication system accuracy is 98.1 %. Experimental results clearly show that phase congruency methodology has the ability to discriminate similar palmprints. The future work is to explore the quality of phase congruency for different illumination and contrast.

References

1. Jain, A.K., Ross, A., Prabhakar, S.: An introduction to biometric recognition. *IEEE Trans. Circuits Syst. Video Technol.* 14(1), 4–20 (2004)
2. Pavešić, N., Ribarić, S., Ribarić, D.: Personal authentication using hand-geometry and palmprint features – the state of the art. In: *Proc. Workshop: Biometrics – Challenges Arising from Theory to Practice*, Cambridge, pp. 17–26 (2004)
3. Kumar, A., Wong, D.C.M., Shen, H.C., Jain, A.K.: Personal authentication using hand images. *Pattern Recognit. Lett.* 27(13), 1478–1486 (2006)
4. Kumar, Zhang, D.: Combining fingerprint, palmprint and handshape for user authentication. In: *Proceedings of ICPR*, vol. 4, pp. 549–552
5. Kumar, A., Zhang, D.: Personal authentication using multiple palmprint representation. *Pattern Recognit.* 38(10), 1695–1704 (2005)
6. Zhang, D., Kong, W.K., You, J., Wong, M.: Online palmprint identification. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(9), 1041–1050 (2003)
7. Wu, X.Q., Wang, K.Q., Zhang, D.: Wavelet based palmprint recognition. In: *Proc. Int. Conf. Machine Learning and Cybernetics*, Beijing, pp. 1253–1257 (2002)
8. Wu, X.Q., Wang, K.Q., Zhang, D.: Palm line extraction and matching for personal authentication. *IEEE Transactions on Systems, Man and Cybernetics* 36(5), 978–986 (2006)
9. Wu, X.Q., Wang, K.Q., Zhang, D.: Palmprint Recognition using Directional Line Energy feature. In: *Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004* (2004)
10. Wu, X., Wang, K., Zhang, D.: Palmprint authentication based on orientation code matching. In: Kanade, T., Jain, A., Ratha, N.K. (eds.) *AVBPA 2005. LNCS*, vol. 3546, pp. 555–562. Springer, Heidelberg (2005)
11. Huang, D.S., Jia, W., Zhang, D.: Palmprint verification based on principal lines. *Pattern Recognit.* 41(4), 1316–1328 (2008)
12. Lu, G., Zhang, D., Wang, K.: Palmprint recognition using eigenpalm features. *Pattern Recognit. Lett.* 24(9-10), 1463–1467 (2003)
13. Shang, L., Hunag, D.S., Du, J.X., Zheng, C.H.: Palmprint recognition using FastICA algorithm and radial basis probabilistic neural network. *Neurocomputing* 69(13-15), 1782–1786 (2006)
14. Hud, Fengg, Zhou: Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition. *Pattern Recognit.* 40(1), 339–342 (2007)
15. Savič, T., Pavešić, N.: Personal recognition based on an image of the palmar surface of the hand. *Pattern Recognit.* 40(11), 3152–3163 (2007)
16. Han, C.C., Cheng, H.L., Lin, C.L., Fan, K.C.: Personal authentication using palm-print features. *Pattern Recognit.* 36(2), 371–381 (2003)
17. Marr, D., Hildreth, E.C.: Theory of edge detection. *Proc. R. Soc. Lond. B* 207, 187–217 (1980)

18. Canny, J.F.: A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 8(6), 112–131 (1986)
19. Kovese, P.: Image features from phase congruency. *Videre J. Comput. Vis. Res.* 1(3), 1–26 (1999)
20. Punsawad, Y., Wongsawat, Y.: Palmpoint Image Enhancement Using Phase Congruency. In: *Proceedings of the 2008 IEEE International Conference on Robotics and Biometrics*, Bangkok, Thailand, February 21–26 (2009)
21. Struc, V., Pavesic, N.: Phase congruency features for palmpoint verification. *IET Signal Processing* 3(4), 258–268 (2008)
22. The PolyU Palmpoint Database,
<http://www4.comp.polyu.edu.hk/biometrics/>
23. Wong, K.Y.E., Sainarayanan, G., Chekima, A.: Palmpoint Identification Using SobelCode. In: *Malaysia-Japan International Symposium on Advanced Technology 2007*, Kuala Lumpur, Malaysia (2007)

Design and Implementation of 3D DWT for 4D Image Based Noninvasive Surgery

P.X. Shajan¹, N.J.R. Muniraj², and John T Abraham³

¹ Research Scholars, KarpagamUniversity, Coimbatore, India
shajanpx@gmail.com

² Professor & Head, Department of ECE, Karpagam college of Engineering, Coimbatore, India
njrmuniraj@yahoo.com

³ Director, Mount Zion Engineering College, Kerala, India
johntabraham@yahoo.com

Abstract. In this paper a 3D DWT algorithm is proposed for 4D imaging. In this proposed algorithm original image is preprocessed to 8X8X8 sub-blocks and is processed using Simulink model. The algorithm results obtained similar MSE and PSE compared to the referenced models.

Keywords: 3D DWT, 4D.

1 Introduction

High speed, high resolution 4-D imaging systems could help surgeons account for the relative location of target tissue during involuntary movements such as respiration, heartbeat, and digestive motion in noninvasive surgery. Although the field has been the subject of research for several years, many areas for innovation exist and could apply to a range of noninvasive surgical therapies. The vision for Non-invasive Surgery is simple: the destruction of undesirable tissue in the human body without physically penetrating the body. As it is impossible for the surgeon to remove what he cannot see, capturing useful images of the undesirable tissue and the progress of its destruction during surgery is imperative. Different three dimensional (3-D) imaging modalities usually provide complementary medical information about patient anatomy or physiology. Four-dimensional (4-D) medical imaging is an emerging technology that aims to represent a patient's movements over time. A 3-D medical image contains a sequence of parallel two dimensional (2-D) slice images representing anatomic or physiologic information in 3-D space. The smallest element of a 3-D image is a cubic volume called a voxel. A 4-D medical image contains a temporal series of 3-D images. 4-D images can be used for monitoring respiratory and cardiac motion. In 4-D imaging technologies, progress has been made combining image registration and classification to provide intrinsic simplification and cross verification. 3-D images have been used clinically since CT was invented. Ultrasound imaging, MRI, PET, and SPECT have been developed, providing 3-D imaging modalities that complement CT. Among the most recent advances in clinical imaging, helical multi-slice CT provides

improved image resolution and capacity of 4-D imaging, but is too slow for use during non-invasive surgery. Other advances include mega-voltage CT (MVCT), cone-beam CT (CBCT), functional MRI, open field MRI, time-of-flight PET, motion-corrected PET, various angiography, and combined modality imaging, such as PET/CT, and SPECT/CT. Some preclinical imaging techniques have also been developed, including parallel multichannel MRI. Post-imaging analysis (image processing) is required in many clinical applications. Image processing includes image filtering, segmentation, registration, and visualization, which play a crucial role in medical diagnosis/treatment, especially in the presence of patient motion and/or physical changes. In this paper, we have considered 3D DWT for medical image processing.

The 3D wavelet decomposition is computed by applying three separate 1D transforms along the coordinate axes of the video data. The 3D data is usually organized frame by frame. The single frames have again rows and columns as in the 2D case, x and y direction often denoted as “spatial co-ordinates” where as for video data a third dimension z for “time” is added. In the case of 2D decompositions, it does not matter in which order the filtering is performed. A 2-D filtering frame by frame with subsequent temporal filtering, three 1-D filtering along y, t, and x axes, one decomposition step results in 8 frequency sub bands out of which only the approximation data is processed further in the next decomposition step. The paper “Low-Power And High-Speed VLSI Architecture For Lifting-Based Forward And Inverse Wavelet Transform”[1] proposed by Xuguang Lan and Nanning Zheng, presents the low-power, high-speed architecture which performs two-dimension forward and inverse discrete wavelet transform (DWT) for the set of filters in JPEG2000 using line based and lifting scheme. Anirban Das, Anindya Hazra, and Swapna Banerjee have proposed the architecture of the lifting based running 3-D discrete wavelet transform (DWT), which is a powerful image and video compression algorithm in the paper “An Efficient Architecture for 3-D Discrete Wavelet Transform”[2]. The proposed design is one of the first lifting based complete 3-DDWT architectures without group of pictures restriction. Chin-Fa Hsieh, Tsung-Han Tsai, Neng-Jye Hsu, and Chih-Hung Lai, proposed[3] a novel, efficient VLSI architecture for the implementation of one-dimension, lifting-based discrete wavelet transform (DWT). Both folded and the pipelined schemes are applied in the proposed architecture the former scheme supports higher hardware utilization and the latter scheme speed up the clock rate of the DWT. Jen-Shiun Chiang, and Chih-Hsien Hsia have proposed a highly efficient VLSI architecture for 2-D lifting-based 5/3 filter discrete wavelet transform (DWT) in “An Efficient VLSI Architecture for 2-D DWT using Lifting Scheme”[4] paper. The architecture is based on the pipelined and folding scheme processing to achieve near 100% hardware utilization ratio and reduce the silicon area. Proposed efficient 2-D lifting-based DWT VLSI architecture uses lossless 5/3 filter and pipelined processing. A low bit rate three dimensional decomposition algorithm for video compression with simple computational complexity is proposed [5] by Awad Kh. Al-Asmari and Abdulaziz Al-Rayes. The algorithm performs the temporal decomposition of a video sequence in a more efficient way by using 4-tap short symmetric kernel filter (Haar filters) with decimation factor of 4:1 instead of 2:1 used in the classical 3D-wavelet algorithms. M.F. L’opez, S.G. Rodr’iguez, J.P. Ortiz, J.M. Dana, V.G. Ruiz and I. Garc’ya have

proposed “Fully Scalable Video Coding with Packed Stream” [6] where Scalable video coding is a technique which allows a compressed video stream to be decoded in several different ways. This ability allows a user to adaptively recover a specific version of a video depending on its own requirements. Video sequences have temporal, spatial and quality scalabilities. In the paper[7], “3D Discrete Wavelet Transform VLSI Architecture for Image Processing” Malay Ranjan Tripathy, Kapil Sachdeva, and Rachid Talhi have proposed an improved version of lifting based 3D Discrete Wavelet Transform (DWT) VLSI architecture which uses bi-orthogonal 9/7 filter processing. This is implemented in FPGA by using VHDL codes. The lifting based DWT architecture has the advantage of lower computational complexities transforming signals with extension and regular data flow. “An Efficient Architecture For Lifting-Based Forward and Inverse Discrete Wavelet Transform” [8] is proposed by Aroutchelvame, S.M. and K. Raahemifar where architecture performs both forward and inverse lifting-based discrete wavelet transform.

Haw-Jing Lo, Heejong Yoo and David V. Anderson [2008], proposed a reusable Distributed Arithmetic Architecture for FIR Filtering, with 16-bit precision and 16-tap FIR filter. This design could not address higher order filters. Chen Jing and Hou Yuan Bin [2008] proposed Efficient Wavelet Transform on FPGA based on Distributed Arithmetic, with 8 bit input and 28 word LUT based implementation. Memory occupied was very large thus is not suitable for high speed applications. Wang Sen, Tang Bin, Zhu Jun [2007], proposed Distributed Arithmetic for FIR Filter Design on FPGA, with 70-tap low pass filter, operating at 40MHz sampling freq., and 12 – 13 bit precision. Patrick Longa and Ali Miri [2006] proposed Area-Efficient FIR Filter Design on FPGA's using Distributed Arithmetic, with 4 input -LUT based FPGA implementation of DA-FIR. This architecture is hardware efficient and better throughput performance useful for handheld applications that require low-power consumption. Most of the DA architectures strive for reduced memory and do not concentrate on speed and power. In this paper, we propose a modified architecture for FPGA implementation of DA based DWT.

In this work, we propose a modified 3D DWT architecture based on 5/3 lifting scheme architecture. In this architecture, predict and update phase of lifting steps consisting multipliers and adders are reused based on pipelined architecture to reduce the area and hence reduces power. Section II discusses a overview of 3D DWT architecture, section III presents the hardware model for the proposed 3D DWT architecture and section IV presents the results.

2 3D Discrete Wavelet Transform for 4D Imaging

The 3D DWT can be considered as a combination of three 1D DWT in the x , y and z directions, as shown in Fig. 1. The preliminary work in the DWT processor design is to build 1D DWT modules, which are composed of high-pass and low-pass filters that perform a convolution of filter coefficients and input pixels. After a one-level of 3D discrete wavelet transform, the volume of frame data is decomposed into HHH, HHL, HLH, HLL, LHH, LHL, LLH and LLL signals as shown in the Fig. 1.

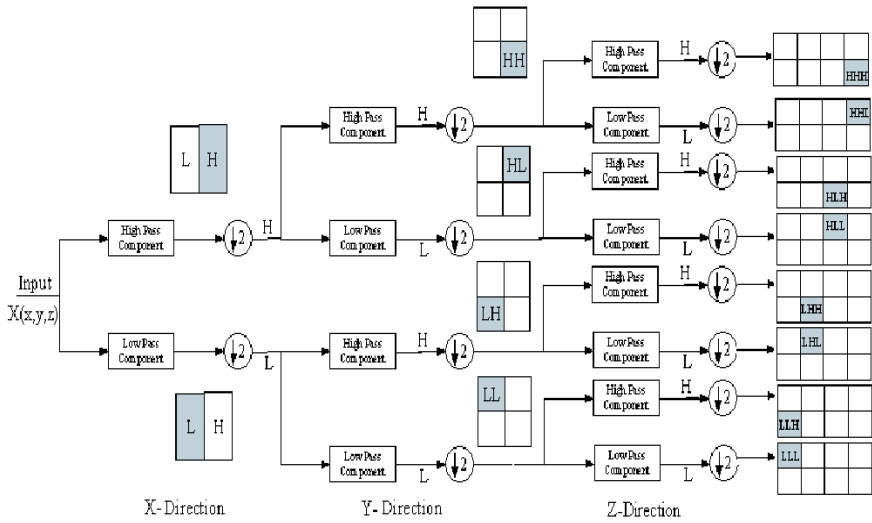


Fig. 1. One-level 3D DWT structure [3]

Fig. 2. shows the data flow diagram of 3D wavelet architecture for 4D medical imaging application. The input image is first read row wise and processed using 1D DWT algorithm, the row processed data is then column processed and then in the last stage the frames are processed in the temporal domain.

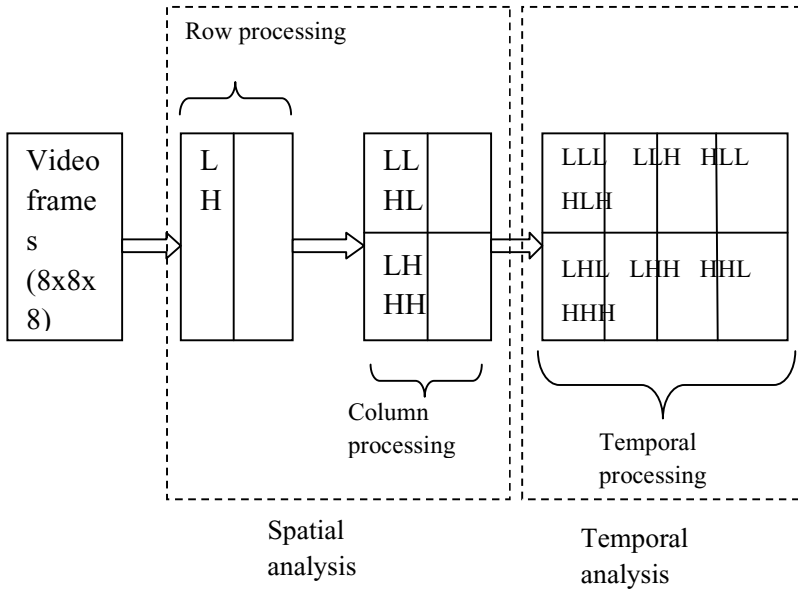


Fig. 2. Block diagram of 3D 1 level DWT decomposition

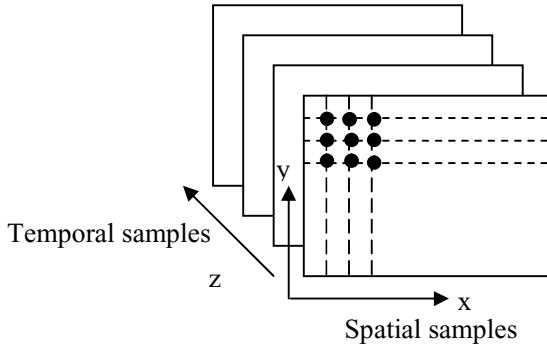


Fig. 3. Spatial and temporal sampling of a video sequence

The frame rate is 30 frames per second, thus the time interval for processing every frame is 33ms. 2D DWT should be performed in less than 10ms as another 20ms will be required for encoding. Also to perform 3D DWT, 8 frames of 512x512 images is required to be processed within less than 0.25 seconds. For video compression it is required to perform atleast three levels of decomposition, thus within a time of less than 0.25 seconds it is required to perform three stages of 3D DWT, each 2D DWT requires a time of less than 10ms. Fig. 3. shows the block level representation of 3D DWT processing logic.

3 Proposed 3D DWT for 4D Imaging

In order to improve the processing speed of 3D DWT, a modified algorithm is proposed in this section. The modified algorithm is modeled using Simulink model and 3D DWT and 3D IDWT architecture is verified using known set of MRI data. The comparison of MSE and PSNR is carried out with reference design and the computation time is also estimated to prove the advantages of the proposed design. Fig. 4. and fig. 5. shows the 2D DWT processing logic.

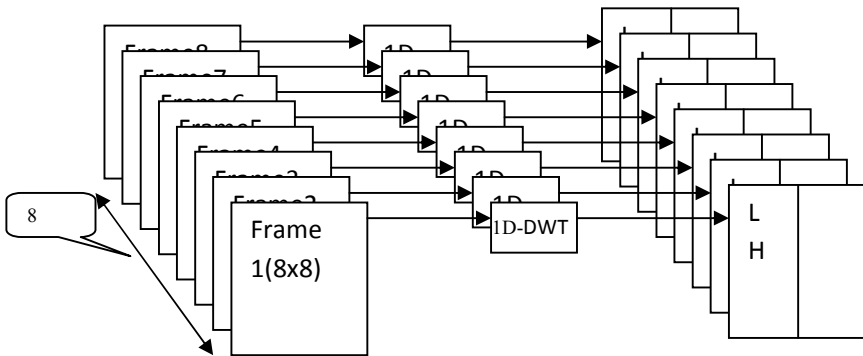


Fig. 4. Block diagram of Row processing

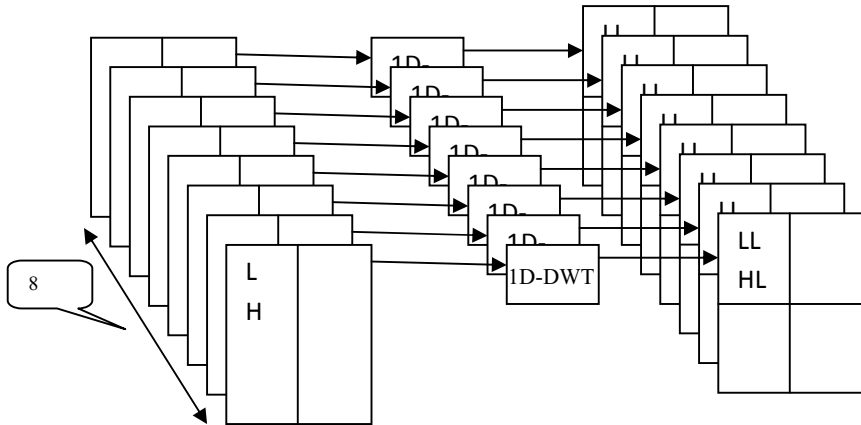


Fig. 5. Block diagram of Column processing

In the proposed design, all the row elements of 8 frames are independently processed in parallel using eight different 1D DWT algorithm. Further the row processed DWT output is further transformed and then processed in column wise, using eight 1D DWT. The parallel operations of row processor and column processor reduce the processing time by $1/8^{\text{th}}$ the original time required. The cost of hardware required is increased as they need additional 8 different 1D DWT units in every stage of processing.

4 Simulink Model Design of 3D-DWT

The proposed parallel processing algorithm for 3D DWT algorithm is modeled using Simulink. An input video sequence of MRI data is read and is first preprocessed. The video sequence consisting of 20 frames is subdivided into three slots of 8 frames. Each slot of eight frames is processed using the parallel algorithm proposed in the previous section. The input video frames are also subdivided into $8 \times 8 \times 8$ frames from the original $512 \times 512 \times 8$. Every sub slot consisting of $8 \times 8 \times 8$ set of data is processed using the proposed algorithm. Fig. 6. shows the top level Simulink model for proposed 3D DWT algorithm.

5 Result and Discussion

MRI data of human brain is used as a test case to verify the developed Simulink model. Video data is captured at 2 frames per second, with 240×320 resolutions. The captures frames are accumulated in the framer unit and the video is converted to $240 \times 320 \times 8$ frames per second, and is sub divided into $8 \times 8 \times 8$ frames. The video data is processed using the proposed algorithm and the results are shown in fig. 7.

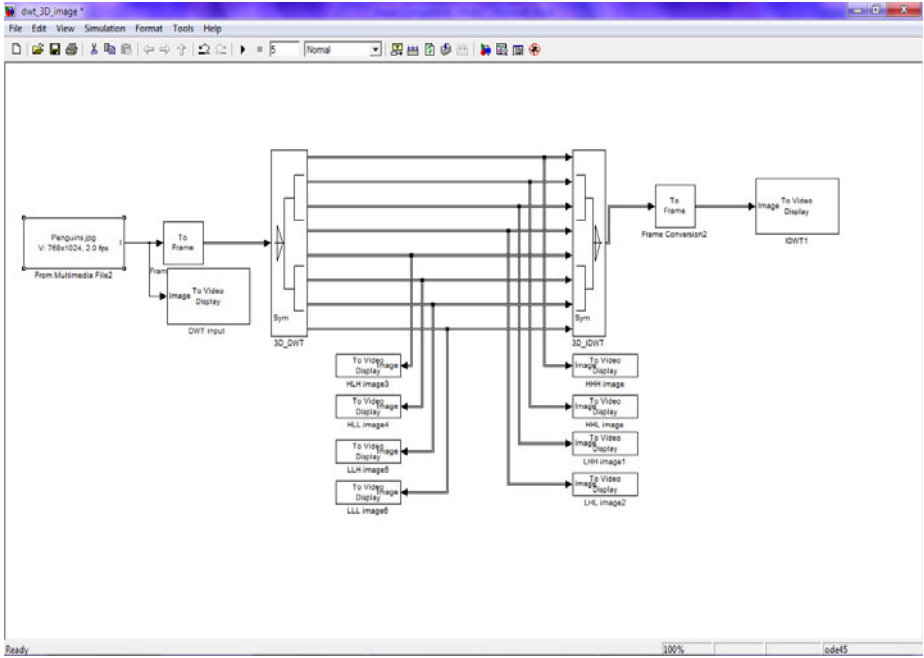


Fig. 6. Top level block diagram of Simulink model for 3D DWT

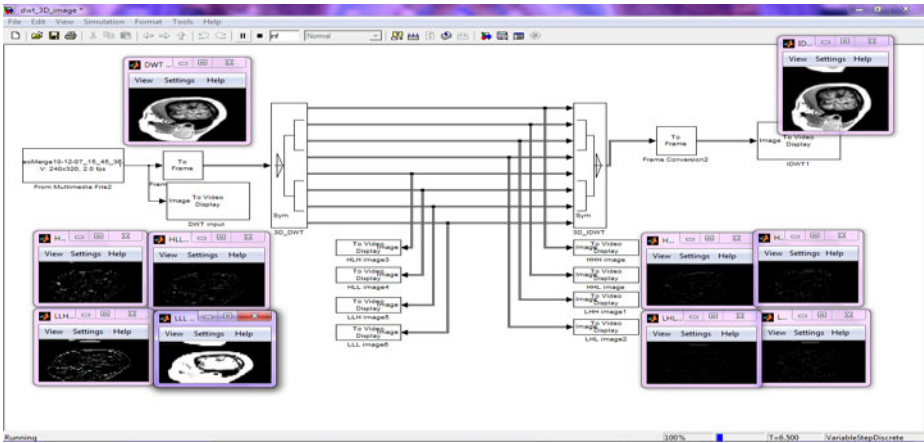


Fig. 7. Simulation results of 3D DWT using the proposed algorithm

Fig. 8 to Fig. 10 presents the simulation results of original image, decomposed image and reconstructed image using the proposed algorithm. Table 1. below shows the MSE and PSNR for 8 frames of the input video sequence obtained using the proposed model.

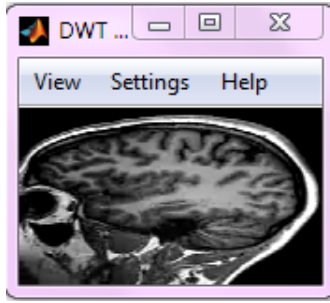


Fig. 8. Original Image

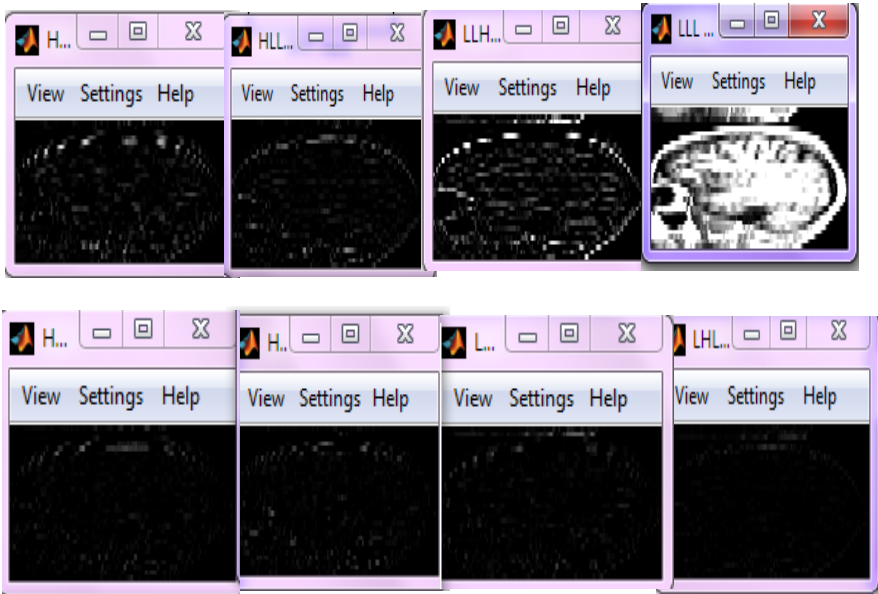


Fig. 9. Images after passing into 3D- DWT block

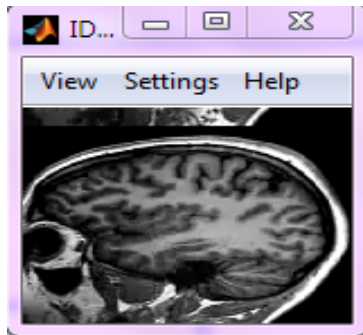


Fig. 10. Reconstructed images

Table 1. Shows the MSE and PSNR for 8 frames of the input video sequence obtained using the proposed model

Frames	MSE (Reference)	MSE (Proposed)	PSNR(Reference)	PSNR(Proposed)
Frame 1	12.2	12.1	43	43
Frame 2	12.7	12.7	42	42
Frame 3	16.3	16.2	39	39
Frame 4	11.2	11.3	49	49
Frame 5	13.3	13.3	41	40
Frame 6	14.6	14.7	42	42
Frame 7	11.8	11.9	48	47
Frame 8	12.6	12.5	43	42

Table 2. Presents the comparison of computation time for 3D DWT. The proposed algorithm is faster than the existing design by more than 60%.

Computation time	Time in seconds Iteration 1	Time in seconds Iteration 2	Time in seconds Iteration 3
Reference design	45	63	12
Proposed design	14	32	5

6 Conclusion

A 3D DWT algorithm is proposed for 4 D imaging. In this algorithm 8 row processors and 8 column processors are designed to compute the row processing and column processing of input image in parallel. The original input image is preprocessed to 8x8x8 sub blocks and is processed using the Simulink model. MSE and PSNR are computed and are compared with the reference model. The results obtained shows the proposed design achieves similar MSE and PSNR compared to the reference design. The computation time of the proposed design is very less compared to the reference design, due to parallel processing. The cost of hardware increases due to higher computation complexity. However, suitable VLSI techniques can be adopted to minimize the computation complexity.

References

1. Lan, X., Zheng, N., Liu, Y.: Low-Power And High-Speed VLSI Architecture For Lifting-Based Forward And Inverse Wavelet Transform. IEEE Senior Member (December 3, 2004)
2. Das, A., Hazra, A., Banerjee, S.: An Efficient Architecture for 3-D Discrete Wavelet Transform. IEEE Transactions on Circuits and Systems for Video Technology 20(2) (February 2010)

3. Hsieh, C.-F., Tsai, T.-H., Hsu, N.-J., Lai, C.-H.: A Novel, Efficient Architecture for the 1D, Lifting-Based DWT with Folded and Pipelined Schemes. IEEE Trans. on Department. of Electronics Engineering, China Institute of Technology, Taipei, Taiwan and Department. of Electrical Engineering, National Central University, Chung-Li, Taiwan (2004)
4. Chiang, J.-S., Hsia, C.-H.: An Efficient VLSI Architecture for 2-D DWT using Lifting Scheme. In: IEEE ICSS 2005 International Conference On Systems & Signals, Department of Electrical Engineering, Multimedia IC Design Lab. Tamkang University, Taipei (2005)
5. Al-Asmari, A.K., Al-Rayes, A.: Low bit rate video compression algorithm using 3-D Decomposition. The Arabian Journal for Science and Engineering 29(1B) (April 2004); Electrical Engineering Department, King Saud University
6. López, M.F., Rodfíguez, S.G., Ortiz, J.P., Dana, J.M., Ruiz, V.G., Garçya, I.: Fully Scalable Video Coding with Packed Stream. IEEE trans. on Computer Architecture and Electronics Dept. University of Almería, Almería, Spain (2003)
7. Tripathy, M.R., Sachdeva, K., Talhi, R.: 3D Discrete Wavelet Transform VLSI Architecture for Image Processing. In: PIERS Proceedings, Department of Electronics and Communication Engineering Jind Institute of Engineering and Technology, Jind, Haryana, India, Moscow, Russia, August 18-21 (2009)
8. Arouchelvame, S.M., Raahemifar, K.: An Efficient Architecture For Lifting-Based Forward And Inverse Discrete Wavelet Transform. Dept. of Electrical & Computer Engg. Dept. of Ryerson University, Toronto. IEEE, Los Alamitos (2005), 0-7803-9332-5/05/\$20.00 ©, 2005
9. Adami, N., Signoroni, A., Leonardi, R.: State-of-the-art and trends in scalable video compression with wavelet-based approaches. IEEE Trans. Circuits and Systems for Video Technology 17, 1238–1255 (2007)
10. Shapiro, J.: Embedded image coding using zerotree of wavelet coefficients. IEEE Trans. Signal Processing 41, 3445–3462 (1993)
11. Said, A., Pearlman, W.A.: A new, fast and efficient image codec based on set partitioning in hierarchical trees. IEEE Trans. Circuits and Systems for Video Technology 6, 243–250 (1996)
12. Campisi, P., Gentile, M., Neri, A.: Three dimensional wavelet based approach for a scalable video conference system. In: IEEE International Conference on Image Processing (ICIP 1999), vol. 3, pp. 802–806 (1999)
13. Kim, B.J., Xiong, Z., Pearlman, W.A.: Low bit-rate scalable video coding with 3-D set partitioning in hierarchical trees (3-D SPIHT). IEEE Trans. Circuits and Systems for Video Technology 10, 1374–1387 (2000)
14. He, C., Dong, J., Zheng, Y.F., Gao, Z.: Optimal 3-D coefficient tree structure for 3-D wavelet video coding. IEEE Trans. Circuits and Systems for Video Technology 13, 961–972 (2003)

Stratified SIFT Matching for Human Iris Recognition

Sambit Bakshi, Hunny Mehrotra, and Banshidhar Majhi

Department of Computer Science and Engineering,
National Institute of Technology Rourkela
Odisha, India
sambitbakshi@gmail.com,
{hunny,bmajhi}@nitrkl.ac.in
<http://www.nitrkl.ac.in>

Abstract. This paper proposes an efficient three fold stratified SIFT matching for iris recognition. The objective is to filter wrongly paired conventional SIFT matches. In Strata I, the keypoints from gallery and probe iris images are paired using traditional SIFT approach. Due to high image similarity at different regions of iris there may be some impairments. These are detected and filtered by finding gradient of paired keypoints in Strata II. Further, the scaling factor of paired keypoints is used to remove impairments in Strata III. The pairs retained after Strata III are likely to be potential matches for iris recognition. The proposed system performs with an accuracy of 96.08% and 97.15% on publicly available CASIAV3 and BATH databases respectively. This marks significant improvement of accuracy and FAR over the existing SIFT matching for iris.

Keywords: Iris Recognition, Stratified SIFT, Keypoint, Matching.

1 Introduction

Iris is the sphincter having unique flowery random pattern around the pupil. It is an internal organ with complex unique features that are stable throughout the lifetime of an individual. There has been significant research done in the area of iris recognition using global features [1,2,3,4]. However, these approaches fail to possess invariance to affine transformations, occlusion and robustness to unconstrained iris images. Thus, there is a stringent requirement to develop iris recognition system suitable for non-cooperative images. Keypoint descriptors are invariant to affine transformation as well as partial occlusion. Scale Invariant Feature Transform (SIFT) is a well known keypoint descriptor for object recognition [5]. Due to inherent advantages, SIFT is capable of performing recognition using non-cooperative iris images [6]. In SIFT matching approach, the difference of Gaussian images are used to identify keypoints at varying scale and orientation. The orientation is assigned to each detected keypoint and a window is taken relative to direction of orientation to find the descriptor vector. During recognition, the keypoints are detected from gallery and probe images and matching is performed using nearest neighbour approach. The challenge with conventional SIFT matching when applied to iris recognition is to find texture similarity between same regions of two iris. SIFT fails as it does not consider spatial information of the keypoints. To make SIFT

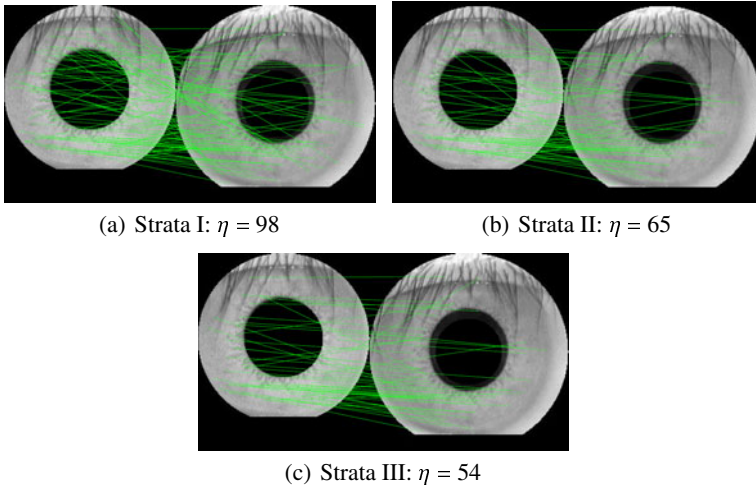


Fig. 1. Matches (η) at different strata obtained between two instances of same individual taken from CASIAV3

functional for iris, a novel matching approach has been developed that combines spatial information along with local descriptor of each keypoint.

The organization of the paper is as follows. Conventional SIFT matching approach for local descriptors is discussed in Section 2. Proposed stratified SIFT matching is explained in detail in Section 3. The job of filtering is done in two sequential steps: Gradient based filtering and Scale based filtering. Section 4 provides the results obtained using proposed approach. Finally, conclusions are given at the end of the paper.

2 Conventional SIFT Matching

The matching algorithm plays a significant role in any biometric system as it acts like a one way gateway through which only genuine matches (if two images are from same subject) will pass and imposter matches (if two images are from different subjects) are blocked. In local feature matching, the total number of paired keypoints is used to find the authenticity of an individual. Let I be the set of all images available in the iris database. For understanding, I_m be a gallery iris image and I_n be a probe iris image where $I_m, I_n \in I$. The matching algorithm validates I_n against I_m . In the conventional SIFT matching, for each keypoint in I_m the Euclidean distance is found with every keypoint in I_n . The nearest neighbour approach pairs the i^{th} keypoint in I_m with j^{th} keypoint in I_n , iff the descriptor distance between the two (after multiplying with a threshold) is minimum [5]. The two keypoints are paired and removed from the set of keypoints detected from I_m and I_n . This process is iterated for the remaining keypoints until any two keypoints can be matched. This approach performs moderately well for unconstrained iris recognition [6]. However, as SIFT determines image similarity using 128-dimensional local features only, hence it may wrongly pair (impairment) some

keypoints for iris. Thus, the existing approach is modified using two strata which removes impaired matches using spatial information of the matching keypoints and contributes in achieving better recognition accuracy.

3 Stratified SIFT Matching

In the proposed paper an effort is made to improvise the conventional SIFT matching. The pupil and iris circles are assumed to be concentric, hence all localized images have pixel size $2r \times 2r$, where r is the radius of iris. The pupil center as well as iris centre are located at (r, r) . Therefore the localized images do not have transformation due to translation. However, there is a possibility of iris images being transformed due to rotation (tilt of subject's head), scaling (change in camera to eye distance) or both [6]. The SIFT matching algorithm matches keypoints that have similarity between the local descriptors (as discussed in Section 2) but fails to conform to spatial relationship. The removal (filtering) of impairments by the proposed approach retains only those matches that are more probable to be potential matches. Let K_m be the set of m_1 keypoints found in I_m and K_n be the set of n_1 keypoints found in I_n by applying SIFT detector. These sets of keypoints are used to comprehend the stratified SIFT matching.

3.1 Strata I: SIFT Matching

Let R be the ordered set containing the matches between K_m and K_n by conventional SIFT matching as discussed in Section 2. Hence, R contains only those pairs (i, j) where i^{th} keypoint in K_m is matched with j^{th} keypoint in K_n as shown in Fig. 1(a). Let η be the number of matches found where $\eta \in [0, \min(m_1, n_1)]$. As set R is generated solely on the basis of local descriptor property it may wrongly pair keypoints from different regions of iris. Hence, there is a need to combine spatial information with local descriptor to filter out impaired keypoints as discussed in subsequent strata.

3.2 Strata II: Gradient Based Filtering

In this strata gradient based filtering is performed to remove impairments from R . To compute gradient for each pair of keypoints (i, j) in R , the angles are obtained from respective image centers (r, r) . Thus, θ_i is computed from I_m and ϕ_j is computed from I_n . The angle of rotation for k^{th} pair is calculated as $\gamma_k = (\phi_j - \theta_i) \bmod 360^\circ$ (depicted in Fig. 3(a)). Considering SIFT to be completely *flawless* (due to robustness property, no false match is found) and *efficient* (due to property of repeatability, all possible matches are found) [5]; the value of γ_k derived should be same $\forall k$. But in practice, SIFT does not give such precise matches. Thus, it is difficult to obtain unique value of γ even when I_m and I_n belong to the same subject. Rather a distribution of γ is obtained. A histogram is plotted with x-axis comprising bins with a range of values of γ , and y-axis comprising number of matches falling in a particular bin as shown in Fig. 2. The number of bins in the histogram (*nobins*) is subject to implementation issue. In proposed system, *nobins* is taken as 10. The distribution of γ gives a single peak in case the two iris images (I_m and I_n) are from the same subject. In contrast, no distinct peak should be found in case the two iris images are from different subjects. There may be error due to discretization

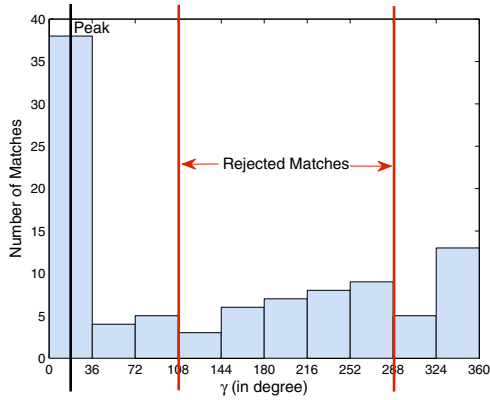
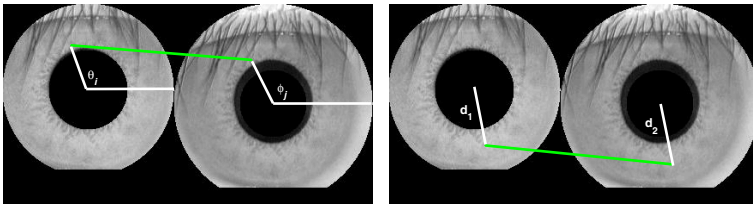


Fig. 2. Distribution of γ for number of matches between two instances of same individual from CASIAV3



$$(a) \gamma = (\phi_j - \theta_i) \bmod 360^\circ$$

$$(b) \psi = \frac{d_2}{d_1}$$

Fig. 3. (a) Gradient computation in Strata II, (b) Local scaling factor computation in Strata III

of bins, so two adjacent bins of the peak are combined to improve peak density (number of matches). The idea is to find whether the density of the peak exceeds the boundary criteria. It is inferred that a peak is strong if the density exceeds certain higher bound ($hp\%$ of total number of matches).

Likewise peak is weak if the density is less than lower bound ($lp\%$ of total number of matches). If a strong peak is found, an angular range is specified around the peak. Those matches in R for which γ are not within the angular range are found to be impaired and removed from R to generate R_{inter} .

For example, as shown in histogram in Fig. 2, the peak is found at 0^{th} bin which represents gradient value of 0° to 36° with a central value of 18° . Hence only those pairs having angular range between $(18 \pm 90) \bmod 360^\circ$ are retained. Thus, it is evident that $R_{inter} \subseteq R$ after removing some impairments. Fig. 1(b) shows paired keypoints in R_{inter} with considerable reduction of η . If no strong peak is found it is inferred that all matches in R are faulty, and removed. As a result R_{inter} becomes empty.

Table 1. Performance measures for stratified SIFT matching

Databases →	CASIAV3				BATH			
Approach ↓	FAR	FRR	ACC	d'	FAR	FRR	ACC	d'
Strata I (Conventional SIFT)	17.48	10.91	85.81	1.20	1.57	4.35	97.04	2.73
Strata II (Gradient based Filtering)	2.49	5.45	96.03	2.46	0.97	6.09	96.47	2.81
Strata III (Scale based Filtering)	2.39	5.45	96.08	2.20	1.34	4.35	97.15	2.90

3.3 Strata III: Scale Based Filtering

In this strata further filtering of *Rinter* is performed on the basis of global and local scaling factor between the gallery and probe images. The global scaling factor (sf) between two images is defined as ratio of probe iris radius (r_n) to gallery iris radius (r_m). A scale range with certain tolerance around sf is empirically taken to handle aliasing artifact. From implementation perspective, the scale range is taken as $(sf \pm 0.2)$. In ideal case, if gallery and probe belong to same individual the scaling factor between all paired keypoints should be unique. However, this does not hold good in practical scenarios.

During filtering, for each element in *Rinter*, two Euclidean distances are calculated- (a) d_1 : distance of i^{th} keypoint of I_m from its center and (b) d_2 : distance of j^{th} keypoint of I_n from its center. Local scaling factor (ψ) for each element of *Rinter* is calculated as $\psi = d_2/d_1$ (shown in Fig. 3(b)). Matches having ψ within scale range discussed above qualifies to be potential and stored in *Rnew*, else are labeled as faulty and filtered. Fig. 1(c) shows paired keypoints in *Rnew* after further reduction of η .

4 Experimental Results

The proposed stratified SIFT matching is tested on publicly available BATH [8] and CASIAV3 (CASIAV3) [9] databases. Database available from BATH University includes images from 50 subjects (20 images per subject from both the eyes). CASIAV3 (CASIAV3) comprises 249 subjects with total of 2655 images from both the eyes. The experiments are carried out on 2.13GHz Intel(R) CPU using Matlab. To validate the system performance some standard error measures [7] are used. The results are carried out in three different strata as given in Table 1. In *Strata I*, the two iris images are matched using conventional SIFT approach. This approach performs with an accuracy of 85.81% on CASIAV3 database. Likewise, for BATH database an accuracy of 97.04% is obtained. To improve the performance of the system, the objective of the proposed research is to reduce false acceptances. In *Strata II*, the impairments are removed using gradient filtering which significantly increases the separability measure between false and genuine matches as indicated by d' values given in the Table 1. Further improvement in separability and accuracy are brought by scale filtering in *Strata III*. The accuracy values are plotted against change in number of matches as shown in Fig. 4(a).

The Receiver Operating Characteristic (ROC) curves [7] for three different strata are shown in Fig. 4(b). The distribution of genuine and imposter scores after Strata

¹ FAR: False Acceptance Rate, FRR: False Rejection Rate, ACC: Accuracy, d' : d-prime value.

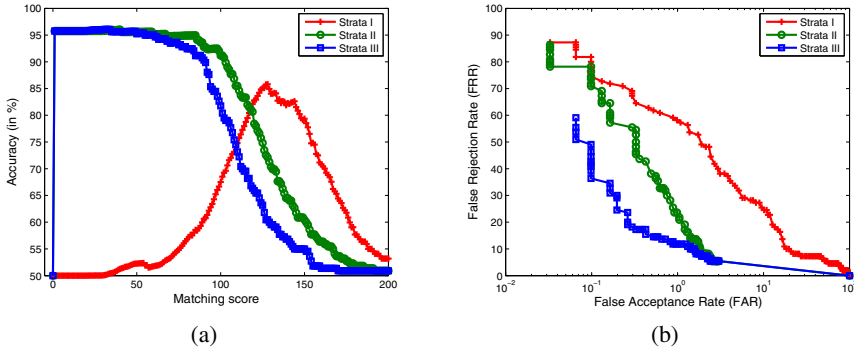


Fig. 4. (a) Accuracy curve for three strata, (b) ROC curve for three strata on CASIAV3

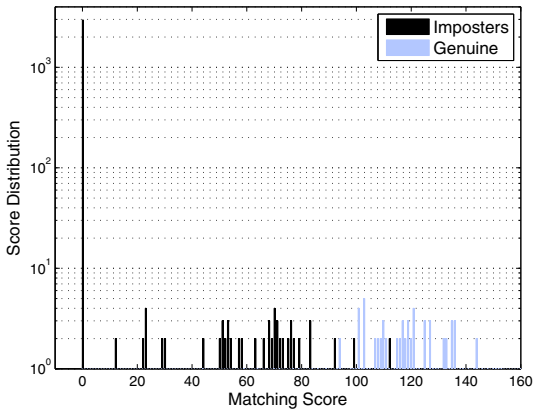


Fig. 5. Final histogram of scores on CASIAV3

III is shown in Fig. 5. All graphical results are obtained for CASIAV3 and similar observations are made for BATH database.

5 Conclusions

In this paper, a novel stratified SIFT matching technique is proposed that improves conventional SIFT by removing wrong pairs. This approach provides boost in accuracy due to considerable reduction in FAR. The FAR is reduced by 15.09% and 0.23% for CASIAV3 and BATH respectively. From the results it has been observed that the proposed algorithm is completely flawless, i.e., matches removed are guaranteed to be wrong matches whereas it is not completely efficient, i.e., all impairments by SIFT are not guaranteed to be filtered. However, the gain in accuracy is substantial which marks its applicability for unconstrained iris recognition.

References

1. Daugman, J.: The importance of being random: Statistical principles of iris recognition. *Pattern Recognition* 36(2), 279–291 (2003)
2. Kim, J., Cho, S., Choi, J., Marks, R.J.: Iris Recognition Using Wavelet Features. *The Journal of VLSI Signal Processing* (38), 147–156 (2004)
3. Monro, D.M., Rakshit, S., Zhang, D.: Dct-based iris recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 29(4), 586–595 (2007)
4. Tisse, C., Martin, L., Torres, L., Robert, M.: Person identification technique using human iris recognition. In: *Proc. of Vision Interface*, pp. 294–299 (2002)
5. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 6(2), 91–110 (2004)
6. Belcher, C., Du, Y.: Region-based sift approach to iris recognition. *Optics and Lasers in Engineering* 47(1), 139–147 (2009)
7. Jain, A., Flynn, P., Ross, A.: *Handbook of Biometrics*. Springer-Verlag New York, Inc., Secaucus (2007)
8. Bath University Database,
<http://www.bath.ac.uk/elec-eng/research/sipg/irisweb>
9. Chinese Academy of Sciences' Institute of Automation (CASIA) Iris Image Database V3.0,
<http://www.cbsr.ia.ac.cn/english/IrisDatabase.asp>

Quality Index Based Face Recognition under Varying Illumination Conditions

K.T. Dilna and T.D. Senthilkumar

Department of Electronics and communication
K.S.R College of Technology, Tiruchengode
{dilnageethanand,tdsenthilece}@gmail.com

Abstract. Face recognition is one of the most popular biometric techniques for automatically identifying or verifying a person from a video or digital image. The face recognition accuracy can be affected by intraclass variations and interclass variations. A change in lighting condition is one of the intraclass variations. Preprocessing is an approach to normalize the intraclass variations of light varying image. Histogram equalization (HE) is one of the techniques to normalize the variations in illumination. But it is not suitable for well lighted images. Image quality based adaptive face recognition is used for well lighted face image recognition. The multiresolution property of wavelet transforms is used in face recognition to extract facial feature descriptors. Low and high frequency wavelet subbands are extracted and fusion of match scores from the subband is used to improve the recognition accuracy under varying lighting conditions. For face recognition, 2DPCA (2D Principle Component Analysis) method is used and can be verified with illumination variant face images. 2DPCA is based on 2D image matrices rather than 1D vector so the image matrix does not need to be transformed into a vector prior to feature extraction.

Keywords: Biometrics, Face Recognition, Quality Measure, Wavelet Transform, 2DPCA.

1 Introduction

Biometrics is an application of image processing. Biometrics [3] refers to the automatic identification of a person based on his or her physiological or behavioral characteristics. Various types of biometric systems are being used for real-time identification. The most popular are based on face recognition and fingerprint matching. Face recognition [3] is measuring facial characteristics. One of the ways to do this is by comparing selected facial features from the image and a facial database. It is typically used in security systems. It can be done for one-dimensional, two dimensional and three dimensional images

Face recognition can be done in different conditions. A human face is not only 3D object, it is also a non rigid body. Moreover, facial images are often taken under natural environment. That is, the image background could be very complex and the illumination condition could be extreme. Face recognition algorithms identify faces by extracting landmarks, or features from the face image. For example, an algorithm

may analyze the relative position, size, and/or shape of the eyes, nose, cheekbones, and jaw. These features are then used to search for other images with matching features. Other algorithms normalize a gallery of face images and then compress the face data, only saving the data in the image that is useful for face detection. A probe image is then compared with the face data. One of the earliest successful systems is based on template matching techniques applied to a set of salient facial features, providing a sort of compressed face representation. There are many face recognition algorithms present such as PCA, LDA, and ICA to extract the facial features. PCA method is extended into 2DPCA method which has many advantages over PCA method. In the PCA-based face recognition technique, the 2D face image matrices must be previously transformed into 1D image vectors. But 2DPCA is based on 2D matrices rather than 1D vector. That is, the image matrix does not need to be previously transformed into a vector.

Numerous methods are proposed for normalization of face image and thus recognition, including Local Binary Pattern (LBP) [8], Self- Quotient Image (SQI) [9], Logarithmic Total Variation (LTV) model [10], Quotient Image [12]. In [5] authors proposed a method for illumination normalization, using Discrete Cosine Transform (DCT) in logarithmic domain. In this method an appropriate number of DCT coefficients are truncated to minimize variations under different lighting conditions. Another method is gradientfaces [6] which is illumination insensitive features for face recognition under varying lighting. In [7] it presents a solution for illumination invariant face recognition for indoor, cooperative-user applications using an active Near Infrared (NIR) imaging system which is able to produce face images of good condition regardless of visible lights in the environment.

Face recognition under varying illumination condition is very difficult. Under this condition, preprocessing of the image is needed to normalize the effect of illumination variation. Typical methods employed to address varying illumination conditions could be categorized as feature-based methods, generative methods, and holistic methods [1]. Histogram Equalization (HE) is the most commonly used normalization method which include histogram equalization, histogram matching, gamma intensity correction, and quotient image. But HE of well light image reduces the image quality and thus recognition accuracy. Thus to improve the recognition accuracy, image quality based adaptive recognition is used.

2 Image Quality Index Based Face Recognition

Image quality based adaptive recognition scheme is shown in Fig .1. Face images under different illumination conditions are given as input.

Quality measurement [1], [11] of an image, point out the importance of preprocessing, which is a normalization technique under varying illumination conditions. The quality indexes of different images under varying illumination conditions are varied. Loss of correlation, Luminance distortion, and Contrast distortion are the three factors of image distortion. The illumination quality of a given face image is to be defined in terms of its luminance distortion in comparison to a known reference image.

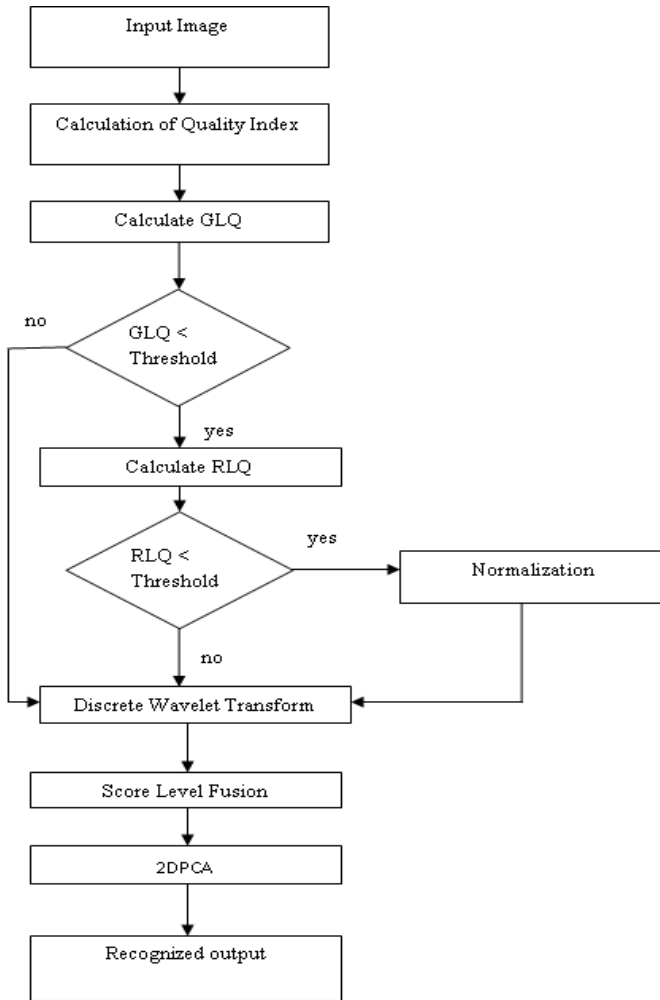


Fig. 1. Flow diagram of proposed method

Here let x ($i=1\dots N$) be the test image and y ($j=1\dots N$) be the reference image. The universal Quality Index can be calculated as

$$Q = \frac{(4\sigma_{xy} \bar{x} \bar{y})}{((\sigma_x^2 + \sigma_y^2)[(\bar{x}^2) + (\bar{y}^2)])} \quad (1)$$

where \bar{x} , \bar{y} , σ_x and σ_y are given by

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad (2)$$

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i \quad (3)$$

$$\sigma_x^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (4)$$

$$\sigma_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2 \quad (5)$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}) \quad (6)$$

From this universal quality index, Luminance distortion factor can be written as

$$LQ = \frac{\overline{2xy}}{((\bar{x})^2 + (\bar{y})^2)} \quad (7)$$

This Luminance Quality factor is used to find the luminance distortion of the test image by comparing with the reference face image.

Normalization can be applied either globally or regionally. First calculate the Global Luminance Quality Index (GLQ) which is similar to the calculation of the single Q value. GLQ is calculated by applying LQ for the whole face image. In order to improve the accuracy Regional Quality Index (RLQ) is calculated. For the calculation of RLQ the face image is partitioned into several regions [13]. This partitioning can be done according to different local regions. However this local region partitioning is very difficult for complex regions. So in this work face image is simply partitioned into four regions.

The normalization is done according to the value of GLQ and RLQ. GLQ and RLQ values are compared with a predefined threshold. If the GLQ value is less than the predefined threshold, RLQ value is calculated. Then regional quality based normalization is carried out. Histogram equalization is done for normalization. Find the RLQ value for each four region separately, and compare with the redefined threshold. If the RLQ value is less than the threshold then histogram normalization is applied. This is referred as Regional Quality based Histogram Equalization (RQbHE).

Multiresolution property [14] of Discrete Wavelet Transform (DWT) [4] is used to reduce the dimensionality. DWT decomposes an image into low and high frequency subbands and it provides a compact representation of a signal in time and frequency. Low frequency subband (LL) is suitable [15] for face descriptor for recognition, but it is affected by varying illumination. So fusion of low (LL) and high (HH) frequency subbands are using for recognition. Face features are extracted from each subband by applying 2DPCA.

In the PCA-based face recognition technique, the 2D face image matrix must be transformed into 1D image vectors. The resulting image vectors of faces usually lead to a high dimensional image vector space. In this vector space, it is difficult to find

out the covariance matrix accurately due to its large size. 2DPCA [2] is based on 2D matrices rather than 1D vectors. That is, the image matrix does not need to be previously transformed into a vector. Instead, an image covariance matrix can be constructed directly using the original image matrices and the size of the image covariance matrix using 2DPCA is much smaller than PCA. The covariance matrix can be calculated as

$$G_t = \frac{1}{M} \sum_{j=1}^M (A_j - \bar{A})^T (A_j - \bar{A}) \quad (8)$$

where \bar{A} the mean of the input image and M is the number of samples images.

Project the image onto a unitary matrix and find m dimensional projected vector. Find the eigen value and eigen vector of the covariance matrix G_t . Then find the optimum projection vector and the optimum projection axes. Using these values, feature vectors or principal component vectors can be obtained. It is used to form a matrix which is called the feature matrix or feature image of the image. The projected image can be calculated by

$$Y = AX \quad (9)$$

where X is the eigen vector of covariance matrix.

After a transformation by 2DPCA, a feature matrix is obtained for each subband. Fusion score from each subband is used for recognition. By using Euclidean distance score measure, the matching is tested. Euclidean distance can be measured as

$$d(B_i, B_j) = \sum_{k=1}^d \left\| Y_k^i - Y_k^j \right\| \quad (10)$$

where $B = [Y_1, \dots, Y_d]$, i and j are represented by two images.

The image which is given the minimum Euclidean distance is taken as the recognized image. That is

$$d(B, B_k) = \min d(B, B_j) \quad (11)$$

where k is the number of training samples.

3 Results and Discussion

The input images under different illumination conditions are taken from face database. Dark and well lighted images of original images are given as input as shown in Fig.4.

The GLQ and RLQ value should be compared with a predefined threshold. The predefined threshold used in this work is found as 0.8 from the LQ value table (1) as shown. After applying the Fig 4 (b), that is dark image it finds that the GLQ value is less than the threshold. So to find the RLQ, input image is partitioned into four as shown in Fig.5 (a).

After the partitioning, RLQ value is calculated separately for four regions and compared with threshold. It is shown that, it is less than threshold and thus normalization is done separately for the four regions as shown in Fig.5 (b). After the normalization of four regions, it again combined to get the original image sizes shown in Fig 5(c). This image is given to the 2DPCA algorithm and thus the face is recognized.

If the well light image which is shown in Fig.4(c) is given, GLQ value is calculated. This GLQ value is higher than the threshold and thus normalization is not done.

Table 1. Identification of Threshold Value

LQ VALUES	
DARK IMAGE	BRIGHT IMAGE
0.3644	0.8145
0.4668	0.8372
0.5048	0.8423
0.5358	0.8602
0.5594	0.8651
0.5873	0.8715
0.6105	0.8767
0.6388	0.8830
0.6869	0.9219
0.7980	0.9248



Fig. 3. Subset Face images

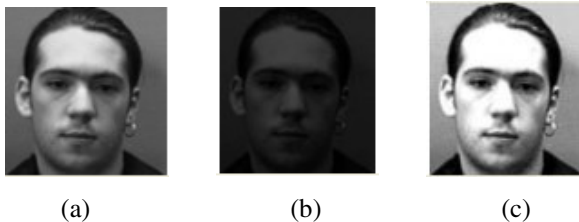


Fig. 4. Input face images (a) Original Image, (b) & (c) Illumination varied image

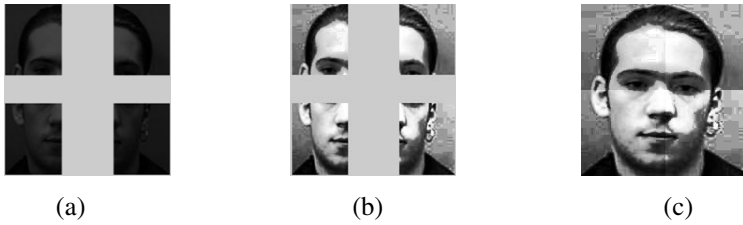


Fig. 5. (a) Partitioned Image of Fig.3 (b), (b) Normalized image of Fig. 4 (a), (c) RQbHE image

4 Conclusion

This method presents Image quality index based normalization method for normalizing the face images under varying illumination conditions and in turn for face recognition. This is done through the calculation of image quality index. It is found that dark and bright images are categorized by fixing LQ threshold as 0.8. If the LQ threshold value is above and below 0.8 images is recognized as bright images and dark images respectively. Thus image quality based recognition method recognizes dark and bright face images.

References

1. Sellahewa, H., Jassim, S.A.: Image-Quality-Based Adaptive Face Recognition. *IEEE Transactions on Instrumentation and Measurement* 59, 805–813 (2010)
2. Yang, J., Zhang, D., Frangi, A.F.: Two Dimensional PCA A New Approach to Appearance Based Face representation and recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 26, 131–137 (2004)
3. Jain, A.K., Ross, A., Prabhakar, S.: An Introduction to Biometric Recognition. *IEEE Trans. On Circuits and Systems for Video Technology* 14 (2004)
4. Chien, J.T.: Discriminant wavelet faces and nearest feature classifiers for face recognition. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 24, 1644–1649 (2002)
5. Chen, W., Meng, J.E., Wu, S.: Illumination Compensation and Normalization for Robust Face Recognition Using Discrete Cosine Transform in Logarithm Domain. *IEEE Transactions on Systems, Man, and Cybernetics* 36, 458–466 (2006)
6. Zhang, T., Tang, Y.Y., Fang, B., Shang, Z.: Face Recognition under Varying Illumination Using Gradientfaces. *IEEE Transactions on Image Processing* 18, 2599–2606 (2009)
7. Li, S.Z., Chu, R.F., Liao, S.C., Zhang, L.: Illumination Invariant Face Recognition Using Near-Infrared Images. *IEEE Transactions on Pattern Analysis And Machine Intelligence* 29, 627–639 (2007)
8. Vázquez, H.M., Reyesand, E.G., Molleda, Y.C.: A New Image Division for LBP Method to Improve Face Recognition under Varying Lighting conditions. *IEEE, Los Alamitos* (2008)
9. Wang, H., Li, S.Z., Wang, Y.S.: Face Recognition under Varying Lighting Conditions Using Self Quotient Image. In: *IEEE International Conference on Automatic Face and Gesture Recognition* (2004)
10. Chen, T., Yin, W., Zhou, X.S.: Total Variation Models for Variable Lighting Face Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28 (2006)

11. Wang, Z., Bovik, A.C.: A Universal Image Quality Index. *IEEE Signal Processing Letters* 9, 81–84 (2002)
12. Shashua, A., Riklin-Raviv, T.: The Quotient Image: Class-Based Re-Rendering and Recognition with Varying Illuminations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 129–139 (2001)
13. Shan, S., Gao, W., Cao, B., Zhao, D.: Illumination normalization for robust face recognition against varying lighting conditions. In: *Proc. IEEE Int. Workshop Anal. Model. Faces Gestures*, pp. 157–164 (2003)
14. Ekenel, H.K., Sankur, B.: Multiresolution face recognition. *J. Image and Vision Computing* 23, 469–477 (2005)
15. Sellahewa, H., Jassim, S.A.: Illumination and expression invariant face recognition: Toward sample quality-based adaptive fusion. In: *Proc. 2nd IEEE Int. Conf. Biometrics, Theory, Appl. Syst.*, pp. 1–6 (2008)

Noise Adaptive Weighted Switching Median Filter for Removing High Density Impulse Noise

Madhu S. Nair and P.M. Ameeramol

Department of Computer Science
University of Kerala, Kariavattom, Thiruvananthapuram - 695 581, Kerala, India
madhu_s_nair2001@yahoo.com, ameeramol@gmail.com

Abstract. This paper proposes a new efficient noise adaptive weighted switching median filter for the restoration of images that are corrupted by high density of impulse noise. The proposed method consists of two phases- noise detection and filtering. In our proposed method, the filtering window size is chosen adaptively depending on the percentage of noise that corrupts the image. Noise detection is done by using Boundary Discriminative Noise Detection method proposed by P.-E Ng et.al and then filtering is applied to only the corrupted pixels in the noisy image. Each detected noisy pixel is replaced by a weighted median value of uncorrupted pixels in the filtering window. Weight value assigned to each uncorrupted pixel depends on its closeness to the central pixel.

Keywords: Image denoising, Impulse noise, Noise adaptive weighted switching median filter, Salt and pepper noise.

1 Introduction

Digital Images are often contaminated by noise during acquisition, storage and transmission, thereby degrading the quality of images; therefore a common problem in applied science and engineering is the restoration of the corrupted images. Impulse noise is a specific type of noise which causes alteration of the pixels in the images so that their gray values do not exhibit compatibility with their local neighborhood. Usually images are degraded by impulse noise of short duration and high energy because of the errors caused by noisy sensors or transmission channels [1].

There are many methods for removal of impulse noises from the images. Usage of linear filters such as averaging filters produces blurring of the images. Non linear filters such as median filters are the most popular technique for removing impulse noise because of its good denoising power and computational efficiency. However most of the median filters are implemented uniformly across the image and thus tend to modify both noisy and noise free pixels. Consequently the effective removal of impulse noise is often accomplished at the expense of blurred and distorted features thus removing fine details in the image. Switching median filters are shown to be simple and yet more effective than uniformly applied methods such as median filters [2] [11]. In the case of switching median filters it will identify the possible noisy pixels in the image and then replace them by using median filters and its variants

while leaving all other uncorrupted pixels unchanged. Impulse detection therefore becomes crucial to subsequent filtering in switching median filters.

The earlier developed switching median filters were commonly found, being non adaptive to a given, but unknown, noise density and prone to yielding pixel misclassifications especially at higher noise density interference. There are different methods for impulse noise detection: fuzzy approaches [3-5], neural approaches [6] and boundary based approaches [7]. Among the three categories boundary based approach [7] is preferred due to its simplicity compared to computational complexity and system structure of other two categories. To address pixel misclassification issue in switching median filters at high density noise, a noise adaptive soft switching median filter (NASM) was proposed [8], which consists of a three level hierarchical soft switching process. The boundary based approach called boundary discriminative noise detection [7] is very good in detecting impulse noises of various densities. BDND [7] can handle image corruption with even up to 90% noise density.

In noisy pixel replacement stage, many median based schemes have been proposed. The NASM [8] noise replacement strategy gives robust performance in removing impulse noise while preserving signal details across a wide range of noise densities, ranging from 10% to 50%. However, for those corrupted images with noise density greater than 50%, the quality of the recovered images become significantly degraded, due to the sharply increased number of misclassified pixels. The modified NASM [7] along with BDND detector gives better performance for high density impulse noises compared to filtering in NASM [8] but causes blurring of edges and loss of finer details in the image. To improve the visual quality of the restored image a noise adaptive weighted median filter is proposed in this paper.

The proposed method uses BDND [7] detection method for determining the corrupted pixels in the noisy image and then a noise adaptive weighted median filter is applied on each corrupted pixels. The filtering window size is chosen adaptively and depends on the percentage of noise that corrupts the image. The initial filtering window size is taken as 3×3 and maximum window size is chosen depending on the percentage of noise that corrupts a local region around the detected noisy pixel in the image. The noisy pixels are replaced by the weighted median value of uncorrupted pixels in the filtering window. The proposed method is giving significant improvement in visual quality of the image especially in the case of high density impulse noises and the thinner edges and finer details in the images are restored more effectively than previous methods.

This paper is organized as follows. In section 2 there is a brief introduction about the impulse noise models used in the paper. The section 3 describes the proposed method in detail. Section 4 describes various performance measures used. The section 5 contains the results and discussion and section 6 contains conclusion.

2 Noise Models

In this section four different random impulse noise models are discussed. These models [8] are considered for extensively examining the performance of the proposed filter considering practical situations.

Noise Model 1: Noise is modeled as salt-and-pepper impulse noise, pixels are randomly corrupted by two fixed extreme values, 0 and 255 (for gray level image), generated with the same probability. That is, if P is noise density, then P_1 , the noise density of salt is $P/2$ and P_2 , the noise density of pepper is also $P/2$.

Noise Model 2: This is similar to Noise Model 1, but here each pixel might be corrupted by either pepper noise or salt noise with unequal probabilities. That is $P_1 \neq P_2$.

Noise Model 3: Instead of two fixed values, impulse noise could be more realistically modeled by two fixed ranges that appear at both ends with a length of m each, respectively. That is, $[0, m]$ denotes salt and $[255-m, 255]$ denotes pepper. Here for noise density P , $P_1 = P_2 = P/2$.

Noise Model 4: This is similar to Noise Model 3 but here probability densities of low intensity impulse noise and high density impulse noise are different. That is, $P_1 \neq P_2$

3 Noise Adaptive Weighted Switching Median (NAWSM) Filter

The proposed method NAWSM filter is a double stage filter, where initially it will perform a detection stage for identifying the noisy pixel positions. When a noisy pixel is detected it is subjected to the second stage known as filtering stage. For detection purpose we are using the Boundary Discriminative Noise Detection method proposed by P.-E Ng et.al [7].

3.1 Noise Detection Stage

The basic strategy of BDND [7] is to examine each pixel in its neighborhood from coarse to fine. If current pixel is categorized into “corrupted” in both “coarse” and “fine” stages, then it is considered to be contaminated. So the most critical part of this paradigm is the determination of the decision boundaries. Here two decision boundaries b_1 and b_2 are calculated, and depending on that, pixels are classified in to three clusters – low intensity impulse noise, uncorrupted pixels and high density impulse noise. If $x(i,j)$ is the intensity of noisy image pixel under consideration and if $x(i,j) \leq b_1$, it is classified as low intensity impulse noise, if $x(i,j) > b_2$, it is classified as high intensity impulse noise and otherwise it is classified as uncorrupted pixel.

The following steps are used in the detection stage [9].

- 1) For each pixel $x(i,j)$ in the image, impose a 21×21 window centers on $x(i,j)$.
- 2) Sort the pixels lie in the window region to an ordered vector v_o and find the median med .
- 3) Compute the difference vector of v_o .
- 4) Find the pixels which correspond to the maximum differences in the intervals of $[0, med]$ and $(med., 255]$. And set these two pixels’ intensities as the decision boundaries b_1 and b_2 respectively.

5) Classify the pixels in current window into three clusters according to the rule - if $x(i,j) \leq b1$, it is classified as low intensity impulse noise, if $x(i,j) > b2$, it is classified as high intensity impulse noise and otherwise it is classified as uncorrupted pixel.

6) Validate the noisy candidates by imposing a 3x3 window, and repeat steps 2)–6).

After such an application to the entire a two-dimensional binary decision map is formed at the end of the noise detection stage, with ‘1s’ indicating the positions of ‘uncorrupted’ pixels and ‘0s’ for those of ‘corrupted’ ones.

$$N(i, j) = \begin{cases} 0, & x(i, j) = \text{noisy pixel} \\ 1, & \text{otherwise} \end{cases}$$

This noise mask generated is used in the filtering stage by the proposed Noise Adaptive Weighted Switching Median filter (NAWSM).

3.2 Noise Filtering Stage

The proposed NAWSM filter is an improvement over modified NASM (MNASM) filter explained along with BDND method [7]. The restored image obtained by the application of MNASM filter suffers from smoothening of edges and loss of finer details in the image. But the proposed method uses a weighted median filter which produce restored image with more prominent edges and with retained finer details in the image.

The NAWSM filter is applied to each pixel identified as noisy during detection stage. The noise mask or the binary decision mask generated by the BDND [7] detection scheme will contain a value ‘0’ in the noisy pixel position. First step in the filtering stage is to select appropriate window size for the filtering to achieve better results.

The proposed NAWSM uses a square filtering window $W_F \times W_F$ with odd dimension as follows.

$$W_F(i, j) = W_{2s+1}(i, j) = \{X(i+m, j+n)\}, \text{ where } m, n \in (-s, \dots, 0, \dots, s).$$

In order to determine the filter’s window size $W_F \times W_F$, the limit of the maximum filtering window size $W_D \times W_D$ is to be determined first. It is determined by considering local noise distribution around each corrupted pixel. In the case of MNASM filter it was determined globally from the binary decision map obtained from the detection stage. The local estimation of noise density is found to be giving an optimum value for $W_D \times W_D$.

For determining the limit of the maximum filtering window size $W_D \times W_D$, a window of size $K \times K$ around each corrupted pixel is considered and percentage of noise that corrupts the image in that local window is calculated. This can be calculated by counting the number of zeros in the local region of the binary decision map that corresponds to noisy image and dividing it by total pixels in the $K \times K$ local window. The calculation of noise density p , for a local window $K \times K$ is given as shown below

$$p = \frac{\sum_{i,j=1}^k Zc_k(i, j)}{K^2}$$

where Zc_k is the zero count in $K \times K$ window and is given as

$$Z_{ck}(i, j) = \begin{cases} 1, & N_k(i, j) = 0 \\ 0, & \text{Otherwise} \end{cases}$$

where N_k is the binary decision mask that corresponds to $K \times K$ window in noisy image.

After calculating noise density p , the limit of the maximum filtering window size, $W_D \times W_D$ is determined according to Table 1. Table 1 is empirically established based on multiple test images, in which different window sizes are suggested for different noise-density levels of corruption estimated.

Table 1. Window size for different noise densities

Noise Density	$W_D \times W_D$
$0\% < p \leq 20\%$	3×3
$20\% < p \leq 40\%$	5×5
$p > 40\%$	7×7

After finding out the maximum window size $W_D \times W_D$, filtering is done as follows. Initial filtering window size $W_F \times W_F$ is set as $L \times L$. The filtering window is then iteratively extended outwards by one pixel in all the four sides of the window, provided that the number of uncorrupted pixels (denoted by N_c) is less than half of the total number of pixels (denoted by $Sin = 1/2[W_F \times W_F]$) within the filtering window, while $W_F \leq W_D$. Filtering window will also be extended when the number of uncorrupted pixels is equal to zero in that window. That is, while (($N_c < Sin$) and ($W_F \leq W_D$)) or $N_c=0$), window will be extended outwards in all the four sides of the window. Then noisy pixels are replaced by a weighted median value of the uncorrupted pixels in the filtering window. Weight value assigned to each uncorrupted pixel depends on its closeness to the central pixel.

The weight value $w_{s,t}$ assigned to each uncorrupted pixel, that corresponds to the position (s, t) of the filtering window $W_F \times W_F$ is defined as follows.

$$w_{s,t} = \begin{cases} 3, & (s,t) \in \Omega^3 \\ 2, & (s,t) \in \Omega^5 \wedge (s,t) \notin \Omega^3 \\ 1, & \text{otherwise} \end{cases}$$

where Ω^n , denotes the $n \times n$ neighbourhood around the central pixel and is defined as below

$$\Omega^n = \{(s,t) / -n \leq s, t \leq n\}$$

If the filtering window is reached to its maximum size and N_c , the number of uncorrupted pixels in the filtering window is equal to zero ($N_c = 0$), then the central noisy pixel is replaced by performing a conventional median filtering on already restored pixels in the filtering window $W_F \times W_F$.

So each noisy pixel $X(i,j)$ in the image is restored by a value $Y(i,j)$ defined as follows.

$$Y(i, j) = \begin{cases} \text{median}\{w_{s,t} \hat{\Delta} X(i-s, j-t) / (s, t) \in W\} & \text{with } N(i-s, j-t) = 1, \quad N_c > 0 \\ \text{median}\{X'(i-s, j-t) / (s, t) \in W\}, & N_c = 0 \end{cases}$$

where $W = \{(s, t) / -(W_F - 1) / 2 \leq s, t \leq (W_F - 1) / 2\}$ and $X'(i, j)$ represents already restored pixel at (i, j) th location of the filtering window.

4 Performance Measures

The performance of the restoration process is quantified using peak signal-to-noise ratio (PSNR), structured similarity index (SSIM) and image enhancement factor (IEF), is defined as follows.

$$PSNR = 10 * \log_{10}(255^2 / MSE)$$

$$MSE = \sum_{m,n} [O(m, n) - R(m, n)]^2 / (M * N)$$

$$SSIM = L(O, R) * C(O, R) * S(O, R)$$

$$L(O, R) = (2\mu_O \mu_R + C_1) / (\mu_O^2 + \mu_R^2 + C_1)$$

$$C(O, R) = (2\sigma_O \sigma_R + C_2) / (\sigma_O^2 + \sigma_R^2 + C_2)$$

$$S(O, R) = (\sigma_{OR} + C_3) / (\sigma_O \sigma_R + C_3)$$

$$C_1 = (K_1 * G)^2, C_2 = (K_2 * G)^2, C_3 = C_2 / 2$$

$$G = 255; K_1, K_2 \gg 1, (K_1 = 0.001, K_2 = 0.002)$$

$$IEF = \left(\sum_{m,n} [P(m, n) - O(m, n)]^2 \right) / \left(\sum_{m,n} [R(m, n) - O(m, n)]^2 \right)$$

where O is the original Image, R is the restored image, P is the corrupted image, MSE is the mean square error, $M \times N$ is the size of the image, L is the luminance comparison, C is the contrast comparison, S is the structure comparison, μ is the mean and σ is the standard deviation.

In this paper, we also used a qualitative-based performance measure named image quality index (IQI) to prove the efficiency of our proposed NAWSM algorithm. This universal objective image quality index was proposed by Wang and Bovik [10], which is easy to calculate and applicable to various image processing applications. IQI is designed by modelling any image distortion as a combination of three factors: loss of correlation, luminance distortion and contrast distortion.

$$IQI_j = Corr(O_w, R_w) * Lum(O_w, R_w) * Cont(O_w, R_w)$$

IQI is first applied to local regions using a sliding window approach with size 8×8 . At the j th step, the local quality index IQI_j is computed within the sliding window using the formula given above. O_w and R_w represent the sliding window of original and restored images, respectively. If there are a total of M steps, then the overall image quality index is given by $IQI = (1/M) \sum_j IQI_j$, where j varies from 1 to M . The dynamic range of IQI is $[-1, 1]$, and the best value 1 is achieved if and only if restored image R is equal to the original image O .

5 Results and Discussion

The performance of NAWSM filter has been evaluated qualitatively and quantitatively through experimental analysis. Although extensive simulations were carried out using standard test images, only performance evaluation using images such as Lena image of size 512×512 , Peppers image of size 384×512 Goldhill image of size 512×512 and Boat image of size 512×512 are explained in this section. The Fig.1 shows the test images along with their edge maps obtained using Sobel operator.

All the noise models described in the section 2 are used for measuring the performance of the proposed algorithm. For Noise Model 1 and Noise Model 3 images are corrupted with salt and pepper noise with different noise densities ranging from 10% to 90%. For Noise Model 2 and Noise Model 4 we consider different noise densities for salt and pepper. For Noise Model 3 and Noise Model 4 the low intensity values in the range 0-9 and high intensity values in the range 246-255 is used. In our experiment, we have taken value of K as 7, value of L as 3 and maximum value of W_F as 21.



Fig. 1. The First row shows standard test images Peppers, Lena, Boat and Goldhill used for performance evaluation of our proposed NAWSM filter. The second row shows their edge maps obtained by Sobel operator.

The Boat image corrupted by Noise Model 1 along with the restored images using modified NASM filter (MNASM) [7] and our proposed NAWSM filter for various noise densities are shown in Fig. 2. The Fig. 2 also contains the edge maps of restored images obtained by using Sobel operator. Fig. 3 and Table 2 shows various performance measures graphically and quantitatively. The corrupted Peppers image along with the restored images produced by modified NASM filter (MNASM) [7] and our proposed NAWSM filter for Noise Model 2 is shown in Fig.4. Table 3 shows quantitative measures of performance metrics. The noisy images generated using Noise Model 3 along with the restored images using modified NASM filter

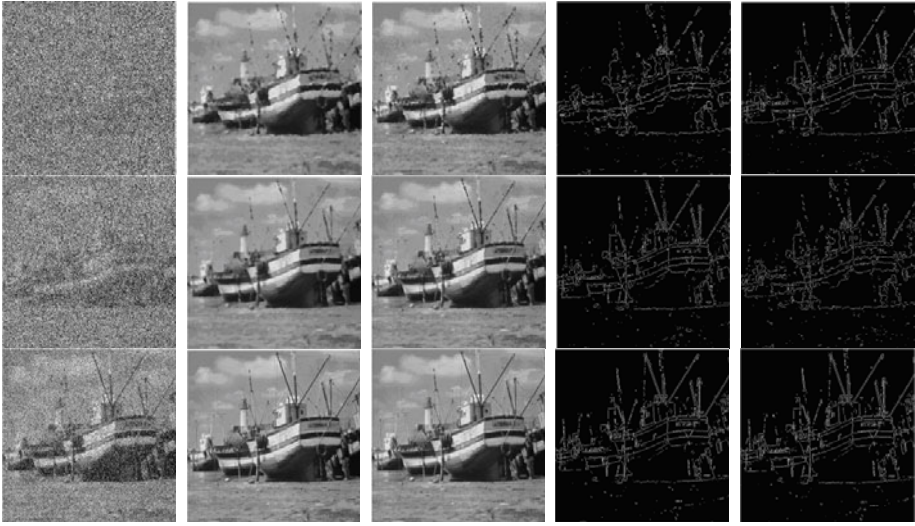


Fig. 2. The first column contains Boat image corrupted by Noise Model 1 with noise densities 90%, 70% and 30% respectively. The second column contains images restored by MNASM filter and third column contains images restored by proposed NAWSM filter. The fourth and fifth column shows edge maps (using Sobel operator) of restored images by MNASM and NAWSM respectively.

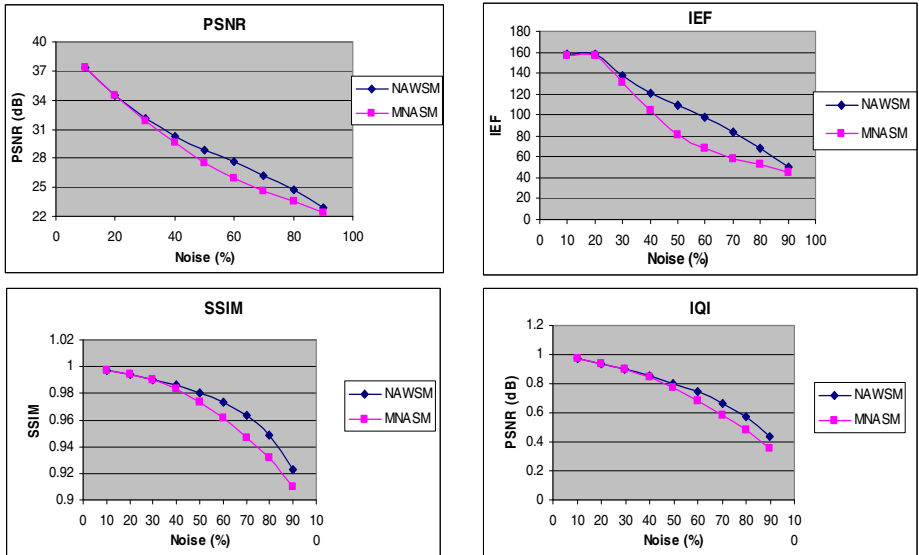


Fig. 3. Graphical representation of various performance measures for Boat image corrupted by Noise Model 1

(MNASM) [7] and our proposed NAWSM filter various noise densities are shown in Fig.5. The Fig.6 graphically demonstrates performance of NAWSM over MNASM in terms of quantitative measures such as PSNR, IEF, SSIM and IQI. The values of these measurements for various noise densities are given in the Table 4. . The corrupted Goldhill image generated using Noise Model 4 along with the restored images produced by the modified NASM filter (MNASM) [7] and our proposed NAWSM filter for various noise densities are shown in Fig.7. The comparison of various performance measures for restored images using MNASM filter and proposed NAWSM filter is shown in Table 5. It can be seen that in all cases there is a significant improvement in performance for NAWSM over MNASM for all the four noise models considered.

Table 2. The performance measures obtained for Boat image corrupted by Noise Model 1

Noise (%)	PSNR		IEF		SSIM		IQI	
	NAWSM	MNASM	NAWSM	MNASM	NAWSM	MNASM	NAWSM	MNASM
90	22.9743	22.4115	50.4965	44.3593	0.9226	0.9101	0.4391	0.3584
80	24.7634	23.6205	68.1491	52.3802	0.9488	0.9322	0.5698	0.48
70	26.2207	24.6718	83.1641	58.2167	0.9634	0.9471	0.6661	0.5854
60	27.6087	25.9988	97.979	67.6308	0.9736	0.9614	0.7413	0.6835
50	28.8722	27.5804	108.9475	80.9163	0.9803	0.9733	0.8021	0.7728
40	30.2706	29.6215	120.8469	104.0705	0.9858	0.9834	0.8523	0.8439
30	32.0875	31.903	137.2288	131.5206	0.9906	0.9902	0.897	0.8955
20	34.4578	34.4342	157.5554	156.6994	0.9946	0.9945	0.9364	0.9362
10	37.4172	37.4059	157.645	157.2348	0.9972	0.9972	0.9683	0.9683

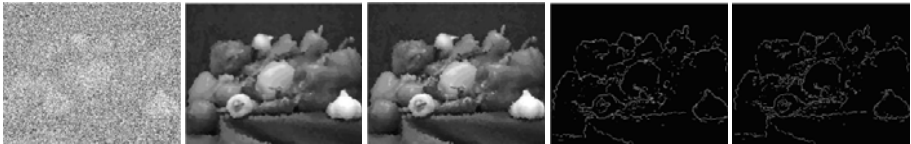


Fig. 4. First figure is the Peppers image corrupted by the Noise Model 2 (20% salt and 60% pepper), second and third figure shows the images restored by MNASM and NAWSM filter respectively and fourth and fifth figures shows their corresponding edge maps (obtained by Sobel operator).

Table 3. Performance comparison of MNASM filter with Proposed filter for Peppers image and Noise Model 2

Noise (%)	PSNR		IEF		SSIM		IQI	
	MNASM	NAWSM	MNASM	NAWSM	MNASM	NAWSM	MNASM	NAWSM
(Salt,Pepper)								
(20%,60%)	25.8187	26.6778	123.8491	150.942	0.9608	0.968	0.4621	0.5242
(15%,45%)	31.835	33.8044	369.455	581.429	0.99	0.9937	0.8023	0.8555
(10%,30%)	36.6686	37.3533	753.92	882.657	0.9967	0.9972	0.9155	0.9235
(5%,15%)	40.685	41.0711	956.193	10451	0.9987	0.9988	0.9667	0.9673

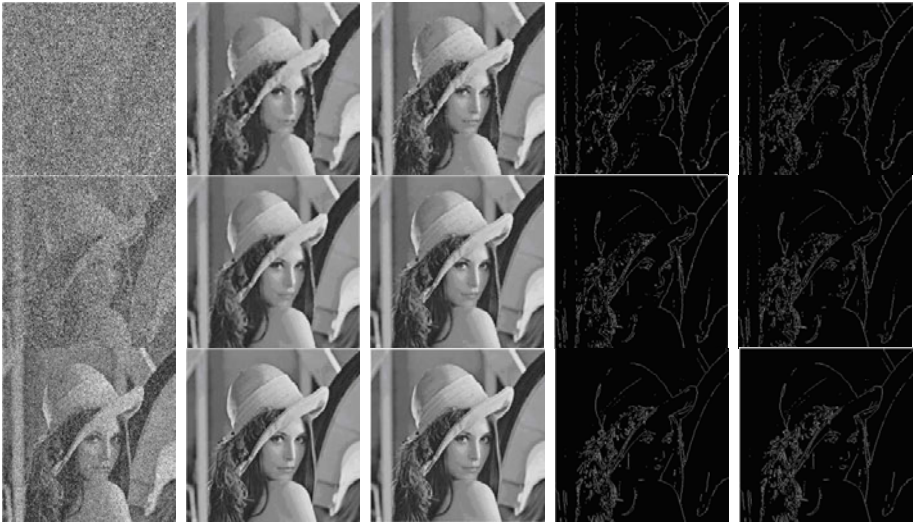


Fig. 5. First column contains Lena image corrupted by 90%, 70% and 30% impulse noise of Noise Model 3, second column shows image restored by MNASM filter, third column shows image restored by proposed NAWSM filter, fourth and fifth column shows edge map of restored images by MNASM and NAWSM respectively.

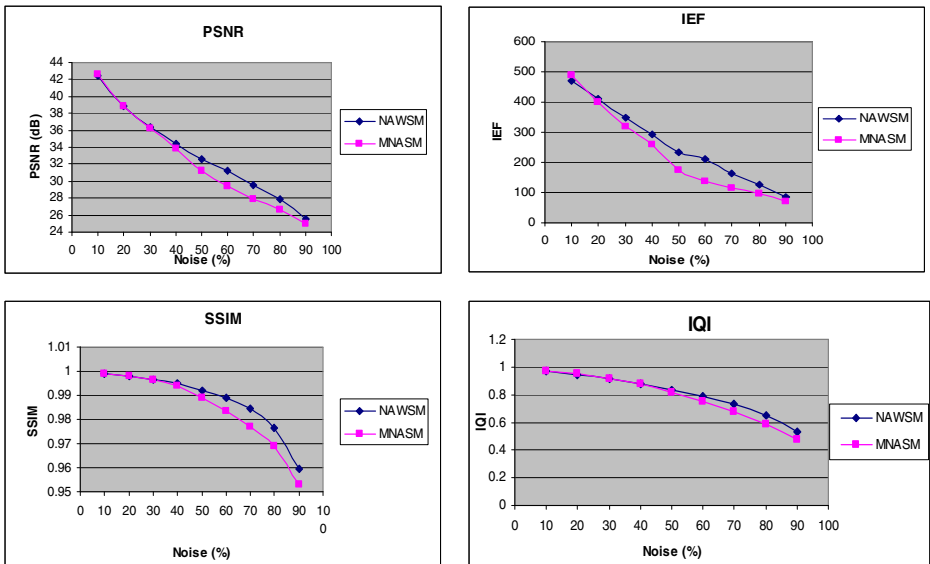


Fig. 6. Graphical representation of performance measures of Lena image corrupted by Noise Model 3

Table 4. Performance measures of Lena image corrupted by Noise Model 3

Noise (%)	PSNR		IEF		SSIM		IQI	
	NAWSM	MNASM	NAWSM	MNASM	NAWSM	MNASM	NAWSM	MNASM
90	25.5197	24.9038	86.033	70.9897	0.9597	0.9532	0.5268	0.4756
80	27.8906	26.6817	124.6788	97.0024	0.9767	0.969	0.6491	0.589
70	29.6107	27.9578	164.3794	115.6737	0.9843	0.9769	0.7295	0.6745
60	31.1646	29.427	210.7852	138.7926	0.989	0.9835	0.79	0.7493
50	32.5749	31.195	233.6758	172.5273	0.9921	0.9891	0.8374	0.8193
40	34.4142	33.8186	292.0192	260.1284	0.9948	0.994	0.8816	0.8758
30	36.3585	36.1799	346.8399	318.4747	0.9967	0.9965	0.9173	0.9162
20	38.9164	38.8374	410.1349	399.6762	0.9981	0.9981	0.948	0.9481
10	42.5845	42.4369	488.0409	470.5334	0.9992	0.9991	0.9747	0.9743



Fig. 7. First figure is the Goldhill image corrupted by the Noise Model 4(20% salt ,60% pepper),second and third figure shows the images restored by MNASM and NAWSM filter respectively and fourth and fifth figure shows the corresponding edge maps (obtained by using Sobel operator).

Table 5. Performance comparison of MNASM filter with Proposed filter for Pepper image and for Noise Model 4

Noise (%)	PSNR		IEF		SSIM		IQI	
	MNAS M	NAWS M	MNAS M	NAWS M	MNAS M	NAWS M	MNAS M	NAWS M
(20%, 60%)	23.755	24.1048	59.5747	64.5718	0.9433	0.9482	0.3601	0.4312
(15%, 45%)	28.1656	29.5656	118.095	163.018	0.9792	0.985	0.7052	0.7682
(10%, 30%)	31.8601	32.3137	183.207	203.376	0.9912	0.992	0.8666	0.8758
(5%, 15%)	36.1983	36.2206	261.116	262.462	0.9967	0.9967	0.947	0.9474

The efficiency of the proposed algorithm is proved using standard images and has been found that NAWSM produces better results than MNASM in terms of both quantitative measures such as PSNR, SSIM, IEF and qualitative measures such as image quality index (IQI). On analysing the various performance measures and edge maps of the restored images, obtained by applying both the filters, it is clearly evident that, our proposed NAWSM filter outperforms MNASM [7] filter.

6 Conclusion

In this paper, we proposed a new efficient Noise Adaptive Weighted Switching Median filter for the restoration of images that are corrupted with high density of

impulse noises based on different noise models. NAWSM is an improved switching median filter in which the filtering is applied only to corrupted pixels in the image while the uncorrupted pixels are left unchanged. The corrupted pixels are identified using BDND detection. The advantage of NAWSM filter is that it is retaining the edge information in the case of high density impulse noises.

References

1. Zhou, Y.-T., Chellappa, R., Vaid, A., Jenkins, B.K.: Image Restoration Using a Neural Network. *IEEE Transactions on Acoustics, Speech and Signal Processing* 36(7), 1141–1151 (1989)
2. Civicioglu, P.: Using Uncorrupted Neighborhoods of the Pixels for Impulsive Noise Suppression With ANFIS. *IEEE Transactions on Image Processing* 16(3) (March 2007)
3. Luo, W.: Efficient removal of impulse noise from digital images. *IEEE Transaction on Consumer Electronics* 52(2), 523–527 (2006)
4. Schulte, S., Nachttegael, M., Witte, V.D., Weken, D.V., Kerre, E.E.: A Fuzzy Impulse Noise Detection and Reduction Method. *IEEE Transactions on Image Processing* 15(5), 1153–1162 (2006)
5. Schulte, S., Witte, V.D., Nachttegael, M., Weken, D.V., Kerre, E.E.: Fuzzy Two-Step Filter for Impulse Noise Reduction From Color Images. *IEEE Transactions on Image Processing* 15(11), 3567–3578 (2006)
6. Nair, M.S., Raju, G.: A new fuzzy-based decision algorithm for high-density impulse noise removal. *Signal Image and Video Processing*, doi:10.1007/s11760-010-0186-4
7. Ng, P.-E., Ma, K.-K.: A Switching Median Filter With Boundary Discriminative Noise Detection for Extremely Corrupted Images. *IEEE Transactions on Image Processing* 15(6), 1506–1516 (2006)
8. Eng, H.-L., Ma, K.-K.: Noise adaptive soft-switching median filter. *IEEE Trans. Image Process.* 10(2), 242–251 (2001)
9. Duan, F., Zhang, Y.-J.: A Highly Effective Impulse Noise Detection Algorithm for Switching Median Filters. *IEEE Signal Processing Letters* 17(7) (July 2010)
10. Wang, Z., Bovik, A.C.: A universal image quality index. *IEEE Signal Processing Letters* 9(3), 81–84 (2002)
11. Nair, M.S., Revathy, K., Tatavarti, R.: An Improved Decision-Based Algorithm for Impulse Noise Removal. In: *International Congress on Image and Signal Processing - CISP 2008*, vol. 1, pp. 426–431. IEEE Computer Society Press, Los Alamitos (2008), doi:10.1109/CISP.2008.21

SMT-8036 Based Implementation of Secured Software Defined Radio System for Adaptive Modulation Technique

Sudhanshu Mehta^{*}, Surbhi Sharma^{**}, and Rajesh Khanna^{***}

Department of Electronics and Communication Engineering
Thapar University, Patiala, Punjab, India
sudhanshu123_mehta@yahoo.co.in,
{surbhi.sharma, rkhanna}@thapar.edu

Abstract. In this paper, we present the implementation and verification of an secured adaptive modulation system on a Software Defined Radio (SDR) development platform. The SDR platform is supported by highly productive development system emphasizing the value of a structured development process that takes the developer through modeling, system analysis, implementation and verification by bit error rate analysis. SDR platform is a massively parallel processor in which transmitter and a part of receiver is implemented on DSP processor and demodulator is implemented on FPGA. For adaptive modulator-demodulator architecture, a bit error rate analysis is presented. Result shows that for same BER, BPSK requires minimum SNR.

Keywords: SMT 8036, Adaptive SDR system.

1 Introduction

Wireless communication devices are composed of three main entities; signaling, physical hardware, and its functionalities [1]. These three main streams, which complement each other, have evolved since the invention of the radio transmission by Guglielmo Marconi. The primitive communications devices had very simple signaling, analog hardware, and limited functionality.

In time, each of these entities evolved significantly but J Mitola [2] had given a wonderful idea that if components that have typically been implemented on hardware (i.e. mixers, filters, amplifiers, modulators/demodulators, detectors. etc.) are instead implemented using software on a personal computer or other embedded computing devices which give birth to software defined radio. Simply Software Defined Radio is defined as "Radio in which some or all of the physical layer functions are software defined". The ultimate objective of Software Defined Radio (SDR) is to replace the entire analog signal processing in the wireless transceivers with digital signal

^{*} Student.

^{**} Assistant Professor.

^{***} Professor.

processing [3]. This had changed the approach of the communication engineers for designing the communication equipments and develop number of ways to remove the threats for software of software defined radios [4]. SDR system is a useful and adaptable future-proof solution to cover both existing and emerging standards e.g. While dealing with SDR, frequency band, air interface protocol and functionality can be upgraded with software download and update instead of a complete hardware replacement [5]. In this paper, adaptive software defined radio system for different modulation scheme is designed. In this receiver can select its demodulator in accordance with the signal transmitted by transmitter. The signal to be transmitted will depend upon operator's choice or it depends upon channel conditions [6]. For carrier frequency, fractional 'N' frequency synthesizer [7] is used.

Modulator circuit is implemented on DSP processor while demodulator circuit is implemented on FPGA processor because demodulator circuit is complex and we prefer complex circuit to be implemented on FPGA because component to component difference in FPGA is very less so it will take less time to demodulate.

This paper is organized as follows: System architecture is defined in section 2. Random number generator based Multi-modulation SDR transmitter and adaptive SDR receiver has been presented in section 2.1 and section 2.2 respectively. Platform used has been presented in section 3 and system level simulations and results are discussed in section 4 and finally conclusion and future scope is given in section 5.

2 System Architecture

System architecture includes a transmitter (which can transmit signal with modulation scheme like (FSK, BPSK, QPSK, 8PSK) and a receiver (which can detect the modulation scheme of received signal depending upon the frequency of received signal and detect the received signal). The brief explanation of transmitter and receiver is given below in next sections.

2.1 Random Number Generator Based Multi-modulation Software Defined Radio Transmitter

Figure (1) shows the block diagram of transmitter which is implemented on the DSP processor. Data input to transmitter is random generated bits, which is fed to data processing task. A random number generation block generates four values (one value after 10 sec) based upon which modulation scheme is selected as shown in table 1. There are two output lines of random number generation block (which has same values on it). Both output lines are used at the same time by the use of Threads in programming.

Table 1. Value of random number generation block and corresponding selected modulation scheme

S. No.	Value at the output of Random number generator	Modulation scheme
1	0	FSK
2	1	BPSK
3	2	QPSK
4	3	8-PSK

One of the output line of Random number generation block is given to selection switch, which selects the modulation scheme and data is given to data processing task where data is channel coded. After processing data is further modulated [8] (FSK, BPSK, QPSK, 8PSK) converted into analog signal. Modulation scheme changes dynamically by the change in the value of random number generation block. The other output line is connected to the software control up-converter which up-converts the modulated analog signal. Up-conversion is done for different modulation schemes in different RF frequency bands. These bands are specified in table 2. After baseband to RF conversion signal is passed through RF filter and then transmitted.

Table 2. Modulation scheme and their frequency bands

S.No.	Modulation scheme	Frequency band MHz
1	FSK	$\leq 2400 - 2425$
2	BPSK	$< 2425 - 2450$
3	QPSK	$< 2450 - 2475$
4	8PSK	$< 2475 - 2500$

This up-conversion is completed basically in three steps.

- 1) Conversion of baseband signal to first IF signal in the range 70 ± 8 MHz
- 2) Conversion of first IF signal to second IF signal in the range 374 ± 43 MHz
- 3) Then finally 2nd IF signal to the RF signal in range 2.4 to 2.5 GHz.

2.2 Adaptive SDR Receiver

Block diagram of adaptive software defined radio receiver is shown in figure (2). The signal will be received from tunable receiver antenna, which is fed to the DSP processor through ADC-2 where the frequency of the received signal will be detected. Based on the received signal frequency, DSP processor modifies the parameters of software tunable (ST) band pass filter, ST down-converter, ST amplifier and sampling rate of analog to digital convertor (ADC-1).

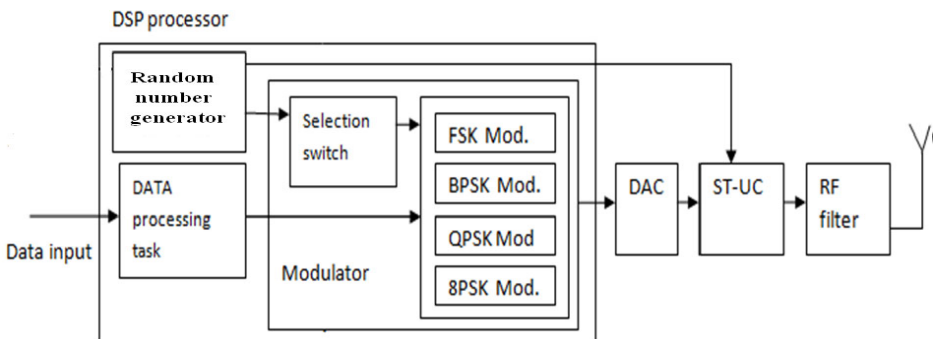


Fig. 1. Random number generator based Multi-modulation software defined radio transmitter

For the same received signal frequency, DSP processor activates the demodulator circuit (which is implemented on FPGA vertex (II)). The detection of received signal is done through selected demodulator circuit.

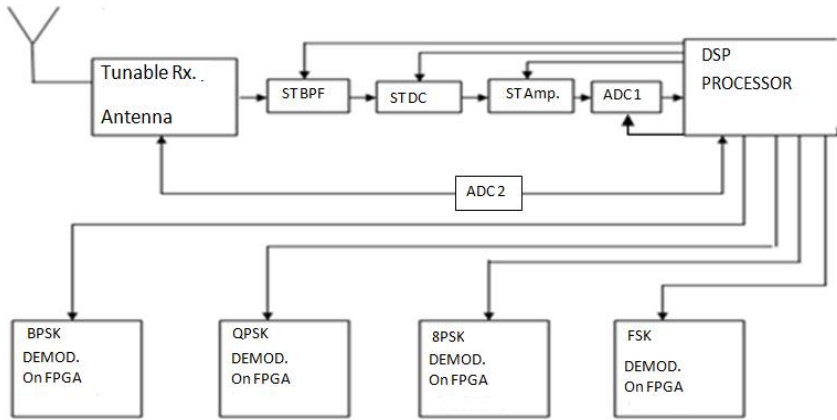


Fig. 2. Adaptive software defined radio receiver structure



Fig. 3. SMT 8036 module

3 Platform Used

There are many SDR Platforms given in the literature in which SDR is implemented on general purpose processors. Above described Adaptive software defined radio system for any modulation scheme is developed using hardware produced by Sundance SMT 8036 kit shown in figure 3. It includes C6416 (DSP Processor) and FPGA virtex II 2000 based module ‘SMT-365’, combined with a virtex II 1000 and

dual high-speed ADC/DAC module SMT-370, both plugged on a carrier board SMT310Q. Receiver calculations are complex so they are implemented on FPGA, which will give results very fast as compared to general purpose processors. It also has an additional IF/RF front end SMT-349 module with two omni-directional antennas working at 2.4GHz with 100MHz bandwidth. The benefits of system based on this PCI carrier board approach include low-cost, scalable system configuration, flexible task allocation and above all expandable system which allow future expansions.

The programming environment for SMT-8036 is given by 3L Diamond where as Programming language supported is 'C' on DSP and 'VHDL' on FPGA.

4 System Level Simulations

In order to characterize the performance of the adaptive modulation techniques, a SDR based transmitter - receiver model has been constructed. The channel is assumed to be Gaussian and the noise considered as white. To evaluate the BER for different modulation schemes for each value of SNR, a frame consisting of 20000 bits is sent from the transmitter having random values 0 and 1. The results below have been obtained from 100 independent iterations of each such frame.

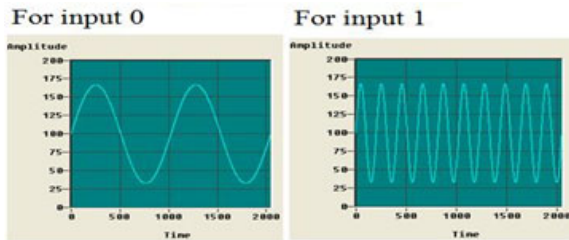


Fig. 4. FSK modulation (baseband)

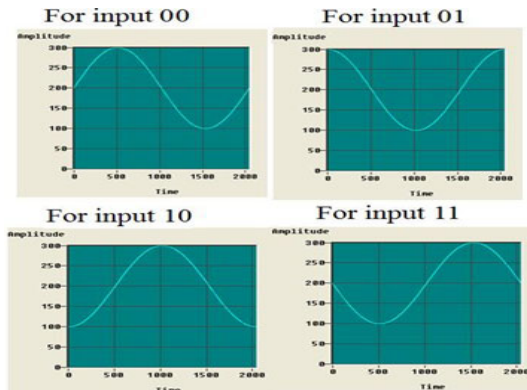


Fig. 5. QPSK modulation (baseband)

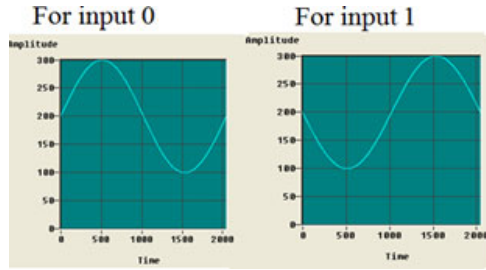


Fig. 6. BPSK modulation (BPSK)

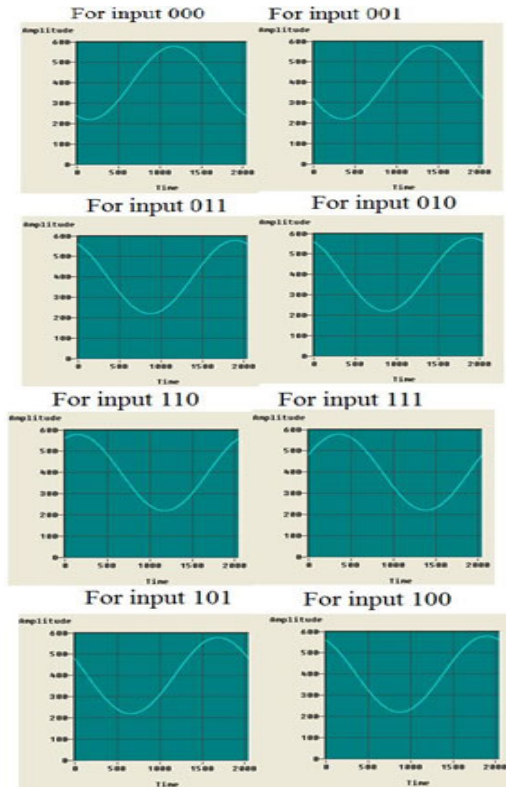


Fig. 7. 8PSK modulation (baseband)

Initially, random number is generated by random number generation block and modulation scheme is selected. After that if modulation scheme selected is FSK then 1 bit per symbol period is selected to transmit. Then if the input data is 0 then signal of less frequency will be transmitted and if input data is 1 then signal of high frequency will be transmitted. These frequencies are corresponding to the mark and space frequencies. This can be seen in figure 4.

In QPSK, two bits per symbol period is selected to transmit. Then if the input data is 00 then signal of phase shift 0 will be transmitted and if input data is 01 then signal with phase shift 90 will be transmitted. Similarly we have the outputs of other 2 two inputs (i: e for 10 and 11). This can be seen in figure 5.

In BPSK, one bit per symbol period is selected to transmit. Then if the input data is 0 then signal of phase shift 0 will be transmitted and if input data is 1 then signal with phase shift 180 will be transmitted. This can be seen in figure 6.

Similarly for 8-PSK, three bits per symbol period is selected to transmit. So we have 8 signals of different phase shifts. This can be seen in figure 7.

At the receiver end, the coded signal is received at the tunable received antenna. The signal is divided into decision regions, one for each transmitted symbol. An error occurs if a symbol is received in a different region from the one it was transmitted. Thus, the maximum allowed error is equal to half the distance between two adjacent symbols. Recovered signal is sent to the Adaptive software defined radio receiver structure where decoding of signal is done. After decoding, the uncoded signal is considered as a reference signal and the decoded signal is compared with it and BER is evaluated. Figure 8 shows the adaptive SDR receiver based BER performance for different modulation schemes. From the figure 8 it is clear that for same BER, BPSK requires minimum SNR.

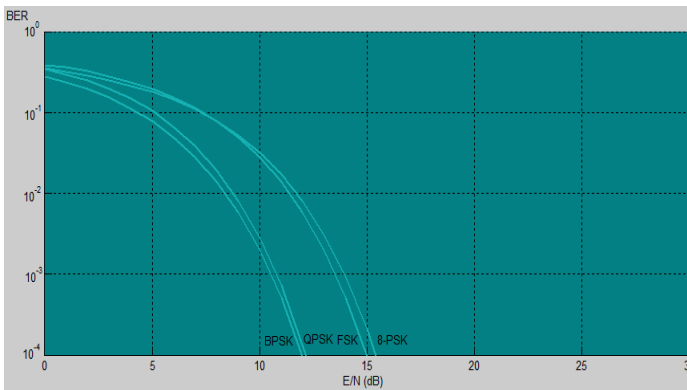


Fig. 8. BER Comparison for different modulation schemes in Adaptive SDR receiver

5 Conclusion and Future Scope

The SMT-8036 based implementation of secured Software Defined Radio system for adaptive modulation technique is proposed and simulated. System here presented is highly secured because Random number generator block will generate number after every ten seconds. This can be detected only by the adaptive SDR receiver purposed in section 2.2 which selects a different demodulator circuit dynamically depending upon modulation of signal received. Other receiver circuits can not demodulate this because they cannot detect the correct modulation scheme selected at specific time. Simulation studies are conducted to evaluate the performance of the proposed system.

This product will be able to detect and demodulate the signals with adaptive modulation schemes. The above described SDR system can be used to transmit secret information in military applications with very high security.

References

- [1] Arslan, H.: *Cognitive Radio, Software Defined Radio, and Adaptive Wireless Systems*. Springer, Heidelberg (2007)
- [2] Mitola, J.: *The Software Radios—Survey, Critical Evaluation And Future Directions*. In: *Proceedings of National Telesystems Conference (NTC 1992)*, pp. 15–23. IEEE, Los Alamitos (1992)
- [3] Vinod, A.P., Lai, E.M.-K., Omondi, A. (eds.): *Special Issue On Signal Processing For Software Defined Radio Handsets*. *Journal of Signal Processing System* 42(8), 2 (2009)
- [4] Murotake, D., Martin, A.: *A High Assurance Wireless Computing System (Hawcs™) For Software Defined Radio*. In: *Proceeding of the SDR 06 Technical Conference and Product Exposition*, pp. 116–118 (2006)
- [5] Saravanan, M., Ravi, S.: *SDR Based Multiple Access Technique For Multimedia Applications*. In: *Proceedings of Communications and Information Technologies*, pp. 491–494 (2007)
- [6] Tarniceriu, A., Iordache, B., Spiridon, S.: *An Analysis On Digital Modulation Techniques For Software Defined Radio Applications*. In: *Proceedings of Semiconductor Conference*, pp. 571–574 (2007)
- [7] Osmany, S.A., Herzel, F., Christoph Scheytt, J.: *An Integrated Fractional-N Frequency Synthesizer For Software-Defined Radio Applications*. In: *Proceedings of Silicon Monolithic Integrated Circuits in RF Systems (SiRF)*, pp. 243–246 (2010)
- [8] Proakis, J.G.: *Digital Communications*, 4th edn. McGraw-Hill, New York

Abstraction of Exudates in Color Fundus Images

Richu Paul and S. Vasanthi

Department of Electronics and communication
K.S.R College of Technology, Tiruchengode
richupaul1@yahoo.com, vasanthiramesh@gmail.com

Abstract. Diabetic retinopathy is a major cause of blindness. Earliest signs of diabetic retinopathy are damage to blood vessels in the eye and then the formation of lesions in the retina. This project presents a method for the detection of abnormalities in the retina such as the exudates in retinopathy images using computational intelligence techniques. The color retinal images are segmented using fuzzy c-means clustering following color normalization and contrast enhancement. For classification of these segmented regions into exudates and nonexudates, a set of initial features such as color, size, edge strength, and texture are extracted. A genetic based algorithm is used to rank the features and identify the subset that gives the best classification results. Using a multilayer neural network classifier, the selected feature vectors are then classified.

Keywords: Fuzzy c-means (FCMs), Kmeans algorithm, Gabor filters, Neural Networks (NNs).

1 Introduction

Diabetic retinopathy is the most common diabetic eye disease and a leading cause of blindness. It is caused by changes in the blood vessels of the retina. Unfortunately, in many cases the patient is not aware of any symptoms until it is too late for effective treatment. Early detection of DR through screening can help prevent blindness and vision loss. For this, computer aided diagnosis technology is essential. As more than 90% of screening images in a typical screening program do not contain abnormalities, the focus given here is on the development of a DR pre-screening system. This system should select only those images that might contain abnormalities associated with DR and present these to an ophthalmologist for further evaluation.

Diabetic retinopathy can lead to several retinal abnormalities, including microaneurysms, haemorrhages, cotton wool spots, and exudate. Here main focus is on exudates because it provides information about early diabetic retinopathy. The main cause of exudate is proteins and lipids leaking from the bloodstream into the retina through damaged blood vessels. In retinal images, exudate manifests as hard white or yellowish localized regions with varying sizes, shapes, and locations. In this paper, the retinal images with exudates are to be automatically discriminated from normal images. This system can be used by non experts to identify exudates and can realize diabetic retinopathy in its earliest stage itself.

1.1 Related Works

Gardner [2] used a neural network (NN) classifier to identify the exudates in grey level images. The authors reported a sensitivity of 93.1%. This was the result of classifying whole 20×20 regions rather than a pixel level classification. One novelty of the proposed method is that here exudates are located at pixel resolution rather than estimate for regions. Phillips investigated exudates detection and identification [1]. This paper provides lesion-based sensitivity between 61% and 100% based on 14 images. This method had a drawback that other bright lesions (such as cotton wool spots) could also be identified as exudates.

By using a minimum-distance discriminant classifier to identify the retinal bright lesions such as exudates and cotton wool spots, Wang [3] addressed the same problem. This work, which used image-based diagnosis, has 100% sensitivity and 70% specificity based on a dataset of 150 images. But in this work, some of other abnormalities were also incorrectly classified as exudates. Hunter [4] conducted another NN-based exudates detection work. The NN was trained to distinguish exudates from drusen based on 16×16 pixel patches. The reported performance was based on whether each 16×16 pixel patch contains exudates, and no image-based and pixel-level validation was reported.

Using selected threshold values, Sinthanayothin [5] applied a recursive region growing technique in gray-level images. In this paper, the processed retinal images are only including exudates, and other bright lesions were not considered. The author reported a lesion-based accuracy of 88.5% sensitivity and 99.7% specificity for the detection of exudates.

Exudates were identified from green channel of the retinal images according to their gray-level variation by Walter [6]. This algorithm also did not discriminate exudates from cotton wool spots. Niemeijer [7] extracted the bright lesions, i.e., exudates, cotton wool spots, and drusen from color retinal images. The author reported a pixel resolution accuracy of 70.0% sensitivity, and 88.0% specificity for the detection of exudates,

Past works based on detection of exudates mainly relied on gray-level information, and most of them were not evaluated on large datasets or failed to give good results for large numbers of images as encountered in a screening process. The majority of these methods were concentrated only in terms of either lesion-based [1]–[2], [4]–[5] or image-based [3] criterion. There were lots of errors reported in all those papers which were based on patch resolution accuracies due to the small areas, which some exudates could occupy. The diagnostic accuracy of the proposed method is in terms of both pixel resolution and image-based. It will overcome the disadvantages of past works which was based on lesion-based accuracy as it cannot find exudates occupy in borders. The proposed method also overcomes the drawback of the effects of misclassification errors for individual patches when the performance is based on patch resolution.

2 Proposed Method

2.1 Image Acquisition

The input images were captured using 24 bit per pixel at a resolution of 800×602 pixels by a Cannon nonmydriatic CR6-45NM camera. Of the 60 images in the dataset,

30 are of patients with no pathologies (normal) and the rest of the images are abnormal (contain pathologies such as exudates, cotton woolspots, microaneurysms, and hemorrhages).

2.2 Preprocessing

Input images are made to undergo two pre-processing steps before commencing the detection of exudates as the input images will be have wide variations in the color of fundus from different patients based on their iris color and race. Image pre-processing involves contrast enhancement and color normalization. The first step is to normalize the color of the retinal images across the dataset. The histogram specification [9] technique was independently applied to each individual RGB channel to match the shapes of three specific histograms of the reference image. This is to modify the values of each image such that its frequency histogram matched that of reference image. Here, the reference histograms were taken from an image, which represents a frequent retinal pigmentation color among the image dataset. In the next preprocessing step, the contrast between the exudates and the retina background was enhanced to facilitate later segmentation.

Here local contrast enhancement is applied to distribute the values of pixels around the local mean. In this way, a pixel p in the center of a small running window w was changed to a new value p_n and Max and Min are the maximum and minimum values in the whole image, while μ_w and σ_w indicate the local window mean and standard deviation, respectively.

$$P_n = 255 * \left(\frac{[\varphi_w(p) - \varphi_w(\text{Min})]}{\varphi_w(\text{Max}) - \varphi_w(\text{Min})} \right) \tag{1}$$

Where

$$\varphi_w(p) = \frac{1}{1 + \exp\left(\frac{\mu_w - p}{\sigma_w}\right)} \tag{2}$$

The contrast enhancement improves the contrast of exudates lesions and may also enhance the contrast of some non exudate background pixels so that these pixels can wrongly be identified as exudate lesions. So a 3x3 median filter was applied to suppress the noise.

2.3 Retinal Image Segmentation

Retinal images are then segmented using a two segmentation algorithm based on Gaussian-smoothed histogram analysis and FCMs clustering. The first task in color image processing is to choose an appropriate representation using a color space definition. In a suitable color space, color pixels of interest can be clustered into well defined, less overlapping groups, which are easily bounded by segmentation algorithms in the color space.

At the image segmentation stage, the algorithm utilizes the histogram information of the three 1-D color components to estimate the number of valid classes by a

statistical evaluation of the histograms. The convolution of a signal (histogram) f and a smoothing kernel $g(x,\sigma)$ has to be done, then the smoothed signal at the scale σ can be written as:

$$F(x, \sigma) = f(x) * g(x, \sigma) \tag{3}$$

where $*$ denotes convolution. Having smoothed the image’s histograms based on Gaussian function, the coarse stage begins to segment the image using the located thresholds. Find histograms of each pixels and peak and valley thresholds are to be calculated. The valleys are obtained by computing the first and second derivative of each 1-D histogram, and find the locations that satisfy the following equation:

$$F_g(x, \sigma) = f(x) * \frac{dg(x, \sigma)}{dx} = 0, F_{xx}(x, \sigma) > 0 \tag{4}$$

The histogram peaks represent the number of clusters, and are localized as follows:

$$F_g(x, \sigma) = f(x) * \frac{dg(x, \sigma)}{dx} = 0, F_{xx}(x, \sigma) < 0 \tag{5}$$

A zero crossing in $f(x)$ denotes a local maximum (peak) of $f(x)$ when the F_{xx} is negative) and a local minimum (valley) when the F_{xx} is positive. By projecting each color image onto its three color components (L , u , and v), one can obtain the number of peaks in each colour component, i.e., N_L , N_u , and N_v .

Kmeans algorithm starts with some clusters of pixels in the feature space, each of them defined by its center. The first step consists in allocating each pixel to the nearest cluster. In the second step the new centers are computed with the new clusters. These two steps are repeated until convergence. By using Kmeans clustering, all peaks of 3 channels are clustered into one cluster. Similarly all valleys and intermediate pixels into separate clusters. Separating Luv space into several haexahedra can be obtained as Cartesian products of peak intervals for each color component is given as:

$$\text{Cluster}(n) = \begin{matrix} T \text{ High}(L_i) & T \text{ High}(u_i) & T \text{ High}(v_i) \\ \sum L(x, y)\Lambda & \sum u(x, y)\Lambda & \sum v(x, y) \\ L=T_low(L_i) & u=T_low(u_i) & v=T_low(v_i) \end{matrix} \tag{6}$$

where n is an index for hexahedral partitioned clustered (i.e., $n = 1, \dots, N_{max}$), L , u , and v are the image’s color components, and Λ is a logical AND operator

In the fine stage of segmentation, FCM assigns any remaining unclassified pixels (pixels from ambiguous regions) to the closest cluster based on a weighted similarity measure between the pixels in the image and each of cluster centers. Local extrema of this objective function are indicative of an optimal clustering of the image.

The objective function [10] is minimized when high membership values are assigned to pixels whose values are close to the centroid for its particular class, and low membership values are assigned when the pixel data are far from the centroid.. Hence, in this paper, there is no need to recompute the class centers for exudate and nonexudate classes by using, and instead they are considered as sufficiently well approximated within the coarse segmentation phase. At this stage, pixels from ambiguous regions are assigned to the remaining clusters.

2.4 Feature Selection

Once our color retinal images are segmented, each image is represented by its corresponding segmented regions. These regions, however, need to be identified in terms of exudates and nonexudates. This is attempted, in a bottom-up approach, by extracting a set of features for each region and classifying the regions based on the generated feature vectors. A 2-D Gabor function g is the product of a 2-D Gaussian and a complex exponential function. When the Gaussian part is symmetric, we obtain the following isotropic Gabor function:

$$g_{\theta, \lambda, \sigma}(x, y) = \exp\{-x^2 + y^2 / 2\sigma^2\} \exp\{j\pi/\lambda(x \cos \theta + y \sin \theta)\} \quad (7)$$

A Gabor filter [12] bank is used. This bank contains 72 filters, 6 orientations (θ spanning from 0° up to 180° at steps of 15°), 3 wavelengths ($\gamma = 1.5, 2.5, 3.5$), and 4 scales ($\sigma = 3, 5, 7, 9$). The segmentation scheme begins by Gaussian filtering of Luv histograms of this image to derive the number of clusters and approximate the location of their centers. The segmentation algorithm requires two parameters, i.e., scale parameter (σ) which controls the smoothing extent of the histograms' fine structures and the secure-zone parameter which has a direct effect in evaluating the lower and higher threshold values. Here, the optimum and secure-zone values are very much experimentally tuned according to prior knowledge of retinal image characteristics and the conducted experiments of different values. The response of such a filter bank to an input image is a set of filtered images.

2.5 Region Level Classification

For the classification of the segmented regions, NNs discriminative classifiers are used. Here the optimum feature set is selected and a corresponding 65-D feature vector is computed for each segmented region. Experiment with hidden layers through a range of 2–20 hidden units to find the optimum architecture where all networks are trained using back propagation (BP) learning method. This BP uses scaled conjugate gradient algorithm for training images. A single output node gives the final classification probability, and the sigmoid activation functions are used in the hidden and output layers.

3 Results

Input images of about 30 abnormal images and 30 normal images were collected. Detection of exudates in a sample image is shown here. Both reference image [Fig. 1] and abnormal input image [for example, Fig. 2] are matched together forms the color normalization output given in Fig. 3(a). Contrast enhancement was done to improve contrast between foreground and background which is given in Fig. 3(b). Gaussian function shown in fig.3(c) has applied to smooth the curve. Then coarse segmentation and then fine segmentation is done. Fig. 3(d) is shown as classification of exudates/cotton wools, blood vessels/red lesions etc. In fine segmentation figure, it is given as the remaining unclassified pixels are assigned to the valid classes.



Fig. 1. Reference image

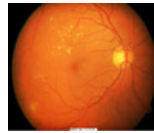
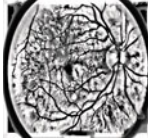


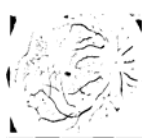
Fig. 2. Abnormal image



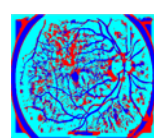
(a)



(b)



(c)



(d)



(e)



(f)

Fig. 3. a) Color normalized output (b) Contrast Enhancement output (c) Gaussian Smoothing output (d) Coarse segmentation output (e) Fine segmentation output (f) NN-based identified exudates

Table 1. Classification results

Classes	Number of pixels	Assigned color
Exudates	911.00	red
Non-exudates	477214.00	blue

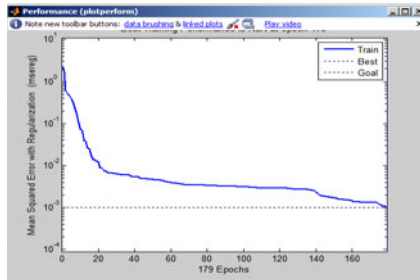


Fig. 4. Performance curve of training algorithm

All 30 images with $\sigma = 5$ and secure zone=0.4 gives the most accurate segmented results. From the table I, it is shown that there were total 478125 pixels. After fine segmentation, it is classified as 911 exudates pixels which are denoted as red and 477214 non-exudates pixels which are denoted as blue in color.

Among several NNs containing varying numbers of hidden units that are trained, a network, with 2 hidden units performs best in terms of the overall generalization ability. The proposed method can identify affected retinal images with 95.5% sensitivity (correct classification of 28 out of 30 abnormal images) while it recognizes 90.4% specificity (correct classification of 27 out of 30 normal images).

The performance curve of mean squared error with regularization with number of epochs shown as fig.4 is plotted.

4 Conclusion

Here a study of detecting retinal exudates using FCM segmentation and a back propagation classifier is presented. The best diagnostic accuracy for the identification of patients with evidence of retinopathy was 95.5% sensitivity and 90.4% specificity in terms of image based classification. The results shown here indicate that automated diagnosis of exudates in diabetic retinopathy based on color retinal image analysis was successful. Hence, the system could be used to evaluate digital retinal images which can be used in screening programmes for diabetic retinopathy. The main use of this paper is that even non-experts can identify exudates from its early stage and can take necessary treatments. The diagnostic accuracy of method in terms of both pixel-resolution and image-based criteria are evaluated. Here large majority of exudates were detected. The proposed system can correctly identify most of our normal images.

Acknowledgment. The authors would like to thank Dr. Freddy George who is working at Dr.Tonys' Superspeciality Eye Hospital, Kerala, for providing data base images.

References

1. Phillips, R., Spencer, T., Ross, P., Sharp, P., Forrester, J.: Quantification of diabetic maculopathy by digital imaging of the fundus: Eye, vol. 5, pp. 130–137 (1991)
2. Gardner, G., Keating, D., Williamson, T., Elliott, A.: Automatic detection of diabetic retinopathy using an artificial neural network: A screening tool. Brit. J. Ophthalmol. 80, 940–944 (1996)
3. Wang, H., Hsu, H., Goh, K., Lee, M.: An effective approach to detect lesions in retinal images. In: Proc. IEEE Conf. Comput. Vis. Pattern Recogn., Hilton Head Island, SC, vol. 2, pp. 181–187 (2000)
4. Hunter, A., Lowell, J., Owens, J., Kennedy, L.: Quantification of diabetic retinopathy using neural networks and sensitivity analysis. Proc. Artif. Neural Netw. Med. Biol., 81–86 (2000)
5. Sinthanayothin, C.: Image analysis for automatic diagnosis of diabetic retinopathy: Ph.D. dissertation, King's College of London, London, U.K (1999)

6. Walter, T., Klein, J., Massin, P., Erginary, A.: A contribution of image processing to the diagnosis of diabetic retinopathy, detection of exudates in colour fundus images of the human retina. *IEEE Trans. Med. Imag.* 21(10), 1236–1243 (2002)
7. Niemeijer, M., Ginneken, B.V., Russell, S.R., Suttorp, M., Abramoff, M.D.: Automated detection and differentiation of drusen, exudates and cotton-wool spots in digital color fundus photographs for diabetic retinopathy diagnosis. *Invest. Ophthalmol. Vis. Sci.* 48, 2260–2267 (2007)
8. Osareh, A., Mirmehdi, M., Thomas, B., Marham, R.: Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks. In: *Proc. Medical Image Understanding Analysis Conf.*, pp. 49–52 (2001)
9. Rafael, C.G., Richard, E.W.: *Digital Image Processing*, 3rd edn., pp. 443–445, 598–605. University of Tennessee (1992)
10. Dougman, J.: High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans. Pattern Anal. Mach. Intell.* 15(11), 1148–1161 (1993)
11. Osareh, A., Shadgar, B., Markham, R.: A Computational-Intelligence-Based Approach for Detection of Exudates in Diabetic Retinopathy Images. *IEEE Trans. Inform Tech. in Biomedicine* 13(4), 535–545 (2009)
12. Stelios, K., Vassilios, C.: A robust fuzzy local information c-means clustering algorithm. *IEEE Trans. Image Processing* 19(5), 1328–1338 (2010)

A Histogram Adaptation for Contrast Enhancement

Lisha Thomas and K. Santhi

Department of Electronics and communication
K.S.R College of Technology, Tiruchengode
lisha_pala@yahoo.co.in, santhiksr@gmail.com

Abstract. Histogram equalization is the common methods used for improving contrast in image processing application. But this technique is not well suited for implementation in consumer electronics such as television as it introduces unnecessary visual deterioration such as the saturation effect. It causes changes in the brightness of the input image. Thus, for the implementation of contrast enhancement it should be able to maintain the original input brightness in the output image. By adapting the input histogram, input brightness can be preserved. The adapted histogram can then be accumulated to map input pixels to output pixels. By introducing designed penalty terms, the level of contrast enhancement can be adjusted. Thus it is possible to generate a modified histogram which is closer to uniform histogram. Experimental results show a comparison of various quantitative measurements.

Keywords: Contrast enhancement, histogram equalization, histogram modification, image processing.

1 Introduction

Histogram equalization (HE) is a technique for enhancing the contrast of an image. This method maps the gray levels based on the probability distribution of the input gray levels. It flattens and stretches the dynamics range of the image histogram and resulting in overall contrast improvement. This can be applied in various fields such as medical image processing and radar image processing. HE is not commonly used in consumer electronics such as TV because it significantly changes the brightness of an input image and cause undesirable artifacts known as saturation effect. The brightness of the histogram-equalized output image is always lies in the middle gray level. Thus it does not preserve the input brightness.

Contrast enhancement provides a more visually pleasing image. There are many contrast enhancement techniques to improve the contrast of an image. These are classified into two groups such as direct methods and indirect methods. In direct methods includes a contrast measure and try to improve it. This method establishes a criterion of contrast enhancement and enhance image by improving contrast measure. Contrast can be measured globally or locally.

In indirect method modify the histogram by assigning new values to the original intensity levels. This method improves the contrast through using the under-utilized regions of the dynamic range without defining a specific contrast term. Indirect

method can enhance the contrast globally or locally. Indirect method can enhance the contrast globally or locally. In global histogram equalization method a single mapping function derived from input image is used. In local histogram equalization method the neighborhood of each pixel is used to obtain local mapping function. Although histogram equalization is an effective method, without adaptation result in excessive enhancement.

Various methods have been proposed for limiting the level of enhancement, most of which are obtained through adaptation on HE. In HE produces images with mean intensity approximately in the middle of the dynamic range. Thus it does not preserve original image brightness. To avoid this problem the first work is proposed by Kim in 1997 known as brightness preserving bi histogram equalization BBHE [1]. In this, two separate histograms from the same image are created and equalized independently. The first is the histogram of intensities that are less than the mean intensity and the second is the histogram of intensities that are greater than the mean intensity. Then equalize the two sub image separately. By doing this, the mean brightness of the resultant image will lie between the input mean and the middle gray level. In 1999, dualistic sub-image histogram equalization (DSIHE) has been proposed by Wan et al [2]. In which the two separate histograms were created using the median intensity instead of the mean intensity. DSIHE is better than BBHE in term of preserving an image's brightness and entropy.

Minimum mean brightness error bi histogram equalization (MMBEBHE) is proposed by Chen and Ramli to maintain the mean brightness [3]. In this method the differences between the mean brightness of the input and the mean brightness of the outputs that is produced from every separating intensity values are calculated. Then, a value is chosen that can produce the minimum difference between input and output means. The actual output image is obtained by bi-histogram equalize the input image with this value.

Chen and Ramli proposed an enhancement scheme known as recursive mean separate histogram equalization (RMSHE) [4]. The technique separates the histogram into two parts based on the average input brightness. Then, the mean of each piece of the resultant sub-histograms is calculated. These sub-histograms are again divided into two parts based on the average brightness values. This process is repeated for r times, where the value of r is set by the user. Then, each sub-histogram is equalized independently. RMSHE is good brightness preservation technique as the value of r is large, output mean converges to the input mean brightness.

There is also method for separating the histogram based on the shape of the histogram. Wongsritong et al [5], proposed multipeak histogram equalization with brightness preserving (MPHEBP). In this the histogram is first smoothed with one dimensional smoothing filter. Then, the histogram is divided based on the local maximums of the smoothed histogram. Each sub histogram is then independently equalized using histogram equalization.

In brightness preserving dynamic histogram equalization (BPDHE) partitions the histogram based on the local maximums of the smoothed histogram [6]. Before the histogram equalization taking place, this method maps each partition to a new dynamic range. As the change in the dynamic range cause the change in mean brightness, this method involves the normalization of the output intensity. So, the average intensity of the resultant image will be same as the input.

Although these methods are visually good than HE, these techniques cannot adjust the level of enhancement. These are not robust to noise, which become a problem when the histogram has spikes. To deal with histogram spikes, method known as histogram low-pass filtering [7] can be applied. Another method is proposed to adjust the cumulative function of the histogram thus can alter the level of enhancement [8]. The modified contrast enhancement techniques perform well on some images but they can create problems when a sequence of images is enhanced, or when the histogram has spikes. To avoid this problem it incorporates additional penalty terms into the bi-criteria optimization. It also presents a content-adaptive algorithm with low computational complexity.

2 Histogram Adaptation

To utilize the existing dynamic range, HE creates a uniformly distributed output histogram. For this it use a cumulated histogram as its mapping function. But HE produces excessively enhanced unnatural looking images. The problem with histogram equalization is due to large backward-difference values of transfer function. In order to avoid this input histogram should be adapted. The adapted histogram can be accumulated to map input pixels to output pixels. The resultant histogram is closer to a uniformly distributed histogram.

The adapted histogram is a solution of a bi-criteria optimization problem. This histogram is represented as \hat{h} that is closer to uniform histogram represented as u , make the residual $h-h_i$ small. In this h_i is the input histogram. The bi-criteria optimization problem can be formulated as a weighted sum of the two objectives as (1)

$$\min \|h-h_i\| + \lambda \|h-u\| \tag{1}$$

where h, h_i, \hat{h} and $u \in \mathbb{R}^{256 \times 1}$ and λ is a problem parameter. The various levels of contrast enhancement can be achieved by varying λ .

3 Methodologies

3.1 Adjustable Histogram Equalization

A systematic solution to bi-criteria optimization problem as in (1) obtained when the squared sum of the Euclidean norm is used as in (2).

$$\hat{h} = \operatorname{argmin} \|h-h_i\|_2^2 + \lambda \|h-u\|_2^2 \tag{2}$$

which results in the quadratic optimization problem

$$\hat{h} = \operatorname{argmin} (h-h_i)^T (h-h_i) + \lambda (h-u)^T (h-u) \tag{3}$$

The solution to (3) is

$$\hat{h} = (h_i + \lambda u) / (1 + \lambda) = (1 / (1 + \lambda)) h_i + (\lambda / (1 + \lambda)) u \tag{4}$$

The modified histogram \hat{h} , to be a weighted average of h_i and u , by changing the value of λ the level of enhancement can be modified.

When the value of λ is zero, the adjusted histogram is equal to the input histogram which is similar as the standard HE applied. The resulting image is over-enhanced, with many unnatural details. When λ value is increased to one, the penalty term is introduced and the enhanced image is similar to the original image. For the value of $\lambda = 2$, the level of enhancement is further decreased and the details are mostly preserved. Even if the stage of enhancement is decreased with increasing λ , the slope of the mapping at the fixed point is large.

The problem of having a large slope arises from spikes in the input histogram. One way to deal with histogram spikes is to use norm for the histogram approximation term. Hence, the problem in (2) is changed to

$$\mathfrak{H} = \operatorname{argmin} \|h - h_i\| + \|h - u\|^2 \tag{5}$$

To transform this mixed norm problem into a constrained quadratic programming problem, the first term can be expressed as a sum of auxiliary variables as (6)

$$\mathfrak{H} = \operatorname{argmin} [t^T I + \lambda(h-u)^T(h-u)] \tag{6}$$

subject to $-t \leq (h - h_i) \leq t$

Where $t \in \mathbb{R}^{256 \times 1}$ represents the auxiliary variables and $1 \in \mathbb{R}^{256 \times 1}$ of ones. But it has high computational complexity. Another way to deal with the histogram spikes in the input histogram is to use one more penalty term to measure the smoothness of \mathfrak{H} , which reduces the modified histogram's sensitivity to spikes

3.2 Histogram Smoothing

To avoid spikes that lead to strong repelling fixed points, a smoothness constraint is added. The backward difference of the histogram is used to measure its smoothness. A smooth modified histogram has fewer spikes as they are essentially abrupt changes in the histogram. The difference matrix is $D \in \mathbb{R}^{256 \times 256}$ bi-diagonal as given in (7)

$$D = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & -1 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix} \tag{7}$$

With the additional penalty term for smoothness, the optimal trade-off is obtained as in (8)

$$\min \|h - h_i\|_2^2 + \lambda \|h - u\|_2^2 + \gamma \|Dh\|^2 \tag{8}$$

The solution of this three-criterion problem is

$$\mathfrak{H} = ((1 + \lambda) I + \gamma D^T D)^{-1} (h_i + \lambda u) \tag{9}$$

While (4) results in a weighted average of h_i and u . It is further smoothed this weighted average to avoid spikes in (9). The first term in (9) is given as in (10).

$$S^{-1} = ((1 + \lambda) I + \gamma D^T D)^{-1} \tag{10}$$

It corresponds to a low-pass filtering operation on the averaged histogram. This can be expressed as (11)

$$S = ((1 + \lambda)I + \gamma D^T D) \tag{11}$$

where S is a tridiagonal matrix. Although histogram smoothing is good in avoiding histogram spikes, S^{-1} has to be computed for each image as γ to be adjusted based on the magnitude of the histogram spikes.

Instead of using (9), a low-pass filtering on the histogram can also be performed. But the number of taps and the transfer function must also be adaptive. An approach that is less computationally complex is to use a weighted error norm for the approximation error $h - h_i$.

3.3 Weighted Histogram Approximation

Histogram spikes occur because of the existence of large number of pixels with exactly the same gray-level values as their neighbors. This results in an input-output transformation that maps a narrow range of pixel values to a much wider range of pixel values. A large number of pixels having exactly the same gray-levels are due to large smooth areas in the image. So the average local variance of all the pixels with the same gray-level can be used to weight the approximation error. Histogram approximation error at the corresponding bin will be weighted with a smaller weight. Thus the modified histogram bin will not closely follow the input histogram's spike bin to minimize the approximation error. The objective function with the weighted approximation error is given in (12)

$$\min [(h - h_i)^T W (h - h_i) + \lambda (h - u)^T (h - u)] \tag{12}$$

Where $W \in \mathbb{R}^{256 \times 256}$ is the diagonal error weight matrix, and measures the average local variance of pixels with gray level. The solution of (12) is

$$\bar{h} = (W + \lambda I)^{-1} (W h_i + \lambda u) \tag{13}$$

This is computationally simpler than (9). Only simple division operations for the diagonal elements are needed to compute its inverse.

3.4 Low-Complexity Histogram Modification Algorithm

Low-complexity histogram modification algorithm deals with histogram spikes, performs B&W stretching, and adjusts the level of enhancement adaptively. In this dynamic range is better utilized which handle the noise visibility and the natural look requirements. Using histogram smoothing or weighted histogram approximation is computationally complex. Histogram smoothing requires either solving (9) or low-pass filtering with adaptive filter length and transfer function. Weighted approximation with solution given in (13) requires division operation.

3.4.1 Histogram Computation

To deal with histogram spikes instead of smoothing or weighting the input histogram, change the way a histogram is computed. Histogram computation can be modified so as to take pixels that have some level of contrast with their neighbors. For a good

contrast enhancement, the histogram should be modified in such that the modified histogram represents the conditional probability of a pixel, given that it has a contrast with its neighbors.

It can obtain by counting only those pixels that have contrast. To obtain the histogram, the local variation of each pixel can be used to decide if a pixel has sufficient contrast with its neighbors. For this use a horizontal variation measure by taking advantage of the row-wise pixel processing architecture. A horizontal one-lagged difference operation is a high-pass filter, which will measure noise. The horizontal two-lagged difference operation is a band-pass filter which will attenuate high-frequency noise signals. Histogram is created using pixels with a two-lagged difference that has a magnitude larger than a given threshold. The number of pixels included in the histogram is also counted for proper normalization.

3.4.2 Adjusting the Level of Enhancement

It is feasible to adjust the level of histogram equalization to achieve natural looking enhanced images. The modified histogram is a weighted average of the input histogram and the uniform histogram. The contribution of the input histogram in the modified histogram is $K^* = 1/(1+\lambda)$. The level of histogram equalization adjusted depending on the input image contrast. Low contrast images have narrow histograms and with histogram equalization noise created. Therefore k is computed to measure the input contrast using the aggregated outputs of horizontal two-lagged difference operation. This multiplied by a user-controlled parameter, is then normalized to the range $[0, 1]$ to get K . By choosing the maximum value that g_k can take on as a power of two, the normalization is done using a bit-shift operation. To ensure that h_i and u have the same normalization, u is obtained using the number of pixels that are included in the histogram. u_{\min} ensure that very low bin regions of the histogram will not result in very low slope in the mapping function.

Black and White stretching is then performed. Parameters b , w and α can be adapted with the image content, b and w is usually derived from the histogram as the minimum and maximum intensities. For noise robustness b should be chosen as the minimum gray-level that is bigger than some predefined number of pixel intensities, w can be chosen similarly. The stretching parameter should also be adapted with image content. For dark images white stretching can be favored, while for bright images black stretching can be favored.

4 Results and Discussion

Histogram equalized images result in the best utilization of the dynamic range of the pixel values for maximum contrast. For this low contrast image is given as input. First histogram equalization is applied on the input image. Histogram equalization creates a uniformly distributed output histogram by using a cumulated histogram as its mapping function. But as shown in the Fig.1 (b) it creates an excessively enhanced output image.

Adjustable histogram equalization is also applied on the input image to avoid the problem caused by HE. By varying the value of λ various level of contrast enhancement can be achieved. If the value of $\lambda = 0$ is similar as the output obtained in standard histogram equalization. As the value of λ increases it converge to preserve original image brightness in the output.

Examples of image and enhanced images using adjusted histogram equalization with three different values of λ (0, 1, and 2) are shown in Fig.1(c)-(e). The resulting image is over-enhanced, with many losses of details when normal HE is applied. When λ is increased to one, the penalty term comes into play and the enhanced image looks more like the original image, thus original image brightness preserved. For $\lambda=2$, the level of enhancement is further decreased and the details are mostly preserved.

Histogram smoothing is also applied on the input image. It is useful in avoiding histogram spikes in the input histogram is shown in the Fig.1 (f) .But in this method parameter value is adjusted with spikes in the input image.

Weighted histogram approximation for the input image is shown in the Fig.1 (g). This method is less computationally complex. Since the first term is a diagonal matrix, taking matrix inverse is avoided. Simple division operations for the diagonal elements are needed to compute its inverse.

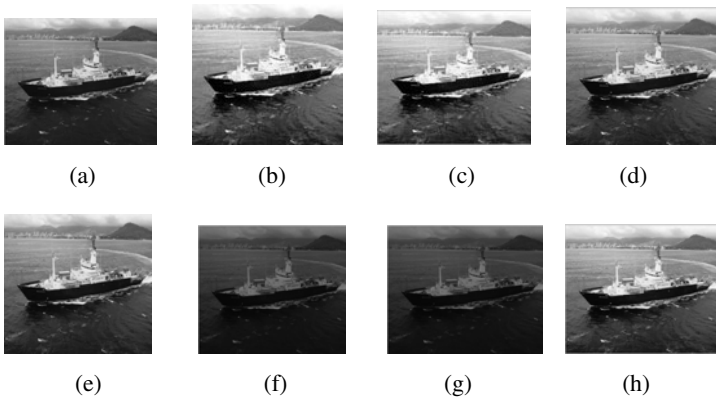


Fig. 1. (a)originalimage (b)equalizedimage (c)adjustable equalized with $\lambda=0$ (d) $\lambda=1$ (e) $\lambda=2$ (f) smoothing (g) weighted histogram approximation (h) low complexity method for boat image

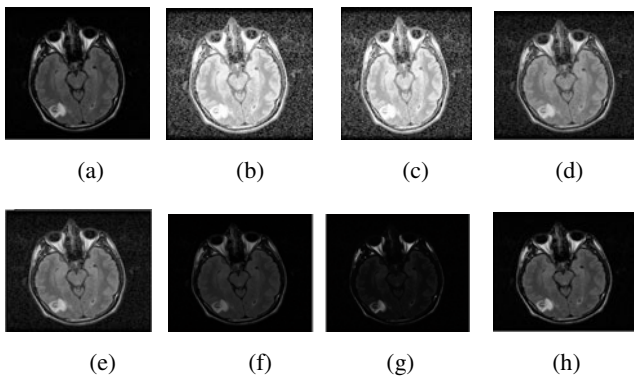


Fig. 2. (a)originalimage (b)equalized image (c)adjustable equalized with $\lambda=0$ (d) $\lambda=1$ (e) $\lambda=2$ (f) smoothing (g) weighted histogram approximation (h) low complexity method for boat image

The proposed method also can also be applied to medical images as shown in Fig.2. It shows that histogram adaptation scheme will achieve very good performance. The given input is a magnetic resonance image of axial normal brain of PNG type. It has pixel size of 256*256.

The histogram of various methods are shown in Fig(3)

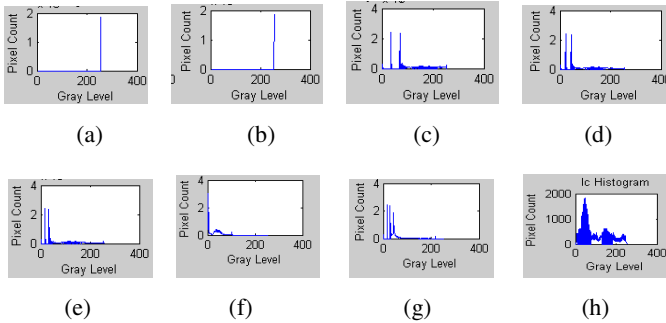


Fig. 3. (a)originalimage (b)equalized image (c)adjustable equalized with $\lambda=0$ (d) $\lambda=1$ (e) $\lambda=2$ (f) smoothing (g) weighted histogram approximation(h) low complexity method for boat image

Table (1) shows a comparison of various quantitative measurements.

Table 1. Comparison of Various Quantitative Measurements

Comparison of Quantitative Measurements for Boat Image			
	PSNR	AMBE	H
Histogram equalization	15.380966	38.408205	7.076154
Adjustable histogram equalization	21.835893	18.267295	7.224068
Histogram smoothing	13.526498	43.588128	6.404968
Weighted histogram approximation	14.285480	35.145796	6.142456
Low Complexity Algorithm	23.136483	14.541140	7.167812

5 Conclusion

A general structure for image contrast enhancement is presented. It designed penalty terms to adjust the various aspects of contrast enhancement. Hence, the contrast of the image can be improved without introducing visual artifacts. The proposed method avoids complex calculations and obtained images are visually pleasing and natural looking.

The proposed method is applicable to a wide variety of images including medical images also. It also offers a level of controllability and adaptively through which different levels of contrast enhancement can be achieved. Quantitative measurement can be obtained through measurements like PSNR which denotes Peak Signal to Noise Ratio, AMBE Absolute Mean Brightness Error and H entropy. From the values it shows that Low Complexity algorithm is working well.

For a low contrast image given as an input, output image with various level of enhancement is obtained. Such as normal equalized image, adjustable histogram equalized with various values of λ , histogram smoothing, weighted histogram approximation and finally Low Complexity algorithm is obtained. Output image have high level of contrast comparable to input image.

References

1. Kim, Y.T.: Enhancement using brightness preserving bi-histogram equalization. *IEEE Transaction on Consumer Electronics* 43(1), 1–8 (1997)
2. Wang, Y., Chen, Q., Zhang, B.: Image enhancement based on equal area dualistic sub image histogram equalization method. *IEEE Trans. Consumer Electronics* 45(1), 68–75 (1999)
3. Chen, S.-D., Ramli, A.R.: Minimum mean brightness error bihistogram equalization in contrast enhancement. *IEEE Trans. Consumer Electronics* 49(4), 1310–1319 (2003)
4. Chen, S.-D., Ramli, A.R.: Contrast enhancement using recursive mean-separate histogram equalization for scalable brightness preservation. *IEEE Trans. Consumer Electronics* 49(4), 1310–1319 (2003)
5. Wongsritong, K., Kittayaruasiriwat, K., Cheevasuvit, F., Dejhan, K., Somboonkaew, A.: Contrast enhancement using multipeak histogram equalization with brightness preserving. In: 1998 IEEE Asia-Pacific Conference on Circuit and System, pp. 455–458 (1998)
6. Ibrahim, H., Kong, N.S.P.: Brightness Preserving Dynamic Histogram Equalization for Image Contrast Enhancement. *IEEE Transactions on Consumer Electronics* 53(4), 1752–1757 (2007)
7. Stark, J.A.: Adaptive image contrast enhancement using generalizations of histogram equalization. *IEEE Trans. Image Process.* 9(5), 889–896 (2000)
8. Chen, Z.Y., Abidi, B.R., Page, D.L., Abidi, M.A.: Gray-level grouping (GLG): An automatic method for optimized image contrast enhancement—Part I: The basic method. *IEEE Transaction on Image Processing* 15(8), 2290–2302 (2006)
9. Gonzalez, R.C., Woods, R.E.: *Digital Image Processing*, 3rd edn., pp. 443–445, 598 – 605. University of Tennessee (1992)

Evaluating the Performance of a Speech Recognition Based System

Vinod Kumar Pandey and Sunil Kumar Kopparapu

TCS Innovation Labs - Mumbai,
Tata Consultancy Services, Pokharan Road 2
Thane 400 601, Maharashtra, India
{vinod.pande,sunilkumar.kopparapu}@tcs.com

Abstract. Speech based solutions have taken center stage with growth in the services industry where there is a need to cater to a very large number of people from all strata of the society. While natural language speech interfaces are the talk in the research community, yet in practice, menu based speech solutions thrive. Typically in a menu based speech solution the user is required to respond by speaking from a closed set of words when prompted by the system. A sequence of human speech response to the IVR prompts results in the completion of a transaction. A transaction is deemed successful if the speech solution can correctly recognize all the spoken utterances of the user whenever prompted by the system. The usual mechanism to evaluate the performance of a speech solution is to do an extensive test of the system by putting it to actual *people use* and then evaluating the performance by analyzing the logs for successful transactions. This kind of evaluation could lead to dissatisfied test users especially if the performance of the system were to result in a poor transaction completion rate. To negate this the Wizard of Oz approach is adopted during evaluation of a speech system. Overall this kind of evaluations is an expensive proposition both in terms of time and cost. In this paper, we propose a method to evaluate the performance of a speech solution without actually putting it to *people use*. We first describe the methodology and then show experimentally that this can be used to identify the performance bottlenecks of the speech solution even before the system is actually used thus saving evaluation time and expenses.

Keywords: Speech solution evaluation, Speech recognition, Pre-launch recognition performance measure.

1 Introduction

There are several commercial menu based ASR systems available around the world for a significant number of languages and interestingly speech solution based on these ASR are being used with good success in the Western part of the globe [5], [3], [6], [2]. Typically, a menu based ASR system restricts user to speak from a pre-defined closed set of words for enabling a transaction. Before

commercial deployment of a speech solution it is imperative to have a quantitative measure of the performance of the speech solution which is primarily based on the speech recognition accuracy of the speech engine used. Generally, the recognition performance of any speech recognition based solution is quantitatively evaluated by putting it to actual use by the people who are the intended users and then analyzing the logs to identify successful and unsuccessful transactions. This evaluation is then used to identifying any further improvement in the speech recognition based solution to better the overall transaction completion rates. This process of evaluation is both time consuming and expensive. For evaluation one needs to identify a set of users and also identify the set of actual usage situations and perform the test. It is also important that the set of users are able to use the system with ease meaning that even in the test conditions the performance of the system, should be good, while this can not usually be guaranteed this aspect of keeping the user experience good makes it necessary to employ a wizard of Oz (WoZ) approach. Typically this requires a human agent in the loop during actual speech transaction where the human agent corrects any mis-recognition by actually listening to the conversation between the human user and the machine without the user knowing that there is a human agent in the loop. The use of WoZ is another expense in the testing a speech solution. All this makes testing a speech solution an expensive and time consuming procedure.

In this paper, we describe a method to evaluate the performance of a speech solution without actual people using the system as is usually done. We then show how this method was adopted to evaluate a speech recognition based solution as a case study. This is the main contribution of the paper. The rest of the paper is organized as follows. The method for evaluation without testing is described in Section 2. In Section 3 we present a case study and conclude in Section 4.

2 Evaluation without Testing

Fig. 1 shows the schematic of a typical menu based speech solution having 3 nodes. At each node there are a set of words that the user is expected to speak and the system is supposed to recognize. In this particular schematic, at the entry node the user can speak any of the n words, namely W_1 or W_2 or \dots or W_n ; n is usually called the perplexity of the node in the speech literature. The larger the n the more the perplexity and higher the confusion and hence lower the recognition accuracies. In most commercial speech solutions the perplexity is kept very low, typically a couple of words. Once the word at the entry node has been recognized (say word W_k has been recognized), the system moves on to the second node where the active list of words to be recognized could be one of $W_{k1}, W_{k2}, W_{k3}, \dots, W_{kp}$ if the perplexity at the k^{th} node is p . This is carried on to the third node. A transaction is termed successful *if and only if* the recognition at each of the three nodes is correct. For example, typically in a banking speech solution the entry node could expect someone to speak among

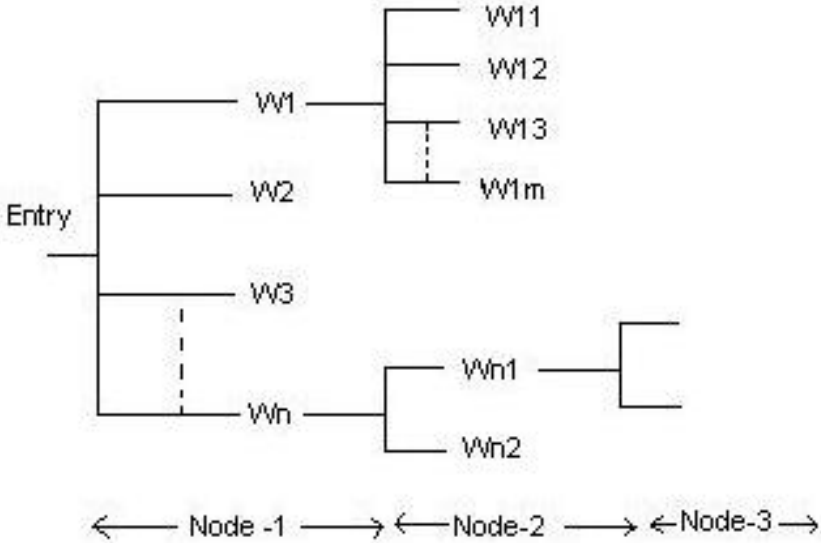


Fig. 1. Schematic of a typical menu based ASR system (W_n is spoken word)

*/credit card/*¹, */savings account/*, */current account/*, */loan product/*, */demat/*, and */mutual fund transfer/* which has a perplexity of 6. Once a person speaks, say, */savings account/* and is recognized correctly by the system, at the second node it could be */account balance/* or */cheque/* or */last 5 transactions/* (perplexity 3) and at the third node (say, on recognition of */cheque/*) it could be */new cheque book request/*, */cheque status/*, and */stop cheque request/* (perplexity 3).

Note 1. Though we will not dwell on this, it is important to note that an error in recognition at the entry node is more expensive than a recognition error at a lower node.

Based on the call flow, and the domain the system can have several nodes for completion of a transaction. Typical menu based speech solutions strive for a 3 - 5 level nodes to make it usable. In any speech based solution (see Fig. 2) first the spoken utterance is hypothesized into a sequence of phonemes using the acoustic models. Since the phoneme recognition accuracy is low, instead of choosing one phoneme it identifies l -best (typically $l = 3$) matching phonemes. This phone lattice is then matched with all the expected words (language model) at that node to find the best match. For a node with perplexity n the constructed phoneme lattice of the spoken utterance is compared with the phoneme sequence representation of all the n words (through the lexicon which is one of the key

¹ We will use // to indicate the spoken word. For example $/W_1/$ represents the spoken equivalent of the written word W_1 .

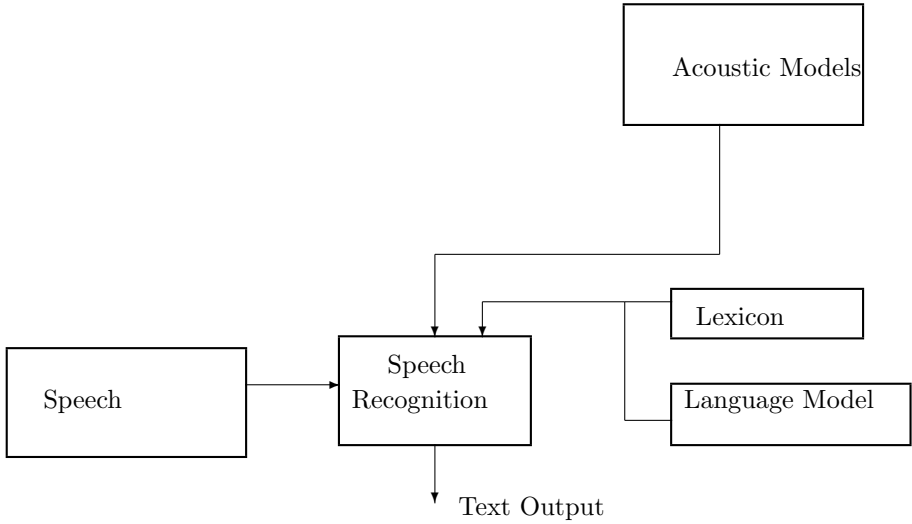


Fig. 2. A typical speech recognition system. In a menu based system the language model is typically the set of words that need to be recognized at a given node.

components of a speech recognition system). The hypothesized phone lattice is declared one of the n words depending on the closeness of the phoneme lattice to the phoneme representation of the n words.

We hypothesize that we can identify the performance of a menu based speech system by identifying the possible confusion among all the words that are active at a given node.

Note 2. If active words at a given node are phonetically similar it becomes difficult for the speech recognition system to distinguish them which in turn leads to recognition errors.

We used Levenshtein distance [1], [4] a well known measure to analyze and identify the confusion among the active words at a given node. This analysis gives a list of all set of words that have a high degree of confusability among them; this understanding can be then used to (a) restructure the set of active words at that node and/or (b) train the words that can be confused by using a larger corpus of speech data. This allows the speech recognition engine to be equipped to be able to distinguish the confusing words better. Actual use of this analysis was carried out for a speech solution developed for Indian Railway Inquiry System to identify bottlenecks in the system before its actual launch.

3 Case Study

A schematic of a speech based Railway Information system, developed for Hindi language is shown in Fig. 3. The system enables user to get information on five

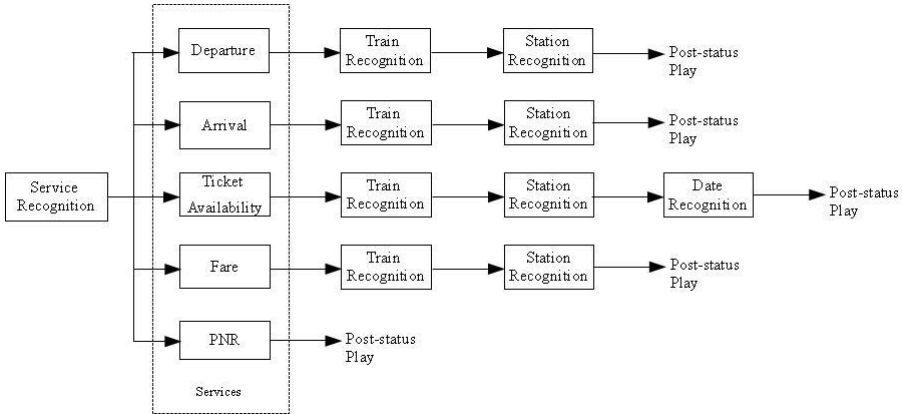


Fig. 3. Call flow of Indian Railway Inquiry System (W_n is spoken word)

different services, namely, (a) ARRIVAL of a given train at a given station, (b) DEPARTURE of a given train at a given station, (c) TICKET AVAILABILITY on a given date in a given train between two stations, and class, (d) FARE in a given class in a given train between two stations, and (e) PNR STATUS. At the first recognition node (node-1), there are one or more active words corresponding to each of these services. For example, for selecting the service FARE, the user can speak among */kiraya jankari/*, */kiraya/*, */fare/*. Similarly, for selecting service TICKET AVAILABILITY, user can speak */upalabdhata jankari/* or */ticket availability/* or */upalabdhata/*.

Note 3. Generally the perplexity at a node is greater than or equal to the number of words that need to be recognized at that node.

In this manner each of the services could have multiple words or phrases that can mean the same thing and the speaker could utter any of these words to refer to that service. The sum of all the possible different ways in which a service can be called (d_i) summed over all the 5 services gives the perplexity (\mathcal{N}) at that node, namely,

$$\mathcal{N} = \sum_{i=0}^5 d_i \tag{1}$$

The speech recognition engine matches the phoneme lattice of the spoken utterance with all the \mathcal{N} words which are active. The active word (one among the \mathcal{N} words) with highest likelihood score is the recognized word. In order to avoid low likelihood recognitions a threshold is set so that even the best likelihood word is returned only if the likelihood score is greater than the predefined threshold. Completion of a service requires recognitions at several nodes with different perplexity at each node. Clearly depending on the type of service that the user is wanting to use; the user has to go through different number of recognition nodes. For example, to complete the ARRIVAL service it is required to

pass through 3 recognition nodes namely (a) selection of a service, (b) selection of a train name and (c) selection of the railway station. While the perplexity (the words that are active) at the service selection node is fixed the perplexity at the station selection node could depend on the selection of the train name at an earlier node. For example, if the selected train stops at 23 stations, then the perplexity at the station selection node will be ≥ 23 .

For confusability analysis at each of the node, we have used the Levenshtein distance [4] or the edit distance as is well known in computer science literature. We found that the utterances /*Sahi*/ and /*Galat*/ have 100% recognition. These words *Sahi* is represented by the string of phonemes in the lexicon as S AA HH I and the word *Galat* is represented as the phoneme sequence G L AX tT in the lexicon. We identified the edit distance between these two words *Sahi* and *Galat* and used that distance measure as the threshold that is able to differentiate any two words (say T). So if the distance between any two active words at a given recognition node is lower than the threshold T , then there is a greater chance that those two active words could get confused (one word could be recognized as the other which is within a distance of T). There are ways in which this possible misrecognition words could be avoided. The easiest way is to make sure that these two words together are not active at a given recognition node.

Table 1 shows the list of active word at the node 1 when the speech application was initially designed and Table 2 shows the edit distance between all the active words at the node service given in Fig. 3. The distance between words *Sahi* and *Galat* was found to be 5.7 which was set at the threshold, namely $T = 5.7$. This threshold value was used to identify confusing active words. Clearly, as seen in the Table the distance between word pairs *fare*, *pnr* and *pnr*, *prasthan* is 5.2 and

Table 1. List of Active Words at node 1

W. No.	Active Word	Phonetic
W_1	kiraya_jankari	K I R AA Y AA J AA tN K AA R I
W_2	kiraya	K I R AA Y AA
W_3	fare	F AY R
W_4	aagaman_jankari	AA G AX M AX tN J AA tN K AA R I
W_5	aagaman	AA G AX M AX tN
W_6	arrival_departure	AX R AA I V AX L dD I P AA R CH AX R
W_7	upalabdhata_jankari	U P AX L AX B tDH AX tT AA J AA tN K AA R I
W_8	ticket_availability	tT I K EY tT AX V AY L AX B I L I tT Y
W_9	upalabdhata	U P AX L AX B tD AX tTH AA
W_{10}	arrival	AX R AA I V AX L
W_{11}	prasthan	P R AX S tTH AA tN
W_{12}	departure	dD I P AA R CH AX R
W_{13}	pnr_jankari	P I EY tN AA R J AA tN K AA R I
W_{14}	pnr	P I AX tN AA R

Table 2. Distance Measurement for Active Words at Node 1 of the Railway Inquiry System

	W ₁	W ₂	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀	W ₁₁	W ₁₂	W ₁₃	W ₁₄
W ₁	0	8.4	14.2	8.5	14.1	15.3	11	20	17.1	13	13	13.2	6.2	11.2
W ₂	8.4	0	7.2	13.8	8.5	14.7	17.8	15.7	11	7.2	7.8	7.7	11.2	6.2
W ₃	14.2	7.2	0	14.2	7.2	14.8	18.2	15.8	11.2	8.2	8.2	7.8	14.2	5.8
W ₄	8.5	13.8	14.2	0	8.4	15.9	9.6	19.7	14.3	13	13	14.7	8.2	11.2
W ₅	14.1	8.5	7.2	8.4	0	13.2	16.7	15.7	9.7	7.2	6.7	8.2	14.1	7.1
W ₆	15.3	14.7	14.8	15.9	13.2	0	18.7	18.9	15.5	9.4	13.7	7	17.3	11.8
W ₇	11	17.8	18.2	9.7	16.7	18.7	0	20	11.2	17.1	15.6	17.8	10.2	13.8
W ₈	20	15.7	15.8	19.7	15.7	18.9	20	0	14.5	15.8	17.5	16.5	18.6	15.7
W ₉	17.1	11	11.2	14.3	9.7	15.5	11.2	14.5	0	10	9.2	11.1	16.3	9.7
W ₁₀	13	7.2	8.2	13.1	7.2	9.4	17.1	15.8	10	0	8.5	8	13	7.8
W ₁₁	13.1	7.8	8.2	13.1	6.7	13.7	15.7	17.5	9.2	8.5	0	8.4	11.7	5.2
W ₁₂	13.2	7.7	7.8	14.7	8.2	7	17.8	16.5	11.1	8.1	8.4	0	12.1	6.8
W ₁₃	6.2	11.2	14.2	8.2	14.1	17.3	10.2	18.6	16.3	13.1	11.7	12.1	0	9.8
W ₁₄	11.2	6.2	5.8	11.2	7.1	11.8	13.8	15.7	9.6	7.8	5.2	6.8	9.8	0

Table 3. Recognition of Active Words at Node 1 of the Railway Inquiry System (before restructuring active words)

	W ₁	W ₂	W ₃	W ₄	W ₅	W ₆	W ₇	W ₈	W ₉	W ₁₀	W ₁₁	W ₁₂	W ₁₃	W ₁₄
/kiraya jankari/	40	0	0	0	0	0	0	0	0	0	0	0	0	0
/kiraya/	0	40	0	0	0	0	0	0	0	0	0	0	0	0
/fare/	0	4	20	0	0	0	0	0	0	0	0	0	0	16
/aagaman jankari/	0	0	0	40	0	0	0	0	0	0	0	0	0	0
/aagaman/	0	0	0	0	40	0	0	0	0	0	0	0	0	0
/arrival departure/	0	0	0	0	0	40	0	0	0	0	0	0	0	0
/upalabdhatta jankari/	0	0	0	0	0	0	40	0	0	0	0	0	0	0
/ticket availability/	0	0	0	0	0	0	0	40	0	0	0	0	0	0
/upalabdhatta/	0	0	0	0	0	0	0	0	40	0	0	0	0	0
/arrival/	0	0	0	0	0	0	0	0	0	40	0	0	0	0
/prasthan/	0	0	0	0	0	0	0	0	0	0	29	0	0	11
/departure/	0	0	0	0	0	0	0	0	0	0	0	40	0	0
/pnr jankari/	0	0	0	0	0	0	0	0	0	0	0	0	40	0
/pnr/	0	0	17	0	0	0	0	0	0	0	19	0	0	4

5.8 respectively, which is very close to the threshold value of 5.7. This can cause a high possibility that /fare/ may get recognized as pnr and vice-versa. One can derive from the analysis of the active words that fare and pnr can not coexist as active words at the same node. The result of the analysis was to remove the active words fare and pnr at that node.

Table 4. Recognition of Active Words at Node 1 of the Railway Inquiry System (after restructuring active words)

	W_1	W_2	W_4	W_5	W_6	W_7	W_8	W_9	W_{10}	W_{11}	W_{12}	W_{13}
<i>/kiraya jankari/</i>	40	0	0	0	0	0	0	0	0	0	0	0
<i>/kiraya/</i>	0	40	0	0	0	0	0	0	0	0	0	0
<i>/aagaman jankari/</i>	0	0	40	0	0	0	0	0	0	0	0	0
<i>/aagaman/</i>	0	0	0	40	0	0	0	0	0	0	0	0
<i>/arrival departure/</i>	0	0	0	0	40	0	0	0	0	0	0	0
<i>/upalabdhata jankari/</i>	0	0	0	0	0	40	0	0	0	0	0	0
<i>/ticket availability/</i>	0	0	0	0	0	0	40	0	0	0	0	0
<i>/upalabdhata/</i>	0	0	0	0	0	0	0	40	0	0	0	0
<i>/arrival/</i>	0	0	0	0	0	0	0	0	40	0	0	0
<i>/prasthan/</i>	0	0	0	0	0	0	0	0	0	40	0	0
<i>/departure/</i>	0	0	0	0	0	0	0	0	0	0	40	0
<i>/pnr jankari/</i>	0	0	0	0	0	0	0	0	0	0	0	40

Table 3 presents result of actual recognition at Node 1 before restructuring the active words at that node. The testing was done by 2 male and 3 female users. Each of the 14 active words at that node was spoken 8 times by each user, such that each active word was spoken 40 times in all by all the users. Analysis was done before restructuring (see Table 3) and after restructuring (see Table 4) of the active words at Node 1. As seen in the Table 3, 16 and 4 out of 40 */fare/* was recognized as *pnr* (W_{14}) and *kiraya* respectively. While 11 instances out of a total of 40 instances of */prasthan/* was recognized as *pnr*, similarly only 4 instances of a total of 40 instances of */pnr/* was recognized as *pnr*. This confusion is expected as can be seen from the edit distance analysis of the active words in Table 2. Table 4 presents result of the recognition at Node 1 after restructuring active word list (removal of *fare* and *pnr*). As can be seen in Table 4 restructuring increased the recognition accuracy at Node 1 (Fig. 3) from 88% to 100%. A similar analysis was carried out at other recognition nodes and the active word list was suitably modified to avoid possible confusion between active word pair. This analysis and modification of the list of active words at a node resulted in a significant improvement in the transaction completion rate.

4 Conclusion

In this paper we proposed a methodology to identify words that could lead to confusion at any given node of a speech recognition based system. We used edit distance as the metric to identifying the possible confusion between the active words. We showed that this metric can be used effectively to enhance the performance of a speech solution without actually putting it to people test. There is a significant saving in terms of being able to identify recognition bottlenecks in a menu based speech solution through this analysis because it does not require actual people testing the system. This methodology was adopted to restructuring

the set of active words at each node for better speech recognition in an actual menu based speech recognition system that caters to masses.

References

1. Gusfield, D.: Algorithms on Strings, Trees, and Sequences: Computer Science and Computational Biology. Cambridge University Press, Cambridge (2001)
2. Kim, C., Stern, R.: Feature extraction for robust speech recognition based on maximizing the sharpness of the power distribution and on power flooring. In: IEEE International Conference on Acoustics Speech and Signal Processing, pp. 4574–4577 (2010)
3. Lua, X., Matsudaa, S., Unokib, M., Nakamura, S.: Temporal contrast normalization and edge-preserved smoothing of temporal modulation structures of speech for robust speech recognition. *Speech Communication* 52, 1–11 (2010)
4. Navarro, G.: A guided tour to approximate string matching. *ACM Computing Surveys* 33, 31–88 (2001)
5. Sun, Y., Gemmeke, J., Cranen, B., Bosch, L., Boves, L.: Using a DBN to integrate sparse classification and GMM-based ASR. In: Proceedings of Interspeech 2010 (2010)
6. Zhao, Y., Juang, B.: A comparative study of noise estimation algorithms for VTS-based robust speech recognition. In: Proceedings of Interspeech 2010 (2010)

Unified Approach in Food Quality Evaluation Using Machine Vision

Rohit R. Parmar^{1,*}, Kavindra R. Jain^{2,**}, and Chintan K. Modi^{3,***}

EC Department, G.H. Patel College of Engineering & Technology,
Vallabh Vidyanagar, India

rohitrparmar@gmail.com, {kavi_gcet,chintankmodi}@yahoo.com

Abstract. The paper presents a unified approach for quality evaluation of food using image processing and machine vision. In this paper basic tool is combination of computer and machine vision for image analysis and processing through which fast and accurate quality is achieved that too with the help of non-destructive method. Machine vision in food has broadened its range of applications from grains, cereals, fruits to vegetables including processed products as well as spices in which there is a high degree of quality achieved as compared to human vision inspection. In this paper we quantify the qualities of various food products and figure out features which are directly or inversely affect the quality of the food product. Based on these features a generalized formula of quality is proposed to be used for quality evaluation of any type of food product.

Keywords: Quality, Machine vision, Image processing, Fruit, Vegetables, Grains, Spices, Unification approach.

1 Introduction

The agricultural industry is probably too oldest and most widespread industry in the world. Quality control is of major importance in the food industry. After harvesting a food product based on quality parameter a food product has been sorted and graded in different grades. Also a continuous development in mechanical harvesting system and need for automatic grading system has arisen. In recent years due to rapid industrialization and massive rural-urban migration the necessity of agricultural mechanization needs to be grown faster.

Non-destructive quality evaluation of food products is an important and very vital factor in food/agricultural industry. Various parameters which define quality of these products (e.g. size, shape, color, texture, external defects, smell etc.) are evaluated by human inspectors [35]. Together with the high labor costs, inconsistency and variability associated with human inspection accentuates the need for objective measurements systems. Efforts are being geared towards the replacement of human operator with automated systems, as human operations are inconsistent and less efficient [14].

* Research Scholar.

** Lecturer.

*** Professor.

Rapid advances in hardware and software for digital image processing have motivated several studies on the development of computer vision systems to evaluate quality of diverse raw and processed foods. Advancement in computer technology leads to use these in the domain of food processing like grading, sorting and quality inspection [32]. Computer vision systems are recognized as the integration of devices for non-contact optical sensing, and computing and decision processes which receive and interpret automatically an image of a real scene. Recently many different features like size, shape, color and texture of food products are combined together for their applications in the food industry. Normally, by increasing the features used, the performance of the methods proposed can be increased. However, in literature there is a need of a common evaluation scheme. In this paper we propose a unification approach to evaluate the quality of food products.

In this paper we present a unified approach to evaluate the quality of food products using machine vision. Section 2 discusses basic machine vision technique for quality evaluation of food. Section 3 discusses background of various food products in which machine vision is used. We propose a unified approach for evaluation of quality of food products in section 4. Section 5 concludes the paper.

2 Quality Evaluation Using Machine Vision

Computer & Machine vision system for quality evaluation of food products contains standard hardware configuration as shown in figure 2.1. It consists of:

1. Some means of presenting the object to be inspected to the camera
2. Light source for proper illumination
3. CCD Camera to acquire image
4. A frame-grabber, to perform analog-to-digital conversion.
5. Software, if a computer is used for image processing

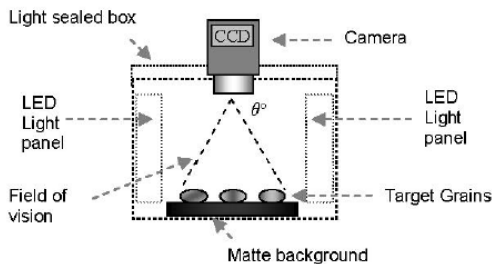


Fig. 2.1. Imaging setup of computer vision system [5]

Illumination plays a vital role in quality of image. So here proper illumination is adjusted with the help of available light sources. Image acquisition is done (or image formation) to produce digital images of the food products. Image is captured by high quality CCD camera. Pre-processing is used to enhance, clean, and improve key parts

of the input images [15], [16]. During the process of image acquisition, captured images may suffer from severe noise and external disturbances this can be removed with the help of filtering process.

Image can be converted from color to gray scale depending upon whether color parameter is of importance or not. Segmentation is done to identify relevant parts of a food product, as disjoint and non-overlapping regions separated from the background. To partition an image, thresholding is applied on image. Object recognition and measurement are done to quantify key features of the selected objects. Here different parameters based on size, shape etc. is calculated. Edge operators detect discontinuities in grey level, color, texture, etc. Region segmentation involves the grouping together of similar pixels to form regions representing single objects within the image. In interpretation, where the features extracted from the products are interpreted using application domain knowledge.

3 Assessment of Food Products

Food product images captured by vision system can be used to identify, analyze and quality assessment of food products.

(1) Peanuts

Peanut is a mass consumption item and is used for extraction of oil, for making butter, chikkies and chocolates, as an ingredient in making several food and various snacks. India is one of the largest producers along with the USA; China and Argentina. Gujarat, Andhra Pradesh, Tamilnadu and Maharashtra are the main cultivating states. Peanuts from the Saurashtra region of Gujarat are famous all over the world on account of their big size, nutty flavor and crunchy taste. Peanuts having a wide range of seed size and maturity are often obtained at harvest. Due to non-uniform maturity levels size variation has been seen in peanuts like a smaller sized seed would be immature seed and a big seed would be mature as shown in figure 3.1 [3] [11].



Fig. 3.1. Partial pictures of peanut in different varieties' [11]

In 1978, J. I. Davidson derived a mathematical relationship for seed size distribution to describe relation between different seed (*arachies hypogaea* L.) varieties. A step further in 1989, F. E. Dowell, who developed an automatic peanut grading system based on machine vision system with the integration of mechanical component to

grade peanuts based on size or defect [11]. In 2009, H. Zhong-zhi and Z. You-gang compared 48 different peanut varieties. Using principal component analysis they developed an artificial neural network to establish a seed recognition model which was made up of 49 distinct appearance characteristics referred as shape, texture, color etc. Here they got variety recognition rate and quality recognition rate up to 91.2% and 93.0% respectively [10]. In 2010, H. Zhong-zhi et al developed a neural network, which was based on 52 appearance features. They got success up to 95.6% in their results. This method has been used for appraisal of peanut quality in China.

Seeing to the above studies on peanut the conclusion can be drawn that there are few quality parameters affecting the production of peanuts. For instance color (x_1), texture (x_2) are directly affecting the quality & non uniform maturity level i.e. size (major (y_1) and minor (y_2) axis) are inversely proportional to quality of peanuts.

(2) Rice

Rice is one of the leading food crops of the world and more than half of the world's population relies on rice as the major daily source of calories and protein [4]. The quality of rice has distinct effect, so the proper inspection of rice quality is very important. There are two main factors for checking the quality of rice kernel: percentage of broken rice and percentage of the purity of rice as shown in figure 3.2.

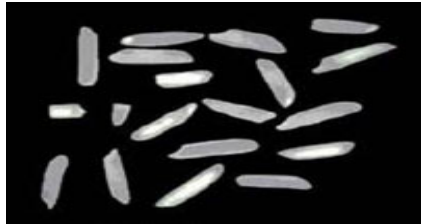


Fig. 3.2. Picture of rice characteristics [5]

In 2002, Y. N. Wan developed an automatic grain kernel handling system in which 1296 singularized kernel images per minute were taken for machine vision inspection. He developed a Windows-based software program for rice quality inspection. [33]. Later on S, Sansomboonsuk and N. Afzulpurkar (2008) developed image processing algorithms and used to extract features for kernels [30]. They measured and calculated area, perimeter, circularity and shape compactness as criteria in Fuzzy logic for classifying each kernel. From testing the image analysis algorithms, the results are obtained, they found accuracy averaging 92% for both of the broken rice and the purity of rice compared with human inspection.

In 2009, B. Emadzadeh, S.M.A. Razavi, and R. Farahmandfar [9] compared three Iranian rice varieties, namely Tarom Mahalli, Fajr and Neda, determined parameters using micrometer and image processing methods. Comparison of the results obtained by both procedures showed that the geometric characteristics (length, width, height and projected area) of all three varieties decrease and sphericity increases after removing the outer and the brownish layers. It was found that the values of micrometer data are having lower for all the geometric factors and that the true size and sphericity.

In 2010, S. Shantaiya, U. Ansari developed digital image analysis algorithm based on color, morphological and textural features to identify the six varieties rice seeds in Chhattisgarh region [29]. Nine color and nine morphological and textural features were used for analysis. They developed a neural network-based classifier to identify the unknown grain types. In the test dataset, the classification accuracies were 90.00%, 88.00%, 95.00%, 82.00%, 74.00%, 80.00% respectively.

Seeing to the above studies on rice the conclusion can be drawn that there are few quality parameters affecting the production of rice. For instance area (x_1), perimeter (x_2), circularity (x_3), sphericity (x_4), shape compactness (x_5) are directly affecting the quality & non uniform broken rice (y_1) are inversely proportional to quality parameter of rice.

(3) Bakery Products

Appearance of baked products is an important quality attribute which influences the visual perceptions of customers and hence potential demands of the products. The appearance of the internal and external features contributes to the overall impression of the products quality. Consequently, the inspection of the bakery products occupies a major role in the manufacture process as it affects the taste, texture and appearance of the products. Computer vision system has been used to measure characteristics before packing. Variety of biscuits on which quality evaluation is done is show in figure 3.3.

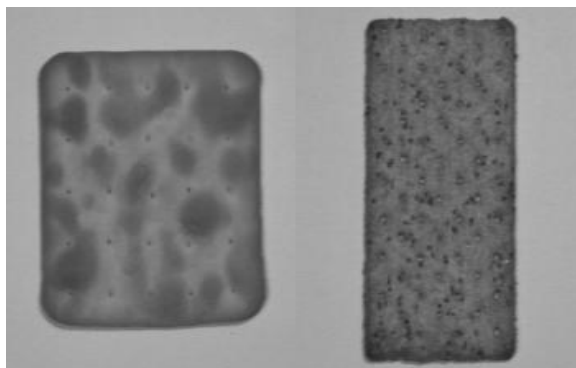


Fig. 3.3. 1Variety of biscuit [17]

L. Hamey et al (1993) developed a prototype computer system for inspection of biscuit [17]. A system employs a monochromatic image and histogram of image to identify defect and classify a product as under backed, correctly backed and over backed [17]. J. C. H. Yeh et al (1995) [12] present the implementation of the inspection system with a hybrid neural network of self-organizing maps and feed forward neural networks. They have tested system and its grading performance on biscuit. The bake levels were evaluated and compared with trained human inspector. They found that the proposed color system with a hybrid neural network performed significantly better than the human inspector. They also found that the cross-validation technique can prevent over-training and preserve generalization of an ANN. Raji et al

(2000) developed a program in FORTRAN using the principle of edge detection in image analysis to determine the edge of sliced breads and biscuits (round and rectangular) with a view to detecting defects (breakage) [27].

The crispness of biscuit measured in acoustic sound and not in texture of biscuit. In crispness of biscuit or snacks the distraction could happen due to environment. So, this new initial method is purposed to analyze the texture to determine the crispness of biscuit. Here they have take care of all samples of the same size and thickness because the humidity can influence if these two criteria are different. The normal texture stored as a reference to the irregular texture image with lightning condition. They applied wavelet transforms in this analyzing together with image processing from machine vision. All the texture stored in a database as a comparison to determine the good quality of biscuit [22].

A prototype-automated system for visual inspection of muffins was developed by Zaid Abdullah et al. [36]. The color of 100 light brown and 100 dark brown muffins was evaluated using the vision system and discriminants analysis compared with visual examination. The automated system was able to correctly classify 96% of pregraded and 79% of ungraded muffins. [1]

For bulk sugary products classification and identification developed by B. S. Anami et al (2009). [2]. they used gray level co-occurrence method for texture analysis and feature extraction and developed a neural network model to classify 10 different varieties of bulk sugary food products. In the classification of Indian food products like Apple cake, Bundeladu, Burfi, Doodhpada, Jamun, Jilebi, Kalakand, Ladakiladu, Mysorepak, Suraliholige, they got accuracy of 90%.

Seeing to the above studies on bakery biscuits the conclusion can be drawn that there are few quality parameters affecting the baked biscuit. For instance area (x_1), crispness (x_2), thickness (x_3) directly affect the quality & color variation (y_1), less amount of ingredients (like amount of chocolate chips in chocolate cookies) (y_2) are inversely proportional to quality of baked product.

(4) Fruits and Vegetables

Computer vision has been widely used for the quality inspection and grading of fruits and vegetables. The fresh fruit and vegetable postharvest sector is dynamic and largely to increasing consumer demand for quality product. The condition of fruit and vegetable at the time of harvest has an important effect on the consumer's level of satisfaction at consumption. Computer vision offers the potential to automate manual grading practices and thus to Standardize techniques and eliminate tedious inspection tasks.

Color provides information about estimating the maturity and also in examining the freshness of fruits & vegetables [18]. P. Sudhakara Rao et al uses HSI for color representation, which provides an efficient method for color discrimination. With the help of median density of Hue as a grading criterion, the image processing system achieved around 98 % accuracy in color inspection of Apples. They proposed new method namely Improved Radius signature [31]. This method can be effectively used for comparing many samples against a reference shape for the purpose of sorting and grading using correlation coefficient technique, Graphical analysis or Fourier transformation technique. Du and Sun (2004), shape features can be measured independently (for example, by Fourier descriptors of the planar image boundary,

invariant moments) or by combining size measurements (for example, circularity, aspect ratio, compactness, eccentricity, roundness). So with, the determination of fruit size parameters allows simple shape sorting [6].



Fig. 3.4. Pictures of Apples

Color is the important parameter regarding freshness of fruits and vegetables. H. M. W. Bunnik et al developed a system which quantifies the color of entire object using mean, standard deviation, skew and kurtosis [19].

Seeing to the above studies on fruits & vegetables the conclusion can be drawn that there are few quality parameters affecting the quality. For instance area (x_1), shape (x_2), size (x_3), color (x_4), shape compactness (x_5) are directly affecting the quality & black moles (y_1), brown undubbed shapes (y_2) are inversely proportional to quality parameter of fruits & vegetables.

(5) Fennel and Cummin seeds

The booming global spice market also possesses good opportunities for the Indian spice industry especially from Gujarat (Unjha) and Tamil Nadu to provide quality spices at competitive prices. Unjha faces stiff competition from China, Malaysia and Pakistan in terms of pricing of the products. Therefore the need of the manufacturers there is to ensure consistency in supply, product quality, pricing and marketing strategy to increase the share in exports. The major one out of all is of quality [14] [15].

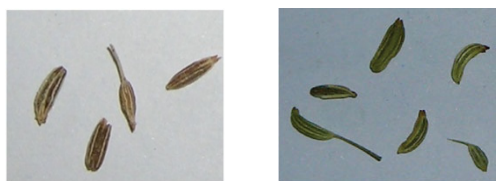


Fig. 3.5. figures of cummin & fennel seeds [14] [15]

K. Jain et al [14] [15] proposed to use machine vision techniques for quality evaluation of cummin and fennel seeds. The approach was based on finding minor axis length, major axis length and area of the seeds. The classification of good and bad quality was done on finding the number of seeds with pedestals and foreign elements available in bulk of samples as shown in figure 3.5. Here, the quality was thought to be inversely proportional to number of seeds with pedestals (x_1) and number of foreign elements (x_2) present in the samples.

4 Proposed Unified Approach for Quality Evaluation

General review in the field of quality depicts that there is a requirement of a unified approach for the quality evaluation of food products which widely range from raw to processed product. The old rule of thumb says that acquire an image, pre process it, segment it, then extract the features and finally find its quality based on those features by comparing it with the standard one. The features to be extracted would vary for each and every food product as discussed above. So it can be concluded in general that those features would affect the quality of specific food product. In general, some features (x_i) will be directly proportional to the quality of the food product while other features (y_j) will be indirectly proportional to the quality. Hence a general formula of quality of any food product based on machine vision extracted features can be given by

$$Q = \frac{c_1 \sum_{i=1}^n w_i x_i}{c_2 \sum_{j=1}^m w_j y_j} \quad (1)$$

Where, w_i and w_j are weights for features x_i and y_j respectively. The weights can be set as per the food quality requirement. The constants c_1 and c_2 plays a major role in defining final grades from the quality. Example of grading can be found in [14], [15].

5 Conclusion

In this paper various food products quality evaluation is discussed in detail. For every product features are figured out for which the quality is either directly or inversely varied. A formula of quality is devised out which try to generate a unification approach for quality evaluation of all types of food products. The weighted sum of features based formula suggests using proper weights for each and every features on which the quality is dependent on.

References

- [1] Abdullah, Z.M., Aziz, A.S., Dos-Mohamed, A.M.: Quality inspection of bakery products using a color-based machine vision system. *Journal of Food Quality* 23(1), 39–50 (2000)
- [2] Anami, B.S., Burkpalli, V.C.: Texture based Identification and Classification of Bulk Sugary Food Objects. *ICGST-GVIP Journal* 9(4), 9–14 (2009) ISSN: 1687-398X
- [3] Hong, C., Chuanhua, Z., Youchun, D.: Peanut Shape Recognition Based on Fourier Descriptor. *IEEE, Los Alamitos* (2009) 978-1-4244-3894-5/09/
- [4] Cheng, F., Liu, Z.-Y., Ying, Y.-B.: Machine Vision Analysis of Characteristics and Image Information Base Construction for Hybrid Rice Seed. *Rice Science* 12(1), 13–18 (2005)
- [5] Hobson, D.M., Carter, R.M., Yan, Y.: Characterization and Identification of Rice Grains through Digital Image Analysis. In: *Instrumentation and Measurement Technology Conference - IMTC Warsaw, Poland, May 1-3. IEEE, Los Alamitos* (2007)

- [6] Du, C.-J., Sun, D.-W.: Recent developments in the applications of image processing techniques for food quality evaluation. *Trends in Food Science & Technology* 15, 230–249 (2004)
- [7] Du, C.-J., Sun, D.-W.: Learning techniques used in computer vision for food quality evaluation: a review. *Journal of Food Engineering* 72(1), 39–55 (2006)
- [8] Emadzadeh, Razavi, S.M.A., Farahmandfar, R.: Monitoring geometric characteristics of rice during processing by image analysis system and micrometer measurement. *Institute of Agro physics, Polish Academy of Sciences* 24, 21–27 (2010)
- [9] Graves, M., Batchelor, B.G.: *Machine Vision for the Inspection of Natural Products*. ISBN 1852335254. Springer, Heidelberg (2003), 1943-670.2'8563
- [10] Han, Z.-Z., Zhao, Y.-G.: Quality Grade-Testing of Peanut Based on Image Processing. In: *Second International Conference on Information and Computing Science*, pp. 21–24. IEEE, Los Alamitos (2009)
- [11] Davidson Jr., J.I., Blankenship, P.D., Chew, V.: Probability Distributions of Peanut Seed Size. *Peanut Science* 5, 91–96 (1978)
- [12] Yeh, J.C.H., Hamey, L.G.C., Westcott, S. T., Sung, S.K.Y.: Color Bake Inspection System Using Hybrid Artificial Neural Networks. IEEE, Los Alamitos (1995) 0-7803-2768-3/95
- [13] *Journal of Scientific Research and Development*. Invited Overview VII (February 2005)
- [14] Jain, K., Modi, C.K., Pithadiya, K.: Non Destructive quality evaluation in spice industry with specific reference to *Cuminum Cyminum L (Cumin)* seeds. In: *International Conference on Innovations & Industrial Applications, Malaysia*, pp. 25–26. IEEE, Los Alamitos (2009)
- [15] Jain, K., Modi, C.K., Patel, J.J.: Non Destructive quality evaluation in spice industry with specific reference to *Cuminum Cyminum L (Cumin)* seeds. In: *International Conference on Emerging Trends in Engineering and Technology, ICETET, Nagpur*. IEEE Computer Society, Los Alamitos (2009)
- [16] Chai, K.H., Lee, K.-J., Kim, G.: Non-destructive Quality Evaluation Technology for Fruits and Vegetables Using Near-Infrared Spectroscopy
- [17] Hamey, L.G.C., Watson, A.J., Tasman Westcott, C.: Machine inspection for biscuit bake. In: *Appears in Proc. Digital Image Computing: Techniques and Applications, Sydney, December 8-10*, pp. 124–129 (1993)
- [18] Malamasa, E.N., et al.: A survey on industrial vision systems, applications and tools. *Image and Vision Computing* 21, 171–188 (2003)
- [19] Bunnik, M.W., Bailey, D.G., Mawson, A.J.: Objective color measurement of tomatoes and lime
- [20] Hobson, M., Carter, R.M., Yan, Y.: Characterization and Identification of Rice Grains through Digital Image Analysis. In: *Instrumentation and Measurement Technology Conference – IMTC*. IEEE, Los Alamitos (2007) 1-4244-0589-0/07
- [21] Narendra, V.G., Hareesh, K.S.: Prospects of Computer Vision Automated Grading and Sorting Systems in Agricultural and Food Products for Quality Evaluation. *International Journal of Computer Applications* (0975 – 8887) 1(4) (2010)
- [22] Ahmad, N.S.E.b., Sulaiman, R.: Texture Analysis for Biscuit Using Wavelet. In: *International Conference on Electrical Engineering and Informatics, Selangor, Malaysia, August 5-7*. IEEE, Los Alamitos (2009)
- [23] Sudhakara Rao, P., Gopal, A., Revathy, R., Meenakshi, K.: Color Analysis of Fruits Using Machine Vision System for Automatic Sorting and Grading. *J. Instrum. Soc. India* 34(4), 284–291
- [24] *Process Operations in Nigeria*. *Agricultural Engineering International: the CIGR. Journal of Scientific Research and Development*. Invited Overview VII, 1-12

- [25] Mereda, P., Ortiz-canavate, J., Garcia-Ramos, F.J., Ruiz-Altisent, M.: Nondestructive technologies for fruits and vegetable size determination- a review
- [26] Raji, Alamutu, A.: Prospects of Computer Vision Automated Sorting Systems in Agricultural Process Operations in Nigeria. *Agricultural Engineering International: the CIGR. Journal of Scientific Research and Development VII* (2005)
- [27] Raji, A.O., Fagboun, A.A., Dania, M.K.: An Approach to Detecting Defects in Food Products. In: *Proceedings of the First International Conference of the Nigerian Institution of Agricultural Engineers*, pp. 36–39 (2000)
- [28] Arivazhagan, S., Newlin Shebiah, R., Selva Nidhyandhan, S., Ganesa, L.: Fruit Recognition using Color and Texture Features. *Journal of Emerging Trends in Computing and Information Sciences* 1(2), 90–94 (2010)
- [29] Shantaiya, S., Ansari, U.: Identification of Food Grains and Its Quality Using Pattern Classification. In: *International Conference [ICCT], Special Issue of IJCCT*, December 3-5, vol. 2(2,3,4), pp. 70–74 (2010)
- [30] Sansomboonsuk, S., Afzulpurkar, N.: Machine vision for Rice Quality Evaluation. In: *Technology and Innovation for Sustainable Development Conference (TISD)*, pp. 343–346 (2008)
- [31] Sudhakara Rao, Gopal, A., Iqbal, S.M., Revathy, R., Meenakshi, K.: Classification of Fruits Based on Shape Using Image-Processing Techniques. *J. Instrum. Soc. India* 34(4), 227–239
- [32] Brosnan, T., Sun, D.-W.: Improving quality inspection of food products by computer vision-a review. *Journal of Food Engineering* 61, 3–16 (2004)
- [33] Wan, Y.-N.: Kernel handling performance of an automatic grain quality inspection system. *Trans. ASAE* 45(2), 369–378 (2002)
- [34] Wan, Y.-N., C.-M. Lin and J.-F. Chiou. Rice quality classification using an automatic grain quality inspection system. *Trans. ASAE* 45(2), 379–388. (2002)
- [35] Zheng, C., Sun, D.-W., Zheng, L.: Recent applications of image texture for evaluation of food qualities - a review. *Trends in Food Science and Technology* 17(3), 113–128 (2006)
- [36] Zaid Abdullah, M., Abdul Aziz, S., Dos-Mohamed, A.M.: Quality inspection of bakery products using a color-based machine vision system. *Journal of Food Quality* 23(1), 39–50 (2000)

Mach-Zehnder Interferometer Based All-Optical Peres Gate

G.K. Maity¹, J.N. Roy², and S.P. Maity³

¹ Calcutta Institute of Technology, Uluberia, Howrah, W.B. India

² Department of Physics, National Institute of Technology,
Agartala, Jirania, 799055, Tipura, India

³ Department of Information Technology, Bengal Engineering and Science University,
Shibpur, Howrah, 711103, India

{goutam123_2005, jnroys}@yahoo.co.in,
santipmaity@it.becs.ac.in

Abstract. Conservative and reversible logic gates are widely known to be compatible with revolutionary computing paradigms such as low-power CMOS, nanotechnology, optical and quantum computing. A conservative reversible logic gate is the Peres Gate (PG). This gate is also known as New Toffoli Gate (NTG), combining Toffoli Gate and Feynman Gate. This paper presents an optical circuit for realization of Peres Gate in all-optical domain. Semiconductor optical amplifier (SOA) based Mach-Zehnder interferometer (MZI) can play a significant role in this field of ultra fast all optical signal processing.

Keywords: Reversible logic, Peres gate, Mach-Zehnder interferometer (MZI).

1 Introduction

Reversible logic is of increasing importance to many future computer technologies. Reversible circuits are those circuits that do not lose information and reversible computation in a system can be performed only when the system comprises of reversible gates [1-5]. These circuits can generate unique output vector from each input vector, and vice versa, that is, there is a one-to-one mapping between input and output vectors. The conventional logic gates such as AND, OR, and EX-OR are not reversible, as they are all multiple-input single output logic gates. A gate is reversible if the gate's inputs and outputs have a one-to-one correspondence, i.e. there is a distinct output assignment for each distinct input. Therefore, a reversible gate must have an equal number of inputs and outputs, inputs can be uniquely determined from its outputs unlike conventional gates,

The output that is not used for further computations is called garbage output. A gate is conservative if the Hamming weight (number of logical one's) at its input equals the Hamming weight at its output. A conservative reversible gate is simultaneously both conservative and reversible. Peres gate (PG), also known as New Toffoli Gate (NTG), combining Toffoli Gate and Feynman Gate is a 3X3 reversible logic gate [1-2]. Peres Gate is equal with the transformation produced by a Toffoli Gate followed by a Feynman Gate.

Design of reversible logic is highly demanding in many applications for lossless data processing. One or two such typical applications may include medical (for correct diagnosis) and military (for security reason through statistical test) data processing that demand high reversibility in order to prevent data loss. To use in future applications, the present work is an attempt for an efficient circuit realization of Peres gate in all-optical domain. Semiconductor optical amplifier (SOA) based Mach-Zehnder interferometer (MZI) can play a significant role in this field of ultra fast all optical signal processing. Simulation has been done with Mathcad-7 and results are reported. The different abbreviations used are as follows:

BC: Beam Combiner; EDFA: Erbium Doped Fiber Amplifier; Wavelength Converter (WC), MZI: Mach-Zehnder Interferometer, CP: control pulse, A, B and C: inputs, X, Y and Z: outputs and (1)(16) are equation numbers.

The rest of the paper is organized as follows. Section 2 describes SOA based MZI switch. Section 3 discusses proposed Peres gate architecture (or realization) with simulation. Section 4 presents simulation and performance evaluation, while concluding remarks are made in Section 5.

2 SOA Based MZI Switch

Mach-Zehnder interferometer (MZI) switch, as shown in Fig.1(a) and (b), is a very powerful optical device to realize ultra fast all-optical switching. In this switch, a semiconductor amplifier (SOA) is inserted in each arm of a MZ interferometer [6-8]. The pulsed signal at the wavelength λ_1 enters the upper arm through coupler C2 such that most power passes through upper arm. At the same time, the incoming signal pulse at the wavelength λ_2 entering port-1 is split equally by this coupler C1 and propagates simultaneously in the two arms. The intensity transmission characteristics at port-3 and port-4 can be expressed as [9]

$$T_3(t) = \frac{1}{4} G_1 \left\{ k_1 k_2 + (1-k_1)(1-k_2) R_G - 2 \sqrt{k_1 k_2 (1-k_1)(1-k_2) R_G} \cos(\Delta\Phi) \right\} \tag{1}$$

$$T_4(t) = \frac{1}{4} G_1 \left\{ k_1(1-k_2) + (1-k_1)k_2 R_G - 2 \sqrt{k_1 k_2 (1-k_1)(1-k_2) R_G} \cos(\Delta\Phi) \right\} \tag{2}$$

where $R_G = G_2/G_1$, G_1 and G_2 are the time dependent gains,

$$\Delta\Phi(t) = -\frac{\alpha}{2} \ln \left(\frac{G_2}{G_1} \right), \quad \alpha \text{ is the line width enhancement factor, (taken 7.5}$$

here), k_1 and k_2 are the ratios of the couplers C1 and C2, respectively. For simplicity we take $k_1 = k_2 = 1/2$. The output signal power at port-3 and port-4 are,

$$P_j(t) = P_{ip}(t) \cdot T_j(t), \quad j = 3, 4. \tag{3}$$

where $P_{ip}(t)$ is the power of the incoming signal pulse. When both beams are present simultaneously (can it be written like this? With the simultaneous presence of both the

beams), the control pulse saturates SOA-1 on change in carrier density inside SOA. The gain of the SOA during this period is [10-12],

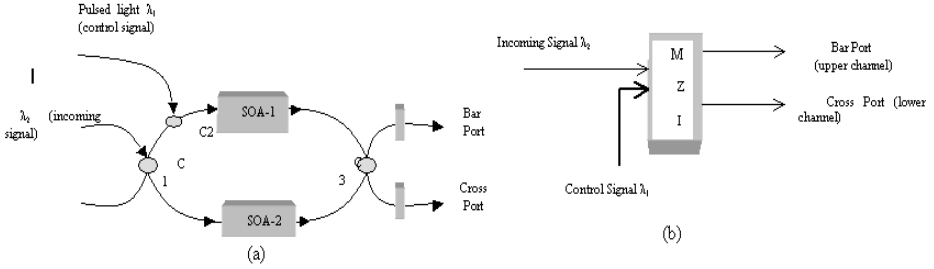


Fig. 1. (a): MZI-based optical switch, (b): Schematic diagram of SOA based MZI optical switch

$$G(t) = \frac{1}{1 - \left(1 - \frac{1}{G_0}\right) \exp\left(-\frac{U_{in}(t)}{U_{sat}}\right)} \tag{4}$$

where U_{sat} is the saturation energy of the SOA and $U_{in}(t) = \int_{-\infty}^t P_{in}(t') dt'$. Here we consider a Gaussian pulse $P_{in}(t) = \frac{E_{in}}{\sigma\sqrt{\pi}} \exp\left(-\frac{t^2}{\sigma^2}\right)$ as control signal, where E_{in} is the input pulse energy, σ is the full width at half maximum (taken 2.8 here).

The gain recovery is achieved in SOA-1 with time constant τ_e . The momentarily (is this term is used widely in the domain or can be called as ‘instantaneous’) gain during this time is [11]

$$G(t) = G_0 \left[\frac{G(t_s)}{G_0} \right]^{\exp[-(t-t_s)/\tau_e]} , \quad t > t_s \tag{5}$$

where, $G(t_s)$ is the gain after saturation of SOA-1. We show the gain change for SOA-1 in the Fig-2(a)(all the simulation and calculation are done with Mathcad-7). Here we take G_0 = unsaturated amplifier gain = 29.6 dB, τ_e = 95 ps, $E_{in}/U_{sat} = 0.1$. From the graph in the Fig-2(a), we find $t_s = 5.5$ ps and $G(t_s) = 7.969$ dB. The beam in the lower arm experiences the unsaturated amplifier gain G_0 (as there is no strong optical pulse to saturate the SOA-2) i.e. $G_1 \neq G_2$, recombines at the coupler C3, So

that $\Delta\Phi \approx \pi$. Hence all one bits are directed toward the bar port (upper port-3 in Fig. 1(a)). In the presence of control pulse, the output pulse at port-3 and port-4 is shown in the Fig-2(b). The transmitted intensity for both the ports are also shown in Fig-2(c). However, in the absence of the λ_1 beam, both the incoming signal beams in two arms experience the same unsaturated amplifier gain G_0 in both SOA (i.e. $G_1 = G_2$), recombine at the coupler C3, So that $\Delta\Phi = 0$.

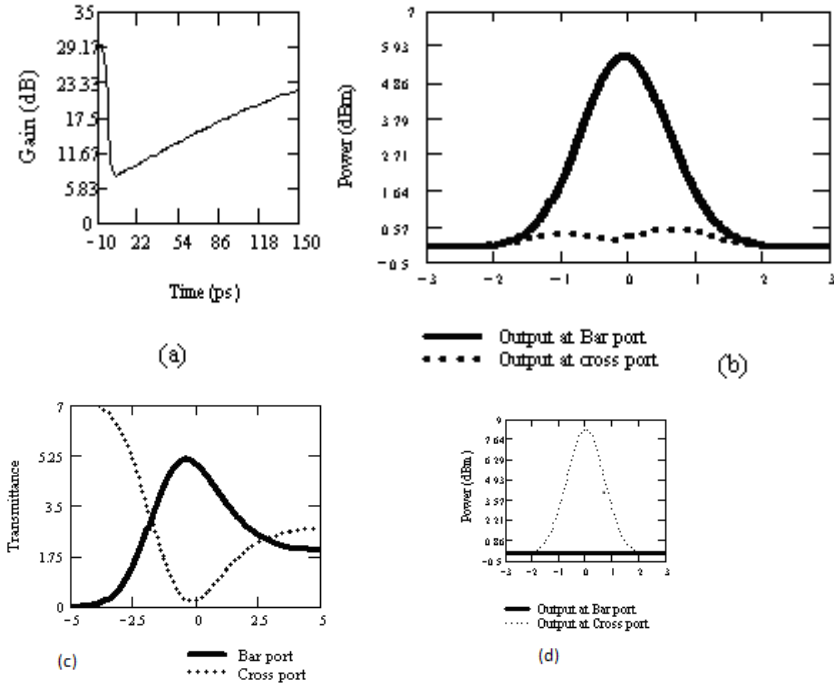


Fig. 2. (a) : Gain Variation in SOA, (b) Output pulse at Bar and cross port when Control pulse=1, (c): Intensity Transmission characteristics for port-3 and port-4 when Control pulse=1, (d): Output pulse at Bar and cross port when Control pulse=0

It can be stated from equation (3) that $P_3(t) = 0$ and the pulse only exits at the cross port (lower port-4 in Fig. 1(a)). In the absence of control pulse, the output pulses at port-3 and port-4 are shown in Fig. 2(d). Optical filters are placed in front of the output ports for blocking the λ_1 signal. The MZ scheme is preferable over cross-gain saturation as it does not reverse the bit pattern and results in a higher ON-OFF contrast, simply because nothing exits from bar port during 0 bits. Now, it is clear that in the absence of control signal (λ_1), the incoming signal (λ_2) exits through cross-port (lower channel) of MZI as shown in Fig.1(a), while no light is present in the bar-port (upper channel). On the other hand, in the presence of control signal, the incoming signal exits through bar port of MZI as shown in Fig. 1(a), while no light is present in

the cross port. In the absence of incoming signal, bar-port and cross- port receive no light as the filter blocks the control signal. Schematic block diagram of MZI is shown in Fig.1 (b).

3 Peres Gate (PG) Architecture and Its Simulation

Peres gate is a (3*3) conservative reversible gate [1, 2, 4, 5]. It has three inputs (A, B, C) and three outputs (X, Y, Z) satisfy the relations as follows:

$$X = A$$

$$Y = A \oplus B$$

$$Z = A.B \oplus C \quad (6)$$

The truth table is in the Table-1. Schematic diagram is given in Fig. 3(a).

Table 1. Truth table of Peres Gate

Inputs			Outputs		
A	B	C	X	Y	Z
0	0	0	0	0	0
0	0	1	0	0	1
0	1	0	0	1	0
0	1	1	0	1	1
1	0	0	1	1	0
1	0	1	1	1	1
1	1	0	1	0	1
1	1	1	1	0	0

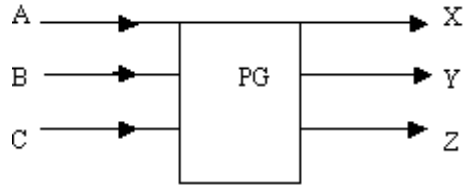


Fig. 3. (a): Schematic diagram of Peres Gate

3.1 Peres Gate (PG) Configuration Using MZI

The MZI based circuit for all optical reversible Peres gate is given in Fig. 4. Light from input A is incident on beam splitter (BS) and is split into two parts. One part is directly connected with MZI-2 and acts as incoming signal for MZI-2. Another part is connected with MZI-1 through wavelength converter (WC) and erbium doped fibre amplifier (EDFA) so that they can act as control signal to MZI-1. In a similar way light signal from input B is connected with MZI-1 and MZI-2 as shown in Fig.4. The bar port of MZI-1 (B_1) and MZI-2(B_2) and cross port of MZI-2 (C_2) are combined by a beam combiner BC-1 to get the output X. Light from cross port of MZI-1 (C_1) and MZI-2(C_2) are combined by BC-2 to get the final output Y. Light signal from MZI-2(B_2) is connected to MZI-3 that acts as incoming signal of MZI-3. The control signal of MZI-3 is taken from the input 'C'. The input C is also connected with MZI-4 as incoming signal. MZI-4 gets its control signal from B_2 . The cross ports of MZI-3 and MZI-4 are combined by a beam combiner (BC-3) to get the output 'Z'.

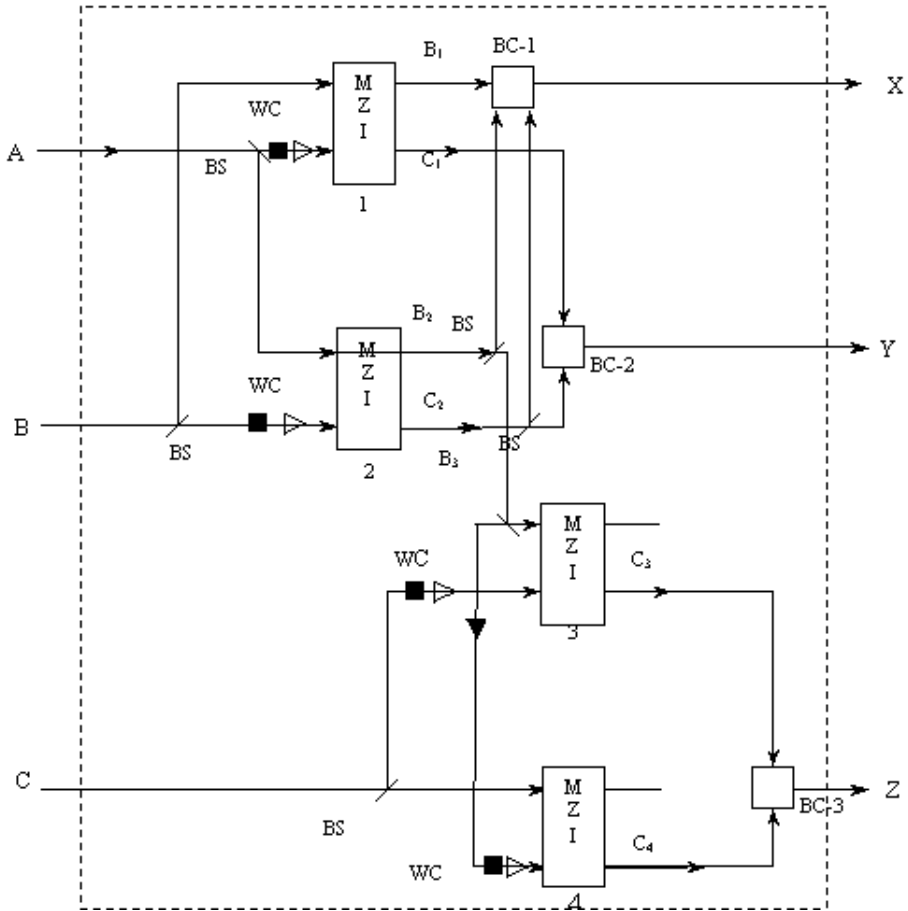


Fig. 4. Circuit for all-optical Peres Gate (PG)

3.2 Functional Operations

Let us discuss the operation in details. Here, the presence of light is taken as 1 state and absence of light is taken as 0 state.

- (1) When $A=B=C=0$ i.e. no light is present at input, the final outputs show no light condition i.e. $X=Y=Z=0$.
- (2) When $A=B=0$ and $C=1$, incoming signal is absent at MZI-1 and MZI-2. So no light comes out through cross or bar port of MZI-1 and MZI-2. Incoming signal is present at MZI-4 but control is absent. Hence, the light beam emerging out through C_4 gives the output $Z=1$. Therefore, when $A=B=0$ and $C=1$ then $X=Y=0, Z=1$. It satisfies the second row of the truth table-1.

- (3) When $A=C=0$ and $B=1$, then incoming signal is absent on MZI-2 and MZI-4 but incoming signal is present at MZI-1. So no light comes out through cross or bar port of MZI-2 and MZI-4 but signal is present at cross port of MZI-1. Therefore, the conditions $X=0, Y=1$ and $Z=0$ are satisfied.
- (4) When $A=0$ and $B=C=1$, then the MZI-2 receives no light in the incoming signal, so $B_2=C_2=0$, and subsequently $C_3=0$. Now for the MZI-4 the incoming signal receives light but the control signal receives no light, hence, $C_4=1$. This indicates $X=0, Y=Z=1$.
- (5) When $A=1$ and $B=C=0$, then MZI-2 receives light in its incoming signal, but control signal is absent. So $B_2=0, C_2=1$ are obtained, which indicates $C_3=0$. Again as the input $C=0$, the incoming signal of MZI-4 is zero, so $C_4=0$. $X=1, Y=1$ and $Z=0$ are satisfied.
- (6) When $A=1, B=0$ and $C=1$, then MZI-2 receives light as incoming signal but does not receive control light signal. Again light from input C acts as incoming signal of MZI-4. Hence $X=1, Y=1$ and $Z=1$ are satisfied.

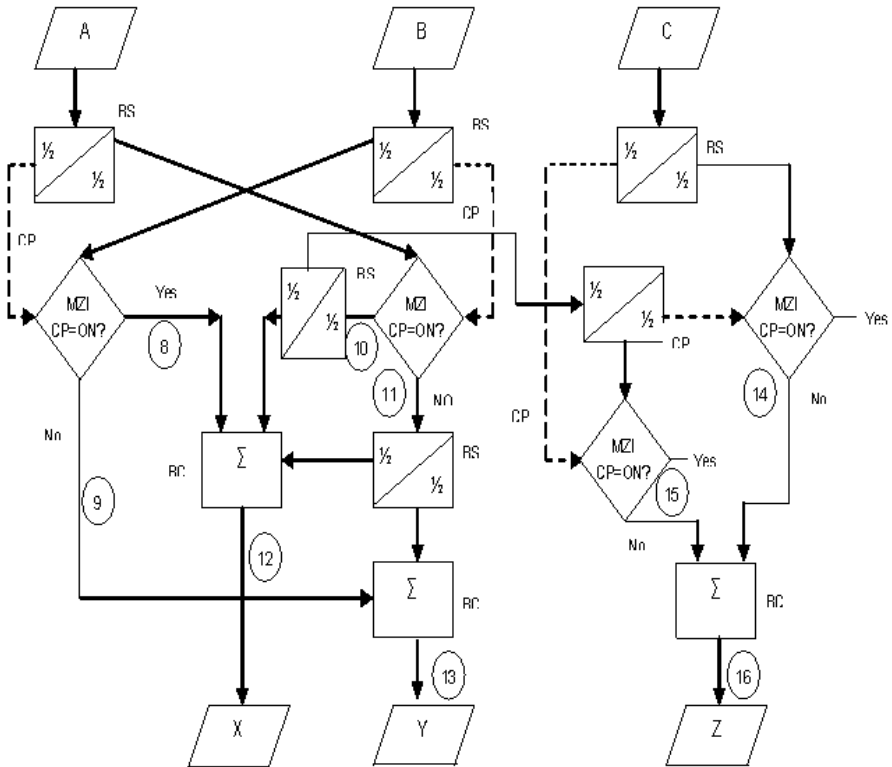


Fig. 5. Flow chat for simulation of Peres Gate, BS: 50:50 beam splitter, BC: beam combiner, MZI: Mach-Zehnder Interferometer, CP: control pulse, A, B and C: inputs, X, Y and Z: outputs and (8)(16) are equation numbers

- (7) When $A=1$, $B=1$ and $C=0$, then MZI-1 and MZI-2 receive light pulse as both the incoming and control signals. MZI-3 receives the incoming light pulse, but does not receive light in its control. The incoming signal of MZI-4 is absent, which means $X=1$, $Y=0$ and $Z=1$.
- (8) When $A=1$, $B=1$ and $C=1$, then all the MZI switches receive light pulse as incoming and control signals. The consequent results are $B_1=1$, $C_1=0$, $B_2=1$, $C_2=0$, $B_3=1$, $C_3=0$, $B_4=1$, and $C_4=0$, which indicates $X=1$, $Y=0$ and $Z=0$.

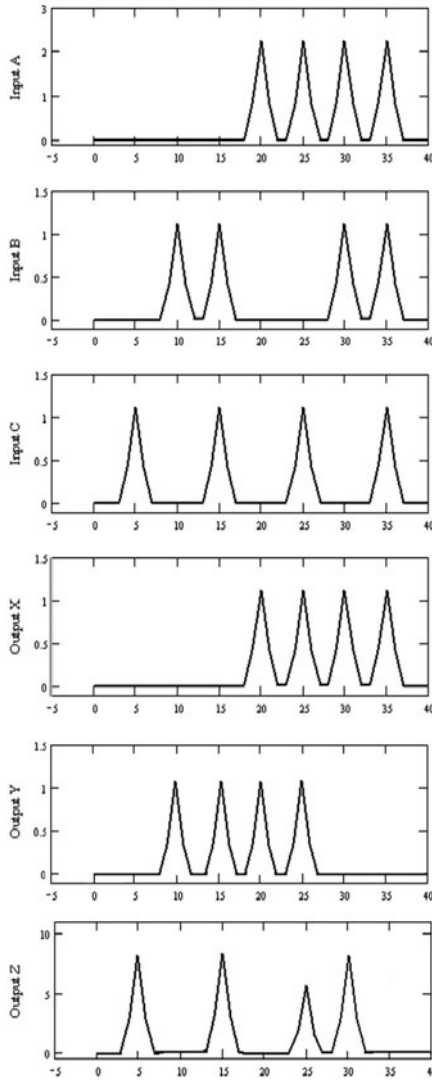


Fig. 6. Simulation result of Fig. 4 [x-axis: Time (ps) & y-axis: Power(dBm)]

4 Simulation and Performance Evaluation

Simulation is done in Mathcad-7 taking the same values used in the section-2. The power of the input pulse is taken $A=1.13$ dBm, $B=2.26$ dBm and $C=1.13$ dBm. Here we use 50:50 beam splitters. Power in different ports can be calculated from the equations given below.

$$P_{B1}(t) = \frac{1}{2} P_B(t) T_3(t) \quad (7)$$

$$P_{C1}(t) = \frac{1}{2} P_B(t) T_4(t) \quad (8)$$

$$P_{B2}(t) = \frac{1}{2} P_A(t) T_4(t) \quad (9)$$

$$P_{C2}(t) = \frac{1}{2} P_A(t) T_4(t) \quad (10)$$

$$P_X(t) = P_{B1}(t) + P_{B2}(t) + P_{C2}(t) \quad (11)$$

$$P_Y(t) = P_{C1}(t) + \frac{1}{2} P_{C2}(t) \quad (12)$$

$$P_{C3}(t) = \frac{1}{4} P_{B2}(t) T_4(t) \quad (13)$$

$$P_{C4}(t) = \frac{1}{2} P_C(t) T_4(t) \quad (14)$$

$$P_Z(t) = P_{C3}(t) + P_{C4}(t) \quad (15)$$

The flow chart of this simulation is shown in Fig. 5. The equation numbers are also given. The simulated result is shown in the Fig. 6.

5 Concluding Remarks

In this paper, MZI based circuit has been proposed and described for realization of Peres gate. It is all-optical in nature. Simulations are done. As Peres gate is the conservative reversible logic gate so the different arithmetic and logic operations in reversible system can be performed. While constructing reversible circuits, some restrictions should be strictly maintained:

- Fan-out is not permitted.
- Loops are not permitted.

References

1. Majid, H., Somayyeh, J.J., Keivan, N., Omid, H.: Design of a Novel Reversible Multiplier Circuit Using HNG Gate in Nanotechnology. *World Applied Sciences Journal* 3(6), 974–978 (2008)
2. Perkowski, M., Kerntopf, P., Buller, A., Chrzanowska-Jeske M., Song, X., Massery B.: Regular realization of symmetric functions using reversible logic, pp. 245-252. *IEEE, Los Alamitos* (2001) 0-7695-1239-9/01
3. Saso, T., Kinoshita, K.: Conservative logic elements and their Universality. *IEEE Trans. on Computers* 28(9), 682–685 (1979)
4. Shamir, J., Caulfield, H.J., Micelli, W., Seymour, R.J.: Optical Computing and Fredkin gates. *Applied Optics* 25(10), 1604–1607 (1986)

5. Vasudavan, D.P., Lala, P.K., Parkerson, J.P.: Reversible λ -logic design with online testability. *IEEE Trans. on Instrumentation and Measurement* 55(2), 406–414 (2006)
6. Agrwal, G.P.: Applications of nonlinear fibre optics. Academic press, India (2001) (an imprint of Elsevier, San Diego, USA)
7. Chattopadhyay, T., Roy, J.N.: Design of SOA-MZI based all-optical programmable logic device (PLD). *Optics Communication* 283(12), 2506–2517 (2010)
8. Schreieck, R., Kwakernaak, M., Jackel, H., Melchior, H.: Ultrafast switching dynamics of Mach-Zehnder Interferometer switches. *IEEE Photonics Technology Letters* 13(6), 603–605 (2001)
9. Zhang, M., Zhao, Y., Wang, L.: Design and analysis of all-optical XOR gate using SOA-based Mach-Zehnder Interferometer. *Optics Communications* 223, 301–308 (2003)
10. Agrawal, G.P., Olsson, N.A.: Self-phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers. *IEEE Journal of Quantum Electronics* 25(11), 2297–2306 (1989)
11. Eiselt, M., Pieper, W., Weber, H.G.: SLALOM: Semiconductor Laser Amplifier in a Loop mirror. *Journal of Lightwave Technology* 13(10), 2099–2112 (1995)
12. Taraphdar, C., Chattopadhyay, T., Roy, J.N.: Mach-Zehnder Interferometer based all-optical Reversible Logic Gate. *Optics and Laser Technology* 42(2), 249–259 (2010)

Using PSO in Image Hiding Scheme Based on LSB Substitution

Punam Bedi¹, Roli Bansal¹, and Priti Sehgal²

¹ Department of Computer Science, University of Delhi,
Delhi, India

² Department of Computer Science, Keshav College,
University of Delhi, Delhi, India
punambedi@ieee.org, rolibansal1@rediffmail.com,
psehgal@keshav.du.ac.in

Abstract. With the massive growth in internet applications, there is a continuous need of efficient steganography techniques for the purpose of secret data communication and for the authentication and ownership identification of host data. This paper presents an efficient image hiding scheme using Particle Swarm Optimization (PSO) in the spatial domain of digital images. The proposed technique uses PSO to find the best pixel locations in an image where the secret image pixel data can be embedded. This PSO algorithm uses the Structural similarity Index (SSIM) as the objective function which is based on the simple visual effect of the human visual perception capability. As a result, the pixel positions generated by the proposed method, when used for embedding secret image data, result in minimum distortion of the host image. The results of the proposed technique have been analyzed qualitatively and quantitatively and also compared with some recent LSB techniques. The results show better stego image quality along with high embedding capacity.

Keywords: Steganography, spatial domain, particle swarm optimization, image hiding.

1 Introduction

With the enormous growth of the internet and digital media technology, there is a large increase in digital communication. This has given rise to a number of issues related to data security in digital communication, invisible communication via digital media, copyright protection of digitized properties etc. Steganography [1] is the technique of encoding secret information in a communication channel in such a way that the very existence of information is concealed. The secret and host data may be image, text, audio or video. While embedding secret data in a host image, the emphasis is on producing a stego image with an acceptable level of quality so that the distortion resulted due to data embedding is imperceptible.

A number of steganography techniques [2] are available for embedding information in an image. These can be broadly classified as spatial domain techniques [3-12] and transform domain techniques [13-15]. In the spatial domain, the simplest

technique is to embed the data in the Least Significant Bits (LSB) of each pixel in the cover image. LSB based embedding algorithms are vulnerable to frequency attacks but more robust to geometric attacks. Transform domain techniques work by computing the Discrete Cosine Transform (DCT) or the Discrete Wavelet Transform (DWT) or the Discrete Fourier Transform (DFT) of the host image and embed the message by modulating coefficients in the corresponding domain. Transform domain techniques are more resilient towards frequency based attacks but are more susceptible to geometric attacks as the distortions introduced by the embedded data into the transform coefficients will spread over all the pixels in the spatial domain. As a result the changes introduced in these pixel values are visually less significant but the overall image quality is reduced.

Generally spatial data hiding schemes can be categorized as LSB substitution, LSB matching and Pixel Value Differencing (PVD) methods. LSB substitution is the most commonly used methods directly replacing LSB's of pixels in the cover image to obtain the corresponding stego image [3]. In LSB matching [4], if one secret bit does not match the LSB of the cover image, then another one will be randomly added or subtracted from the cover pixel value. PVD based methods [5, 6, 7] calculate the difference of two consecutive pixels to determine the depth of embedded bits. Some techniques also used an Optimal Pixel Adjustment Process [8] (OPAP) to improve efficiency and enhance visual quality of the stego image. The above mentioned techniques replace the same length bits of each original pixel. However not all pixels can tolerate changes without noticeable distortion and hence the resulting stego image is of poor quality.

Authors in [9,10] employed a genetic algorithm to generate a substitution table for transforming the value of the secret data to be embedded into each host pixel to another value in advance which is closer to the original value of the host pixel. However, it is possible that even if the substitution table is good, it may not be optimal. Chang et. al. [11] proposed their dynamic programming strategy to efficiently pick the best from all substitution tables. This paper proposes an efficient data hiding method to hide an image into a gray level image imperceptibly without degrading the quality of the resultant image in which the secret image is to be hidden. The proposed data hiding scheme uses Particle Swarm Optimization to obtain the best pixel positions in an image where secret image pixel data can be hidden. PSO was used in [12] to embed a message in an image using LSB. In the proposed work, PSO is used to hide an image within another image using LSB technique. The paper also proposes a technique to hide MSBs of the image in the LSBs of the cover image. The proposed techniques gives better results than the standard LSB technique and the techniques based on genetic algorithms and dynamic programming as it is based on the simple visual effect of the human visual perception capability. The better values of quality metrics like Peak signal Noise Ratio (PSNR), Mean Square Error (MSE) and Structural Similarity Index (SSIM) depict the effectiveness of the proposed technique in maintaining imperceptibility and improving stego image quality along with high embedding capacity. The rest of the paper is organized as follows: Section 2 explains Particle Swarm Optimization. The proposed technique for secret image embedding is explained in section 3. Section 4 explains the method for extraction of the secret image. Section 5 presents the experimental results with discussion and finally, section 6 concludes the paper.

2 Particle Swarm Optimization

Particle Swarm Optimization (PSO) [16, 17] is a population based stochastic optimization technique, inspired by social behavior of bird flocking or fish schooling. A PSO algorithm maintains a swarm of particles, where each particle represents a potential solution. A swarm is similar to a population, while a particle is similar to an individual. The particles are “flown” through a multidimensional search space, where the position of each particle is adjusted according to its own experience and that of its neighbors. Let $x_i(t)$ denote the position of particle i in the search space at time step t . The position of the particle is changed by adding a velocity $v_i(t)$ to the current position, i.e.

$$x_i(t+1) = x_i(t) + v_i(t+1) . \quad (1)$$

It is the velocity vector that drives the optimization process, and reflects both the experiential knowledge of the particle and socially exchanged information from the particle’s neighborhood. The experiential knowledge of a particle is generally referred to as the cognitive component, which is proportional to the distance of the particle from its own best position found since the first time step. The socially exchanged information is referred to as the social component of the velocity equation. PSO algorithms can be categorized as local best PSO and global best PSO. The local best PSO uses a ring social network topology where small neighborhoods are defined for each particle. For the global best PSO, the neighborhood for each particle is the entire swarm. The social network employed by the gbest PSO reflects the star topology. Here, all particles are interconnected and each particle can therefore communicate with every other particle. The social component of the particle velocity update reflects information obtained from all the particles in the swarm. In this case, the social information is the best position found by the swarm, referred to as $\hat{y}(t)$. Each particle is attracted towards the best solution found by the entire swarm. For gbest PSO, the velocity of particle i is calculated as:

$$v_{ij}(t+1) = v_{ij}(t) + c_1 r_{1j}(t)[y_{ij}(t) - x_{ij}(t)] + c_2 r_{2j}(t)[\hat{y}_j(t) - x_{ij}(t)] \quad (2)$$

where $v_{ij}(t)$ is the velocity of particle i in dimension $j = 1, \dots, n_x$ at time step t . $x_{ij}(t)$ is the position of particle i in dimension j at time step t , c_1 and c_2 are positive acceleration constants used to scale the contribution of cognitive and social components respectively. $r_{1j}(t)$, $r_{2j}(t) \sim U(0,1)$ are random values in the range $[0,1]$, sampled from a uniform distribution. The personal best position y_i associated with particle i is the best position the particle has visited since the first time step. Considering minimization problems, the personal best position at the next time step $t+1$ is calculated as:

$$y_i(t+1) = \begin{cases} y_i(t) & \text{if } f(x_i(t+1)) \geq f(y_i(t)) \\ x_i(t+1) & \text{if } f(x_i(t+1)) < f(y_i(t)) \end{cases} \quad (3)$$

where f is the objective/fitness function which measures how close the corresponding solution is to the optimum, i.e. it quantifies the performance, or quality of a particle (or solution). The global best position $\hat{y}(t)$ at time step t , is defined as:

$$\hat{y}(t) \in \{y_0(t), \dots, y_{ns}(t)\} \mid f(\hat{y}(t)) = \min\{f(y_0(t)), \dots, f(y_{ns}(t))\} \quad (4)$$

where, ns is the total no. of particles in the swarm.

The PSO process is iterative. After generating an initial swarm the value of the fitness function is evaluated in every iteration and the velocity and position of each particle is updated accordingly. The algorithm is terminated when one of the following occurs:

1. Maximum number of iterations has been reached.
2. An acceptable solution has been found.
3. No improvement is observed over a number of iterations.

3 Proposed Technique

The proposed data hiding algorithm (Fig.1) aims at searching optimum locations adaptively in the cover image so that the stego image thus produced is less distorted and is therefore of good quality. The method consists of the following steps:

Step 1: Read the cover image I and the secret image S which is to be hidden.

Step 2: Convert the secret image S to a sequence of binary digits that can be hidden in the cover image. For this purpose two approaches can be used.

- 1) Convert each pixel value of S into its eight bit representation and concatenate them to obtain a sequence of bits.
- 2) Extract four MSB's of each pixel in S and concatenate them to obtain a sequence of bits to be hidden in I .

In the former case the extracted secret image would be of the same quality as the input secret image but the quality of stego image would be lower, whereas in the latter case, the extracted image would be marginally lower in quality but the stego image quality would be much better. Moreover at the same level of distortion in the stego image, a much larger secret image could be hidden into the host cover image. Let the constructed bit stream of the secret image be of length len .

Step 3: Divide the $I(m \times n)$ into a number of blocks each of size pxq .

Step 4: In order to obtain the best pixels for each block, set certain parameters for the PSO function. The number of iterations can be set according to the level of accuracy required. Specify the number of particles in the swarm depending on the bit stream size and the size of the cover image. Each particle has a dimensionality equal to the number of pixels that must be located for hiding which in turn depends on the number of bits to be hidden in each pixel.

For hiding 1 bit in a pixel the number of pixels required from the cover image is len .

For hiding 2 bits in a pixel the number of pixels required is $len/2$.

For hiding 3 bits in a pixel the number of pixels required is $len/3$.

For hiding 4 bits in a pixel the number of pixels required is $len/4$.

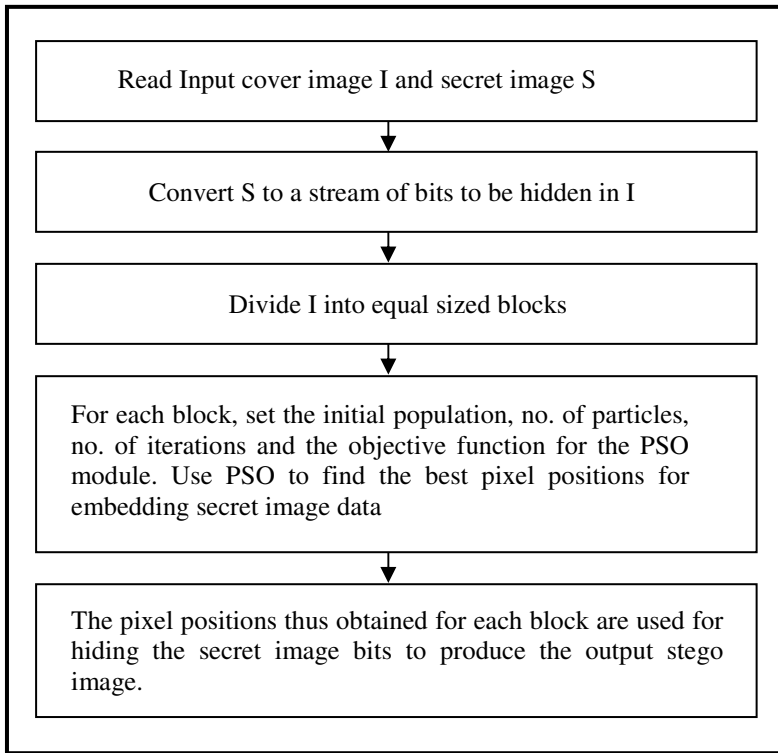


Fig. 1. Block diagram of the proposed technique

Step 5: Set the objective function. The proposed method uses the Structural similarity index (SSIM) as the quality metric to be used as the objective function. It is the generalized form of the Universal Image Quality Index (UQI) [18] defined as follows:

$$Q = 4\sigma_{xy} \hat{x}\hat{y} / (\hat{x}^2 + \hat{y}^2)(\sigma_x^2 + \sigma_y^2) \quad (5)$$

where, x and y are corresponding windows of the same size of the original and stego images and \hat{x} and \hat{y} are the corresponding averages of x and y respectively. σ_x^2 and σ_y^2 are the corresponding variances of x and y and σ_{xy} is the covariance of x and y . The quality index takes the Human Visual system (HVS) into consideration. It models any distortion as a combination of three different factors: the loss of correlation, the luminance distortion and the contrast distortion. In order to understand this, its definition can be rewritten as a product of three components:

$$Q = (2\sigma_{xy} / \sigma_x \sigma_y) \times (2\hat{x}\hat{y} / (\hat{x}^2 + \hat{y}^2)) \times (\sigma_x \sigma_y / (\sigma_x^2 + \sigma_y^2)). \quad (6)$$

The first component measures the degree of linear correlation between x and y with dynamic range $[-1, 1]$. The second component measures how close the mean luminance is between x and y with dynamic range $[0, 1]$. The last component with range $[0, 1]$ measures how similar the contrasts of the images are. Therefore, the

dynamic range of Q is $[-1, 1]$ with the best value of 1 and the worst value of -1. The UQI produces unstable results when either of the terms in the denominator are very close to zero. To avoid this problem the measure has been generalized to SSIM. Structural Similarity Index (SSIM) measures the similarity between two images and is defined as:

$$SSIM(x, y) = (2\hat{x}\hat{y} + c_1)(2\sigma_{xy} + c_2) / (\hat{x}^2 + \hat{y}^2 + 1)(\sigma_x^2 + \sigma_y^2 + c_2) \quad (7)$$

where, x and y are corresponding windows of the same size of the original and stego images and \hat{x} and \hat{y} are the corresponding averages of x and y respectively. σ_x^2 and σ_y^2 are the corresponding variances of x and y and σ_{xy} is the covariance of x and y . c_1 and c_2 are appropriate constants. Although the SSIM has a relatively simple mathematical definition, experiment shows that it outperforms the PSNR under different types of image distortion. The following objective function will be minimized by PSO.

$$Fitness(I, S) = 1 - SSIM(I, I') \quad (8)$$

where, I and I' are original and stego images respectively.

Step 6: After setting all the parameters, PSO is executed for each block to minimize our objective function. The output gives us an array of locations in that block which are best suitable for hiding. The block which produces the minimum value of the objective function can be used to hide more bits than other blocks. In this way we can hide variable number of bits in different blocks. For this purpose, the output array of each block is used and the secret image bits are hidden in 1, 2, 3 or 4 LSB's of the cover image pixels to obtain the required stego image.

4 Extraction of Secret Image

The hidden image can be extracted by the reverse process. A secret key is used at the receiver end to extract the secret image data from the stego image. The secret key can be send to the receiver separately in a secure manner. This secret key consists of a set of ordered pairs of the image block number and the number of bits of the hiding scheme applied in that block. For example: if the secret key generated is [2 3; 4 1; 3 2; 1 3], then it signifies that the second image block applies 3 bit LSB substitution, fourth image block applies 1bit LSB substitution and so on. Using this secret key, the receiver can move to the corresponding image blocks and compute the indexes where the secret image bits are embedded in that block. The receiver then extracts the appropriate number of LSB's as indicated by the secret key, from the set of best pixel locations. These bits are concatenated to get the bit stream of the original secret image S .

5 Experimental Results

The performance of the proposed technique has been evaluated and compared with the simple LSB technique, the Chang et al. [11] technique and the Khodaei et al.[10]

technique for hiding an image in another image. Matlab 8 has been used for the implementation of our technique. In the implementation, four standard images (Fig 2) of size 256x256 have been taken. All these images have been used to hide smaller images (Fig 3) of size 128x128 using the simple LSB and the proposed techniques, where the input image is divided into 4 blocks and PSO is run block wise to obtain best pixel locations in each block. 4 LSBs of one cover image pixel are used to hide 4 bits of the input bit stream of the secret image data. The objective quantitative measures used for comparison between the original and the stego images of dimension $M \times N$ are as follows:-

- The simplest measure of distortion is the Mean Square Error (MSE) and is defined as :

$$MSE = 1/MN \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [O(i, j) - W(i, j)]^2 \quad (9)$$

where, $O(i,j)$ and $W(i,j)$ are corresponding pixel intensities of the original and stego images respectively.

- Peak Signal Noise Ratio (PSNR) defined as:

$$PSNR = 10 * \log_{10} \left\{ \frac{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} 255^2}{\sum_{i=0}^{M-1} \sum_{j=0}^{N-1} [O(i, j) - W(i, j)]^2} \right\} \quad (10)$$

where, $O(i,j)$ and $W(i,j)$ are corresponding pixel intensities of the original and stego images respectively.

- Structural Similarity Index (SSIM) as already defined in eq (7) in the previous section.

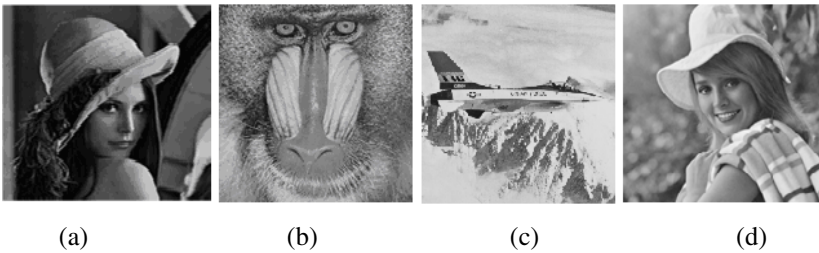


Fig. 2. Standard test covers of size 256x256. (a) Lena (b) Baboon (c) Jet (d) Elaine

Table 1 and 2 list the comparative MSE values between cover and stego images for all the techniques. Table 1 used the Boats image as secret image and table 2 used the Tank image as secret image. Similarly, Table 3 and 4 list the comparative PSNR values between cover and stego images for all the techniques. Table 3 used the Boats image as secret image and table 4 used the House image as secret image. The results for Chang method and Khodaei method have been taken from Khodaei et al. [10].

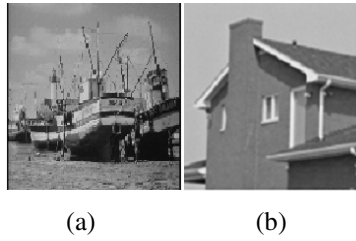


Fig. 3. Standard test images used as secret images of size 128X128. (a) Boats (b) House

Table 1. Values of the quality measure MSE obtained after embedding the Boats image

Host Image	Simple LSB Method	Chang et al [11] Method	Khodaei et al [10] Method	Proposed Method using PSO	Proposed Method using PSO hiding only 4 MSB's of the secret image
Lena	18.74	15.62	16.92	15.26	8.2152
Baboon	17.83	14.28	15.78	14.07	7.65
Jet	17.47	15.12	16.01	14.84	7.82
Elaine	18.10	15.71	16.24	15.56	7.53

Table 2. Values of the quality measure MSE obtained after embedding the House image

Host Image	Simple LSB Method	Chang et al [11] Method	Khodaei et al [10] Method	Proposed Method using PSO	Proposed Method using PSO hiding only 4 MSB's of the secret image
Lena	18.36	16.27	17.01	15.75	10.85
Baboon	17.51	14.70	15.89	14.59	9.68
Jet	17.32	14.33	15.47	14.42	10.13
Elaine	17.55	13.93	16.09	15.73	9.78

Table 3. Values of the quality measure PSNR obtained after embedding the Boats image

Host Image	Simple LSB Method	Chang et al [11] Method	Khodaei et al [10] Method	Proposed Method using PSO	Proposed Method using PSO hiding only 4 MSB's of the secret image
Lena	35.40	36.19	35.84	36.29	38.98
Baboon	35.61	36.58	36.14	36.64	39.29
Jet	35.70	36.15	36.08	36.41	39.19
Elaine	35.55	36.16	36.02	36.20	39.36

Table 4. Values of the quality measure PSNR obtained after embedding the House image

Host Image	Simple LSB Method	Chang et al [11] Method	Khodaei et al [10] Method	Proposed Method using PSO	Proposed Method using PSO hiding only 4 MSB's of the secret image
Lena	35.49	36.01	35.82	36.15	37.77
Baboon	35.69	36.45	36.11	36.48	38.27
Jet	35.74	36.56	36.23	36.53	38.07
Elaine	35.68	36.69	36.06	36.16	38.22

Table 5. Values of the quality measure SSIM obtained after embedding

Host Image	SSIM when secret image is Boats			SSIM when secret image is House		
	Simple LSB Method	Proposed Method using PSO	Proposed Method using PSO hiding only 4 MSBs of the secret image	Simple LSB Method	Proposed Method using PSO	Proposed PSO Method hiding only 4 MSBs of the secret image
Lena	0.9050	0.9150	0.9515	0.9160	0.9163	0.9399
Baboon	0.9596	0.9695	0.9832	0.9627	0.9691	0.9802
Jet	0.9054	0.9130	0.9557	0.8972	0.9138	0.9473
Elaine	0.9070	0.9183	0.9533	0.9139	0.9166	0.9440

The comparative results of table 1 and 3 are further illustrated by graphs in Fig 4. Table 5 lists the SSIM values between cover and stego images for the simple LSB technique and the proposed techniques. The better values of all the quality metrics depict the effectiveness of the proposed technique in maintaining imperceptibility and improving stego image quality along with high embedding capacity.

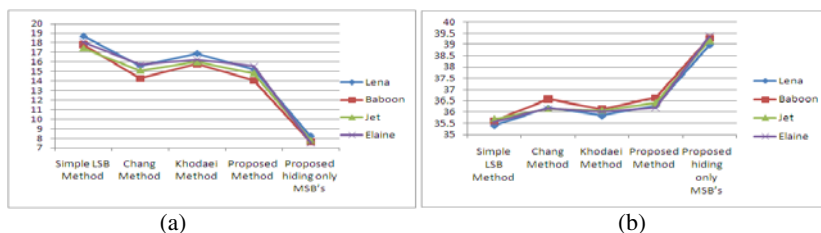


Fig. 4. Comparative (a) MSE and (b) PSNR values obtained from various techniques when Boats image was hidden in different host images

6 Conclusions

This paper presents a steganography technique using Particle Swarm Optimization (PSO) for hiding a secret image in another image. PSO is used for finding the best pixel locations in the cover image for hiding the pixel data of the secret image. The algorithm presented has been applied for hiding image data using all 8 bits of each pixel as well as using 4 MSBs of the image which we want to hide in 4 LSBs of the cover image. The performance of the proposed technique has been evaluated and compared with the simple LSB technique and with other dynamic programming and genetic algorithm based techniques. The results show the effectiveness of this method in producing high quality stego images even at a high embedding rate. Moreover, by hiding only 4 MSB's of the secret image pixels in the host image, the distortion in the stego image reduces considerably thereby increasing its quality.

References

1. Katzenbeisser, S., Petitcolas, F.A.P.: Information hiding techniques for Steganography and Digital Watermarking. Artech House Inc., Boston (2002)
2. Cheddad, A., Condell, J., Curran, K., Mc Kevitt, P.: Digital Image Steganography, Survey and Analysis of Current Methods. Signal Processing 90(3), 727–752 (2010)

3. Chang, C.C., Lin, M., Lu: A fast and secure image hiding scheme based on LSB substitution. *J. Pattern Recognition and Artificial Intelligence* 16(4), 319–416 (2002)
4. Xiaolong, L., Yang, B., Cheng, D., Zheng, T.: A Generalization of LSB Matching. *IEEE Signal Processing Letters* 16(2), 69–72 (2009)
5. Wu, D.C., Tse, W.H.: A Stegnographic method for images by pixel-value differencing. *Pattern Recognition Letters* 24, 1613–1626 (2003)
6. Wang, C.M., Wu, N., Tsai, C.H., Hwang, M.S.: A high quality steganographic method with pixel value differencing and modulus function. *J. Systems and Software* 81(1), 150–158 (2008)
7. Yang, C.H., Weng, C.Y., Wang, S.J., Sun, H.M.: Adaptive data hiding in edge areas of pixels with spatial LSB domain systems. *IEEE Transactions on Information Forensics and Security* 3(3), 488–497 (2008)
8. Chan, C.K., Heng, L.M.: Hiding data in images by simple LSB substitution. *Pattern Recognition* 37, 469–474 (2004)
9. Wang, R.Z., Lin, C.F., Lin, J.C.: Image Hiding by Optimal LSB Substitution and Genetic Algorithm. *Pattern Recognition* 34, 671–683 (2001)
10. Khodaei, M., Faez, K.: Image Hiding by Using Genetic Algorithm and LSB Substitution. In: Elmoataz, A., Lezoray, O., Nouboud, F., Mammass, D., Meunier, J. (eds.) *ICISP 2010*. LNCS, vol. 6134, pp. 404–411. Springer, Heidelberg (2010)
11. Chang, C.C., Chan, C.S., Fan, Y.H.: Image Hiding scheme with Modulus Function and Dynamic Programming Strategy on Partitioned pixels. *Patt. Recog.* 39, 1155–1167 (2006)
12. Bajaj, R., Bedi, P., Pal, S.K.: Best Hiding Capacity Scheme for Variable Length Messages Using Particle Swarm Optimization. In: Panigrahi, B.K., Das, S., Suganthan, P.N., Dash, S.S. (eds.) *SEMCCO 2010*. LNCS, vol. 6466, pp. 230–237. Springer, Heidelberg (2010)
13. Barni, M., Bartiloni, F., Cappellini, V., Pivo, A.: A DCT domain system for robust image watermarking. *Signal Processing* 66, 357–372 (1998)
14. Woo, C.S., Du, J., Pham, B.: Performance Factors Analysis of a Wavelet based Watermarking Method. In: *Australasian Information Security Workshop, AISW 2005* (2005)
15. Li, X., Wang, J.: A Stegnographic method based on JPEG and Particle Swarm Optimization algorithm. *Pattern Recognition* 177, 3099–3109 (2007)
16. Engelbrecht, A.P.: *Fundamentals of Computational Swarm Intelligence*. John Wiley & Sons, Inc., West Sussex (2005)
17. Engelbrecht, A.P.: *Computational Intelligence: An Introduction*. John Wiley & Sons, Inc., West Sussex (2007)
18. Wang, Z., Bovik, A.C.: A Universal Image Quality Index. *IEEE Signal Processing Letters* 9(3)

Matrix Embedding Using Random Linear Codes and Its Steganalysis

P. Harmya, S.S. Anju, Noopa Jagadeesh, and Aishwarya Nandakumar

Centre For Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore, India
{harmya.gopalan, anjuss.cys,
noopajagadeesh, aishwarya.nk12}@gmail.com

Abstract. To improve embedding efficiency, one of the most important steganographic scheme is Matrix embedding, otherwise known as syndrome coding. The schemes that employ matrix embedding is generally more secure because it brings about lower embedding changes.. In this paper matrix embedding scheme based on random linear codes of small dimension is considered. It is suitable for large payloads which makes it appropriate for practical applications. This scheme is efficient for embedding messages close to the embedding capacity. This scheme not only provides high embedding capacity but also enhances the quality of the stego image. The method is proved to withstand attacks like Chi-square and RS steganalysis.

Keywords: Steganography, Binary Linear Codes, Matrix Embedding, Chi-square, RS steganalysis.

1 Introduction

The term steganography is derived from Greek language and means covered writing. It is the art of communication through files in a manner that the very existence of the message is unknown. For a steganographic scheme, the main requirement is statistical undetectability. By undetectability, we understand the inability of an attacker to distinguish between stego and cover objects, given the knowledge of the embedding algorithm and the source of cover media. If two different embedding schemes share the type of cover media and embedding operation, the less detectable one will be that which introduces fewer embedding changes. Embedding efficiency is the number of message bits embedded per one embedding change. Matrix embedding improves efficiency.

2 Basic Concepts

Let F_2^n denote the space of all n -bit column vectors $x=(x_1, \dots, x_n)$. A binary $[n, k]$ code C of block length n and dimension k is a k -dimensional vector subspace of F_2^n , where the sum of two vectors and a multiplication of a vector by scalar are defined using the usual binary arithmetic. The k basis vectors are written as rows of a matrix

form the generator matrix G . The orthogonal complement of an $[n, k]$ code is an $[n, n - k]$ code (called the dual code to C) with an $(n-k) \times n$ generator matrix H with the property that $Hx=0$ for each $x \in C$. The matrix H is the parity check matrix of C . For any $x \in F_2^n, s = Hx \in F_2^{n-k}$ is called the syndrome of x . For each syndrome $s \in F_2^{n-k}$, the set $C(s) = \{x \in F_2^n | Hx = s\}$ is called a coset. Also, $C(0) = C$

.Cosets associated with different syndromes are disjoint with each other. Also, from elementary linear algebra, every coset can be written as $C(s) = x + C$, where $x \in C(s)$ is arbitrary. Thus, there are $2^{(n-k)}$ disjoint cosets, each consisting of 2^k vectors. The hamming distance is defined as the number of places where x and y differ,

$$d_H(x, y) = \left| \{i \in \{1, \dots, n\} | x[i] \neq y[i]\} \right| \tag{1}$$

In the context of steganography, $d_H(x, y)$ is the number of embedding changes.

The hamming weight of $x \in F_2^n$ is $w(x) = d_H(x, 0)$, which is the number of non-zero elements in x . Any member of the coset $C(s)$ with the smallest Hamming weight is called a coset leader and will be denoted as $e_L(s)$.

The ball of radius r centered at x is the set

$$B(x, r) = \{y \in F_2^n | d_H(x, y) \leq r\} \tag{2}$$

The covering radius R of code C is defined as

$$R = \max_{x \in F_2^n} d_H(x, C) \tag{3}$$

where $d_H(x, C) = \min_{c \in C} d_H(x, c)$ is the distance between x and the code. In other words covering, the covering radius is determined by the most distant point x from the code.

Hamming Weight of any coset leader is at most R , the covering radius of C . Consider $x \in F_2^n$ and calculate its syndrome $s = Hx$. Then,

$$R = \max_{z \in F_2^n} d_H(z, C) \geq d_H(x, C) = \min_{c \in C} w(x - c) = w(e(s)) \tag{4}$$

because when c goes through all the codewords, $x-c$ goes through all the members of the coset $C(s)$ [3]. Constructing good covering codes for applications in Steganography is proposed in [5].

3 Matrix Embedding

The matrix embedding theorem provides a recipe for how to turn any linear code into a matrix embedding method. The parameters of the code will determine the properties of the stegosystem, namely its payload and embedding efficiency. Let π be a bit

assignment function $\pi : \mathcal{X} \rightarrow \{0,1\}$ that assigns a bit to each possible value of the cover element from \mathcal{X} . For example, $\mathcal{X} = \{0, \dots, 255\}$ for 8-bit grayscale images and π could be the LSB of the pixels. Thus a group of n pixels can be represented using a binary vector $x \in F_2^n$.

The steganographic scheme is a pair of mappings Emb and Ext,

with the property
$$Emb : F_2^n \times M \rightarrow F_2^n, \tag{5}$$

$$Ext : F_2^n \rightarrow M, \tag{6}$$

$$Ext(Emb(x, m)) = m, \forall x \in F_2^n \text{ and } m \in M \tag{7}$$

Here, $y = Emb(x, m)$ is the vector of bits extracted from the same block of n pixels in the stego image and M is the set of all messages that can be communicated.

Let us suppose that it is possible to embed every message in M using at most R changes,

$$d_H(x, Emb(x, m)) < R, \tag{8}$$

$$\forall x \in F_2^n, m \in M$$

Embedding efficiency is hence

$$e = (\log_2 |M|) / R_a \tag{9}$$

$$R_a = E[d_H(x, Emb(x, m))]$$

where R_a is the average number of embedding changes over all messages and covers. The lower embedding efficiency is $e = \log_2 |M| / R$ (10)

Since $R_a \leq R$ we have $e \leq e [2]$.

4 Matrix Embedding Using Random Linear Codes for Large Relative Payloads

The positive impact of matrix embedding on security is more important for long messages than for shorter ones because longer messages are easier to detect. When considering long messages, the relative payload is large or $\alpha \approx 1$. The most time consuming part in matrix embedding is finding the coset leader. However, for structured codes, such as hamming codes finding a coset leader is particularly simple, for general random codes it is an NP-complete problem whose complexity increases exponentially with n . But, in this matrix embedding using random linear codes we can afford to keep the number of codewords small and find the coset leaders using brute force. For large relative payloads $\alpha = (n - k) / n \approx 1$, the codes in the matrix embedding $[n, n - \alpha n]$ have small dimension $k = n(1 - \alpha)$, and thus a smaller number of codewords. [1]. Assume that the sender wants to embed $n - k$ bits, $m \in \{0, 1\}^{n - k}$, in n

pixels with bits $x \in \{0,1\}^n$. The code parity-check matrix H is generated randomly (from a shared stego key) in its systematic form $H = [I_{n-k}, A]$, where I_{n-k} is an $(n - k) \times (n - k)$ unity matrix and A is a random [4]. The generator matrix is thus $G = [A', I_k]$

4.1 The Algorithm

1. To embed M bits in an N element cover object, first find n such that $\alpha_n \leq M/N < \alpha_{n-1}$.
2. Read the next n bits x from the cover object and read the next message segment m of length $n-k$.
3. Calculate the syndrome H^*x . Find any $\bar{y} \in C(m - Hx)$ that solves $H \bar{y} = m - Hx$. Because H is in its systematic form, for example $\bar{y} = (m - Hx, 0)$, where $0 \in \{0,1\}^k$
4. Find the coset leader $e(m - Hx)$. For this, find the codeword \bar{c} closest to, $\bar{y}, d_H(\bar{y}, \bar{c}) = \min_{c \in C} d_H(\bar{y}, c) = \min_{c \in C} w(\bar{y} - c) = w(e(m - Hx))$

Because the vector $\bar{y} - c$ goes through all the members of the coset.

5. [Embedding modifications] Modify the cover object so that $y = x + e(m - Hx)$. If there are no more message bits to be embedded, stop, otherwise go to 1.
6. [Extraction step] The message bits are extracted by following the same embedding path and calculating $n - k$ bits m from each segment of n bits y of the stego object $m = H^*y$.

5 Experimental Results

This section depicts and analyzes the experimental results by using the matrix embedding method. Random linear codes with varying values of n and k were created and these were used for the embedding purpose. The random linear codes allows the sender to choose the code length n to closely match $\alpha_n = (n - k) / n$ as the relative payload length and thus efficiently use the available embedding space in the cover object. In our experiments, we used about 150 images with size 512×512 to evaluate the performance of the scheme. Our evaluation starts with the well-known images Lena, Baboon and Tiffany which are shown in Figures 1(a)–1(c). Our experiments include embedding efficiency, visual quality, Chi-square attack and RS attack. A pseudorandom number generator with equal probabilities for bits “1” and “0” were

used to generate secret messages in our experiments. The generated stego-images shown in Figures 1(d)-1(f) shows that they contain no artifacts that can be identified by human eyes.

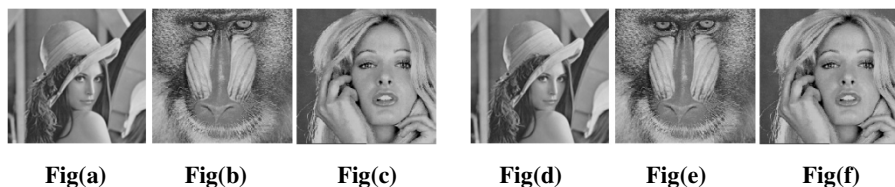


Fig. 1. Cover images (a)-(c) and Stego images (d)-(f) embedded using the linear code [100, 14]

5.1 Embedding Efficiency

Table 1 gives the values of efficiency obtained for class of codes with $n=100$ and $k=10$ and $k=14$.

Table 1. Efficiency for codes with $k=10$ and $k=14$

N	K	A	E
100	10	0.90	2.704
100	14	0.86	2.8202

5.2 PSNR Values

We compared matrix embedding using random linear code with LSB method. The PSNR value is used to compare the quality of stego-images under the same embedding payload. We can see that the matrix embedding has the least distortion.

Table 2. PSNR values for different standard images

Cover images	K=10	K=14	LSB method	PVD
Lena	52.8175	52.90	49.5	46.2
Baboon	52.512	52.6	49.0	46.2
Barbara	52.80	52.85	49.5	46.1
Tiffany	52.80	52.9	49.5	46.1
Pepper	52.51	52.55	49.4	46.2
Airplane	52.55	52.7	49.4	46.1

5.3 Chi-Square Attack

Image formats of type bmp are mainly considered for this type of attack. The embedding percentages that are taken into consideration are 27%, 50%, 90%. Messages were embedded using matrix embedding scheme. Fig.2 shows the experimental result for a stego-image with the matrix embedding. The x-axis gives the percentage of pixels from 1% to 100%. The y-axis gives the probability of a hidden message existing in the image. The Chi-square attack is said to be successful if the probability of embedding is nearly equal to 1[8]. Using the matrix embedding to embed a message, the Chi-square index is always below 0.1, no matter what embedding rate is chosen and hence we conclude that matrix embedding resists Chi-square attack.

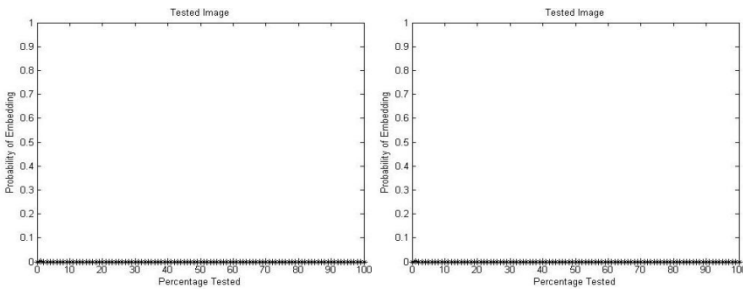


Fig. a.

Fig. b

Fig. 2. Chi-square attack. Fig .a Stego image (50% embedded using matrix embedding) Fig. b Stego image (90% embedded using matrix embedding)

5.4 RS Attack

To check whether the matrix embedding using random linear code can be detected with some newly announced related statistical steganalytic techniques, we tested the stego-images yielded by the method with the dual statistics method proposed by

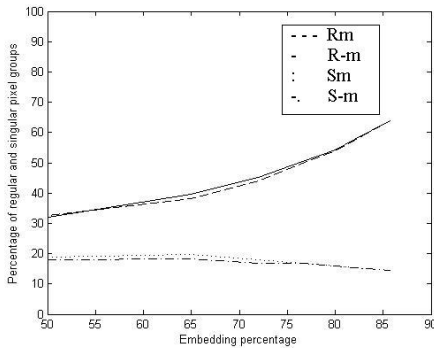


Fig. 3. RS diagram for matrix embedding using random linear codes

Fridrich et al. (2001) as a demonstration[6][7]. The results of RS attack on the matrix embedding scheme as shown by the RS-diagram of Fig. 3 indicate that the stego images seemingly do not contain any embedded data in their LSBs, because the expected value of R_m is seen close to that of $R-m$ and so are the case of S_m and $S-m$, i.e., $R_m \cong R_{-m}$ and $S_m \cong S_{-m}$. This proves that the steganographic method is secure from the perspective of the RS method.

6 Conclusion

Matrix embedding using random linear codes is a novel steganographic scheme that not only keeps high stego-image quality but also conceals large amount of data into cover images for secret communication. Since their relative embedding capacity densely covers the range of large payloads, it makes these codes suitable for practical applications. In terms of payload and stego-image quality, the proposed scheme proves to be better than the simple LSB method and PVD. Experimental results show that both the well-known Chi-square and RS steganalysis attacks are unable to detect the existence of secret messages embedded with the system. In future matrix embedding based on various non-linear codes and Trellis code can be analysed.

Acknowledgements. The authors sincerely thank Jessica Fridrich and David Soukal for their valuable work on Matrix Embedding. We extend our sincere thanks to Department of Cyber Security, Amrita Vishwa Vidyapeetham, Coimbatore for all the help and support for the completion of this work.

References

1. Fridrich, J., Soukal, D.: Matrix Embedding for large payloads. *IEEE Transactions on Information, Security and Forensics* 1(3), 390–395 (2006)
2. Wu, D.C., Tsai, W.H.: A Steganographic Method for Images By Pixel Value Differencing. *Pattern Recognition Letters* 24, 1613–1626 (2003)
3. Williams, F.J.M., Sloane, N.J.: *The Theory of Error-correcting Codes*. North-Holland, Amsterdam (1997)
4. Galand, F., Kabatiansky, G.: Information hiding by coverings. In: *Proc. ITW 2003*, pp. 151–154 (2003)
5. Bierbrauer, J., Fridrich, J.: Constructing Good Covering Codes for Applications in Steganography. In: Shi, Y.Q. (ed.) *Transactions on Data Hiding and Multimedia Security III*. LNCS, vol. 4920, pp. 1–22. Springer, Heidelberg (2008)
6. Fridrich, J., Goljan, M.: Practical Steganalysis-State of the Art. In: *Proc. SPIE Photonics West, Electronic Imaging, Security and Watermarking of Multimedia Contents*, San Jose, California, vol. 4675, pp. 1–13 (January 2002)
7. Fridrich, J., Goljan, M., Du, R.: Reliable detection of LSB steganography in grayscale and color images. In: *Proc. ACM Workshop on Multimedia and Security*, Ottawa, Canada, pp. 27–30 (2001)
8. Westfeld, A., Pfitzmann, A.: Attacks, on Steganographic Systems: Breaking the Steganographic Utilities EzStego, Jsteg, Steganos, and S-Tools—and Some Lessons Learned. In: *3rd International Workshop on Information Hiding* (2000)

An Efficient Directional Weighted Median Switching Filter for Impulse Noise Removal in Medical Images

Madhu S. Nair and J. Reji

Department of Computer Science,
University of Kerala, Kariavattom, Thiruvananthapuram - 695 581, Kerala, India
madhu_s_nair2001@yahoo.com, rejij87@gmail.com

Abstract. This paper proposes a new efficient directional weighted median filter for impulse noise removal in medical images. The proposed method consists of two phases: noise detection and noise filtering. In this method, detection is done by Directional Weighted Median (DWM) detection [1], and filtering is applied to only corrupted pixels in the noisy image. The noise detection stage is based on the differences between the current pixel and its neighbours aligned in the main four directions in the considered window. The weighted median filter is then applied on directional pixel values as well as on uncorrupted pixels in the selected window by giving appropriate weights for the better restoration of corrupted medical images. Extensive experimental analysis shows that the proposed technique can be used for medical images with different impulse type noises. Both quantitative and qualitative analysis shows the superiority of the proposed method over other filters.

Keywords: Directional Weighted Median (DWM), Efficient directional weighted median filter, Medical image denoising, Noise Models, Impulse noise, Salt and pepper noise.

1 Introduction

Medical images are often corrupted by noise due to various sources of interference and other phenomena during their acquisition and transmission that affects the measurement processes in imaging [2]. Preservation of useful diagnostics information and the suppression of noise is a challenging task. Impulse noise is a specific type of noise having random occurrences of energy spikes having random amplitude and spectral content, which results in modification or loss of the image details.

There are many methods for removal of impulse noises from the images. Usage of linear filters such as averaging filters results in blurring of the images. Non linear filters such as median filter are good for removing impulse noise because of its noise suppression capability and computational efficiency [3][11]. Since the Median filtering approach applies the median value to all the pixels in the image, the image details from the uncorrupted pixels are also filtered which results in significant loss of image details. Also, the median filter is more effective in situations where the impulse noise density is low [4]. A solution to this problem is the use of switching median filter, which applies filtering to only the corrupted pixels in the image detected in the

noise detection stage [5-7]. But these filters replace the noisy pixel with a median value in the neighbourhood without taking into account the local features such as edges. The main challenge in the restoration of medical images is the removal of noise and preservation of the edges and texture information, as detail preservation is highly important in medical images. So these methods cannot be used for impulse noise removal in medical images.

The proposed switching median filter uses the impulse detector [1] for determining the corrupted pixels in the noisy image, by making use of the differences between current pixel and its four neighbours. After detection of uncorrupted pixels, we use the information obtained from the four directions to weigh the pixels in the window in order to preserve the details while removing the noise. The robustness of this method is increased by giving more weight to the non-noisy pixel in the considered window. This approach has not been considered in other directional weighted median methods.

This paper is organized as follows. In section 2 there is a brief introduction about the impulse noise models used in the paper. The section 3 describes the proposed method in detail. Section 4 describes various performance measures used. The section 5 contains the results and discussion and section 6 contains conclusion.

2 Noise Models

In this section four different random impulse noise models are discussed. These models [8] are considered for extensively examining the performance of the proposed filter considering practical situations.

Noise Model 1: Noise is modeled as salt-and-pepper impulse noise, pixels are randomly corrupted by two fixed extreme values, 0 and 255 (for gray level image), generated with the same probability. That is, if P is noise density, then P_1 , the noise density of salt is $P/2$ and P_2 , the noise density of pepper is also $P/2$.

Noise Model 2: This is similar to Noise Model 1, but here each pixel might be corrupted by either pepper noise or salt noise with unequal probabilities. That is $P_1 \neq P_2$.

Noise Model 3: Instead of two fixed values, impulse noise could be more realistically modeled by two fixed ranges that appear at both ends with a length of m each, respectively. That is, $[0, m]$ denotes salt and $[255-m, 255]$ denotes pepper. Here for noise density P , $P_1 = P_2 = P/2$.

Noise Model 4: This is similar to Noise Model 3 but here probability densities of low intensity impulse noise and high density impulse noise are different. That is, $P_1 \neq P_2$

3 An Efficient Directional Weighted Median Switching Filter for Impulse Noise Removal in Medical Images

The proposed method EDWMS filter is a double stage filter, which consist of robust impulse noise detection stage and an improved Directional Based Weighted Median (DBWM) [9] filtering process to remove the impulse noise. For detection purpose we are using the Directional based Impulse detector proposed by Yiqiu dong et.al [1].

3.1 Noise Detection Stage

Basically, a noise-free-image consists of locally smooth varying areas separated by edges. The basic strategy of DWM filter [1] is to consider the edges aligned with the main four directions as shown in fig 1.

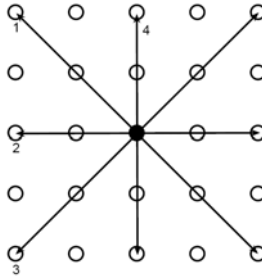


Fig. 1. Directions in the selected window

Let I_n ($n = 1$ to 4) denotes a set of coordinates aligned with the n^{th} direction centered at $(0, 0)$, i.e.,

$$\begin{aligned}
 I_1 &= \{(-2,-2),(-1,-1),(0,0),(1,1),(2,2)\} \\
 I_2 &= \{(0,-2),(0,-1),(0,0),(0,1),(0,2)\} \\
 I_3 &= \{(2,-2),(1,-1),(0,0),(-1,1),(-2,2)\} \\
 I_4 &= \{(-2,0),(-1,0),(0,0),(1,0),(2,0)\}
 \end{aligned}$$

Then, let $I_n^0 = I_n \setminus (0, 0)$ for all n from 1 to 4. In a 5×5 window centered at (i,j) , for each direction, the direction index, $c_{i,j}^{(n)}$ is the sum of absolute differences of gray-level values between the centre pixel and the pixels in the main four direction with $(k,l) \in I_n^0$. If the neighboring pixels are non-edges or non-noisy pixels then their gray scale value will be closer. So the absolute differences between those pixels are given a larger weight, before calculating the sum. If weight is very large, then the value of $c_{i,j}^{(n)}$ is decided by the differences corresponding to weight. Thus,

$$c_{i,j}^{(n)} = \sum_{(k,l) \in I_n^0} w_{k,l} \left| X_{i+k,j+l} - X_{i,j} \right|, \quad 1 \leq n \leq 4 \tag{1}$$

where

$$w_{k,l} = \begin{cases} 2, & (k, l) \in \Omega^3 \\ 1, & \text{otherwise} \end{cases} \tag{2}$$

$$\Omega^3 = \{(k,l) : -1 \leq k, l \leq 1\}$$

The smallest direction index obtained from the four directions is used for the detection of noise in the image:

$$d_{i,j} = \min\{c_{i,j}^{(n)} : 1 \leq n \leq 4\}$$

In the considered 5x5 window, $d_{i,j}$ can take any of the three values:

- 1) When the center pixel is a non-noisy, non- edge pixel, $d_{i,j}$ is small, since the four direction index are small.
- 2) $d_{i,j}$ is small also when the centre pixel is an edge pixel, since at least one of direction index is small.
- 3) $d_{i,j}$ is a large value, if the centre pixel is an impulse pixel, because of the four large direction index.

By applying a suitable threshold T , the pixel can be classified as noisy or noise-free pixel. This decision map is used in the noise filtering stage. Therefore, a pixel:

$$X_{i,j} = \begin{cases} \text{noisy pixel,} & \text{if } d_{i,j} > T \\ \text{noise - free pixel ,} & \text{if } d_{i,j} < T \end{cases} \tag{3}$$

3.2 Noise Filtering Stage

The proposed EDWMS filter is an improvement over DWM filter [1]. The image restored by DWM filter and Fuzzy Logic & Direction based Weighted Median (FL & DWM) filter[9] suffers from smoothening of edges and loss of finer details in the image. But the proposed method uses a weighted median filter which produces restored image with more prominent edges and by retaining finer details in the image.

In the noise filtering stage only the noisy pixels are replaced with the restored value based on the decision map obtained from the detection stage. The proposed method performs filtering in two stages. After impulse detection, most median-based filters replace the noisy pixels by median values in the window, which results in significant loss of information in the image. But in our method, the output of the median filter is improved by utilizing directional information.

Both the filtering stages uses the window of size 5x5 around the center pixel $X(i,j)$. The two stages are illustrated below.

First stage:

Here the Directional Based Weighted Median (DBWM) is calculated. To determine DBWM, we calculate the standard deviation $\sigma_{i,j}^n$ of gray-level values for all $X_{i+k,j+l}$

with $(k,l) \in I_n^0$ ($n=1...4$). The standard deviation describes how tightly all the values are clustered around the mean in the set of pixels. The four pixels aligned with a particular direction will be close to the each other, thereby having a small standard deviation. Therefore to preserve the edge the center value should be close to them. Mathematically it can be represented as:

$$S_{i,j} = \min_n \{ \sigma_{i,j}^{(n)} : n = 1...4 \}$$

Then, the pixels aligned with the direction having minimum standard deviation are repeated in the considered window and the median is calculated. This median is called Directional Based Weighted Median (DBWM) denoted as M_{dir} [9].

Second stage:

The standard deviation is calculated as in the first stage. $S_{i,j}$ shows that the four pixels aligned with direction having minimum standard deviation are closer to each other. Therefore, the centre value should also be close. Thus, we assign a weight w to these pixels, and calculate the median of noisy pixels as:

$med_{i,j} = \text{median}\{w_{k,l} \diamond X_{i+k,j+l} : (k,l) \in \Omega^5\}$, where \diamond denotes the repetition operator.

$$\text{where } w_{k,l} = \begin{cases} w, & (k,l) \in \Omega_{S_{i,j}}^0 \\ 4, & \text{non-noisy pixels in the } 5 \times 5 \text{ window} \\ 1, & \text{otherwise} \end{cases} \tag{4}$$

and

$$w = \begin{cases} 2, & \Omega^3 = (k,l) \in \Omega^3 \\ 1, & \Omega^5 = (k,l) \in \Omega^5 \wedge (k,l) \notin \Omega^3 \\ 1, & \text{for center pixel} \end{cases}$$

$$\Omega^5 = \{(k,l) : -2 \leq k,l \leq 2\}$$

The median values obtained from these two stages are then averaged as follows:

$$m_{i,j} = (med_{i,j} + M_{dir}) / 2.$$

The output of the EDWMS filter is:

$$Z_{i,j} = \Psi_{i,j} X_{i,j} + (1 - \Psi_{i,j}) m_{i,j}$$

where (5)

$$\Psi_{i,j} = \begin{cases} 0, & d_{i,j} > T \\ 1, & d_{i,j} \leq T \end{cases}$$

4 Performance Measures

The performance of the restoration process is quantified using peak signal-to-noise ratio (PSNR), structured similarity index (SSIM) and image enhancement factor (IEF)[8], is defined as follows.

$$PSNR = 10 * \log_{10}(255^2 / MSE)$$

$$MSE = \sum_{m,n} [O(m,n) - R(m,n)]^2 / (M * N)$$

$$SSIM = L(O, R) * C(O, R) * S(O, R)$$

$$L(O, R) = (2\mu_O \mu_R + C_1) / (\mu_O^2 + \mu_R^2 + C_2)$$

$$C(O, R) = (2\sigma_O \sigma_R + C_1) / (\sigma_O^2 + \sigma_R^2 + C_2)$$

$$S(O, R) = (\sigma_{OR} + C_3) / (\sigma_O \sigma_R + C_3)$$

$$C_1 = (K_1 * G)^2, C_2 = (K_2 * G)^2, C_3 = C_2 / 2$$

$$G = 255; K_1, K_2 \gg 1, (K_1 = 0.001, K_2 = 0.002)$$

$$IEF = \left(\sum_{m,n} [P(m,n) - O(m,n)]^2 \right) / \left(\sum_{m,n} [R(m,n) - O(m,n)]^2 \right)$$

where O is the original Image, R is the restored image, P is the corrupted image, MSE is the mean square error, $M \times N$ is the size of the image, L is the luminance comparison, C is the contrast comparison, S is the structure comparison, μ is the mean and σ is the standard deviation.

In this paper, we also used a qualitative-based performance measure named image quality index (IQI) to prove the efficiency of our proposed EDWMS filter algorithm. This universal objective image quality index was proposed by Wang and Bovik [10], which is easy to calculate and applicable to various image processing applications. IQI is designed by modelling any image distortion as a combination of three factors: loss of correlation, luminance distortion and contrast distortion.

$$IQI_j = Corr(O_w, R_w) * Lum(O_w, R_w) * Cont(O_w, R_w) .$$

IQI is first applied to local regions using a sliding window approach with size 8X8. At the j^{th} step, the local quality index IQI_j is computed within the sliding window using the formula given above. O_w and R_w represent the sliding window of original and restored images, respectively. If there are a total of M steps, then the overall image quality index is given by $IQI \approx \frac{1}{M} \sum_j IQI_j$, where j varies from 1 to M . The dynamic range of IQI is $[-1, 1]$, and the best value 1 is achieved if and only if restored image R is equal to the original image O .

5 Results and Discussion

The performance of EDWMS filter has been evaluated qualitatively and quantitatively. Although extensive simulations were carried out using various medical images, only performance evaluation using images such as MRI brain1 image of size 256x256, MRI brain2 image of size 256x256, X-ray mammogram image of size 250x350 and MRI knee image of size 200x200 are illustrated in this section.

For improving the efficiency of the filter, we applied our method iteratively with decreasing threshold. At early iterations with larger threshold values, the detection stage identifies only those pixels that are more likely to be noisy. To consider more noise, in the subsequent iterations, the threshold is decreased. Based on the experiments on a variety of gray scale medical images, the following selection of threshold yields a good result;

$$T_0 = 547 \text{ and } T_{k+1} = T_k \cdot 0.8 \text{ (} k \geq 0 \text{)}$$

Where T_0 is the initial threshold and T_k is the threshold at k^{th} step. The maximum number of iterations is between 6 and 13. Fig .2 shows the original test images along with the edge map obtained using the Sobel operator.

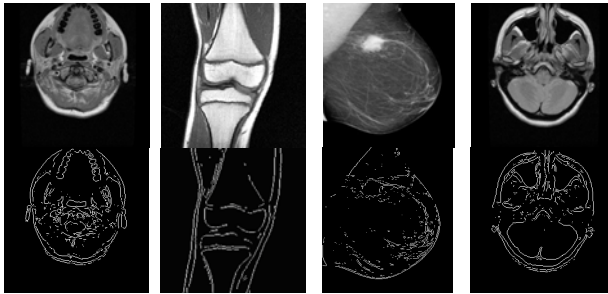


Fig. 2. The First row shows medical images MRI brain1 image, MRI knee image, X-ray mammogram image and MRI brain2 image respectively used for performance measurement of our proposed EDWMS filter. The second row shows their edge maps obtained by Sobel operator.

All the noise models described in the section 2 are used for measuring the performance of the proposed algorithm. For Noise Model 1 and Noise Model 3 images corrupted with salt and pepper noise with different noise densities ranging from 10% to 90% is used. For Noise Model 2 and Noise Model 4 we consider different proportion of noise densities for salt and pepper. For Noise Model 3 and Noise Model 4 the intensity values for the uniformly distributed noise is in the range 0-9 and 246-255 is used.

The MRI brain1 image corrupted by Noise Model 1 and images restored using DWM filter [1], FL & DWM filter [9] and our proposed EDWMS filter for various noise densities are shown below in Fig.3. The Fig.4 contains the edge maps of restored images. Fig.5 and Table 1 shows various performance measures graphically and quantitatively. The corrupted X-ray mammogram image with the restored images and edge maps of FL & DWM filter, DWM filter and our proposed EDWMS filter for Noise Model 2 is shown in Fig.6. Table 2 shows quantitative measures of performance metrics. The noisy image of MRI Knee image generated using Noise Model 3 along with the restored images using FL & DWM filter, DWM filter and our proposed EDWMS filter for various noise densities are shown in Fig.7. Fig.8 contains the edge maps of these restored images. The Fig.9 graphically demonstrates performance of EDWMS over FL & DWM and DWM in terms of quantitative measures such as PSNR, IEF, SSIM and IQI. The values of these measurements for

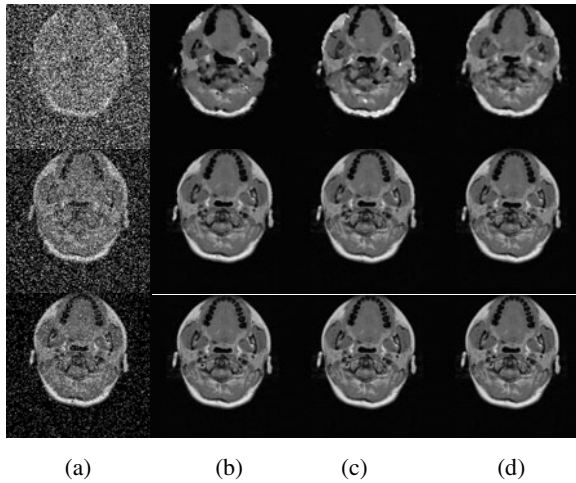


Fig. 3. Column (a) shows MRI brain1 image corrupted by Noise Model 1 with noise densities 60%, 30% and 10% respectively, the restored images by (b) FL & DWM filter, (c) DWM filter and (d) the proposed EDWMS filter

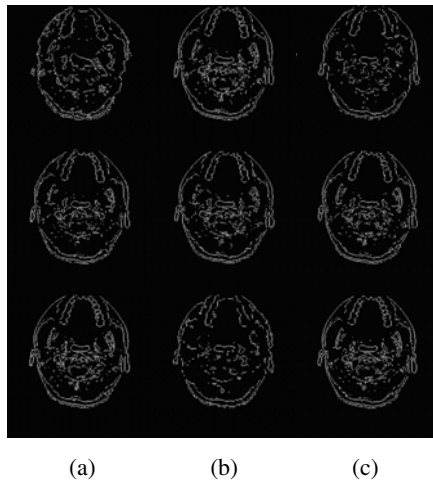


Fig. 4. (a),(b),(c) shows edge map of restored images by FL & DWM, DWM and EDWMS respectively

various noise densities are given in the Table 3. The corrupted MRI brain2 image generated using Noise Model 4 and the restored images and edge maps of FL & DWM filter, DWM filter and our proposed EDWMS filter for various noise densities are shown below in Fig.10. The comparison of various performance measures for restored images using FL & DWM filter, DWM filter and proposed EDWMS filter is shown in Table 4. The EDWMS algorithm is implemented in Pentium IV 3GHz

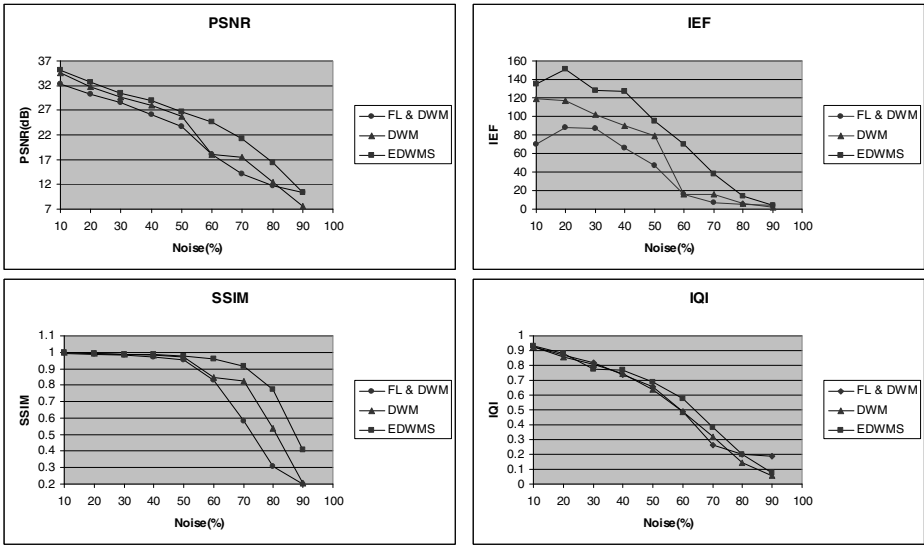


Fig. 5. Graphical representation of various performance measures for MRI brain1 image corrupted by Noise Model 1

Table 1. The performance measures obtained for MRI brain1 image corrupted by Noise Model 1

Noise (%)	PSNR			SSIM			IEF			IQI		
	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS
10	32.326	34.471	35.211	0.993	0.996	0.996	69.761	118.596	135.092	0.926	0.925	0.932
20	30.290	31.671	32.640	0.989	0.990	0.994	88.211	116.741	150.715	0.871	0.853	0.882
30	28.485	29.753	30.419	0.984	0.988	0.990	86.646	102.041	127.600	0.815	0.805	0.773
40	26.062	27.956	28.973	0.972	0.982	0.986	66.077	90.151	127.399	0.738	0.741	0.766
50	23.599	25.835	26.672	0.951	0.971	0.976	47.046	78.640	94.714	0.654	0.636	0.684
60	18.114	18.114	24.549	0.827	0.847	0.960	15.712	15.712	70.098	0.485	0.485	0.574
70	14.064	17.501	21.306	0.583	0.823	0.914	7.234	16.081	38.495	0.259	0.319	0.379
80	11.670	12.489	16.306	0.307	0.537	0.771	4.815	5.814	13.918	0.202	0.143	0.202
90	10.467	7.5534	10.434	0.199	0.204	0.406	4.095	2.098	4.0658	0.184	0.055	0.077

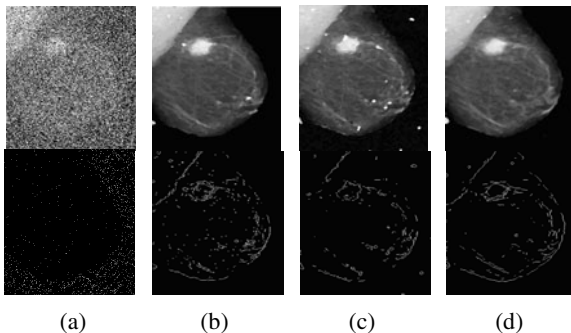


Fig. 6. (a) is the X-ray mammogram image corrupted by the Noise Model 2 (25% salt and 35% pepper), the restored images and its edge maps of (b) FL & DWM (c) DWM (d) the proposed EDWMS

Table 2. Performance comparison of FL & DWMF, DWMF with Proposed filter for the X-ray mammogram image of Noise Model 2

% Noise (Salt, Pepper)	PSNR			IEF			SSIM			IQI		
	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS
(0.4,0.6)	35.194	36.651	37.663	140.402	195.244	245.164	0.996	0.997	0.998	0.959	0.936	0.966
(10,20)	31.450	32.616	33.863	192.239	315.150	331.845	0.992	0.995	0.995	0.884	0.874	0.891
(20,30)	27.215	28.799	29.166	98.013	152.870	155.013	0.981	0.986	0.987	0.680	0.732	0.764
(25,35)	24.944	24.264	27.833	76.267	65.587	148.724	0.968	0.964	0.983	0.622	0.519	0.705

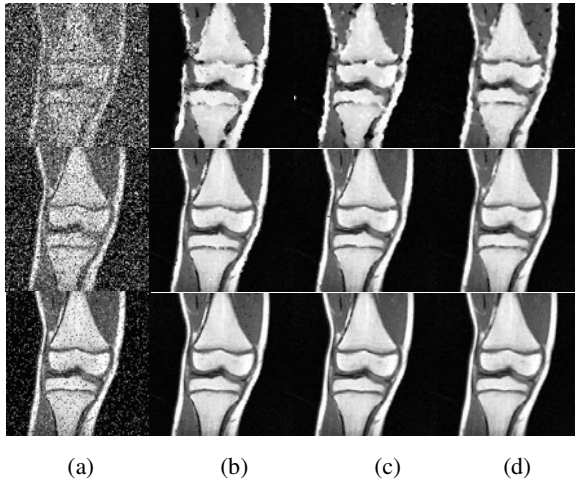


Fig. 7. Column (a) represents MRI knee image corrupted by 60%, 30% and 10% impulse noise of Noise Model 3, the restored images by (b) FL & DWM,(c) DWM and (d) the proposed EDWMS

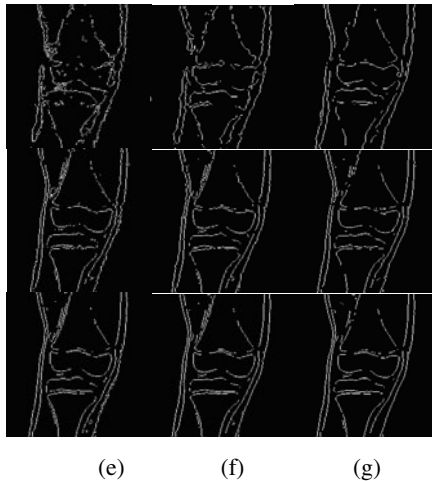


Fig. 8. (e),(f),(g) shows edge map of restored images by FL & DWM,DWM and EDWMS respectively

processor with 512 MB RAM using MATLAB 7.0 package. Table 5 shows the execution time taken for the three methods. It can be seen that in all cases there is a significant improvement in performance for all the noise models considered. From all the above experimentation results, it has been found that EDWMS produces better results than FL & DWM & DWM in terms of both quantitative measures such as PSNR, SSIM, IEF and qualitative measures such as image quality index (IQI).

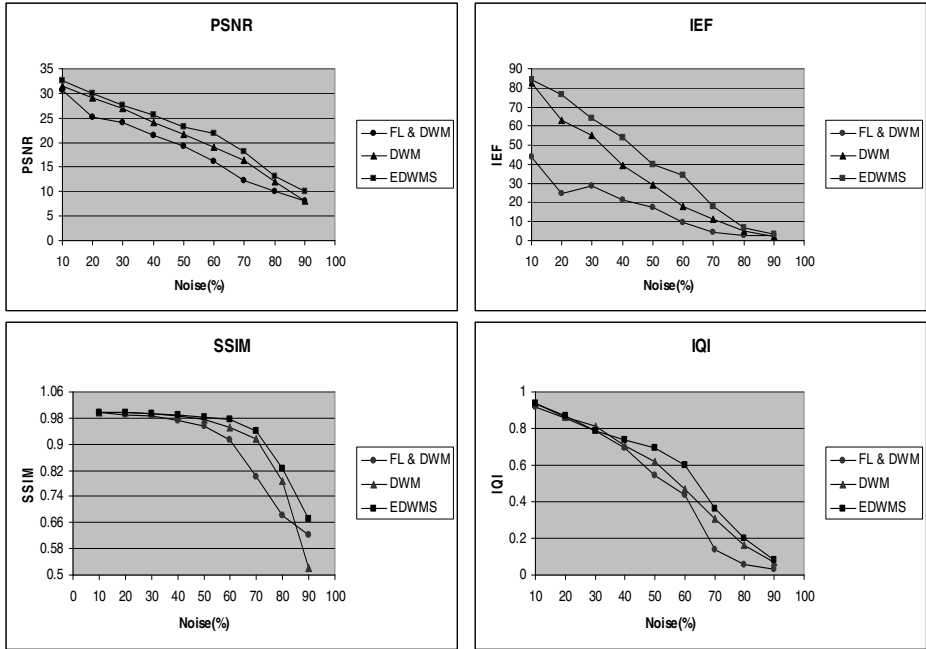


Fig. 9. Graphical representation of performance measures of MRI knee image corrupted by Noise Model 3

Table 3. Performance measures of MRI Knee image corrupted by Noise Model 3

Noise(%)	PSNR			SSIM			IEF			IQI		
	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS
10	30.537	31.443	32.563	0.996	0.998	0.997	43.748	82.621	84.570	0.916	0.938	0.936
20	25.212	29.127	29.995	0.988	0.995	0.996	24.824	62.813	76.541	0.853	0.863	0.867
30	24.095	26.958	27.474	0.985	0.992	0.993	28.793	55.334	64.0527	0.784	0.813	0.788
40	21.378	24.023	25.595	0.972	0.985	0.989	21.146	39.451	53.959	0.692	0.708	0.734
50	19.273	21.749	23.174	0.956	0.974	0.981	17.324	29.391	40.213	0.545	0.618	0.695
60	16.146	18.953	21.863	0.914	0.952	0.975	9.408	17.895	34.070	0.435	0.470	0.599
70	12.316	16.340	18.168	0.801	0.915	0.942	4.674	11.092	17.772	0.134	0.309	0.364
80	10.031	11.955	13.204	0.681	0.786	0.825	3.077	4.796	6.581	0.056	0.163	0.202
90	8.0164	8.118	10.110	0.623	0.521	0.670	2.671	2.208	3.647	0.030	0.068	0.083

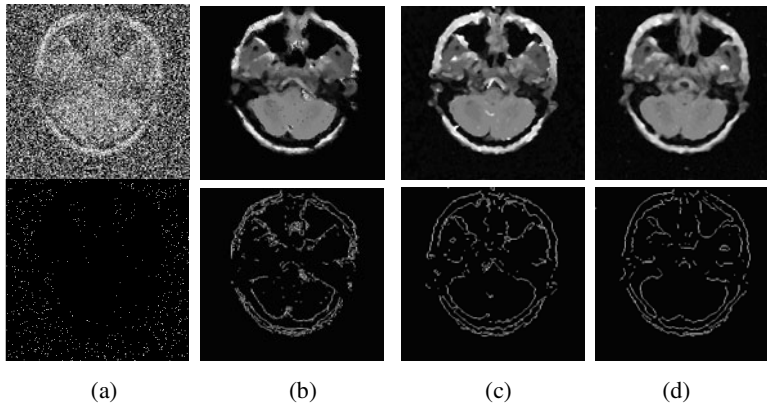


Fig. 10. (a) is the MRI brain2 image corrupted by the Noise Model4 (25% salt and 35% pepper), the restored images and edge maps of (b) FL & DWM (c) DWM(d) the proposed EDWMS filter.

Table 4. Performance comparison of FL & DWM, DWM with Proposed filter for MRI brain2 image for Noise Model 4

% Noise (Salt, Pepper)	PSNR			IEF			SSIM			IQI		
	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS	FL & DWM	DWM	EDWMS
(0.4,0.6)	32.412	33.409	34.520	82.944	98.973	136.390	0.994	0.995	0.996	0.764	0.760	0.771
(10,20)	27.254	29.118	29.937	77.853	94.457	117.331	0.982	0.985	0.988	0.661	0.648	0.645
(20,30)	21.484	23.874	25.144	29.537	55.325	67.265	0.933	0.962	0.970	0.475	0.487	0.511
(25,35)	17.947	20.403	22.750	16.901	28.678	50.880	0.852	0.918	0.948	0.502	0.332	0.375

Table 5. The computational time in seconds for Noise Model 4 corrupted with 20% salt and 30% Pepper for FL & DWM, DWM and EDWMS

Images	FL & DWM	DWM	EDWMS
Brain1	124.563	86.265	51.485
Brain2	130.937	92.547	54.782
Knee	72.813	59.765	34.797
Mammogram	174.953	117.921	63.593

6 Conclusion

In this paper, we proposed a new efficient and improved Directional Weighted Median filter for the restoration of medical images that are corrupted with high density of impulse noises based on different noise models. EDWMS is an improved directional weighted median filter for medical images in which the filtering is applied only to corrupted pixels in the image while the uncorrupted pixels are left unchanged. Also the details and the edges which are highly important in the medical images are preserved. The advantage of EDWM filter is its high edge preservation capability, in contrast to DWM and FL & DWM. Also, the proposed method is computationally efficient than DWM and FL & DWM.

References

1. Dong, Y., Xu, S.: A New Directional Weighted Median Filter for Removal of Random-Valued Impulse Noise. *IEEE Signal Process Lett.* 14(3), 193–196
2. Toprak, A., Güler, I.: Impulse noise reduction in medical images with the use of switch mode fuzzy adaptive median filter. *Digital Signal Processing* 17, 711–723 (2007)
3. Pratt, W.K.: *Digital Image Processing*. John Wiley & Sons, Chichester (1978)
4. Abreu, E., Lightstone, M., Mitra, S.K., Arakawa, K.: A new efficient approach for the removal of impulse noise from highly corrupted images. *IEEE Trans. Image Process.* 5, 1012–1025 (1992)
5. Eng, H.-L., Ma, K.-K.: Noise adaptive soft-switching median filter. *IEEE Trans. Image Process.* 10(2), 242–251 (2001)
6. Zhang, S., Karim, M.A.: A new impulse detector for switching median filters. *IEEE Signal Processing Letters* 9(4), 360–363 (2002)
7. Wang, Z., Zhang, D.: Progressive switching median filter for the removal of impulse noise from highly corrupted images. *IEEE Transactions on Circuits System II* 46(1), 78–80 (1999)
8. Nair, M.S., Raju, G.: A new fuzzy-based decision algorithm for high-density impulse noise removal. *Signal Image and Video Processing*, doi:10.1007/s11760-010-0186-4
9. Hussain, A., Arfan Jaffar, M., Mirza, A.M.: A hybrid image restoration approach: fuzzy logic and directional weighted median based uniform impulse noise removal. Springer-Verlag London Limited, Heidelberg (2009)
10. Wang, Z., Bovik, A.C.: A universal image quality index. *IEEE Signal Process. Lett.* 9(3), 81–84 (2002)
11. Nair, M.S., Revathy, K., Tatavarti, R.: An Improved Decision-Based Algorithm for Impulse Noise Removal. In: *International Congress on Image and Signal Processing - CISP 2008*, vol. 1, pp. 426–431. IEEE Computer Society Press, Los Alamitos (2008), doi:10.1109/CISP.2008.21

An Improved Handwritten Text Line Segmentation Technique

M. Mohammadi¹, S.S. Mozaffari Chaniyani¹,
V.N. Manjunath Aradhya^{2,*}, and G.H. Kumar¹

¹ Dept. of Studies in Computer Science, University of Mysore,
Mysore - 570 006, India

² Dept. of Information Science and Engg,
Dayananda Sagar College of Engineering, Bangalore - 560 078, India
`aradhya1980@yahoo.co.in`
`mohammadi@ctech.ir`
`mozafari@live.com`

Abstract. Document image segmentation to textlines is a decisive stage towards unconstrained handwritten document recognition. In this paper, an improvised scheme for handwritten text line segmentation is proposed. The proposed method is an improvised method to Alaei et al., 2010 [1] by applying block separation and edge detection process. The proposed method was tested on variety of handwritten documents pertaining to English, Persian, and Kannada and the results was remarkable.

Keywords: Document Image Analysis, Unconstrained Handwritten Documents, Skew Estimation, Morphology and Edge Detection.

1 Introduction

Text line extraction is a critical stage towards handwritten document recognition that refers to the segmentation of a document image into distinct entities, namely text lines. The overall performance of a handwritten character recognition system strongly relies on the results of the text line detection process. Text line extraction of handwritten document is much more difficult than that for printed documents. Unlike that printed documents have approximately straight and parallel text lines. Several problems inherent in handwritten documents such as the different in the skew angle between text lines or along the same text line, the existence of adjacent text lines, un-uniformly curved.

There exist several methods for text line segmentation which are roughly categorized as follows. *Smearing methods*: short white runs are filled with black pixels intending to form large bodies of black pixels, which will be considered as text line areas. Smearing methods can't deal well with touching and overlapping components. *Horizontal projections*: a vector containing the sums of each image line is created. The local minima of that vector are assumed to be the projection of white areas in between lines, and the image is segmented accordingly.

* Corresponding author.

Horizontal projections can't deal well with skewed, curved and fluctuating lines. Hough transform considers any image to compose of straight lines. It creates an angle, offset plane in which the local maxima are assumed to correlate with text lines. Hough transform has trouble detecting curved text lines. *Bottom-up approaches*: connected components or even pixels are connected to their close ones based on geometrical criteria to form text lines. Other methods have also been proposed such as: repulsive attractive networks, stochastic methods and text line structure enhancing [7].

Many works have been done on English Chinese, and Arabic. However some preliminary work has also done on Indian scripts. Louloudis et al, 2008 [5], presents text line detection of handwritten documents. The proposed method based on block based Hough transform. A text line segmentation by clustering with distance metric learning is proposed in [11]. The method is based on minimal spanning tree clustering with distance metric learning techniques. A steerable directional local profile technique for extraction of handwritten Arabic text lines is described in [12]. The approach is based generalized adaptive local connectivity map (ACLM) using a steerable directional filter. A steerable filter is used to determine foreground intensity along multiple directions at each pixel while generating the ALCM. Yin and Liu (2009) [10], proposed a variational bayes method for handwritten textline segmentation. The method is based on the Gaussian component and variational bayes technique.

A handwritten document image segmentation into textlines and words is presented in [8]. The presented approach is based on viterbi algorithm. The algorithm tested on the benchmarking datasets of ICDAR-07 handwriting segmentation contents. Du et al (2009) [9], proposed textline segmentation in handwritten document using Mumford shah model. The method used morphing to remove overlaps between textlines and connected broken ones. Liwicki et al (2008) [4], proposed combining diverse on-line and off-line systems for handwritten textline recognition. The approach is based on multiple classifier system(MCS), Hidden Markov Models (HMMs) and bidirectional long short-term memory networks (BLSTM). A segmentation of Bangla unconstrained handwritten text is described in [2]. To extract a text line, horizontal histogram of stripes and the relationship of the minimal values of histogram is used. A text line extraction from multi-skewed handwritten documents is presented in [3]. The method assumes that hypothetical water flows, from both left and right sides of the images, face obstruction from characters of text lines. The stripes of areas left unwetted on the image frame are labeled for extraction of text lines.

Nagabhushan et al (2010) [6], describes tracing and straightening the baseline in handwritten Persian/Arabic text line. The proposed technique identify the candidate points from black and white blocks all along the text line. Using candidate points the algorithm traces the baseline, which is subsequently stretched straight horizontally and subsequently. Recently an elegant method for unconstrained handwritten text line segmentation is proposed in [1]. The approach is

based on painting technique. The painting technique enhances the separability between the foreground and background portions enabling for detection of text lines.

From the above literature survey, it is clear that, many works has been done on Text line extraction for English, Arabic, Chinese, Persian and also some of Indian scripts. Even though existence of many methods, performance does not give compatible result to all languages. Especially south Indian language like Kannada scripts. Kannada scripts text line extraction is a challenging task due to additional modifier-characters, which get plugged in as bottom fixes or top fixes, or as extensions, that remain as disconnected protrusions of a main character. Under these circumstances, there is a need for developing a generalized approach which works for these kind of script.

2 Proposed Method

The proposed method improves the existing algorithm mainly by applying block separation, skew estimation process and edge detection. The proposed algorithm uses the first step of existing method i.e., painting algorithm. The following section briefly summarizes the improved version.

2.1 Painting Algorithm and Stripe Minimization

The idea of painting algorithm is extracted from [1]. Initially, by applying the painting algorithm to input image we will obtain stripe wise painted image (Please refer Figure 1(a) and Figure 1(b)). In painted image there may be dis connectivities and also some of the rectangular white area may have longer height as compared to other white areas. This indicate overlapping/touching portions of adjacent text lines. The longer height of rectangular white area will be greater or equal to twice the size of the average height of white area. Considering the average height of the white area as threshold value which generally denotes overlapping/touching portion. All such bigger white areas are deleted from painted image and the remaining white areas are smoothed using median filter technique which is as shown in the Figure 1(c).

2.2 Line Drawing by Estimated Slope and Adding Words to Blocks

In [1], after the successful completion of painting algorithm and strip connectivity, dilation operation is applied to bridge the connectivity's. This action is

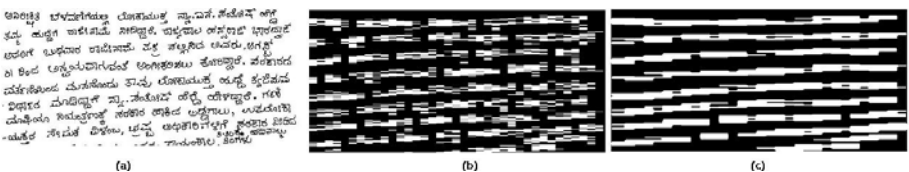


Fig. 1. (a): Kannada handwritten document (b): Result obtained after applying Paint- ing algorithm (c):Result after smoothing operation



Fig. 2. (a): Result after Dilation Operation (b): Result after connecting line to start and end position

alright when we do not have any skewness in images. If we encounter any skewness in images, it will be fail to separate lines from each other and leads to improper result. So in order to cope up this problem, we proposed a method based on slope. In order to find the skewness, we connect the strip squares together horizontally and divide image into two vertical parts. After this operation, we calculated the mean of slope between line blocks in each part of the image separately. The obtained skew is then used for proper dilation operation. Finally we will eliminate Line Blocks whose length is less than $1/5$ of image width and which have critical situation (Please refer Figure 2(a)). We draw the line from both sides (left and right) of each Line Block by the $2 \times \text{slope}$ that we have estimated in each part of image until we touch a next pixel of other Line Blocks or we reach to image borders (Please refer Figure 2(b)).

The next step is put all characters to Line Blocks. For doing so the simplest way is:

- Overlap the two layer images (Image Block and Original document image) together.
- Remove the regions from original image that are under Line Blocks.

The rest area characters in original image have this 3 status :

- A: The labels that are touching just one Line Block in Image Block.
- B: The labels that are between two Line Blocks and they don't touch any Line Block.(like points and signs between lines)
- C: The labels that are touching two Line Blocks.

The next step is to insert some of the labels of original image to Line Block, such that it should not touch two Line Blocks. As we have said there are three cases of this to sort labels: For the case (a) of Figure 3(a) there is no problem of connectivity and we can put them to Image Blocks simply without any fear of connectivity. Case (b) involve points and signs and some other stuff that are out of lines. There may be some problem if it merges to a wrong Line Block. So one good idea is to connect them to the nearest Line Block vertically by a simple line such that its distance to nearest Line Block is more than threshold \mathbf{T} we will ignore that label. Here $\mathbf{T} = 2 \times \text{mean Height of strips squares}$.

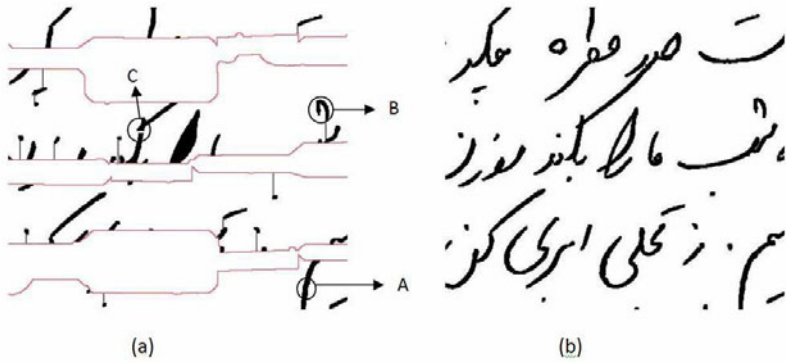


Fig. 3. (a): Image Block Overlapped with Original Image (b): Original Image

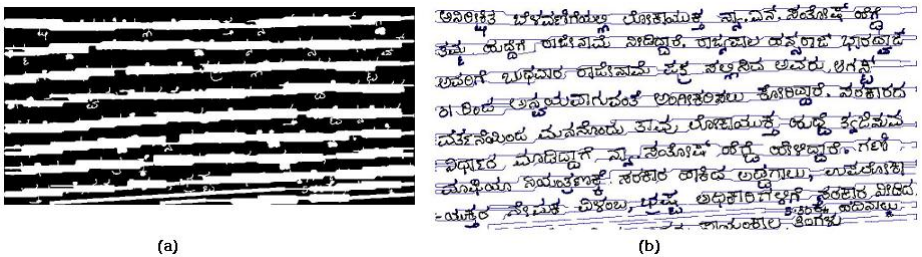


Fig. 4. (a): After inserting components to original image (b): Final Segmentation of Individual Lines

Case (c) this sort of labels are always making problem in hand written line segmentation process, in our paper we have proposed to cut them exactly from center, Figure 3(a) part C.

2.3 Edge Detection Process

After inserting all the components to the Image Block, we may encounter some holes in our result, because of using edge detection for line segmentation these holes will make problem for us. So, in order to get rid of these holes, we used image filling operation. Finally edge detection operator is used for line segmentation. Refer Figure 4(a) and Figure 4(b) for components added and final segmentation result.

3 Experiment Analysis and Comparative Study

This section presents the results of the experiments conducted to study the performance the proposed method. The method has been implemented in MATLAB 8.0 on Pentium IV 2.0 GHz with 1GB RAM. For the experiment purpose, we

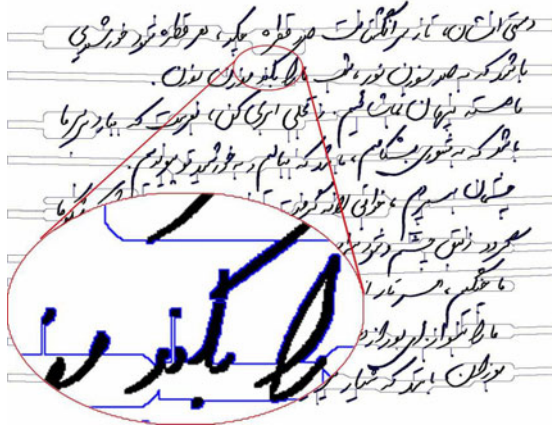


Fig. 5. Final Segmentation of Proposed Method for Persian Script

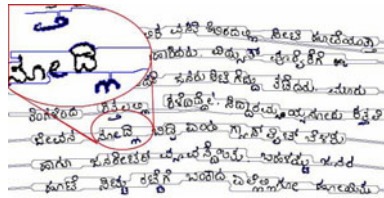


Fig. 6. Final Segmentation of Proposed Method for Kannada Script

have collected around 300 different handwritten documents pertaining to Kannada, Persian and English. It is worth to note that the data sets contain varieties of writing styles. None of the documents include any non-text elements and almost all documents have two or more adjacent text lines touching in several areas. Few documents have variable skew angles among text lines and having text lines with different skew directions as well as document images having text lines with converse skew angles along the same text line. For the experiment we considered single column document pages.

We also compared our method with standard existing method [1]. The detection rate of the proposed method and method by [1] was around 99.1% and 95.6% respectively. Figure 5 and 6 shows example of successful result of text line extraction obtained using proposed model. Final segmetnation of method proposed in [1] is also showed in Figure 7 and Figure 8. From the results it is very clear that the proposed method segments text line accurately compared to method proposed in [1]. Figure 5 and 6 shows how the proposed method accurately segmented the modifiers (pelase see the enlarged portion). The propose model was tested on variety of unconstrained Kannada and Persian documents and paid encouraging results. The proposed method aimed at solving few

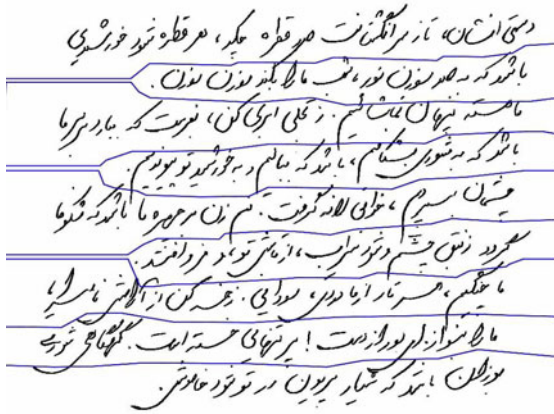


Fig. 7. Final Segmentation of Method II for Persian Script

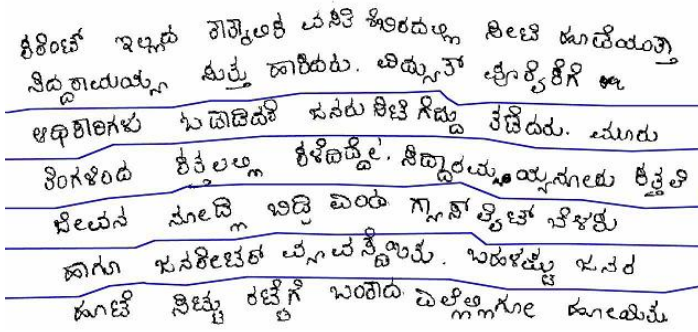


Fig. 8. Final Segmentation of Method II for Kannada Script

challenges such as (i) variability in skew between different texts lines (ii) fluctuating lines (iii) variable intra-word gaps and (iv) influence of author styles (v) touching line segmentation.

4 Conclusion

In this paper, we proposed a, improved text line extraction scheme for unconstrained handwritten documents. The proposed model is an improvised scheme to II. The major highlights of the proposed model are: (i) Application of skew angle estimation for accurate dilation process (ii) Extracting modified components (iii) Edge detection approach for final segmentation stage. The proposed method aimed at solving real challenging issues such as, skew between different texts lines, irregular lines, variable intra-word gaps and different writing

styles. The experiment results was compared with well known existing method and showed encouraging results under different real time documents. There is a need for robust text line extraction algorithms, which needs to concentrate documents corrupted by noise. In future, authors aim at working towards better robust algorithms and to work on different Persian and Indian scripts.

References

1. Alaei, A., Pal, U., Nagabhushan, P.: A new scheme for unconstrained handwritten text-line segmentation. *Pattern Recognition* 44(4), 917–928 (2011)
2. Pal, U., Datta, S.: Segmentation of bangla unconstrained handwritten text. In: *The Proceedings of 7th Intl Conference on Document Analysis and Recognition* (2003)
3. Basu, S., Chaudhuri, C., Kundu, M., Nasipuri, M., Basu, D.k.: Text line extraction from multi-skewde handwritten documents. *Pattern Recognition*, 1–15 (2006)
4. Liwicki, M., Bunke, H.: Combining diverse on-line and off-line systems for handwritten text line recognition. *Pattern Recognition* (2008)
5. Louloudis, G., Gatos, B., Pratikakis, I., Halatsis, C.: Text line detection in handwritten documents. *Pattern Recognition* 41, 3758–3772 (2008)
6. Nagabhushan, P., Alaei, A.: Tracing and straightening the baseline in handwritten persian/arabic text-line: A new approach based on painting -technique. *The Proceeding of Intl Journal on Computer Science and Engineering*, 907–916 (2010)
7. Nicolaou, A., Gatos, B.: Handwritten text line segmentation by shredding text into its lines. In: *In the Proceedings of 10th International Conference on Document Analysis and Recognition*, pp. 626–630 (2009)
8. Papavassiliou, V., Stafylakis, T., Katsouros, V., George: Handwritten document image segmentation into text lines and words. *Pattern Recognition* 43, 369–377 (2010)
9. Du, X., Pan, W., Bui, T.D.: Text line segmentation in handwritten documents using mumford shah model. *Pattern Recognition* (2009)
10. Yin, F., Liu, C.L.: A variational bayes method for handwritten text line segmentation. In: *The Proceeding of 10th Intl Conference on Document Analysis and Recognition*, pp. 436–440 (2009)
11. Yin, F., Liu, C.L.: Handwritten chinese text line segmentation by clustering with distance metric learning. *Pattern Recognition* 42, 3146–3157 (2009)
12. Shi, Z., Setlur, S., Govindaraju, V.: A steerable directional local profile technique for extraction of handwritten arabic text lines. In: *Proceedings of 10th Intl Conference on Document Analysis and Recognition*, pp. 176–180 (2009)

Skew Estimation for Unconstrained Handwritten Documents

V.N. Manjunath Aradhya^{1,*}, C. Naveena¹, and S.K. Niranjana²

¹ Dept. of Information Science and Engg,

Dayananda Sagar College of Engineering, Bangalore - 78

² Dept. of ISE, Maharaja Institute of Technology - Mysore
aradhya1980@yahoo.co.in, naveena.cse@gmail.com

Abstract. Document skew estimation is one of the most important and challenging phase in OCR system. Skew estimation of handwritten documents is still remains challenging in the field of document image analysis due to a non-uniform text line. Hence, in this paper, we present a novel scheme for handwritten documents. The proposed method is based on mixture models. The expectation-maximization (EM) algorithm is used to learn the mixture of Gaussians. Subsequently the cluster means obtained from the individual words is used estimate the skew angle. Experiments on different handwritten documents and documents corrupted by noise shows the effectiveness of the proposed method.

Keywords: Document Image Analysis, Handwritten Documents, Skew Estimation, Mixture Models, EM.

1 Introduction

The skew estimation of document images is particularly crucial among the document processing operations as it affects the subsequent understanding of the document. Estimating skew for handwritten documents is not as straight forward as computing the skew for printed documents. This is due to varying style, non-uniform text line skew and complexity involved in the script. Enormous amount of papers have been proposed in the literature for the estimation and correction of document skew. There are four broad categories of skew estimation approaches: (i) Projection Profile (PP) based, (ii) Hough Transform (HT) based, (iii) Nearest Neighbor Clustering (NNC) based and (iv) Cross Correlation. Many hybrid approaches also have been proposed by combining two or more kinds of these approaches.

Techniques using the well-known Hough transform have been explored by several authors and are based on the observations that a distinguishing feature for text is the alignment of characters and that text lines of a document are usually parallel each other [8]. The horizontal (vertical) projection profile is a histogram of the number of black pixels along horizontal (vertical) scanned lines [6]. The

* Corresponding author.

methods based on nearest neighbor clustering (NNC) aim at exploiting the general assumptions that characters in a line are aligned and close to each other. They are characterized by a bottom up process which starts from a set of objects, connected components or points representative of them, and utilizes their mutual distances and spatial relationships to estimate the document skew [5]. The cross-correlation method proposed in [11] is based on the correlation between two vertical lines in a document image.

From the literature it is clear that, lot of work has been carried out on estimating skew for printed documents. But efforts on estimating skew for handwritten documents is very very less. A skew estimation algorithm for printed and handwritten documents, based on the documents horizontal projection profile and its WignerVille distribution, is presented in [4]. Skew Detection for Chinese Handwriting by Horizontal Stroke Histogram is described in [9]. An approach to skew detection, correction as well as character segmentation for handwritten Bangla words is described in [7]. Skew angle detection of a cursive handwritten Devanagari script is proposed in [3]. The method is based on Shiro-Rekha as inherent feature of Devanagari. Skew angle correction and line extraction from handwritten bengali text is proposed in [2]. Recently an approach based on mixture models is proposed in [10]. This approach is used to estimate skew for printed documents of south Indian scripts. In this paper, we explore the use of mixture model to unconstrained handwritten word documents.

It is very clear from the above observations that, very few attempts have been considered for handwritten skew. Attention to Indian scenario, particularly on south Indian script of Kannada has not shown so far. The proposed method is of first kind to estimate skew for unconstrained handwritten Kannada script. The proposed method is based on mixture models in which Expectation-Maximization (EM) algorithm is used to learn the mixture of gaussians. Subsequently the cluster means are used to estimate the skew angle. The organization of the paper is as follows: section 2 describes about proposed mixture model. Experiment results and comparative study is described in section 3. Finally discussion and conclusion is drawn at the end.

2 Proposed Method

Mixture models use a set of data points as an input for clustering input data. To find a clusters in a set of data points is a considerable problem. To obtain marginalization from joint distribution over observed and latent variables is relatively complex. To solve this problem use of latent variable in a mixture distributions in which the discrete latent variables can be interpreted as defining assignment of datapoints to specific components of the mixture. EM algorithm is a one of the technique for finding maximum likelihood estimation in latent variable.

Considering the problem of identifying groups, or clusters of the datapoints in a multidimensional space. Let x_1, \dots, x_n is a data set consisting of n observation of a random d -dimensional variable. Number of clusters requirement is obtained

by k value. We can represent dataset as an $n * d$ matrix x in which the n^{th} row is given by x_n^T . z is a latent variable matrix of size $n * k$ with rows z_n^T .

The Gaussian mixture of conditional probability of z is given by

$$\gamma(z_k) \equiv p(z_k = 1|x) = \frac{\pi_k \eta(x|\mu_k, \sum_k)}{\sum_{j=1}^k \pi_j \eta(x|\mu_j, \sum_j)} \tag{1}$$

where π_k and $\gamma(z_k)$ is the prior probability of $z_k = 1$ and the quantity of corresponding posterior probability of observed x respectively, μ_k is a mean point associated with k^{th} clusters.

The log likelihood function of i.i.d(independent and identically distributed) for data points is given by

$$\ln p(x|\pi, \mu, \sum) = \sum_{n=1}^N \ln \sum_{k=1}^K \eta(x_n|\mu_k, \sum_k). \tag{2}$$

To obtain Maximum-likelihood function setting derivatives of $\ln p(x|\pi, \mu, \sum)$ w.r.t means μ_k of the Gaussian components to zero

$$0 = - \sum_{n=1}^N \frac{\pi_k \eta(x|\mu_k, \sum_k)}{\sum_{j=1}^k \pi_j \eta(x|\mu_j, \sum_j)} \sum_k (x_n - \mu_k). \tag{3}$$

The posterior probability, or responsibilities, given by (1) appear naturally on the right-hand side

$$\gamma(z_{nk}) = \frac{\pi_k \eta(x|\mu_k, \sum_k)}{\sum_{j=1}^k \pi_j \eta(x|\mu_j, \sum_j)} \tag{4}$$

Multiplying by \sum_k^{-1} and rearranging we obtain

$$\mu_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n \tag{5}$$

where N_k is define as the effective number of points assigned to cluster k .

$$N_k = \sum_{n=1}^N \gamma(z_{nk}) \tag{6}$$

The mean μ_k for the k^{th} Gaussian component is obtained by taking a weighted mean of all the points in the data set, in which the weighting factor for data point x_n is given by the posterior probability $\gamma(z_{nk})$ that component k is responsible for generating x_n .

To obtain a maximum likelihood solution for covariance matrix of single Gaussian is given by

$$\sum_k = \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k)(x_n - \mu_k)^T \tag{7}$$

The mixing coefficient for the k^{th} component is obtained by an average responsibility which that component takes for explaining the data points.

$$\pi_k = \frac{N_k}{N} \quad (8)$$

It is worth to emphasize to use of simple iterative scheme for finding a solution to maximum likelihood problem from equations (5),(7), and (8). First choose some initial values for the means, covariances, and mixing coefficients. Using the alternate between the following two updates that call the E (expectation) step and the M (maximization) step. In the E step, use the values for the parameters to evaluate the responsibilities, or posterior probabilities from equation (1). Then use these probabilities in the M step, to re-estimate the means, covariances, and mixing coefficients using equations (5),(7), and (8). Like this update the parameters resulting from an E step followed by an M step is guaranteed to increase the log likelihood function. The algorithm is converged when the change in the log likelihood function, or alternatively in the parameters, falls below some threshold.

General procedure for EM algorithm

1. Initialize the parameters μ_k , Σ_k and π_k and evaluate the initial value of the log likelihood.
2. **E Step** Evaluate the responsibilities using Eq.(1) with current parameter values.
3. **M Step** Re-estimate the parameters using the current responsibilities:

$$\begin{aligned} \mu_k^{new} &= \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) x_n \\ \Sigma_k^{new} &= \frac{1}{N_k} \sum_{n=1}^N \gamma(z_{nk}) (x_n - \mu_k^{new})(x_n - \mu_k^{new})^T \\ \pi_k^{new} &= \frac{N_k}{N} \end{aligned}$$

where ($N_k = \sum_{n=1}^N \gamma(z_{nk})$)

4. Evaluate the log likelihood:

$$\ln p(X/\mu, \Sigma, \pi) = \sum_{n=1}^N \ln \sum_{k=1}^K \pi_k \eta(x_n/\mu_k, \Sigma_k) \quad (9)$$

and check for convergence of log likelihood. If the convergence criterion is not satisfied iterate from step 2.

Means of k clusters μ_k , $\forall k = 1, \dots, K$, is then used for estimating the skew of a document. Selecting k value is a highly subjective in nature. Therefore, we empirically fix the value of k to 5. Figure 1 depicts the mean points obtained for the input skewed word using mixture-of-gaussians.



Fig. 1. Illustration of mixture-of-gaussians for skewed word

3 Experimental Results and Comparative Study

This section presents the results of the experiments conducted to study the performance of the proposed method. The method has been implemented in MATLAB 7.0 on a Pentium IV 3.0 GHz with 1GB RAM. The proposed method first extracts individual handwritten text lines present in the document image using the method described in [1]. This technique is based on component extension and morphology. The resultant text line obtained from this algorithm is then passed to mixture model to extract mean vector points, which are further used to estimate the skew angle. In this work we compared with different skew estimators such as Linear Regression Analysis(LRA), Slope of a line and Second Order Moments to estimate the skew angle.

Around 30 handwritten Kannada documents is considered for experiment. We applied algorithm, which is described in [1] for extracting text line from the documents. Vertical projection profile is then used to segment into individual word. For experimentation purpose we considered around 500 words. The test set contained various skew angles ranging from 0 to 15 degree. Table 1 shows the accuracy obtained from different estimators. Accuracy is calculated using words with actual angle to the estimated angle. From the table it is ascertained that accuracy using Moments gives best results compared to LRA and Slope of a line. The next best is LRA. By using moments one can study the distribution such as how the distribution is skewed from its mean, or peaked. Another important aspect is that it is not sensitive to outliers. Figure 2 shows the resultant image after skew correction.

To check the robustness of the proposed algorithm, we tested our method on words corrupted by noise. For this, we considered 10 noisy words and each were rotated with four true angles. Here we used Salt-and-Pepper of noise density 0.02. Sample handwritten words contaminated by noise is as shown in Figure 3. Results obtained from the method are reported in Table 2. From the table it is clear that proposed method works efficiently and shows robust performance under noisy condition using Moments.

Table 1. Performance of the proposed method with various estimators

Different Estimators	Accuracy
LRA	94.5
Slope of a line	51.1
Moments	98.8

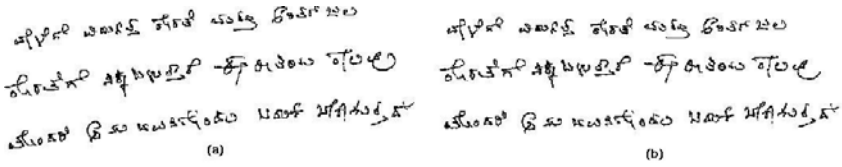


Fig. 2. (a) Original document image (b) Its skewed corrected

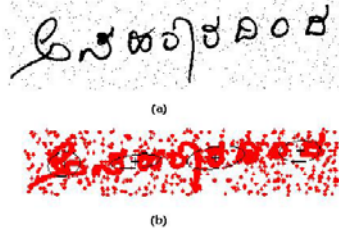


Fig. 3. (a) A Noisy handwritten word (b): Its illustration with mixture of four Gaussians

4 Discussion and Conclusion

Mixture Models are a type of density model, which comprise a number of component functions, usually Gaussian. These component functions are combined to provide a multimodal density. GMM is widely used in data mining, pattern recognition, machine learning, and statistical analysis. In many applications, their parameters are determined by maximum likelihood, typically through iterative learning of the EM algorithm. In this paper, we have effectively used this model in estimating skew of handwritten words. The expectation-maximization (EM) algorithm is used to learn the mixture of Gaussians. Subsequently the cluster means obtained from the words is used to estimate the skew angle. This is the first of its kind in the literature to estimate skew for handwritten Kannada words. Experiments on different handwritten words and words corrupted by noise shows the effectiveness of the proposed method.

Table 2. Mean and Standard Deviation for noisy handwritten Kannada word

True Angle	Noisy Word	
	M	SD
3	3.1	0.22
5	5.05	0.31
10	9.89	0.21
15	15.05	0.19

References

1. Aradhya, V.N.M., Naveena, C.: Text line segmentation of unconstrained handwritten kannada script. In: Proceedings of ACM International Conference on Communication, Computing & Security (ICCCS), Rourkela, India (2011) (accepted for publication)
2. Basu, S., Chaudhuri, C., Kundu, M., Narsipuri, M., Basu, D.K.: Skew angle correction and line extraction from unconstrained handwritten bengali text. In: Fifth International Conference on Advances in Pattern Recognition 2003, pp. 271–274 (2003)
3. Kapoor, R., Bagai, D., Kamal, T.S.: Skew angle detection of a cursive handwritten devanagari script character image. *Journal of Indian Institute of Science* 82, 161–175 (2002)
4. Kavallieratou, E., Fakotakis, N., Kokkinakis, G.: Skew angle estimation for printed and handwritten documents using the wignerville distribution. *Image and Vision Computing* 20, 813–824 (2002)
5. Lu, Y., Tan, C.L.: A nearest neighbor chain based approach to skew estimation in document images. *Pattern Recognition Letters* 24, 2315–2323 (2003)
6. Pavlidis, T., Zhou, J.: Page segmentation by white streams. In: Proceedings of 1st International Conference on Document Analysis and Recognition, pp. 945–953 (1991)
7. Roy, A., Bhowmik, T.K., Parui, S.K., Roy, U.: A novel approach to skew detection and character segmentation for handwritten bangla words. In: Proceedings of the Digital Imaging Computing: Techniques and Applications (DICTA 2005), p. 30 (2005)
8. Srihari, S.N., Govindaraju, V.: Analysis of Textual Images using the Hough Transform. *Machine Vision and Applications* 2, 141–153 (1989)
9. Su, T.H., Zhang, T.W., Huang, H.J., Zhou, Y.: Skew detection for chinese handwriting by horizontal stroke histogram. In: Proceedings of Intl Conf. on Document Analysis and Recognition, pp. 899–903 (2007)
10. Aradhya, V.N.M., Rao, A., Kumar, G.H.: Language independent skew estimation technique based on gaussian mixture models: A case study on south indian scripts. In: International Conference on Pattern Recognition and Machine Intelligence (PReMI), pp. 487–493 (2007)
11. Yan, H.: Skew correction of document images using interline cross-correlation. *Computer Vision, Graphics, and Image Processing* 55, 538–543 (1993)

Recognition of Simple and Conjunct Handwritten Malayalam Characters Using LCPA Algorithm

M. Abdul Rahiman¹ and M.S. Rajasree²

¹ Karpagam University, Coimbatore, Kerala, India
rehman_paika@yahoo.com

² Govt. College of Engg, Trivandrum, Kerala, India

Abstract. This paper mainly focuses on the recognition of both simple and conjunct handwritten characters in Malayalam, a South Indian language. The algorithm proposed recognizes these characters mainly based on the strokes and lines contained in them. Here the input is an image of handwritten Malayalam characters, which undergoes different phases of processing to produce an editable document of Malayalam characters in a predefined format as output. In this paper, detailed description of the methods for character identification is given. The whole OCR process is presented in three different modules: Pre-processing, Skeletonization and Recognition. In Pre-processing, the input image is scanned and subjected to line and character separation. In Skeletonization, the digital image is transformed into a set of original components. In Recognition, the characters are classified based on their features. The feature extraction of the characters is done by the analyzing the position and count of the horizontal and vertical lines. A classification of the simple and conjunct characters is also devised based on the count and position of the horizontal and vertical lines which make up those characters.

Keywords: Malayalam; Optical Character Recognition; Feature Extraction; Wavelet Transform; Neural Networks; HLH Patterns.

1 Introduction

Machine recognition of human reading has been a subject of intensive research for the last many decades. Handwritten Character Recognition systems can improve the Man-machine interaction and integrate computers into the human society. Indian Handwritten Character Recognition is getting much more attention and researchers are contributing a lot in this field. But Malayalam, a South Indian language has very less works in this area and needs further attention.

Reading handwritten texts is a very difficult task because of the diversities that exist in ordinary penmanship. Currently many OCR systems are available for handling handwritten documents of English, European and Asian languages [1][2][3] using methods of defining specifically-sized character boxes and read constrained handwritten entries. As there is no defined set of characters present in South Indian languages, there exist a lot of complex symbols [4][5][6][7][8]. Their characters set include basic characters and complex combined character set. Malayalam is a South

Indian language - which is the principal language of the State of Kerala [9][10]. The Malayalam script uses both old and new scripts for depicting characters. The old script concatenates various characters whereas new script separates the characters with a special character. The modern Malayalam script contains 13 vowel letters, 36 consonant letters, and a few other symbols including vowel signs.

Vowels are known as Svaram or Svarakshrangal in Malayalam. The vowel signs may be placed to the left (e, ē, ai) or right (ā, i, ī) of a consonant letter to which it is attached. The vowel signs o and ō may go to the left of a consonant letter or to the right of it. The figure 1, Set: 1 represents the vowels used in Modern Malayalam script.

Consonants in Malayalam are known as Vyanjanam or Vyanjanaksharangal. A consonant letter, does not represent a pure consonant, but represents a consonant + a short vowel /a/ by default. In figure 1, Set: 2 represents the consonants used in Modern Malayalam script.

A vowel sign is a diacritic attached to a consonant letter to indicate that the consonant is followed by a vowel other than /a/. The figure 1, Set: 3 represents the various vowel diacritics of one particular consonant /ka/. Next is conjunct characters shown in figure 1, Set: 4 that formed by the combination of more than one consonant and were widely used in old script.

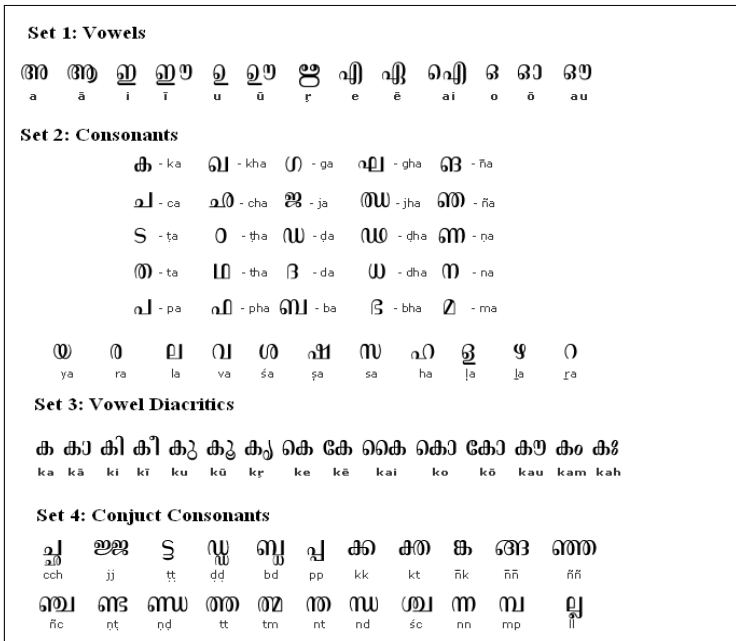


Fig. 1. Malayalam Script

2 Proposed OCR System

The architecture of the method is depicted in figure 2. The method focuses on the identification of handwritten Malayalam characters. The main advantage of this

method is that the identification of different coloured characters written in same document is possible. Characters can also be identified even if they are written on a coloured background.

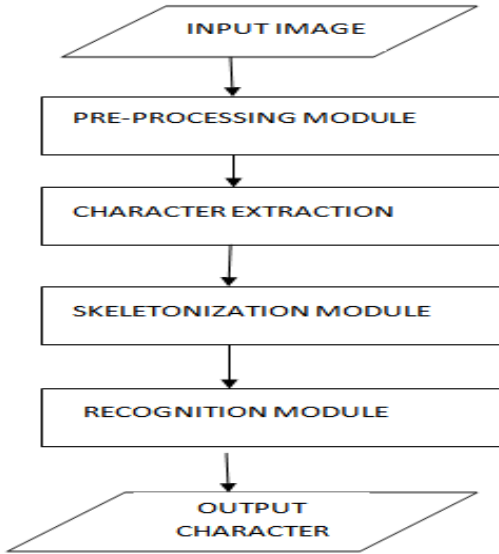


Fig. 2. Architecture of the technique

The method can identify background noises up to a certain extent. The character identification is done through 3 phases: Pre-processing, Skeletonization and Recognition.

2.1 Pre processing Technique

In this technique, the processes done before the actual identification[11][12] is performed. In this phase, colour of the background of the text and the written characters is checked. If they are of different colours, methods are applied to produce uniformly coloured characters. Background noises up to certain intensities can be identified and removed. In order to identify the characters, the characters are segmented first to produce individual units of characters. The scanned text is first subjected to line separation process where the written document is separated into line of characters. After line separation, each character in a line is subjected to the character separation process. The figure 3 illustrates line separation and character separation of both simple and conjunct characters.

2.2 Skeletonization

Skeletonization is the transformation of a component of a digital image into a subset of the original component. Skeletonization algorithms are the need to compute a reduced

amount of data or to simplify the shape of an object in order to find features for recognition algorithms and classifications. There are different categories of skeletonization methods: one category is based on distance transforms, and a specified subset of the transformed image is a distance skeleton. The original component can be reconstructed from the distance skeleton.

Thinning is the process of peeling off a pattern as many pixels as possible without affecting the general shape of the pattern. The skeleton obtained must be as thin as possible, connected and centered. Individual pixels are either removed in a sequential order or in parallel. Normally, it is implemented by an iterative process of transforming specified contour points into background points. In case of conjunct characters, thinning should be performed carefully as the characters have more loops and complex strokes.

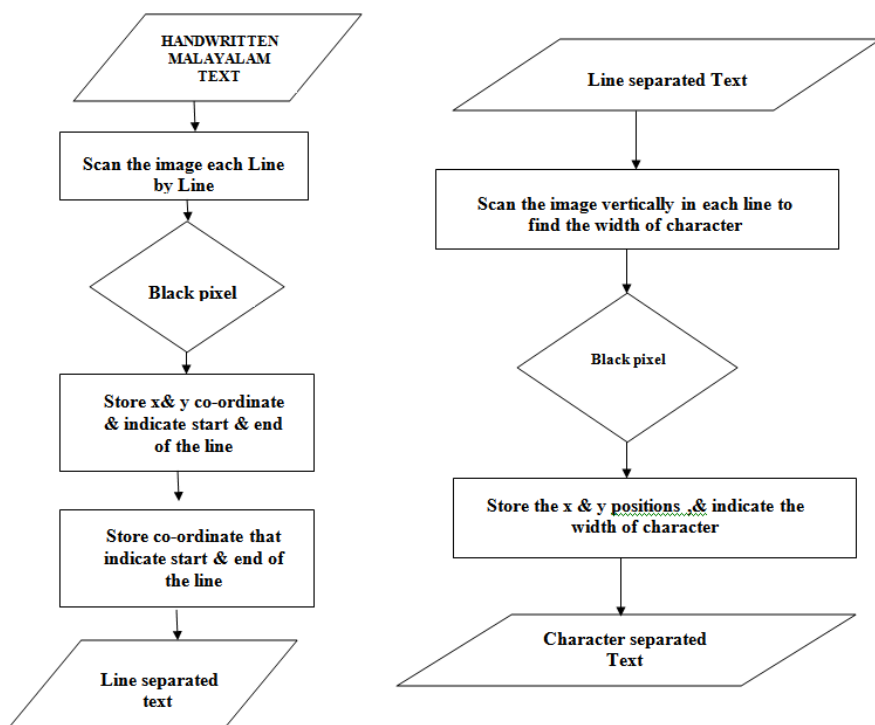


Fig. 3. Line & Character Separation

2.3 Feature Extraction and Recognition

The final phase and the most important phase of the identification of characters involve a series of methods. The skeletonised and segmented characters are made to undergo functions which calculate the number of horizontal and vertical lines which form the features of the characters. Using the count of horizontal and vertical lines, the characters are classified into different groups. For example consider the character 'Ra'. It has two vertical lines and a single horizontal line. So it can be classified into group

of characters having similar features. For recognition of certain characters, the count of horizontal and vertical lines is enough. But for other characters such as ‘La’, ‘Va’, ‘Pa’ etc the position of these lines are also important as these differentiate each other. Hence the positions of these lines are also calculated i.e. whether at the top, bottom, left or right. After calculating the count and position of horizontal and vertical lines, the characters are classified to form different groups. The same approach can be used for identification of conjunct characters but care should be taken while counting the strokes as conjunct characters involve more complex strokes and loops. The figure 4 shows the feature extraction technique.

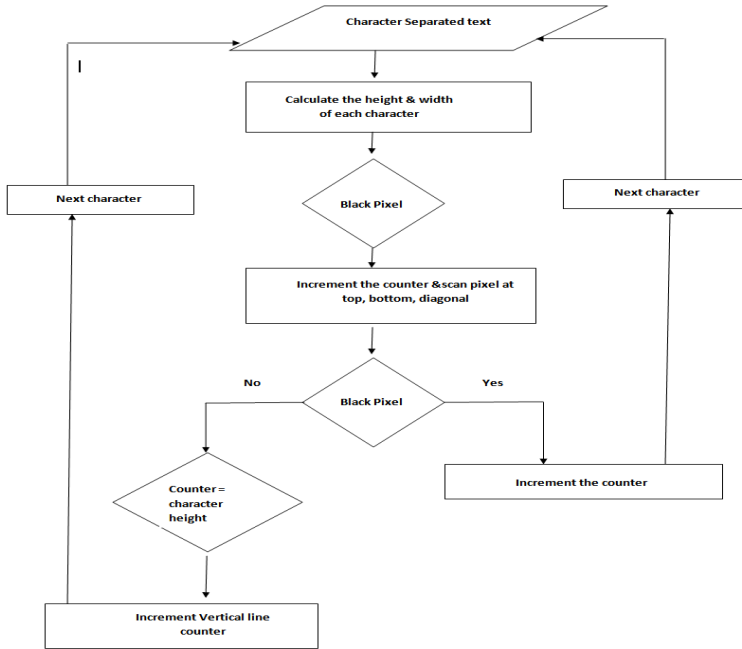


Fig. 4. Feature extraction by counting the number of vertical lines in the characters

Algorithm 1: To count the number of vertical lines in a character

- Step 1: Initially, the height and width of the character is calculated.
- Step 2: The character box is scanned horizontally from the top left position to find the left most black pixel.
- Step 3: When a black pixel is found, count is incremented and the pixels which are present at the top, bottom and diagonal to it are also checked.
- Step 4: If any of those pixels are black, the count is again incremented. This is done to check for slanted or curved vertical lines.
- Step 5: Else if none of these pixels are black, check whether the count has reached the character height. If so a vertical line has been found and the vertical line count is incremented.

Step 6: Repeat the above steps by incrementing the x coordinate till the width of the character.

Step 7: The above process is repeated for each character in the text.

Algorithm 2: To calculate the number & position of horizontal lines in a character

Step 1: Initially, the height and width of the character is calculated.

Step 2: The character box is scanned vertically from the top left position to find the left most black pixel.

Step 3: When a black pixel is found, count is incremented and the pixels which are present at the left, right and diagonal to it are also checked.

Step 4: If any of those pixels are black, the count is again incremented. This is done to check for slanted or curved vertical lines.

Step 5: Else if none of these pixels are black, check whether the count has reached around the character width. If so a horizontal line has been found and the horizontal line count is incremented.

Step 6: Divide the character box into three vertical segments in order to calculate the position of the horizontal lines.

Step 7: If the left end point of the horizontal line lies in left vertical segment and right end point in the right vertical segment, position of the line is in the middle.

Step 8: Else if the left end point of the horizontal line lies in left vertical segment, position of the line is in the left.

Step 9: Else if the right end point of the horizontal line lies in right vertical segment, position of the line is in the right.

Step 10: Else, position of the line is in the middle.

Step 11: Check whether the horizontal line is at the top, bottom or centre of the character box.

3 Classification of Characters Based on Features

When the input data to an algorithm is too large to be processed and it is suspected to be notoriously redundant then the input data will be transformed into a reduced representation set of features or features vector. Transforming the input data into the set of features is called feature extraction. Figure 5 & Figure 6 shows the input and corresponding to the simple and conjunct characters respectively.

Here we have proposed a new method of classification of Malayalam Characters based on their count of lines and the position of component lines. Here the main feature we have found common to all characters is the vertical line. Every single character is made up of more than 2 vertical lines. So the classification technique is developed based on the number of vertical strokes found in the character. The characters can then be divided into various sets based on the horizontal lines and their positions i.e. top horizontal line, middle horizontal line and bottom horizontal line. Figure 7 shows the classification according to vertical lines.



Fig. 5. Input Image of simple characters and the corresponding Character recognition process by calculating number of horizontal and vertical lines

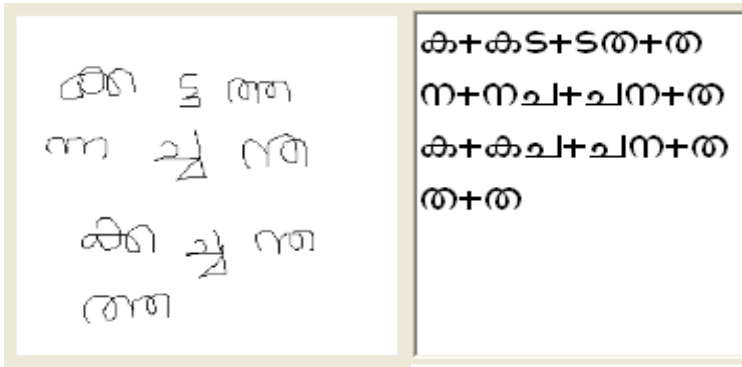


Fig. 6. Input Image of conjuncts and the recognized characters

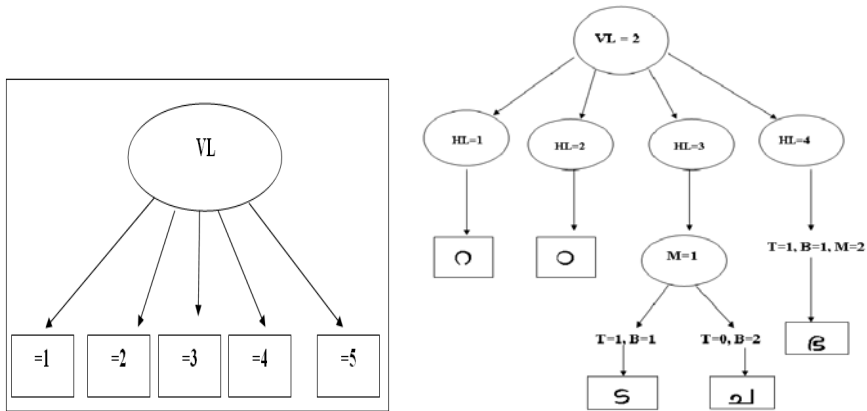


Fig. 7. Decision tree based on classification of number of vertical lines

According to the classification, there are mainly 6 types of characters which are illustrated in Table 1 & Table 2.

Table 1. Details of each type of classification

Type No:	Vertical Bars Count	Example characters
Type 1	VB=2	<u>ട, ച, ഓ, റ</u>
Type 2	VB=3	<u>ഥ, ശ, ദ, ല, ള, മ, റ, റ്റ</u>
Type 3	VB=4	<u>ക, ഛ, ഖ, ത, ശ, ധ, ന, ഴ, ഷ, ററ, യ</u>
Type 4	VB=5	<u>ണ, ഡ, സ, ള, റല</u>
Type 5	VB=6	ഊ, ഒ
Type 6	VB=7	ഠ, ഞ

Table 2. Whole Classified Characters with vertical and horizontal bar count

HB \ VB	1	2	3	4	5
2	ഠ	ഠ	ട, ച	ഭ	
3		ഥ, ശ, മ, ധ, റ, ഴ, ന, ള	മ, ഴ, ല, ര, ദ	ദ	
4		ധ, ന	ഖ, ത, ശ, റ, റ്റ	ക, ഛ	
5			സ, ഡ	യ, റല, ഷ	
6					ഊ, ഒ
7				ണ	ഠ, ഞ

Table 3. Recognition Results

Type	Total Characters	Correctly Recognised
Type 1	1562	1531
Type 2	1400	1344
Type 3	900	819
Type 4	1100	1045
Type 5	640	544
Type 6	380	364

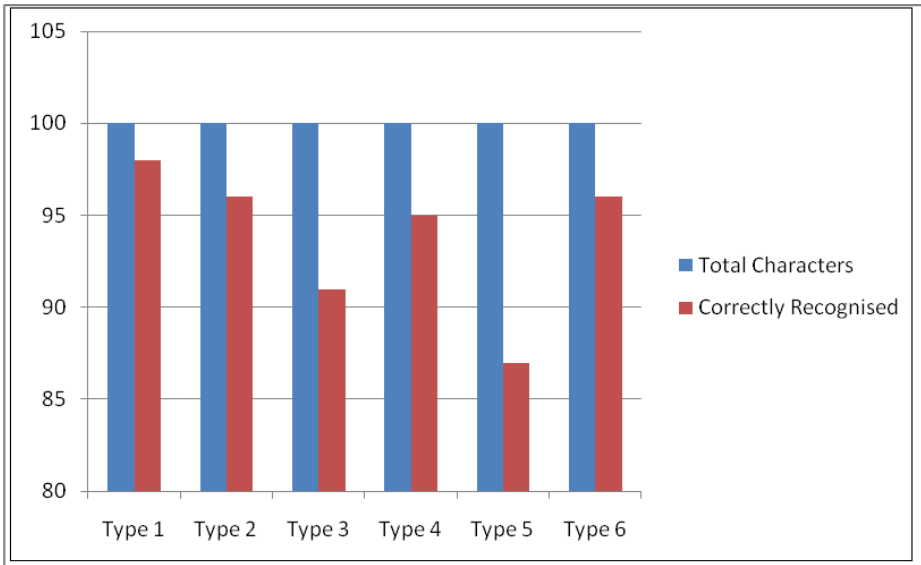


Fig. 8. Recognition results of the technique

4 Experimental Analysis

The experiment was conducted using different lines of text from various sources. We also tried the samples of handwritten characters of 9 different categories of persons. A set of specific characters were selected for sample formation and the character separation implemented. The accuracy of recognition is defined as the number of correct outputs to the total number of input characters that was supplied. Subsequently,

a graph is also drawn based on the experimental results. Table 3 below shows the experimental results and the recognition of characters are illustrated in figure 8 to 11. An accuracy of 93.83% is achieved through this method.

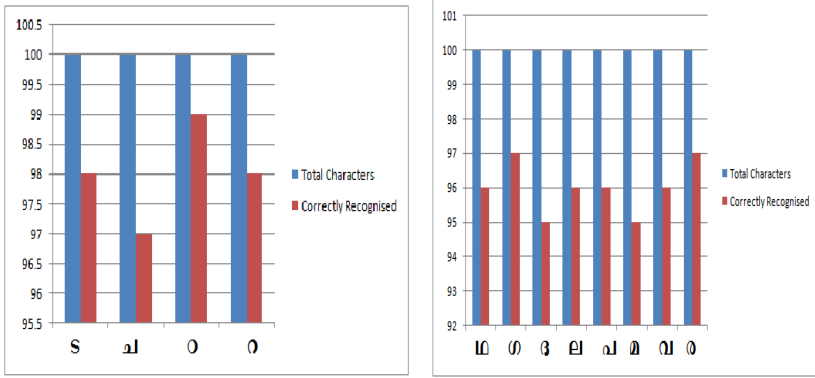


Fig. 9. Recognition of Type 1 & 2 Characters



Fig. 10. Recognition of Type 3 & 4 Characters

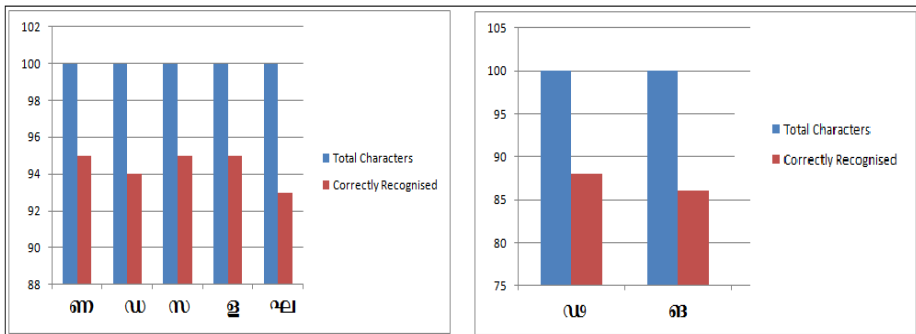


Fig. 11. Recognition of Type 5 & 6 Characters

5 Conclusion

In this paper, we have proposed an algorithm to recognize the simple and conjunct handwritten Malayalam characters on a scanned text to produce an editable document. Till now we have tried to accurately identify the characters written on any different coloured background. Pre-processing is done to improve the accuracy of recognition of characters and to remove the background noises. The major factor hindering the improvement of accuracy is the similarity in character shapes and features of certain Malayalam characters. The refinement of the system is possible by training the OCR Engine to handle commonly encountered errors. Further investigation is underway in order to recognize the old script of Malayalam. The future scope of this paper lies in the recognition of connected characters used in the script.

The connected characters remain a challenge as they are two different characters separated by a space even though, should be considered as a single unit. One technique that can be applied for connected characters is probability distribution technique. Here the probability of the occurrence of a connector next to a main character is calculated and the distribution of black pixel in that distance is evaluated to find out what type of connector is used.

References

1. Trier, D., Jain, A.K., Taxt, T.: Feature Extraction methods for Character Recognition – A Survey. *Pattern Recognition* 29, 641–662 (1996)
2. Srihari, S.N., Yang, X., Ball, G.R.: Offline Chinese Handwriting Recognition: an assessment of current Technology. *Front. Computer Science* 1(2), 137–155 (2007)
3. Amin, A.: Recognition of Printed Arabic Text based on global features and Decision Tree Learning Techniques. *Pattern Trcognition* 33(8), 1309–1323 (2000)
4. Pal, U., Chaudhuri, B.B.: Printed Devanagari script OCR System. *Vivek* 10 (1997)
5. Chaudhuri, B., Pal, U., Mitra, U.: Automatic recognition of printed Oriya script. *Sadhana* 27, Part 1, 23–34 (2002)
6. Seethalakshmi, R., Sreeranjani, T.R., Balachandar, T., Singh, A., Singh, M., Ratan, R., Kumar, S.: Optical Character Recognition for printed Tamil text using Unicode. *Journal of Zhejiang University SCI* 6A(11), 1297–1305 (2005)
7. Lakshmi, C.V., Patvardhan, C.: A multi-font OCR system for printed Telugu text. In: *Proc. of Language Engineering Conference LEC, Hyderabad*, pp. 7–17 (2002)
8. Ashwin, T.V., Sastry, P.S.: A font and size independent OCR system for printed Kannada documents using support vector machines. *Saadhana* 27, Part 1, 35–58 (2002)
9. Abdul Rahiman, M., Rajasree, M.S.: Printed Malayalam Character Recognition Using Back propagation Neural Networks. In: *Proc. of IEEE International Advance Computing Conference (IACC 2009), Patiala*, pp. 1140–1144 (March 2009)
10. *Journal of Language Technology*, Viswabharat@tdil (July 2003)
11. Anuradha, Koteswarra, B.: An efficient Binarization technique for old documents. In: *Proc. of International Conference on Systemics, Cybernetics and Informatics, Hyderabad*, pp. 771–775 (2006)
12. Chaudhuri, B.B., Pal, U.: Skew Angle Detection of Digitized Indian Script Document. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(2) (February 1997)

A Fuzzy Genetic Approach to Impulse Noise Removal

K.K. Anisha and M. Wilsy

Department of Computer Science
University of Kerala, Kariavattom, Thiruvananthapuram 695 581, Kerala, India
anisha.kelamkumarath@gmail.com,
wilsyphilipose@hotmail.com

Abstract. Many practical applications require analysis of digital images. An accurate analysis is possible only from an image free of noise. Image denoising with multiple image filters might produce better results than a single filter, but it is very difficult to find a set of appropriate filters and the order in which the filters are to be applied. In this paper, we propose a Fuzzy Genetic Algorithm to find the optimal filter sets for removing impulse noise from images. Here, a Fuzzy Rule Based System is used to adaptively change the crossover probability of the Genetic Algorithm used to determine the optimal sets of filters from a pool of standard image filters. Fuzzy Genetic Algorithm gives better results than conventional Genetic Algorithm. This method does not require any deep knowledge about the image noise factors; so it can be easily used in any image processing application.

Keywords: Adaptive Genetic Algorithm, Fuzzy Genetic Algorithm (FGA), Fuzzy Rule Based System (FRBS), Genetic Algorithm (GA), Image filters, Impulse noise.

1 Introduction

Digital image processing plays a key role in medical imaging, satellite imaging, underwater imaging, robot vision and many such applications. Since images can be deteriorated during acquisition, storage and transmission, image denoising is a primary precursor for almost all image analysis tasks. Conventional smoothing filters and median filters are the most popular filters for noise reduction in digital images [1]. But, a single smoothing or median filter is not enough for completely removing the noise, especially when the noise level is high. Also, it may not preserve image details such as edges during filtering. Applying a set of denoising and enhancement filters successively on a noisy image may remove noise and preserve image details much more efficiently than a single median or smoothing filter. Such a set of standard filters is called a composite filter. The type of the filters in the composite filter, as well as the order in which the filters are applied must be appropriately chosen for good results.

Jin Hyuk Hong, Sung Bae Cho and Ung Keun Cho proposed a method that used Genetic Algorithm (GA) [2] to determine composite filters that remove different levels of impulse noise from an image [3]. They have extended this method to determine composite filters that performs local and global image enhancement [4]. In

these methods, the GA considers a set of possible filter combinations of a particular length, selects the best combinations among them according to a fitness value assigned to each combination based on a fitness function, and applies genetic operators such as crossover and mutation [2] on the selected combinations to create the next generation of composite filters. This process is repeated, enabling GA to find the optimal composite filters. In this method, GA parameters, which affect the quality of the solutions produced, are kept fixed. If these parameters are not assigned with suitable values, GA may converge to a sub optimal solution, or it may take a long time to converge to the optimal solution. However, choosing the best parameter values is difficult because the parameter values are problem dependent and the interaction of these parameters with the GA is complex.

The performance of GA can be improved by adaptively varying its parameters instead of keeping them fixed. Fuzzy logic [7] based techniques have been used for adaptively selecting GA parameters [5] [8-10]. In Fuzzy Genetic Algorithm (FGA) [5], [8-10], a Fuzzy Rule Based System (FRBS) [7] is used to adapt any of the GA parameters.

The proposed method is an extension of the method in [3]. Here, an FGA is used to determine the optimal filters that can remove different levels of impulse noise from images without relying on deep knowledge about the type of image noise factors. From a pool of standard filters, the GA part of FGA selects several filters and constructs a composite filter. GA analyses such a set of composite filters and determine the optimal filters for removing different levels of impulse noise. The crossover probability [2] of GA, which determines the number of selected solutions that undergo crossover operation, is adapted by the fuzzy part of FGA, where an FRBS determines the amount of variation that should be undergone by the crossover probability value in order to improve the quality of the solutions produced. The proposed method has been tested on benchmark images and its performance has been evaluated using performance metrics such as PSNR value, Tenengrad measure and IQI [6] value. These evaluations clearly show the superiority of the proposed method.

The rest of the paper is organized as follows. Section 2 gives a detailed account of the determination of the optimal composite filters by GA. Section 3 is about the proposed method, where the design and working of the FGA is explained. Section 4 discusses the experimental results. Section 5 provides conclusions.

2 Genetic Algorithm Based Impulse Noise Removal

When there are m filters in the filter pool, optimal composite filters containing l standard filters are to be determined from a total of $(m+1)^l$ filter combinations, where $m+1$ includes the case of not using any filter on the image. Trying all cases to find out the best one is practically impossible, especially when m is large. In this paper, GA is used to find the optimal composite filter, in which the proper type and order of filters are determined [3]. In GA [2], each solution to the problem to be solved is called an individual or a chromosome. GA starts by randomly initializing a set or a population of individuals. This is the first generation of individuals. Each individual is assigned a fitness value based on a fitness function. GA selects those individuals with a good fitness value and applies operations such as crossover and mutation on them to create

the next generation of individuals. This process is repeated until GA satisfies a predefined termination criterion such as the number of generations created, upon which GA is expected to have produced very good individuals.

Table 1 shows the filter pool used in this paper which contains 30 image filters, each indexed by a value from 1 - 30. Value 0 represents the case where no filtering operation is performed. The first 3 filters are histogram brightness measures that adjust the value of the pixel p in the image according to equation (1) for a given scale $(-100 \leq s \leq 100)$.

$$v_{new}(p) = v_{old}(p) + \frac{v_{old}(p) * s}{100} \tag{1}$$

Filters 4 – 7 are histogram contrast measures that adjusts the value of the pixel p for a given scale $(-127 \leq s \leq 127)$ as shown in equation (2).

$$v_{new}(p) = v_{old}(p) + \frac{v_{old}(p) - 128 * s}{128} \tag{2}$$

Filter 8 performs contrast stretching by spanning the range of intensity values in an image, $[c, d]$, to a desired range of values $[a, b]$. It scales each image pixel p according to equation (3).

$$v_{new}(p) = v_{old}(p) - c * \left(\frac{b - a}{d - c} \right) + a \tag{3}$$

Filter 9 equalizes the histogram of the image, thereby improving the image contrast. Filters 10 – 15 are smoothening filters of different sizes and shapes. Filters 16 – 19 are edge enhancement filters of different types. Filters 20 – 29 are standard median filters of different sizes and shapes. Filter 30 is an adaptive median filter with a maximum window size of 7 [1].

Table 1. Description of image filters used in this paper

Group	Filter	Type	Index
Histogram	Brightness	3 values	1~3
	Contrast	4 values	4~7
	Stretch	-	8
	Equalize	-	9
Masks	Smoothening	6 masks	10~15
	Sharpening	4 masks	16~19
	Median, Adaptive Median	10 masks	20~30
None			0

Each composite filter is represented by a string of \mathbf{l} real numbers corresponding to the filter index, where \mathbf{l} is the number of standard filters in the composite filter.

The following procedure describes how GA determines the optimal standard filters [3]. At first, GA randomly initializes a population of composite filters. Then, the fitness of each composite filter is evaluated using the fitness function given in equation (4). Here, the objective of GA is to find the optimal composite filter that can remove impulse noise from all the training images in a given training set. The training images are created by artificially corrupting an image with different levels of impulse noise. In equation (4), n is the number of training images used, MAE_i is the mean absolute error (MAE) of the output image obtained after applying the composite filter x on the i^{th} training image, and MAE_{\max} is the maximum MAE; it would be 255 for 8-bit grayscale images. The fitness value $f(x)$ is assigned to x . From equation (4), it is clear that the composite filter x receives a high fitness value if it can considerably remove the noise from all the training images (thereby producing low MAE_i values).

$$f(x) = \frac{1}{n} \sum_{i=1}^n 1 - (MAE_i / MAE_{\max}) \quad (4)$$

GA then selects the composite filters with high fitness value using Roulette Wheel selection strategy [2], where selection is based on the probability assigned to each composite filter proportional to its fitness value. Then, genetic operators such as crossover and mutation are applied on the selected individuals, to produce the next generation. Elitist-strategy [2] that keeps the best individuals of the previous generation in the current generation is also used. From this current generation, GA produces the next generation using the above procedure. This process is repeated until a predefined termination criterion is satisfied.

In the above method, the GA parameters such as the crossover probability are kept constant. The values assigned to these parameters may not be the best for this problem. Hence, the convergence of GA may be to a sub optimal solution. Even if GA converges to the optimal solution, it may take a large amount of time to do so. The proposed method, which is explained in the following section, avoids the occurrence of these problems by adaptively varying the crossover probability of the GA using an FRBS.

3 The Proposed Method Using Fuzzy Genetic Algorithm

An FGA [5], [8 – 10] is an adaptive GA in which an FRBS is used to adapt one or more of the GA parameters so as to increase the quality of the solutions produced by GA. In FGA, the FRBS accepts one or more measures, which indicate the quality of the outputs produced by GA, as inputs. These values are fuzzified [7] using the corresponding membership functions [7]. From these fuzzified input values, one or more fuzzy outputs are determined by the FRBS using a rule base [7]. These outputs are then defuzzified [7] using the output membership functions. The defuzzified outputs enable the adaptive variation of one or more GA parameters. Thus, the FRBS enables GA to converge to the most optimal solution. It also results in an increase in the speed of convergence of the GA to the best solution.

In the proposed method, an FGA is used to determine the optimal composite filters for impulse noise removal. Here, an FRBS is used to adapt the crossover probability

of GA. The FRBS accepts Genotypic diversity (GD) and Phenotypic diversity (PD) as inputs [5]. GD and PD are two measures that depict the quality of the composite filters produced by GA.

GD represents the genetic diversity of the population and it is evaluated as shown in equation (5).

$$GD = (d - d_{min}) / (d_{max} - d_{min}) \tag{5}$$

Where d , d_{max} and d_{min} are the average, maximum and minimum distances of the chromosomes in the population from the composite filter with the highest fitness value.

PD, as shown in equation (6), is the ratio of average fitness of the population, f_{avg} , to the best fitness f_{best} .

$$PD = f_{avg} / f_{best} \tag{6}$$

The output of the FRBS is a value δp_c that controls the variation of the current crossover probability p_c . δp_c , which ranges from [0, 1.5], determines the degree to which the current p_c value, which is kept within the range [0.25, 0.75], should vary. The variation is carried out by multiplying the δp_c value obtained from the FRBS with the current p_c .

Figures 1(a), (b) and (c) shows the membership functions of GD, PD and δp_c respectively.

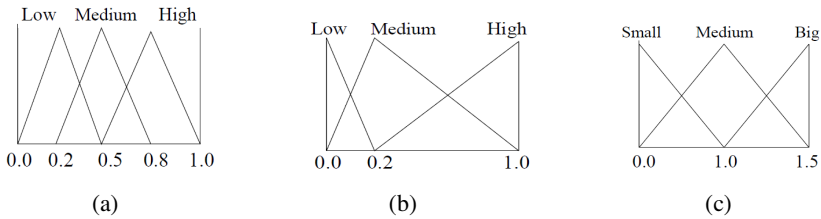


Fig. 1. Membership function of (a) GD (b) PD (c) δp_c

Table 2 shows the rule base using which the FRBS finds the value of δp_c from GD and PD values. GD and PD values range from Low to High, for which the change in δp_c , which ranges from Small to Big, is given in the respective cells. When GD and PD values are ‘Low’, the population is diverse, even if it has not converged to the best solution. In this case, a low crossover probability is desired to prevent loss of this diversity due to crossover. Hence, δp_c is given a ‘Small’ value, to allow as little crossover operations as possible. Similar arguments follow for all the conditions specified in the rule base.

Table 2. Rule base of FRBS

GD	PD		
	Low	Medium	High
Low	Small	Small	Medium
Medium	Big	Big	Medium
High	Big	Big	Medium

In the proposed method, the GA part works as described in section 2. GA creates the first generation of composite filters using a randomly initialized p_c value. The GD and PD values for this generation are fed into the FRBS, which calculates the value of δp_c , which is multiplied with the current p_c value. This adapted p_c value is given to the GA, which uses it to create the next generation to produce a better population of composite filters. The entire process is repeated until GA satisfies its termination criterion.

4 Results and Discussion

The proposed method was implemented and tested using the following GA parameters: population size = 50, number of generations = 50, number of filters in the composite filter = 5, mutation probability = 0.05 and selection rate = 0.9. The initial value assigned to p_c is 0.7. The adapted p_c value obtained after 50 iterations is 0.75.

The well-known Lena, Elaine and Peppers images, all of size 512×512 , were used as benchmark images. These images are shown in figures 2 (a), (b) and (c) respectively. 5 Lena images artificially corrupted by impulse noise factors with corruption rates of 10%, 30%, 50%, 70% and 90% are used as training images. The test images are Lena, Elaine and Peppers images corrupted with various noise levels.

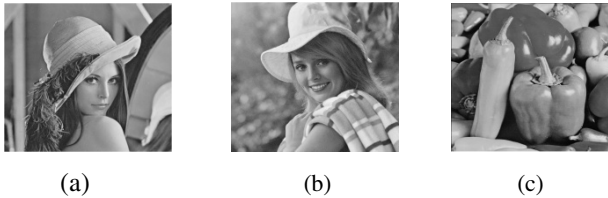


Fig. 2. Benchmark images – (a) Lena, (b) Elaine, (c) Peppers

Table 3 shows the composite filters created by the proposed method for different impulse noise levels by using the Lena training images. From now on, we call these composite filters as FGA filters. Here, the same composite filter, made up of relatively simple standard filters, can be used to remove upto 50% impulse noise. For higher noise levels, separate composite filters are evolved, which are composed of standard filters that either uses the information of larger neighbourhoods or perform further enhancements.

4.1 Performance Evaluation

The performance of FGA filters were compared with that of the composite filters produced by GA [3], using the filter pool in table 1, Lena training images and a p_c value of 0.7. From now on, we call these composite filters as GA filters. FGA filters were also compared with a single 5×5 rectangular median filter, as a median filter is conventionally used for impulse noise removal [1].

Table 3. Composite filters evolved by FGA for the training images

<i>Impulse Noise Level (%)</i>	FGA filter
10	Adaptive median, Adaptive median, Unsharp ($\alpha = 0.5$), 3×3 weighted averaging, 3×1 vertical median
30	Adaptive median, Adaptive median, Unsharp ($\alpha = 0.5$), 3×3 weighted averaging, 3×1 vertical median
50	Adaptive median, Adaptive median, Unsharp ($\alpha = 0.5$), 3×3 weighted averaging, 3×1 vertical median
70	Adaptive median, Adaptive median, Unsharp ($\alpha = 0.5$), 3×3 weighted averaging, 5×5 weighted averaging
90	Adaptive median, Adaptive median, Radius 3 disk median, Radius 5 disk median, 3×3 averaging

Following are the metrics used for performance evaluation:

Peak Signal to Noise Ratio (PSNR)

PSNR value of a denoised image with respect to the original image is calculated as shown in equation (7). This value, represented in dB, denotes the closeness of the denoised image to the original image. A high PSNR value for the denoised image shows its closeness to the original image.

$$PSNR = 10 * \log_{10}(255^2 / MSE) \tag{7}$$

where MSE is the mean squared error.

Tenengrad Measure

Tenengrad measure indicates the amount of edge details present in an image. Higher the value, the more edge details present in the image. Tenengrad method is based on obtaining the gradient magnitude from the Sobel operator. It is calculated as shown in equation (8).

$$TEN = \sum_{x=2}^{M-1} \sum_{y=2}^{N-1} (\nabla S(x, y))^2 \text{ for } \nabla S(x, y) > T \tag{8}$$

Where T is a discrimination threshold value and $\nabla S(x, y)$ is the Sobel gradient magnitude value. Here, T is taken to be zero. When the TEN for the denoised image R is close to the original image O, it shows that the denoising process preserves the edge details in the image. TEN of R is less than TEN of O when the denoising process results in loss of edge details. TEN of R is greater than TEN of O when the denoising process creates false edge details.

Image Quality Index (IQI)

IQI [6] is designed by modelling any image distortion as a combination of three factors: loss of correlation, luminance distortion and contrast distortion. It is calculated as shown in equation (9). The value of IQI ranges from [-1, 1]. A denoised image, which is much similar to the original image, and hence of high quality, has an IQI value close to one.

$$IQI = Corr(O, R) * Lum(O, R) * Cont(O, R) \tag{9}$$

The comparative analysis of FGA, GA and median filters was done for two cases. They are as follows:

Case 1: Using training images for testing the filters

Table 4 shows the values of different metrics obtained for the Lena training images denoised by the median filter, GA filters, and FGA filters in table 3 for respective impulse noise levels. It can be seen that using composite filter for impulse noise removal yields better results than using a single median filter. The difference between the PSNR values of the denoised images created by the median filter and composite filter becomes larger as the noise level in the input image increases. The quality of the denoised image produced by the median filter is less than that of the denoised image produced by the composite filters, as shown by their IQI values. The Tenengrad value for the original Lena image is 22034. For median filter, Tenengrad value is smaller than that for the composite filters for low noise levels, which indicates that loss of edge detail is more when using single filters. For high noise levels, this value is much higher than 22034 and the TEN values for the composite filters, which indicates that median filter creates more false edge details than composite filters. Altogether, composite filter is better than a single median filter for impulse noise removal with edge preservation.

Table 4. Comparing PSNR, Tenengrad measure and IQI obtained for Lena training images

<i>Metrics</i>	<i>Filter (s)</i>	<i>10% noise</i>	<i>30%noise</i>	<i>50%noise</i>	<i>70%noise</i>	<i>90%noise</i>
PSNR (dB)	Median	30.528	27.1045	22.9126	14.0837	7.5496
	GA filter	32.4092	31.5173	30.2289	27.7124	18.5087
	FGA filter	34.7854	33.0935	30.9316	28.3021	19.5968
IQI	Median	0.6500	0.6245	0.5187	0.1495	0.0136
	GA filter	0.6855	0.6704	0.6431	0.5808	0.3134
	FGA	0.7984	0.7755	0.7279	0.614	0.3391
TEN	Median	5157	50688	27580	222462	2820290
	GA filter	9624	14764	45190	28856	63525
	FGA	30141	42917	66133	13801	17930

From table 4, it can be seen that FGA filters yields better results than GA filters. The PSNR values and IQI measures for FGA filter outputs are higher than GA filter outputs for all noise levels. For low noise levels, FGA filters produces false edge details, as shown by their high TEN value. At the same time, in GA filter outputs,

edge details are lost, as indicated by their low TEN values. This situation is changed as the noise level in the image increases. While GA filters introduce a large amount of false details (indicated by high TEN values), FGA filters preserves some of the details without creating false details, as the TEN values for FGA filter outputs for high noise levels are near 22034 (TEN value for the original Lena image).

Case 2: Using other images for testing

Tables 5 and 6 compares the PSNR, TEN and IQI values obtained for denoised images formed from using 5 × 5 median filter, GA and FGA filters, for Elaine and Peppers images respectively. Here too, it can be seen that composite filters produce better results than the single median filter in noise removal and edge preservation.

Here also, the overall performance of FGA is better than GA, as shown by their PSNR values and IQI measures. For Elaine, edge details in GA filter outputs are closer to the original image (TEN of Elaine = 374810) than FGA filter outputs. For Peppers (TEN = 536780), FGA filters result in better edge preservation for low noise levels. For high noise levels, GA filters result in better edge preservation.

Table 5. Comparing PSNR, Tenengrad measure and IQI obtained for noisy Elaine images

<i>Metrics</i>	<i>Filter (s)</i>	<i>10% noise</i>	<i>30%noise</i>	<i>50%noise</i>	<i>70%noise</i>	<i>90%noise</i>
PSNR	Median	31.0845	27.3783	22.6409	14.0808	7.5133
	GA filter	32.3154	31.6214	30.8815	29.1725	19.2634
	FGA	32.9177	31.8163	31.2214	29.2903	20.2065
IQI	Median	0.5766	0.5552	0.4563	0.135	0.0011
	GA filter	0.6093	0.5940	0.5794	0.5190	0.3174
	FGA	0.7175	0.6871	0.6030	0.5444	0.3479
TEN	Median	407341	480621	296487	2063508	22050
	GA filter	386098	422531	340401	302723	327878
	FGA	337736	593210	437009	286386	204864

Table 6. Comparing PSNR, Tenengrad measure and IQI obtained for noisy Peppers images

<i>Metrics</i>	<i>Filter (s)</i>	<i>10% noise</i>	<i>30%noise</i>	<i>50%noise</i>	<i>70%noise</i>	<i>90%noise</i>
PSNR	Median	31.2502	27.5245	22.8211	13.9771	7.3478
	GA filter	30.7426	30.7544	29.3661	26.9013	18.1285
	FGA	32.7587	30.9156	30.0693	27.3942	19.6296
IQI	Median	0.6302	0.6049	0.4983	0.1534	0.0151
	GA filter	0.6489	0.6408	0.6193	0.5646	0.3208
	FGA	0.7364	0.7168	0.6255	0.5743	0.3621
TEN	Median	268845	291954	122690	61623	2459871
	GA filter	284526	268773	260418	420438	29795
	FGA	548998	578615	496616	243475	27729

Figure 3 shows examples of outputs produced by GA filters, FGA filters and 5×5 rectangular median filter from Peppers with 30% noise, Elaine with 50% noise and Lena with 70% noise. While some amount of impulse noise is present in the outputs of 5×5 rectangular median filter especially for high noise levels, noise is almost completely removed by the GA and FGA filters.

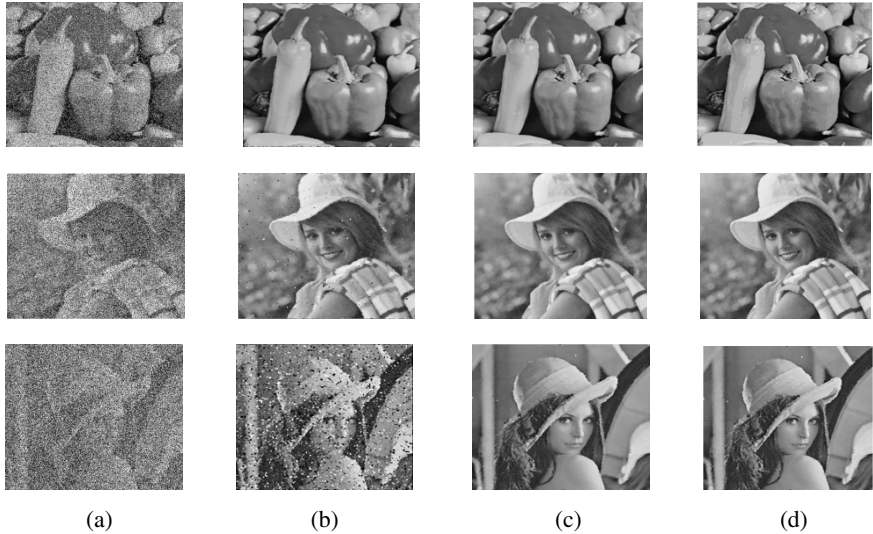


Fig. 3. Column (a): Peppers with 30% impulse noise, Elaine with 50% impulse noise, Lena with 70% impulse noise. Column (b), (c), (d): Outputs produced by median filter, GA filter and FGA filter respectively

5 Conclusion

Median filters are commonly used for impulse noise removal. But, they result in loss of image details. Also, for high noise levels, they are not sufficient in completely removing the noise. In this paper, a FGA is used to determine the optimal composite filters for removal of different impulse noise levels without using deep knowledge about noise factors. Here, an FRBS is used to adaptively change the crossover probability of GA. Experiments on benchmark images shows satisfactory results. As future work, the proposed method can be used for impulse noise removal in satellite and medical images.

References

1. Gonzalez, R., Woods, R.: Digital Image Processing. Addison Wesley, Reading (1992)
2. Goldberg, D.: Genetic Algorithm in Search, Optimization and Machine Learning. Addison Wesley, Reading (1989)
3. Hong, J.H., Cho, S.B., Cho, U.K.: A Novel Evolutionary Method to Image Enhancement Filter Design: Method and Applications. IEEE Transactions on Systems, Man and Cybernetics – Part B, Cybernetics 39(6), 1446–1457 (2009)

4. Cho, U.-K., Hong, J.-H., Cho, S.-B.: Evolutionary Image Enhancement for Impulsive Noise Reduction. In: Huang, D.-S., Li, K., Irwin, G.W. (eds.) ICIC 2006. LNCS, vol. 4113, pp. 678–683. Springer, Heidelberg (2006)
5. Herrera, F., Lozano, M.: Adaptive Genetic Algorithms based on Fuzzy Techniques. In: Proceedings of the Sixth International Conference on Information Processing and Management Uncertainty in Knowledge Based Systems, pp. 775–780. IEEE, Los Alamitos (1996)
6. Wang, Z., Bovik, A.C.: A universal image quality index. *IEEE Transactions on Signal Processing Letters* 9(3), 81–84 (2002)
7. Ross, T.J.: *Fuzzy Logic with Engineering Applications*. McGraw Hill, New York (1995)
8. Herrera, F., Lozano, M.: Adaptive Genetic Operators Based on Coevolution with Fuzzy Behaviours. *IEEE Transactions on Evolutionary Computation* 5(2), 149–165 (2001)
9. Lee, M.A., Takagi, H.: Dynamic Control of Genetic Algorithms using Fuzzy Logic Techniques. In: Proceedings of Fifth International Conference on Genetic Algorithms, Urbana – Champaign, IL, pp. 76–83 (1993)
10. Cordon, O., Herrera, F., Hoffmann, F., Magdalena, L.: Genetic Fuzzy Systems - Evolutionary Tuning and Learning of Fuzzy Knowledge Bases. In: *Advances in Fuzzy Systems — Applications and Theory*, vol. 19. World Scientific Publishing Co. Pte. Ltd., Singapore (2001)

Chain Code Histogram Based Facial Image Feature Extraction under Degraded Conditions

Soyuj Kumar Sahoo, Jitendra Jain, and S.R. Mahadeva Prasanna

Electro Medical & Speech Technology Lab, Indian Institute of Technology Guwahati,
Guwahati-781039, Assam, India

{soyuj,j.jitendra,prasanna}@iitg.ernet.in

Abstract. In this work, we have introduced a feature extraction method based on Chain Code Histogram (CCH) for facial images. A face verification (FV) system has been developed by using CCH feature. The performance of the above system is comparable with that of subspace analysis methods, i.e. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) under degraded condition. All the experimental results are shown upon a subset of IITG MV multi-biometric database which is a real time degraded office environment database. Finally, a better improved verification performance is reported by using the combination of both features which strongly validates the different information representation of both methods.

Keywords: Chain Code Histogram, Face Verification, Degraded Environment, Feature combination, PCA-LDA.

1 Introduction

Chain code histogram (CCH) is one of the most successful feature extraction technique in character recognition task. The directional information captured by the CCH is the key method in identifying the exterior information of any shape or pattern. Hence, it has been widely used in Japanese, Devnagari, Oriya, Arabic, handwritten character recognition applications with the higher recognition rate [1-5]. Here we have introduced the feature extraction technique based on the CCH method for facial image classification.

Face recognition is the most successful application of image processing. Face recognition is the problem of identifying three-dimensional (3D) face objects from two dimensional (2D) face images. Face recognition approaches can be widely classified as feature-based approaches and appearance-based approaches. However research is mainly concentrated in appearance-based approaches. This is because even though feature-based approaches are less sensitive to variations in illumination and viewpoint and to inaccuracy in face localization, the feature extraction techniques needed for feature-based approaches are not still reliable or accurate enough. Among the appearance-based methods, we use subspace methods like principal component analysis (PCA) or eigenfaces followed by linear discriminant analysis (LDA) or fisherfaces [6, 7]. Although PCA is extensively

used technique for dimension reduction, but it is not that much effective in facial pattern classification task. Hence, PCA followed by LDA is one of the classical template matching technique in face recognition.

By considering the importance of appearance-based methods in facial pattern classification, we have presented the feature extraction technique for face recognition task by computing the chain code histogram of edge representation of the facial images. We have also explored the dynamic directional information available in the skeleton image of face by taking the derivative of CCH feature. The different information contained in the PCA-LDA and CCH features are also exploited to make the face verification system more robust and accurate against degraded conditions. Here we have used the well known vector quantization (VQ) based nonparametric modeling and nearest neighbor classification technique for face verification.

The organization of the remaining work is as follows: Section 2 describes the face database used, Section 3 describes the feature extraction for face images using chain code histogram method. All the phases of CCH based face verification system is illustrated in Section 4. Experimental results and discussion are presented in Section 5 and finally, summary and future work are provided in Section 6.

2 Face Database

In this text, we make use of a subset of **IITG MV** multi-biometric database for testing and analysis of facial images [8]. It consists of a total of 94 subjects. Each subject has 40 images split into two sessions of 20 each. The images are mostly frontal with pose, tilt, scale, intensity and face feature variation. Each session is recorded as a video stream using a Logitech CMOS webcam in well illuminated rooms. No restriction is put on the subjects with respect to their posture, dress, facial expression, and hairstyle. The recording is done for a period of 10-20 seconds at a frame rate of 15 fps and a resolution of 640 x 480. Both sessions are recorded at two different periods of time and the database demonstrates many practical degraded-environmental effects as shown in Fig. 1.



Fig. 1. IITG MV face database sample images

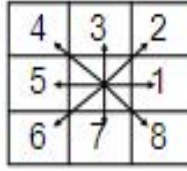


Fig. 2. Chain code directions (adopted from [3])

3 Chain Code Histogram (CCH) for Image Feature Extraction

CCH is an extensively used technique in character recognition tasks. Considering the importance of directional information captured by CCH the same approach we have used in our face verification task. Here we have used the block processing approach in CCH computation. To get the CCH representation of the input binary image, we take the pixel value 1 as the object point and pixel value 0 as the background point. If we consider a 3×3 window surrounded to any object point, then there are 8 directions (shown in Fig. 2) in which neighboring points may exist.

Let us consider a binary image $I_{M \times M}$ of size $M \times M$. Now if we go for block processing approach with block size of $r \times r$ then we get total $B = M^2/r^2$ no. of blocks. For CCH computation we use the following procedure:

1. **Step 1:** Consider an object point in a given block and find out the directional information of its surrounding points as shown in Fig. 2 by considering the following points:
 - If there is an object point in a particular direction then the feature value of that direction is incremented.
 - This procedure is repeated for 8 different directions. The feature indices corresponding to the directions, right side, right upper corner, upper side, left upper corner, left side, left lower corner, lower side, and right lower corner are $\{1, 2, \dots, 8\}$ respectively.
2. **Step 2:** The above step is repeated for all the object pixels of that block. Finally, it gives 8 feature values which represent total no. of object points in those eight directions for that particular block. It is termed as 8-dimensional CCH feature vector in this work.
3. **Step 3:** Step 1 and 2 are repeated for every block and finally, we get B no. of 8-dimensional feature vectors for every image.

The CCH representation of the symbol Σ is shown in Fig. 3. As the major object part of the symbol image Σ is in the horizontal directions, so we find higher values at feature index 1 and 5 i.e. more number of object pixels, in CCH representation of symbol Σ . We also get some object pixels at feature index 2, 4, 6, and 8 as the middle part of the image is in slanted directions. As the object

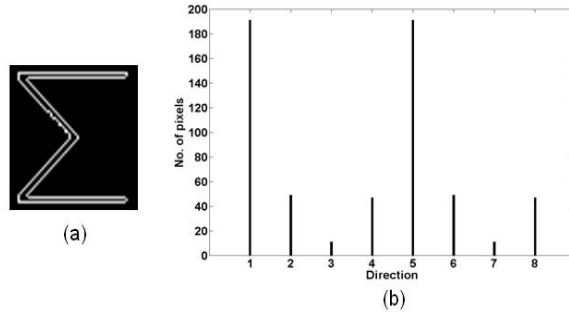


Fig. 3. Chain code histogram for the symbol Σ

part is almost absent in vertical directions so we get a few no. of object pixels at feature index 3 and 7 for the same as shown in Fig. 3(b). This clearly illustrates the effectiveness of CCH in directional information.

4 Face Verification System Using CCH Feature

Face verification (FV) system validates the identity claim of a person. A good FV should accept all the true claims and reject all the false claims. In practical scenario, some of the true trials may be rejected and some false trials may be accepted. The block diagram of CCH based feature extraction for facial images, used in the face verification system is shown in Fig. 4. The different stages used in the CCH based FV system are illustrated in the following sections.

4.1 Preprocessing

In the preprocessing stage, the incoming colored images are resized to a default dimension and converted to gray scale images as shown in Fig. 4. Before finding feature vectors for each image, histogram equalization is done both in training and testing.

4.2 Canny Edge Detection

The preprocessed gray scale image is fed to Canny edge detection algorithm for finding the skeleton image of the facial pattern [9]. Canny edge detection algorithm uses double-thresholding approach for finding the true maxima which makes it a robust method for edge detection. The outcome of the edge detection stage is the binary skeleton image as shown in Fig. 4.

4.3 CCH Feature Extraction

After getting the input binary skeleton image, the complete image is divided into some particular number of blocks depending on the block size. Now

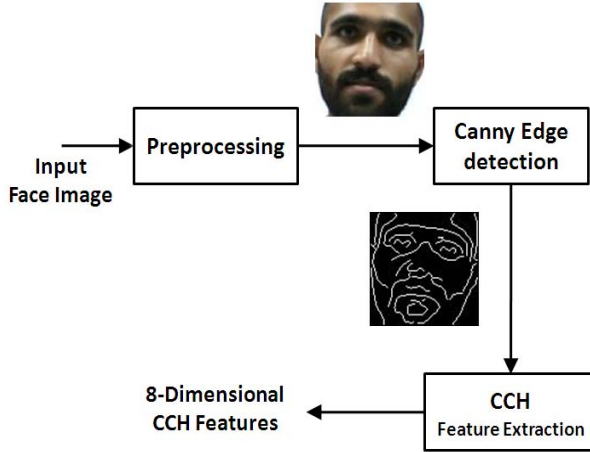


Fig. 4. CCH based feature extraction for facial images

8-dimensional CCH feature is computed for every block using the procedure as described in Sec. 3. Finally, we get same no. of 8-dimensional CCH feature vectors as the no. of blocks available.

4.4 Verification System

In the training phase, models are built for every individual user from the extracted features of their respective facial images. The extracted CCH features of a user are fed to vector quantization (VQ) modeling stage [10]. It builds a defined size of codebook for every user by using binary split and k -mean clustering algorithm.

In testing stage, after finding the test feature set by using the same CCH feature extraction technique, euclidean distances of this feature set are computed from all the code-vectors of the claimed model. Then the test image is classified by using nearest neighbor classifier which works on minimum euclidean distance criterion.

5 Experiments and Outcomes

A robust FV system should distinguish clearly between a positive claim and negative claim. In the mismatched condition the performance of a verification system is measured in terms of false rejection rate (FRR) and false acceptance rate (FAR). Equal error rate (EER) is defined as the error rate at which both FAR and FRR are equal. Hence we have evaluated the performance of FV system in terms of EER using detection error trade-off (DET) curve plot [11]. FAR and FRR are termed as false alarm probability and miss probability, respectively, in case of DET plot. In order to compare the performance of the proposed FV

system, the subspace based i.e. the PCA-LDA based, FV system is termed as the baseline system. To capture the dynamic nature of facial skeleton images, we have used first derivative of CCH feature, which is termed as Δ CCH [12]. Delta (Δ) of CCH is computed using two preceding and two succeeding feature vectors from the current feature vector. The performance of the FV system is computed by using CCH and Δ CCH features and also compared with the PCA-LDA based baseline FV system.

5.1 Experimental Details

For the present work, we consider 94 user set of IITG MV face database which includes 74 male users and 20 female users. For every user, twenty frames collected from the different segments of first session video stream are used in training stage. Ten different frames from the second session video stream of each user are used for testing. Therefore for 94 user set, there are in total 940 test images and each test image is tested against 11 random models, mentioned in detection list, out of which one is genuine model and rest are impostor models.

We have used 96×96 resized facial images for our verification task. For 8-dimensional CCH computation, we have considered block size of 16×16 . Hence, we get total 36 no. of 8-dimensional feature set for CCH and 32 no. of 16-dimensional feature set for Δ CCH in each image, respectively. In the VQ modeling stage, total 94 no. of 64-size codebooks are built for the complete user set.

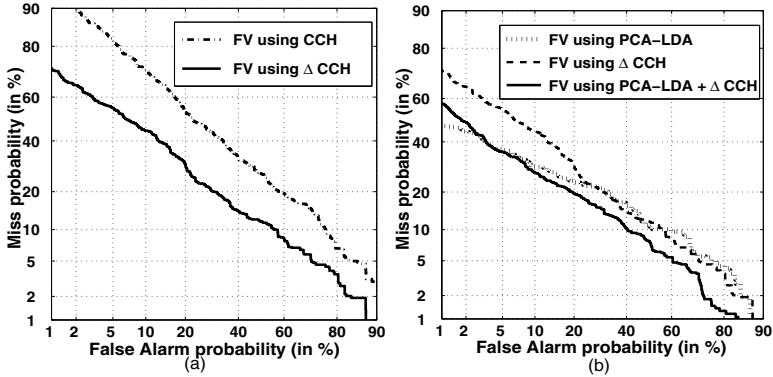
5.2 Experimental Results and Discussions

The performance of face verification system for different feature extraction techniques, in terms of EER is given in Table 1 and their corresponding DET curve plots are shown in Fig. 5. We can clearly observe from the Table 1 that the performance of Δ CCH is very much better than that of CCH based FV system in degraded conditions. In case of Δ CCH, the dynamic nature of skeleton image is captured in an effective manner by computing the first derivative of CCH feature which results a relative improvement of 20% in verification performance. It can be also observed from the DET plots as shown in Fig. 5 (a). The performance of the Δ CCH feature is also comparable with that of subspace based FV system, i.e. PCA followed by LDA. Hence, it shows the strong agreement with this point that dynamic directional information captured by Δ CCH feature effectively classifies the degraded facial patterns.

From the DET plot as shown in Fig. 5 (b), it is observed that the PCA-LDA and CCH methods may contain some different information about the facial patterns. Hence, the combination of both features (i.e. the multimodal approach) should give better performance with respect to their individual performances. This is also proved by the improved verification performance of proposed face verification system which is built using the score level fusion of PCA-LDA and Δ CCH, as shown in Table 1. This results in a relative improvement of 4.13% in EER which can also be observed from Fig. 5 (b). The performance of FV system for all the feature extraction techniques, those mentioned in this work,

Table 1. Performance of face verification system in terms of EER

Feature Extraction Techniques	Face Verification System Performance (in EER)
PCA-LDA	22.87
CCH	36.27
Δ CCH	23.52
PCA-LDA + Δ CCH	19.68

**Fig. 5.** DET curve plots of face verification system using different features; (a) CCH and Δ CCH FV, (b) PCA-LDA, Δ CCH, and PCA-LDA + Δ CCH FV

is not very good enough to implement in a practical scenario due to a lot of variations from training to testing data with respect to illuminance, pose, facial expression, camera distance etc. This is due to the degraded office environmental effects observed in our database as the video streams of facial images are collected in conversational style, which is shown in Fig. 11

6 Summary and Conclusions

We have presented chain code histogram based feature extraction technique for facial pattern classification. A face verification system is developed by using CCH and Δ CCH features and comparable verification performance is reported with that of subspace based methods. All the experimental results show the effectiveness of CCH feature for capturing the dynamic directional information available in skeleton facial images. The different information available in appearance based approach of facial images are exploited by the combination of PCA-LDA and CCH features in degraded conditions. The improved performance justifies the robustness of the proposed system in such a practical real time degraded database.

In future we may try to make the face verification system more robust and accurate for degraded environment by exploring different classification techniques. Future work may include to find a better feature representation of facial images by using chain code histogram method.

References

1. Kimura, F., Wakabayashi, T., Tsuruoka, S., Miyake, Y.: Improvement of handwritten Japanese character recognition using weighted direction code histogram. *Pattern Recognition* 30(3), 1329–1337 (1997)
2. Pal, U., Sharma, N., Wakabayashi, T., Kimura, F.: Off-Line Handwritten Character Recognition of Devnagari Script. In: *Proc. 9th Int. Conf. Document Analysis and Recognition (ICDAR)*, vol. 1, pp. 496–500 (2007)
3. Arora, S., Bhattacharjee, D., Nasipuri, M., Basu, D.K., Kundu, M.: Combining Multiple Feature Extraction Techniques for Handwritten Devnagari Character Recognition. In: *Proc. IEEE Region 10 and the 3rd Int. Conf. Industrial and Information Systems (ICIIS)*, pp. 1–6 (2008)
4. Roy, K., Pal, T., Pal, U., Kimura, F.: Oriya Handwritten Numeral Recognition System. In: *Proc. 8th Int. Conf. Document Analysis and Recognition (ICDAR)*, vol. 2, pp. 770–774 (2005)
5. Lawal, I.A., Abdel-Aal, R.E., Mahmoud, S.A.: Recognition of Handwritten Arabic (Indian) Numerals Using Freeman’s Chain Codes and Abductive Network Classifiers. In: *Proc. 20th Int. Conf. Pattern Recognition (ICPR)*, pp. 1884–1887 (2010)
6. Chen, L., Liao, H., Ko, M., Lin, J., Yu, G.: A new LDA-based face recognition system which can solve the small sample size problem. *Pattern Recognition* 33(10), 1713–1726 (2000)
7. Martinez, A., Kak, A.: PCA versus LDA. *IEEE Trans. Pattern Analysis and Machine Intelligence* 23(2), 228–233 (2001)
8. Sahoo, S.K., Prasanna, S.R.M.: Bimodal Biometric Person Authentication using Speech and Face under Degraded Condition. In: *Proc. National Conference on Communication, NCC* (2011)
9. Canny, J.: A Computational Approach To Edge Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence* 8(6), 679–698 (1986)
10. Linde, Y., Buzo, A., Gray, R.M.: An algorithm for vector quantizer design. *IEEE Trans. On Communications COM-28*, 84–95 (1980)
11. Martin, A., Doddington, G., Kamm, T., Ordowski, M., Przybocki, M.: The det curve in assessment of detection task performance. In: *Proc. Eur. Conf. Speech Communication Technology*, pp. 1895–1898 (1997)
12. Bimbot, F., Bonastre, J.F., Fredouille, C., Gravier, G., Chagnolleau, M.I., Meignier, S., Merlin, T., García, O.J., Delacretaz, P.D., Reynolds, D.A.: A tutorial on text-independent speaker verification. *EURASIP J. Appl. Signal Process.* 4, 430–451 (2004)

Object Recognition Based on Fuzzy Morphological Polynomial Signal Representation

Chin-Pan Huang¹, Ping S. Huang², Chaur-Heh Hsieh¹, and Tsorng-Lin Chia¹

¹ Department of Computer and Communication Engineering, Ming Chuan University

² Department of Electronic Engineering, Ming Chuan University,
333, Taoyuan, Taiwan
hcptw@mail.mcu.edu.tw

Abstract. Fuzzy morphological polynomial (FMP) signal representation provides the merits of compact representation and low computation complexity. Based on FMP representation, a theoretical study for object recognition is presented in this paper. The study indicates that superior properties of FMP can be used to develop an object recognition system. Examples are given to illustrate the effectiveness of FMP representation for object recognition.

Keywords: fuzzy mathematical morphology, object recognition, signal representation.

1 Introduction

Object recognition is a highly important task for image processing and computer vision. An object recognition system [1] can be divided into a training phase and a classification phase as shown in Fig. 1. The goal of object recognition is to match a test object with a template object represented by a vector of measured features or parameters from an object representation. The features are stored in a template library as references and used to be matched with the object identify [2]. Therefore, the key step is to extract the features from the objects and use them to recognize objects uniquely regardless of their location, scale or orientation.

Feature extraction techniques are generally categorized into local and global approaches. If the extracted features involve the whole domain of the object then it is global, otherwise it is local. Examples of global methods are invariant moments [3] and Fourier descriptors [4]. The drawbacks of using global features are intolerance to small distortions and closed object boundaries [5]. Local methods generally use the features of holes and points, line segments, curve segments [6] and sections [5]. Those techniques are less sensitive to distortions but are generally application dependent and computationally expensive. The general criterion used to discriminate the test objects is the concept of minimum distance [5] by using some types of norm. Here, we have reviewed two object recognition methods. Those two methods provide general rationale for using the combination of global and local information and using morphological methods to extract features from the object, respectively.

Liu and his coworkers [5] recently proposed a fuzzy pyramid-based invariant object recognition method. The pyramid data structure is combined with fuzzy

mathematics and the human visual system features are incorporated into their method. Annular and sector windows are used for image segmentation, which is close to the way the human vision system segments an image. They have shown that using annular and sector windows are feasible to maintain the invariant properties (i.e., translation, scaling and rotation). They have also proposed extracting the localized global features to overcome the disadvantages occurred by using only the local or global features. The zero-order moments are adopted as the global features and localized inside the annular and sector windows. Then the extracted features are used to build the fuzzy feature pyramid similar to the conventional feature pyramid. The major purpose is to reduce the computational cost from various image operations after applying the divide-and-conquer principle [7]. This method provides a proper combination of global and local features and avoids the disadvantage of using either feature alone.

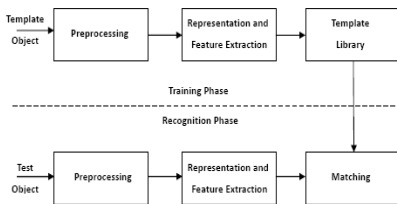


Fig. 1. A general object recognition system

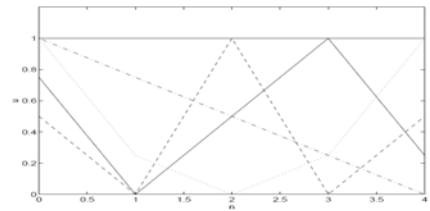


Fig. 2. Fuzzy structuring functions using DLO polynomials for $N=5$

Yu and Venetsanopoulous [6] proposed a local feature extraction method called Scatter Degree Technique (SDT). Geometrical features are extracted by morphological operators and used to measure the relationship between the peaks and the valleys of objects. After combining this technique with neural networks, good results can be achieved in object recognition. This method works on binary images and especially, for the recognition and classification of aircrafts.

Fuzzy morphological polynomial (FMP) signal representation provides the merits of compact representation and low computation complexity as shown in reference [8]. In this paper, a theoretical study for object recognition based on FMP representation is presented. The study indicates that superior properties of FMP can be used to develop an object recognition system. Examples are also given to illustrate the effectiveness of the representation for object recognition. The rest of the paper is organized as follows. The background knowledge including fuzzy mathematical morphology and FMP signal representation is presented in Section 2. FMP object recognition is investigated in Section 3. Examples for one- and two-dimension object recognition are illustrated in Section 4. Finally, conclusions are given in Section 5.

2 Background Knowledge

2.1 Fuzzy Mathematical Morphology

Recently, Sinha and Dougherty [9] proposed using the fuzzy set theory instead of the classical set theory to develop mathematical morphology. They have in fact, presented

a new approach of considering simultaneously binary and multilevel morphology. The concept of “umbra” is no longer needed to develop the multilevel case. Morphological operations are then applied on the “fuzzy” fitting so that for crisp sets the fitting still remains characterized as either 0 or 1, but for fuzzy or non-crisp sets it is possible to have a fitting characterized by a value between 0 and 1. The closer the value is to the unity, the better the fitting of the structuring element is. As in the classical morphology, fuzzy morphology can also transform a fuzzy set into another set. Such a transformation is performed by means of a fuzzy structuring set containing the desired geometric structure.

If we let X be the universe of discourse and x be its generic element, the difference between crisp and fuzzy sets is the characteristic function of a crisp set C defined as $\mu_c : X \rightarrow \{0,1\}$ while the membership function $\mu_f : X \rightarrow [0,1]$ of a fuzzy set F is defined so that $\mu_f(x)$ denotes the degree to which x belongs to the set F . Among those different operations on fuzzy sets, the following operations are described and will be used later:

- (a) *Complement* operation: $\mu_{F^c}(x) = 1 - \mu_F(x)$
- (b) *Translation* of a fuzzy set F by a vector $v \in X$: $\mu_{T(F,v)}(x) = \mu_F(x - v)$
- (c) *Reflection* of a set F : $\mu_{-F}(x) = \mu_F(-x)$
- (d) *Bold union* of two sets F and G :

$$\mu_{F\Delta G}(x) = \min[1, \mu_F(x) + \mu_G(x)] \tag{1}$$

- (e) *Bold intersection* of two sets F and G :

$$\mu_{F\nabla G}(x) = \max[0, \mu_F(x) + \mu_G(x) - 1] \tag{2}$$

The degree of fitting a set A into a set B is measured by an inclusion grade operator:

$$\begin{aligned} I(A,B) &= \inf_{x \in X} \mu_{A^c \Delta B}(x) \\ &= 1 + \min\{0, \inf_{x \in X} [\mu_B(x) - \mu_A(x)]\} \end{aligned} \tag{3}$$

where Δ is the bold union operator. According to the above index the degree of subsethood of two crisp sets A, B is either 0 or 1, while for fuzzy sets C and D $I(C,D) \in [0,1]$. Moreover, if $C \subseteq D$ then $I(C, D) = 1$ and in general $0 \leq I(C,D) \leq 1$. The erosion operation can be defined by using such an index [9] and the dilation, opening, and closing operators can then be obtained. In fact if $f(n)$ is a multilevel and $k(n)$ is a structuring element with supports of F and K and respective membership functions of $\mu_f(x)$ and $\mu_k(x)$ then the following four operations can be defined:

Erosion:
$$\begin{aligned} \mu_{f \ominus k}(n) &= I(T(k; n), f) \\ &= \min_{i \in K} \{ \min[1, 1 - \mu_k(i) + \mu_f(n+i)] \} \end{aligned} \tag{4}$$

Dilation:
$$\begin{aligned} \mu_{f \oplus k}(n) &= \mu_{(f^c \ominus k^c)^c}(n) \\ &= \max_{i \in K} \{ \max[0, \mu_k(i) + \mu_f(n-i) - 1] \} \end{aligned} \tag{5}$$

Opening: $\mu_{f \circ k}(n) = \mu_{(f \ominus k) \oplus k}(n)$

Closing: $\mu_{f \bullet k}(n) = \mu_{(f \oplus k) \ominus k}(n)$

2.2 Fuzzy Morphological Polynomial (FMP) Signal Representation

The FMP representation [8] uses fuzzy morphological opening to obtain a representation similar to a polynomial representation by means of a geometrical decomposition of the signal. One of the difficulties encountered in the process was the selection of the structuring functions, which can be either arbitrary or derived from the signal. In our case, they are chosen from a complete set of ordered real-valued orthogonal polynomials in $0 \leq n \leq N - 1$. In the examples, the discrete Legendre orthogonal (DLO) polynomials are used. Let $\mu_f(n)$ be the membership function of a given signal, and $\{\mu_{k_i}(n) : 0 \leq i \leq N - 1\}$ be one-dimensional fuzzy structuring functions, such that $0 \leq \mu_{k_i}(n) \leq 1$. Let $\{a_i\}$ be adaptive parameters used to make the fuzzy structuring function to be fitted with $\mu_f(n)$ closely. To take all possibilities into consideration, the fuzzy structuring functions $\{\mu_{k_i}(n)\}$ are derived from a shifted and normalized set of orthogonal polynomials $\{\mu_{g_i}(n)\}$ and its complementary functions $\{\mu_{g'_i}(n)\}$. Fig. 2 illustrates the shifted and normalized functions of $\{\mu_{g_i}(n)\}$ when the discrete Legendre orthogonal polynomial of $N = 5$ is considered.

The geometric decomposition for the given membership function $\mu_f(n)$ can be achieved recursively as follows:

- Windowing with $W(n)$:

$$\mu_{z_0}^v(n) = \mu_f(n) \times W(n - vN) \tag{6}$$

- Adaptive recursive approximation of $\mu_{z_0}^v(n)$:

$$\mu_{z_{i+1}}^v(n) = \mu_{z_i}^v(n) - \mu_{z_i \circ a_i k_i}^v(n) \tag{7}$$

where $i = 0, 1, \dots, N-1$, relates to the structuring functions of $a_i \mu_{k_i}(n)$; $v = 0, 1, 2 \dots$ refers to the window W , and a_i are adaptive parameters in $[0, 1]$. Each window is processed similarly.

The term $a_0 \mu_{k_0}(n)$ is essential to the above decomposition as it provides a coarse approximation to the signal membership function while $\{a_i \mu_{k_i}(n), i > 0\}$ gives the fine information of $\mu_{z_0}^v(n)$. After applying equation (8) recursively, we have

$$\mu_{z_0}^v(n) = \sum_{i=0}^{N-1} \mu_{z_i \circ a_i k_i}^v(n) + \mu_{z_N}^v(n) \tag{8}$$

where the last term corresponds to the residual or the part of the signal that cannot be well-represented by N function $\mu_{k_i}(n)$. The above representation can be considerably simplified by choosing values of $\{a_i\}$ such that

$$\mu_{z_i \circ a_i k_i}^v(n) = a_i \mu_{k_i}(n) \tag{9}$$

and equation (9) is converted into

$$\mu_{z_0}^v(n) = \sum_{i=0}^{N-1} a_i \mu_{k_i}(n) + \mu_{z_N}^v(n) \tag{10}$$

This representation is analogous to a polynomial representation of the windowed signal.

3 FMP Object Recognition

In this section, how the FMP signal representation can be used for object recognition is investigated. The object is assumed to be represented by a membership function unless specified otherwise. Let $\mu_f(n)$ and $\mu_g(n)$ be two given signal membership functions of two different objects and $\{\mu_{k_i}(n): 0 \leq i \leq N-1\}$ be one-dimensional ordered fuzzy structuring functions, based on orthogonal polynomials on $0 \leq n \leq N-1$, which are normalized so that $0 \leq \mu_{k_i}(n) \leq 1$. Let $\mu_{z_0}^v(n)$ and $\mu_{w_0}^v(n)$ be the signal membership functions of the v^{th} window of $\mu_f(n)$ and $\mu_g(n)$. $\mu_{z_N}^v(n)$ and $\mu_{w_N}^v(n)$ are the signal residue membership functions of the v^{th} window corresponding to $\mu_f(n)$ and $\mu_g(n)$. Let a_i^v and b_i^v be the fuzzy adaptive parameters of the v^{th} window for $\mu_{z_0}^v(n)$ and $\mu_{w_0}^v(n)$, respectively. If the FMP representations of two different objects are

$$\mu_{z_0}^v(n) = \sum_{i=0}^{N-1} a_i^v \mu_{k_i}(n) + \mu_{z_N}^v(n) \tag{11}$$

$$\mu_{w_0}^v(n) = \sum_{i=0}^{N-1} b_i^v \mu_{k_i}(n) + \mu_{w_N}^v(n) \tag{12}$$

then $\sum_{i=0}^{N-1} a_i^v \mu_{k_i}(n)$ and $\sum_{i=0}^{N-1} b_i^v \mu_{k_i}(n)$ represent the gross approximation of two objects, and $\mu_{z_N}^v(n)$ and $\mu_{w_N}^v(n)$ represent the detailed variations of individual objects. If the objects are simple and have the same gross approximations for all windows, and their detailed variations for all windows are insignificant, then those two objects can be said to resemble each other. The following properties can be easily proven and give the idea of how FMP can be used for object recognition.

Property 1. If two windowed objects of $z_0(n)$ and $w_0(n)$, $0 \leq n \leq N-1$ are related by $z_0(n) = s w_0(n)$, where s is a scaling factor. Then the linearized and fuzzified objects are equal i.e., $\mu_{z_0}^v(n) = \mu_{w_0}^v(n) \forall n$. This property establishes that the FMP would be invariant to scaling factors. Using the FMP method, the objects under consideration are linearized and fuzzified objects, so that they are invariant to scaling factors.

Property 2. For two windowed objects of $\mu_{z_0}(n)$, $0 \leq n \leq N-1$, and $\mu_{w_0}(n)$, $0 \leq n \leq sN-1$, where s is a scaling factor, related as $\mu_{z_0}(n) = \mu_{w_0}(sn)$, $n \in [0, N-1]$, then we have the adaptation coefficients such that $a_i = b_i^s$, $i = 0, \dots, N-1$ and $\mu_{z_N}^v(n) = \mu_{w_N}^v(sn)$, $n \in [0, N-1]$, where $b_i^s = \min_{\substack{0 \leq l \leq N-1 \\ \mu_{k_i}(l) \neq 0}} \left\{ \frac{\mu_{w_i}(sl)}{\mu_{k_i}(l)} \right\}$. This property shows that

we can uniformly take same number of samples and same number of frames with same number of samples in each frame for every object to overcome the problem of different supports in the FMP representation of considering objects.

Property 3. If two windowed objects of $\mu_{z_0}(n)$ and $\mu_{w_0}(n)$, $0 \leq n \leq N-1$ have the relationship $\mu_{z_0}(n) = \mu_{w_0}(n)$, $\forall n$, then the coefficients are such that $a_i = b_i$ for $i=0, \dots, N-1$. This relationship can also be applied to $\mu_{z_N}(n)$ and $\mu_{w_N}(n)$, $\forall n$, and vice versa. This property shows the identity conditions of two objects. The FMP coefficients and its residuals are required to recognize the object uniquely. If the residuals $\mu_{z_N}(n) = \mu_{w_N}(n)$, $\forall n$, are insignificant, then the FMP coefficients can be used to recognize the object.

Property 4. If two windowed objects of $\mu_{z_0}(n)$ and $\mu_{w_0}(n)$, $0 \leq n \leq N-1$ have the relationships of $\mu_{z_0}(n) = \mu_{w_0}(n-t)$, $\forall n$, and a t , then $a_i = b_i$ for $i=0, \dots, N-1$. This relationship can also be applied to $\mu_{z_N}(n) = \mu_{w_N}(n-t)$, $\forall n$, and a t and vice versa. This property establishes that if two objects have the same starting point, then the FMP coefficients and residuals of those two objects are same.

Property 5. If two windowed objects of $\mu_{z_0}(n)$ and $\mu_{w_0}(n)$, $0 \leq n \leq N-1$ have the relationships of $\mu_{z_0}(n) = \mu_{w_0}(-n)$, $\forall n$, then the corresponding FMP coefficients $a_i = b'_i$ for $i=0, \dots, N-1$ and $\mu_{z_N}(n) = \mu_{w_N}(-n)$, $\forall n$. The coefficients of a_i, b'_i and the residues of $\mu_{z_N}(n)$ and $\mu_{w_N}(-n)$ are obtained by using the structuring function and its reflected version, respectively. This property establishes the reflection invariant property for the FMP method.

Those properties indicate that the FMP representation can be used for recognizing object uniquely up to a space invariant factor. Those properties can be used to develop an object recognition system.

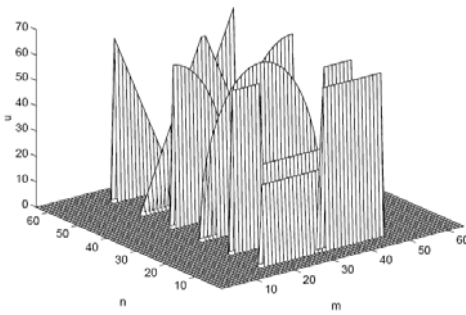


Fig. 3. One-dimensional test objects

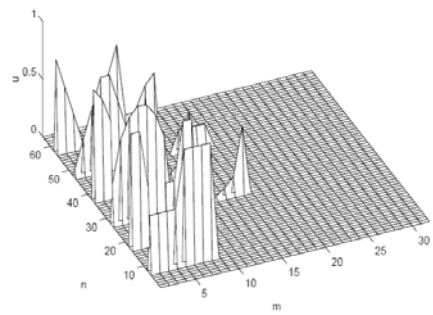


Fig. 4. FMP coefficients

4 Examples

In this section, one- and two-dimensional examples are considered to illustrate the FMP properties. For one-dimensional (1D) object examples, the objects shown in Fig. 3 are used as the template objects. The test objects are the scaled-version of the template objects. FMP representations are applied in the discrimination process for the objects in Fig. 3. Each object is divided into frames with each containing several

points. The starting point is properly chosen and the same structuring function is used for the template and test objects. The FMP coefficients for those 6 objects are calculated and shown in Fig. 4. To measure the similarity of the template object and the test object, with coefficients a_i^v and b_i^v for each frame, we use the distance measure given by

$$d(a,b) = \left(\sum_{v=0}^{M-1} \sum_{i=0}^{N-1} |a_i^v - b_i^v|^2 \right)^{1/2} \tag{13}$$

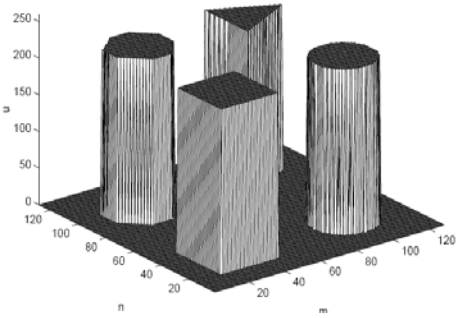


Fig. 5. Two-dimensional test objects

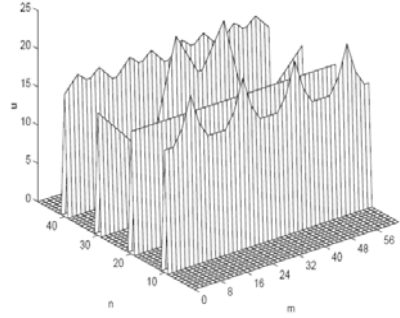


Fig. 6. Transformed one-dimension objects

Table 1. The distances between two objects (1D)

d	Sqr.	S. Sqr.	Circ.	S. Cir.	Tri.	S. Tri.
Sqr.	0.00	1.00	1.40	1.23	1.43	1.43
S. Sqr.	1.00	0.00	1.65	0.87	1.74	1.02
Circ.	1.40	1.65	0.00	1.89	0.91	1.93
S. Cir.	1.23	0.87	1.89	0.00	1.93	0.91
Tri.	1.43	1.74	0.91	1.93	0.00	1.67
S. Tri.	1.43	1.02	1.93	0.91	1.67	0.00

The distances between those template and test objects are shown in Table 1. The column shows the template object and the row shows the test object. It has also shown that each object is discriminated by its FMP's zeroth order coefficients, that is, only the zeroth order coefficients (i.e., $a_0^v, b_0^v, v=0, 1, 2, \dots, M-1$) are used to represent the objects. For two-dimensional (2D) object examples, we can use the above one-dimensional method if we proceed as follows. At first, corresponding binarized images can be obtained by using different threshold values. Then we compute the centroid point of each binarized image and then trace the object boundary to get the radius ($r(n), 0 \leq n \leq L-1$) at each sampled point for the template and test objects. The radii of the sampled points form a one-dimensional object in which we can apply the FMP method as in the previous example. It is convenient to let L be a power of 2 for two-dimensional objects in the same way as for the number of points in one-dimensional objects. As an example, we take the objects in Fig. 5 as test objects. For the threshold value of 0.5 and $L=64$, the results of the transformed

one-dimensional objects are shown in Fig. 6. The same procedures as those performed in one-dimensional objects are done on those objects. The distance is calculated and shown in Table 2. The results show that those objects can be correctly recognized by the FMP method. We also compute the distance between two objects using the zeroth order coefficients only. The results show that those objects can also be correctly recognized. Please note that the starting sampling point in the example is normalized.

Table 2. The distances between two objects (2D)

d	Square	Circle	Triangle	8 Sided
Square	0.00	0.77	0.98	0.47
Circle	0.77	0.00	1.58	0.33
Triangle	0.98	1.58	0.00	1.28
8 Sided	0.47	0.33	1.28	0.00

5 Conclusions

In this paper, a theoretical basis for object recognition based on FMP has been presented. The properties of the FMP representation indicate that the FMP can be used to develop an object recognition system. Examples have been provided to illustrate the effectiveness of the FMP for object recognition.

Acknowledgments. This work is supported in part by National Science Counsel of Taiwan, No. NSC 99-2632-E-130 -001 -MY3.

References

1. Suetens, P., Fua, P., Hanson, A.J.: Computational strategies for object recognition. *ACM Computing Surveys* 24, 5–61 (1992)
2. Avi-Itzhak, H.I.: Multiple subclass pattern recognition: a maximin correlation approach. *IEEE Trans. on Pattern Analysis Mach. Intell.* 17, 418–430 (1995)
3. Bebis, G.N., Papadourakis, G.M.: Object recognition using invariant object boundary representations and neural network. *Pattern Recognition* 25, 25–44 (1992)
4. Persoon, E., Fu, K.S.: Shape discrimination using fourier descriptors. *IEEE Trans. on Systems, Man, and Cybernetics* 7, 170–179 (1977)
5. Liu, X., Tan, S.: Fuzzy pyramid-based invariant object recognition. *Pattern Recognition* 27, 25–44 (1994)
6. Yu, P., Venetsanopoulos, A.N.: Partial pattern recognition and classification using the scatter degree technique and neural network. *Journal of Intelligent and Robotic Systems* 5, 271–282 (1992)
7. Gonzales, R., Woods, R.E.: *Digital Image Processing*, 3rd edn. Prentice-Hall, Englewood Cliffs (2008)
8. Huang, C.P., Chaparro, L.F.: Fuzzy morphological polynomial image representation. *Eurasip Journal on Advances in Signal Processing* 2010, Article 914921 (2010)
9. Shinha, D., Dougherty, E.R.: Fuzzy mathematical morphology. *Journal of Visual Communication and Image Processing*, 286–302 (1992)

Face Detection for Skin-Toned Images Using Signature Functions

H.C. VijayLakshmi and Sudarshan PatilKulkarni

S.J. College of Engineering ,
Mysore, India
vijisjce@yahoo.co.in

Abstract. In this paper, notion of signature function is introduced for a segmented image with face and non-face regions. Then classification is achieved by considering the correlation of signature functions of the template and the segmented image. Proposed method is computationally elegant and results are satisfactory under the constraint of those costs.

Keywords: Signature function, classification, correlation, detection.

1 Introduction

Face detection and localization is the task of checking whether the given input image or video sequence contains any human face, and if human faces are present, returning the location of the human face in the image. The face detection problem involves segmentation, extraction, and classification of the segmented and extracted region as face or non face irrespective of the background. Several researchers are at this task with different approaches, so that the machine detects and locates the faces efficiently as we human beings do in any complex scenarios.

The faces play a major role in identifying and recognizing people. The area of face detection has gained considerable importance with the advancement of human-machine interaction as it provides a natural and efficient way to communicate between humans and machines. Face detection and localisation in image sequences has become a popular area of research due to emerging applications in intelligent human-computer interface, surveillance systems, content-based image retrieval, video conferencing, financial transaction, forensic applications, and many other fields. Face detection is essentially localising and extracting a face region from the background. This may seem like an easy task but the human face is a dynamic object and has a high degree of variability in its appearance [10]. A large number of factors that govern the problem of face detection [4], [6]. The long list of these factors include the pose, orientation, facial expressions, facial sizes found in the image, luminance conditions, occlusion, structural components, gender, ethnicity of the subject, the scene and complexity of image's background. Faces appear totally different under different lighting conditions. A thorough survey of face detection research work is available in [4], [6]. In terms of applications, face detection and good localization is an important pre-processing step in online face recognition. A number of techniques

have been developed by researchers in order to efficiently detect human faces in any given input image or video sequence.

There are several challenges while detecting and locating faces in skin toned regions, refer [9]. In spite of using combination of different colour spaces during segmentation, it is tedious to demarcate region boundaries between skin and pseudo skin regions and also eliminate these regions from searching process. The use of colour space alone sometimes fails to segment the boundary regions of the image. In order to overcome this problem combination of colour spaces for efficient skin segmentation followed by Canny and Prewitt edge detection to demarcate the region boundary is used for input image segmentation [9]. Due to variation in illumination, skin regions may not be properly identified as skin during skin segmentation. Locating faces in these circumstances is more complex as opposed to localizing faces with uniform, non skin-tone background.

Face detection is a classification problem. It calls for classifying the selected segmented region as face or non-face. For efficient classification a robust feature set as well as robust classification method is very much essential. This calls for an efficient feature extraction method followed by efficient classification method. Edges play an important role in extracting prominent facial features. For a detailed analysis of various edge detection algorithms, refer [2], [3].

2 Preprocessing

Preprocessing step is the stage where the input image is processed to extract the regions of interest. The main task of this stage is to avoid processing regions which are not the candidate regions for the algorithm to process and eliminate these regions from further processing.

Segmentation and feature extraction are the two important pre-processing steps that play a vital role in face detection and localisation Segmentation is a process that partitions an image into regions [2] of uniform texture and intensity. In the problem of face detection, segmentation based on skin plays a major role in identifying the candidate face regions [9]. Though there are different segmentation methods, segmentation based on colour is considered. For a detailed survey of colour spaces refer [7].

Segmentation of the input image is the most important step that contributes to the efficient detection and localization of multiple faces in skin tone colour images Skin segmentation of the input image based on skin pixel cue [1] combined with edges help to demarcate region boundaries and segment the image components efficiently [9].

Segmentation of an image based on human skin chromaticity using different colour spaces results in identifying even pseudo skin like regions as skin regions. Hence there is a need for further eliminating these pseudo skin regions. The segmentation using combination of colour spaces combined with Canny and Prewitt edge detection for obtaining the region boundaries segment better [9].

Feature extraction is an important step in pattern classification. Face detection is a two class pattern classification problem classifying whether the given test pattern

belongs to face class or non face class. Feature extraction stage is responsible for finding a set of prominent and robust attributes that would help to correctly decide the object classification boundary.

Facial features such as eyebrows, pupils, and lips appear generally darker than their surrounding facial regions and the edges associated with these regions provide a good feature set in detecting faces. For a detailed study and comparison of edge detecting techniques refer [3].

Wavelets are mathematical functions that initially cut up the data into different frequency components and then study each component with a resolution matched to its scale. Thus wavelets analyze according to scale and self-similarity caused by scales and dilations. Wavelet algorithms process data at different scales or resolutions. The wavelet analysis procedures adopt a wavelet prototype function called an analyzing wavelet or mother wavelet [5]. Prominent facial features are extracted using wavelets and different edge detection techniques. Wavelet decomposition of face images followed by edge extraction for approximation image results in prominent facial edges, which is evident from the images of the facial edges extracted for raw face image “Fig 1a” and “Fig 1b”, edges extracted for the approximation images after first and second level wavelet decomposition “Fig 1c” and “Fig 1d” respectively. The edges extracted for the approximation images eliminate unwanted noisy edges present in the image and reduces the dimension after wavelet decomposition. The column sum of noise free prominent facial edge images of frontal faces, when plotted produce definite facial edge signature “Fig. 2”. This signature function serves as a vital feature in classifying the image as face or non-face.



Fig. 1a. Input Face image



Fig. 1b. Edges of raw input image



Fig. 1c. Edges of first level approximation **Fig. 1d.** Edges of second level approximation

3 Proposed Algorithm

The proposed algorithm is implemented in two stages namely feature extraction and classification stage. For feature extraction frontal faces containing only prominent facial features such as eyes, nose and mouth which fit in to a window of size 80X72 are considered in this experiment.



Fig. 2. Mean Facial Edge Image

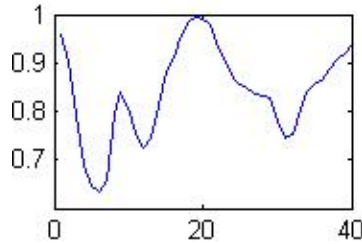


Fig. 3. Mean Face Edge Signature

3.1 Feature Extraction Stage

Step-1: Wavelet decompose each histogram equalized face image containing prominent facial features with slight variation in pose and expression, retain only approximation image and discard the horizontal, vertical and diagonal details .

Step-2: Apply Robert and Sobel edge extraction algorithm separately on the approximation images of each image considered for feature extraction.

Step-3: Combine the edges of the approximation image obtained in Step-2, into a single edge image using pixel by pixel image multiplication. On this edge image perform morphological operations such as erosion and dilation to extract only the prominent facial edge features and remove noisy edges.

Step-4: The extracted prominent facial edges called wavelet edges are stacked into a data edge matrix as column vectors.

Step-5: Compute the mean facial edge image (Fig. 2) of the frontal faces from the wavelet edge matrix generated in **step-4**.

Step-6: Compute the column mean of the mean facial edge image obtained in step-5 and plot the column mean to obtain the signature function of the mean facial edge images shown in Fig. 3.

The correlation coefficient of signature function of each approximation edge face image used for feature extraction with the mean facial edge image signature function is computed to fix up the correlation threshold T1. Also compute the two dimensional correlation coefficient of each approximation window image with the mean facial edge image of Fig. 2. This threshold is fixed considered T2. The following formula is used to compute the correlation.

$$Corr(x, y) = \frac{E[(x - \mu_x)(y - \mu_y)]}{\sigma_x * \sigma_y}$$

where x is the Mean Facial Edge vector and y is the facial edge vectors.

3.2 Testing Phase

Step-1: The given input colour image is skin segmented using the approach proposed by authors in [8]. For skin segmented regions with size larger than the window size considered, use sliding window technique proposed in [8].

Step-2: Histogram equalize each window image followed by wavelet decomposition. Retain only approximation image. Extract the edges of the approximation image using Step-2 and Step-3 of feature extraction phase.

Step-3: For each window approximation edge image compute the column mean obtained in **step-6** of feature extraction phase and plot the vector to generate the signature function.

Step-4: Find the correlation coefficient between the signature functions of column mean of each window image of **step-3** and column mean facial edge image obtained during feature extraction. Also the two dimensional correlation coefficient of each window approximation edge image with the mean facial edge image is computed. The positive correlation values above the threshold values indicate the presence of a face.

4 Results

In this method a window of size 80X70 was used. Color images containing multiple faces are segmented and tested with this algorithm. Around 3000 window contents were checked. Fifty of them were found to be false positives. There were no false negatives. The skin segmented input image is shown in Fig 4a, Fig 4c is the signature function of the window capturing the frontal face shown in Fig. 4b.



Fig. 4a. Skin segmented image



Fig. 4b. Sliding window image

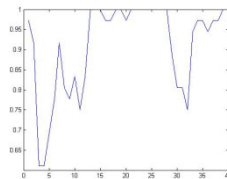


Fig. 4c. Signature function of figure 4b

Table 1.

Method	Distance Measure	Windows Tested	False Positive	False Rejection
Proposed Method	correlation	2000	50	0
Approximation image with Gray Scale Edges	Cityblock	1700	44	5
Approximation Image without Edge Extracion	Bhattacharya Distance	1500	128	20
	Cityblock	1500	44	38

5 Conclusion

The results show that, the proposed method significantly reduces the false rejection rate while the false acceptance rate remains as good as other methods [ref]. Proposed method is computationally elegant since a segmented image $n \times n$ generates a unidimensional signature function of size n and is correlated with a function of same size. however in terms of recognition rate the proposed algorithm has limitations due to significant false acceptance rate.

References

1. Fritsch, F., Lang, S., Kleinhagenbrock, M., Fink, G.A., Sagerer: Improving Adaptive Skin Colour Segmentation by incorporating Results from face detection. In: Proceedings of IEEE International Workshop on Robot and Human Interactive Communication (ROMAN), Berlin, Germany, pp. 337–343 (September 2002)
2. Gonzalez, R.C., Woods, R.E.: Digital Image processing, 2nd edn. Prentice Hall India, Englewood Cliffs
3. Maini, R., Aggarwal, H.: Study and Comparison of Various Image Edge Detection Techniques. International Journal of Image Processing (IJIP) 3(1), 1–11 (2009)
4. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.: Face recognition: A literature survey. ACM Computing Surveys, 399–458 (2003)
5. Frazier, M.W.: An Introduction to Wavelets through Linear algebra. Springer, Heidelberg (1999)
6. Ming-Hsuan, Y., Kreigman, D.J., Ahuja, N.: Detecting Faces in images: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 24, 34–58 (2002)
7. Veznevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: Proceedings Graphicon-2003, pp. 85–92 (2003)

8. VijayLakshmi, H.C., PatilKulakarni, S.: Face Detection and Localization in Skin Toned Color Images Using Wavelet and Edge Detection Techniques. In: International Conference on Advances in Recent Technologies in Communication and Computing 2604-7/06, pp. 231–234. IEEE, Los Alamitos (2010), doi:10.1109/ARTCom.2010.34, 978-0-7695-4201-0/10 \$26.00 © 2010
9. Vijaylakshmi, H.C., PatilKulakarni, S.: Segmentation Algorithm for Multiple Face Detection for Color Images with Skin Tone Regions. In: 2nd International Conference on Signal Acquisition and Processing, Bangalore, February 9-10, pp. 162–166 (2010)
10. Petrescu, D., Gelgon, M.: Face detection from complex Scenes in colour images. *Eurasip /Proceedings/Eusipco2000*, <http://www.eurasip.org/Proceedings/Eusipco/Eusipco2000/sessions/WedAm/PO4/cr1871.pdf>

Recurrent Neural Network Based Phoneme Recognition Incorporating Articulatory Dynamic Parameters

Mohammed Rokibul Alam Kotwal¹, Foyzul Hassan¹, Md. Mahabubul Alam¹,
Abdur Rahman Khan Jehad², Md. Arifuzzaman², and Mohammad Nurul Huda¹

¹ United International University
Dhaka, Bangladesh

² The University of Asia Pacific
Dhaka, Bangladesh

{rokib_kotwal, ripon_wasid, arif_uap}@yahoo.com,
{foyzul.hassan, jehadbd}@gmail.com, mnh@cse.uiu.ac.bd

Abstract. This paper describes a recurrent neural network (RNN) based phoneme recognition method incorporating articulatory dynamic parameters (Δ and $\Delta\Delta$). The method comprises three stages: (i) DPFs extraction using a recurrent neural network (RNN) from acoustic features, (ii) incorporation of dynamic parameters into a multilayer neural network (MLN) for reducing DPF context, and (iii) addition of an Inhibition/Enhancement (In/En) network for categorizing the DPF movement more accurately and Gram-Schmidt orthogonalization procedure for decorrelating the inhibited/enhanced data vector before connecting with a hidden Markov models (HMMs)-based classifier. From the experiments on Japanese Newspaper Article Sentences (JNAS), it is observed that the proposed method provides a higher phoneme correct rate over the method that does not incorporate dynamic articulatory parameters. Moreover, it reduces mixture components in HMM for obtaining a higher performance.

Keywords: Distinctive Phonetic Feature, Multi-Layer Neural Network, Recurrent Neural Network, Inhibition/Enhancement Network, Local Features.

1 Introduction

Articulatory features (AFs) or distinctive phonetic features (DPFs) play an important role in automatic speech recognition (ASR) [1-3]. These features provide a higher word recognition performance in speech recognition in clean and noise corrupted acoustic environment [4-5]. Moreover, a higher phoneme recognition performance in different acoustic environments is also achieved using these features [6-7].

The reason for providing a better recognition performance is the generation of wide margin of acoustic likelihood between two phonemes, which is not affected much by the noisy environments. Besides, these methods incorporated context window of limited size instead of using context sensitive triphone models, which requires a large scale speech corpus and a large number of speech parameters, to resolve coarticulation effects. The context window in multilayer neural network (MLN)-based

speech recognition system reduces coarticulation effect slightly and consequently, provides a reasonable performance at fewer mixture components in the hidden Markov models (HMMs).

On the other hand, a recurrent neural network (RNN) having feedback connections models a context window into several number of frames and shows a better performances [8]. These performances were further improved slightly by adding an MLN in the method proposed by us [9], which reduces DPF fluctuations in phoneme boundaries. The reason not for obtaining a higher performance improvement further is that the second stage MLN has an inability of handing longer context.

In this paper, we propose a phoneme recognition method, which incorporates dynamic articulatory parameters (Δ and $\Delta\Delta$) at second stage, to reduce coarticulation effect further. The method comprises three stages: (i) DPFs extraction using a recurrent neural network (RNN) from acoustic features, (ii) incorporation of dynamic parameters into a multilayer neural network (MLN) for constraining the context, and (iii) addition of an Inhibition/Enhancement network (In/En) network for categorizing the DPF movement more accurately and Gram-Schmidt (GS) orthogonalization procedure for decorrelating the inhibited/enhanced data vector before connecting with HMMs-based classifier. The specialty of this paper is the incorporation of dynamic articulatory parameters to solve the coarticulation effect further.

The paper is organized as follows: Section 2 discusses the articulatory features. Section 3 explains the system configuration of the existing method with the proposed. Experimental database and setup are provided in Section 4, while experimental results are analyzed in Section 5. Finally, Section 6 draws some conclusion and remarks on future works.

2 Distinctive Phonetic Features

A phone can easily be identified by using its unique articulatory features or distinctive phonetic features (DPFs) set [10-11]. The Japanese balanced DPF set [4] for classifying Advanced Telecommunications Research Institute International (ATR) phonemes have 15 elements. These DPF values are mora, high, low, intermediate between high and low <nil>, anterior, back, intermediate between anterior and back <nil>, coronal, plosive, affricate, continuant, voiced, unvoiced, nasal and semi-vowel. Table 1 shows a part of this balanced DPF set. Here, present and absent elements of the DPFs are indicated by “+” and “-” signs, respectively.

Table 1. Japanese balanced DPF-set

DPF/Phone	a	e	...	f	r
mora	+	+	...	-	-
high	-	-	...	+	-
low	+	-	...	-	-
nil	-	+	...	-	-
anterior	-	-	...	+	+
back	+	-	...	-	-
nil	-	+	...	-	-
			...		

3 Phoneme Recognition Methods

3.1 The Existing Method [9]

This method comprises two neural networks: (i) RNN and (ii) MLN, which is called hybrid neural network (HNN) and shown in Fig.1. The RNN represents dynamics in a sequence of acoustic features to resolve coarticulation effects and the MLN reduces fluctuation of DPF patterns. The external input acoustic vector at time t , for the RNN, is formed by taking preceding $(t-3)$ -th and succeeding $(t+3)$ -th frames together with the current t -th frame. Each frame is composed of 25 local features (LFs) [12] that are same as the DPF-based phoneme recognition using MLN [4]. The RNN outputs 45 DPF values of which 15 are for the preceding frame, 15 for the current frame, and the rest for the succeeding frame. Next, the MLN outputs 45 DPF values for the current input frame by reducing DPF fluctuation.

Then, the 45 dimensional DPF vector outputted by the MLN are inserted into In/En network, which will be described in Section 3.2.2, to obtain categorical DPF movements and next, the inhibited/enhanced data vector are decorrelated with each other by using the GS orthogonization procedure [9] before connecting with an MLN.

A fully recurrent neural network (FRNN), which has a hidden layer of 350 units and an output layer, is used for this approach. Each time total input vector is formed by taking the output layer (OL) feedback values and the hidden layer (HL) feedback values together with the external input (25×3) LF values of that time. The feedback values of the hidden layer and the output layer at time t_0 are assumed to be 0.1. The back-propagation through time algorithm is used for training the RNN. Again, the MLN has three layers including two hidden layers and an output layer, and is trained by using the standard back-propagation algorithm. The hidden layers are of 180 and 90 units, respectively.

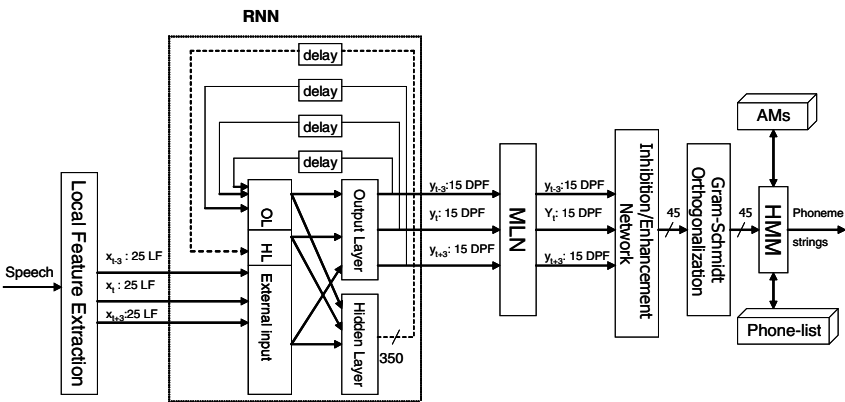


Fig. 1. The existing method [9] without articulatory dynamic parameters

3.2 Proposed Method

The proposed method diagram is depicted in Fig.2 and comprises three stages: (i) DPFs extraction using a recurrent neural network (RNN) from acoustic features, (ii) incorporation of dynamic parameters into a multilayer neural network (MLN) for constraining the context, and (iii) addition of an Inhibition/Enhancement network (In/En) network for categorizing the DPF movement more accurately and Gram-Schmidt (GS) orthogonalization procedure for decorrelating the inhibited/enhanced data vector before connecting with HMMs-based classifier.

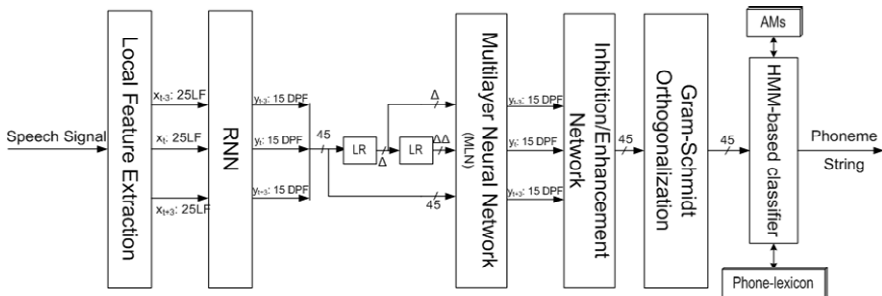


Fig. 2. Proposed method with articulatory dynamic parameters

3.2.1 DPF Extractor

Kindly The RNN, which has same architecture and learning mechanism described in Section 3.1, generates a 45 dimensional DPF vector (15 DPFx3) for the current input frame t . The 45-dimensional context-dependent DPF vector provided by the RNN at time t , and its corresponding Δ and $\Delta\Delta$ vectors calculated by three-point linear regression (LR) are appended into the subsequent MLN with four layers including two hidden layers of 300 and 100 units, respectively. The MLN_{dyn} is trained using the standard back-propagation algorithm and outputs a 45-dimensional DPF vector in which context effects for the current " t "-th frame are reduced.

3.2.2 Inhibition/Enhancement Network

The In/En network is used to obtain modified DPF patterns from the patterns produced by the RNN+MLN. The algorithm for this network is given below:

Step1: For each element of the DPF vectors, find the acceleration ($\Delta\Delta$) parameters by using three-point LR.

Step2: Check whether ($\Delta\Delta$) is positive (concave pattern) or negative (convex pattern) or zero (steady state).

Step3: Calculate $f(\Delta\Delta)$

if pattern is convex,

$$f(\Delta\Delta) = \frac{c_1}{1 + (c_1 - 1)e^{\beta\Delta\Delta}}$$

if pattern is concave,

$$f(\Delta\Delta) = c_2 + \frac{2(1 - c_2)}{1 + e^{\beta\Delta\Delta}}$$

if steady state,

$$f(\Delta\Delta) = 1.0$$

Step4: Find modified DPF patterns by multiplying the DPF patterns with $f(\Delta\Delta)$.

4 Experiments

4.1 Speech Database

The following two clean data sets are used in our experiments:

D1. Training data set

A subset of the Acoustic Society of Japan (ASJ) Continuous Speech Database comprising 4503 sentences uttered by 30 different male speakers (16 kHz, 16 bit) is used [13].

D2. Test data set

This test data set comprises 2379 JNAS [14] sentences uttered by 16 different male speakers (16 kHz, 16 bit).

4.2 Experimental Setup

The frame length and frame rate (frame shift between two consecutive frames) are set to 25 ms and 10 ms, respectively, to obtain acoustic features from an input speech. LFs are a 25-dimensional vector consisting of 12 delta coefficients along time axis, 12 delta coefficients along frequency axis, and delta coefficient of log power of a raw speech signal [12].

Phoneme correct rate (PCR) for D2 data set is evaluated using an HMM-based classifier. The D1 data set is used to design 38 Japanese monophone HMMs with five states, three loops, and left-to-right models. In the HMMs, the output probabilities are represented in the form of Gaussian mixtures, and diagonal matrices are used. The mixture components are set to 1, 2, 4, 8, and 16.

In our experiments of the RNN and MLN, the non-linear function is a sigmoid from 0 to 1 ($1/(1+\exp(-x))$) for the hidden and output layers.

For the In/En network, $C1$, $C2$, and β are set to 4.0, 0.25, and 80, respectively.

To evaluate PCRs using D2 data set for observing the effects of articulatory dynamic parameters (Δ and $\Delta\Delta$), the following six experiments are designed, where input features for HMM-based classifier are DPFs of 45 dimensions for the existing and proposed methods.

- (1) DPF (RNN+Not- Δ .MLN, dim: 45)
- (2) DPF (RNN+Not- Δ .MLN+GS, dim: 45)
- (5) DPF (RNN+ Δ .MLN,dim:45)
- (9) DPF (RNN+Not- Δ .MLN+In/En+GS, dim: 45)
- (i) DPF (RNN+MLN +GS, dim: 45)
- (q) DPF (RNN+MLN+In/En+GS,dim: 45) [Proposed].

5 Experimental Results and Analysis

Figs. 3 and 4 explain the effects of Δ DPF and $\Delta\Delta$ DPF parameters, which are inputted to the second stage MLN of hybrid neural network (HNN)-based phoneme recognizer. From the Fig. 3, in which GS orthogonalization is not used, it is observed that an addition of Δ and $\Delta\Delta$ parameters in the method (1) increase PCR by 1.37% at mixture component 16. Again, an improvement of 2.34% PCR, because of Δ and $\Delta\Delta$ parameters, is shown in Fig. 4 at 16 mixture component by the HNN-based method (q) with the GS orthogonalization procedure.

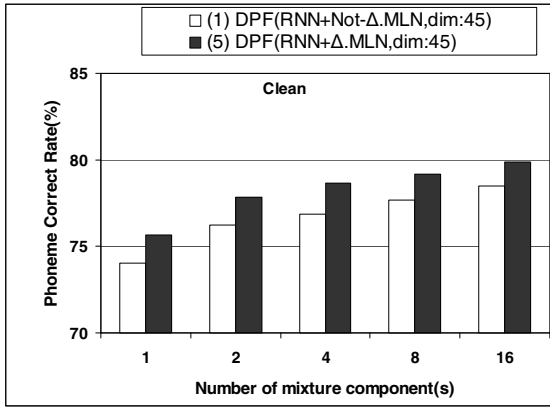


Fig. 3. Effects of articulatory dynamic parameters (Δ and $\Delta\Delta$) on the method (1), DPF(RNN+Not- Δ .MLN,dim:45)

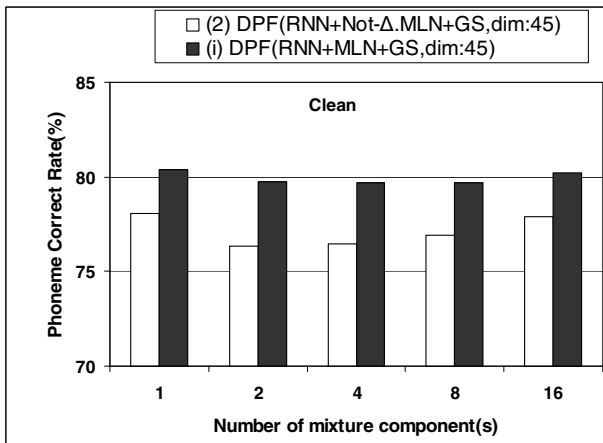


Fig. 4. Effects of articulatory dynamic parameters (Δ and $\Delta\Delta$) on the method (2) containing GS orthogonalization, DPF(RNN+Not- Δ .MLN+GS, dim: 45)

Fig. 5 also shows the effect of using Δ DPF and $\Delta\Delta$ DPF as input to the second stage MLN in the hybrid neural network-based phoneme recognizers with In/En and GS. In the figure, an addition of Δ and $\Delta\Delta$ parameters always increases PCR significantly. For example, at mixture component 16, the proposed method (q) that incorporates Δ and $\Delta\Delta$ parameters improves PCR by 0.73% in comparison with the method (9).

It is claimed that the proposed method reduces mixture components in HMMs and hence computation time. For an example from the Fig. 5, approximately 81.50% phoneme correct rate is obtained by the methods (9) and (q) at mixture components 16 and one, respectively.

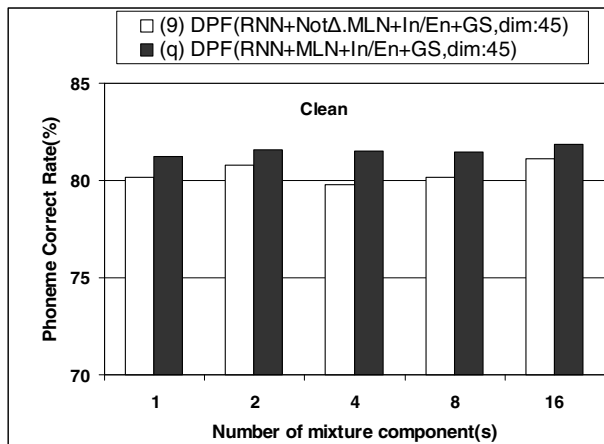


Fig. 5. Effects of articulatory dynamic parameters (Δ and $\Delta\Delta$) on the method (9) containing In/En and GS orthogonalization, DPF(RNN+Not- Δ .MLN+In/En+GS,dim:45) method with articulatory dynamic parameters

6 Conclusion

This paper has presented an articulatory feature based phoneme recognition method using a hybrid neural network for an ASR system, which incorporates articulatory dynamic parameters into it. From the experiments on Japanese Newspaper Article Sentences (JNAS), the following conclusions are drawn:

- i) The proposed method provides a higher phoneme correct rate over the method that does not incorporate dynamic articulatory parameters.
- ii) It reduces mixture components in HMM for obtaining a higher phoneme recognition performance.

In near future, the authors would like to do some experiments for evaluating Bangla phonemes spoken by Bangladeshi People. Moreover, we have intension to evaluate word recognition performance using the proposed method.

References

1. Kirchhoff, K., et al.: Combining acoustic and articulatory feature information for robust speech recognition. *Speech Commun.* 37, 303–319 (2002)
2. Kirchhoffs, K.: Robust Speech Recognition Using Articulatory information. Ph.D thesis, University of Bielefeld, Germany (July 1999)
3. Leung, K.Y., Mak, M.W., Kung, S.Y.: Applying articulatory features to telephone-based speaker verification. In: Proc. IEEE ICASSP 2004, vol. I, pp. 85–88 (2004)
4. Fukuda, T., Yamamoto, W., Nitta, T.: Distinctive Phonetic feature Extraction for robust speech recognition. In: Proc. ICASSP 2003, vol. II, pp. 25–28 (2003)
5. Fukuda, T., Nitta, T.: Orthogonalized Distinctive Phonetic feature Extraction for Noise-Robust Automatic Speech Recognition. The Institute of Electronics, Information and Communication Engineers (IEICE) Transactions on Information and Systems E87-D(5), 1110–1118 (2004)
6. Huda, et al.: Distinctive Phonetic Feature (DPF) based phone segmentation using 2-stage multilayer neural network. In: NCSP 2007, Shanghai, China (March 2007)
7. Ansary, L., et al.: Modeling phones coarticulation effects in a neural network based speech recognition system. In: Proc. Interspeech (2004)
8. Robinson, T.: An application of Recurrent Nets to Phone Probability Estimation. *IEEE Trans. Neural Networks* 5(3) (1994)
9. Huda, M.N., et al.: Phoneme recognition based on hybrid neural network with inhibition/enhancement of distinctive phonetic feature (DPF) trajectories. In: InterSpeech 2008, Brisbane, Australia (September 2008)
10. King, S., Taylor, P.: Detection of Phonological Features in Continuous Speech using Neural Networks. *Computer Speech and Language* 14(4), 333–345 (2000)
11. Eide, E.: Distinctive Features for Use in an Automatic Speech Recognition System. In: Proc. Eurospeech 2001, vol. III, pp. 1613–1616 (2001)
12. Nitta, T.: Feature extraction for speech recognition based on orthogonal acoustic-feature planes and LDA. In: Proc. ICASSP 1999, pp. 421–424 (1999)
13. Kobayashi, T., et al.: ASJ Continuous Speech Corpus for Research. *Acoustic Society of Japan Trans.* 48(12), 888–893 (1992)
14. JNAS: Japanese Newspaper Article Sentences,
<http://www.milab.is.tsukuba.ac.jp/jnas/instruct.html>

Gaussian Noise and Haar Wavelet Transform Image Compression on Transmission of Dermatological Images

Kamil Dimililer and Cemal Kavalcıoğlu

Electrical and Electronic Engineering Department
Near East University, Lefkosa, Mersin 10, Turkey
{kdimililer, ckavalci}@neu.edu.tr

Abstract. Telemedicine provides medical information and services using telecommunication technologies. Teledermatology, is a special part in the medical field of dermatology and one of the most common applications of telemedicine and e-health. Telecommunication technologies are used in Teledermatology to exchange medical information over a distance using audio, visual and data communication. Medical images require compression; Wavelet-based image compression provides substantial improvements in picture quality at higher compression ratios. An ideal image compression system must yield high quality compressed image with high compression ratio; this ratio can be achieved using transform-based image compression, however the contents of the image affects the choice of an optimum compression ratio and the optimum compression method. This paper presents image compression method, Haar wavelet transform, which can be applied to compress dermatology images before the transmission through a communication channel.

Keywords: Telemedicine, Teledermatology, Haar Wavelet Transform, Medical image compression, Adaptive White Gaussian Noise (AWGN), Optimum Image Compression.

1 Introduction

Telemedicine uses telecommunications technologies to obtain medical information and services via safe communication channels. Telemedicine systems can make use of well constituted and emerging wireless technologies and standards such as IEEE 802.11, 802.15, and 802.16-based networks for the reliable delivery of medical information and supply of life-saving services [1].

Dermatology is one of the branch of medicine that deals with the conditions of the skin, along with sweat glands, hair, and other related body parts. Teledermatology, in its simplest terms, is the use of communication information technology to deliver dermatologic care. Typically, technology is applied when a conventional “face-to-face” clinic visit cannot be performed; implying that distance or some other barrier prevents this conventional method of healthcare [2]. In these situations the patient and clinician are separated by a geographic barrier, with technology providing the link. This is actually a restricted view of how Teledermatology may be used in healthcare delivery but is, nonetheless, a useful way to describe the most common

rationale for Teledermatology implementation. A patient and a clinician separated from one another by distance [3], [4].

In recent years there has been an enormous increase in the usage of computers for a variety of tasks. With the invention of digital cameras, the most important problem in the field of storing the data, manipulation of images and transfer of the images through the communication channel has been arisen. The files that comprise these images, however, can be quite large and can quickly take up precious memory space on the computer's hard drive. A gray scale image that is 256 x 256 pixels has 65,536 elements to store, and a typical 640 x 480 color image has nearly a million! The size of these files can also make downloading from internet a lengthy process.

Since the number of patient images increase considering the size and the quality, these medical images should be stored efficiently in digital format in the most economical way and effective data compression technologies are required to minimize mass volume of digital image data produced in the hospitals [5].

There has been a vast development in compression methods to compress huge data files such as images where compression of data in various kinds of applications has become more vital. Efficient methods of compression, to compress and store or transfer image data files while retaining high image quality and marginal reduction in size are needed due to the improvements of technology [6].

Compression methods are being rapidly developed to compress large image data files, where data compression in multimedia applications has lately become more necessary. With the increasing growth of technology and the entrance into the digital age, a vast amount of image data must be handled to be stored in a proper way using efficient methods usually succeed in compressing images, while retaining good image quality and marginal reduction in image size.

Wavelets are a mathematical tool for hierarchically decomposing functions. Image compression using Wavelet Transforms is a powerful method that is preferred by scientists to get the compressed images at higher compression ratios and PSNR values [7].

Image processing and image analysis based on the continuous or discrete image transforms are classical techniques. The image transforms are widely used in image filtering, data description, etc. Nowadays the wavelet theorems make up very popular methods of image processing, denoising and compression. Considering that the Haar functions that are the simplest wavelets, these forms are used in many methods of discrete image transforms and processing.

Digital images needs large amount of memory to store and, when retrieved from the internet, can take a considerable amount of time to download. The Haar wavelet transform provides a method of compressing digital image data so that it takes up less memory [8].

For a Lossy Image Compression method, source encoder is to exploit the redundancies in image data to provide compression. In other words, the source encoder reduces the entropy, which in our case means decrease in the average number of bits required to represent the image.

Wavelets are mathematical functions that were developed by scientists working in several different fields for the purpose of sorting data by frequency. Studying data at different levels allows for the development of a more complete picture. Both small

features and large features are discernable because they are studied separately. Unlike the discrete cosine transform, the wavelet transforms are not Fourier-based providing a better job of handling discontinuities within the data [9].

The Haar wavelet operates on data by calculating the sums and differences of adjacent elements. The Haar wavelet operates first on adjacent horizontal elements and then on adjacent vertical elements [10] [11].

In communications, there are four types noise that can be created within the communication channel. These can be listed as Gaussian Noise, Poisson Noise, Salt & Pepper Noise and Speckle Noise. Gaussian noise will be considered in this application.

An AWGN channel adds white Gaussian noise to a signal that passes through it. The AWGN channel is an applicable model for many satellite and deep space communication links. It is not a good model for most terrestrial links because of multipath, terrain blocking and interference. The channel, which may include other effects besides those of the physical medium, invariably distorts the transmitted modulation signals in random and non-random ways. It is often characterized by the way it distorts the transmitted signals [4]. In the additive white Gaussian noise channel model, zero mean noise having a Gaussian distribution is added to the signal, as shown in Figure 1.

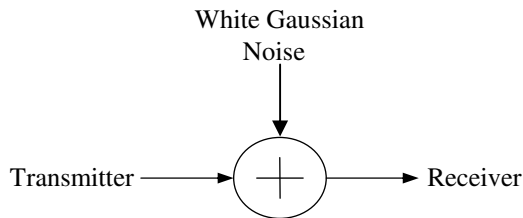


Fig. 1. The Gaussian Noise Channel

The Probability density function of a Gaussian random variable, z is given by;

$$p(z) = \frac{1}{\sqrt{2\pi\sigma}} e^{-(z-\mu)^2/2\sigma^2} \quad (1)$$

Gaussian noise is a very good approximation of noise that occurs in many practical cases. The probability density of the random variable is given by the Gaussian curve shown in figure 2. White noise means constant power spectrum, very crude approximation of image noise.

In this paper, Gaussian noise with the standard deviations of 0.005 and 0.025 has been applied to the dermatological image database and de-noised using Average filter [12]. Image compression has been applied to the same image database with a ratio of 8:10 and the process of transmission has been applied again to see the effect of quality for non-compressed image set and compressed image set. Peak Signal to Noise Ratio has been preferred to be used in the decision of the quality of the reconstructed images [11].

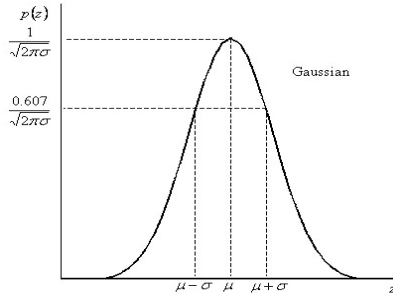


Fig. 2. Gaussian Probability Density Function

The organization of the paper can be given as follows: Section two describes the dermatological image database, section three gives the HWT compression based Teledermatology System (HTS) and section four gives the results, discussions and concludes the system with suggesting the future work.

2 The Dermatological Image Database

The development and implementation of the proposed system for dermatological images uses 80 images from our medical image database, which contains dermatological images in different parts of the body. Figure 3 represents a set of samples from the image database.

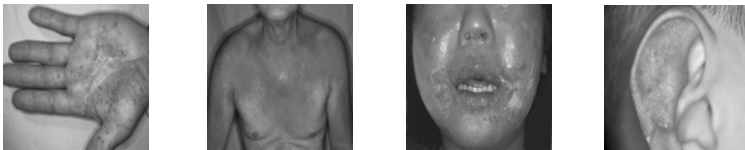


Fig. 3. Dermatological Image Database

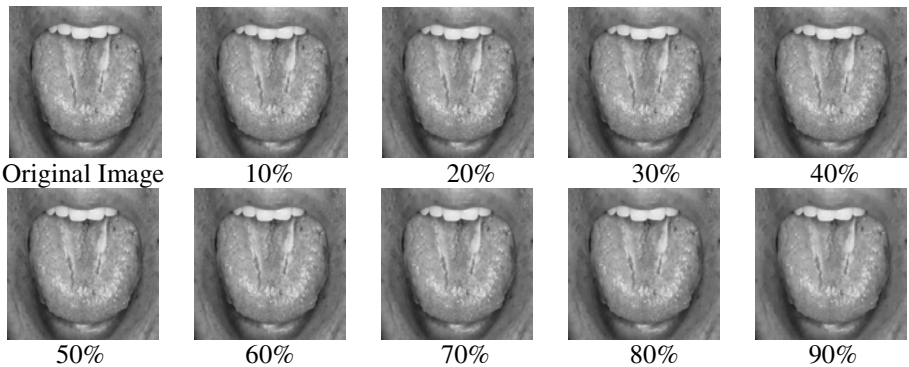


Fig. 4. An original from the database and compressed set of a dermatological Image

Haar Wavelet Transform based image compression has been applied to the dermatological image database using the nine compression ratios (10%, 20%, ..., 90%) as shown in Figure 4.

3 The HWT Compression Based Teledermatology System (HTS)

The suggested dermatological image compression system consists of a Haar Wavelet Transform based image compression system combined with a communication channel which Adaptive White Gaussian Noise has been applied.

Table 1 gives the PSNR values of a dermatological image from the database with Haar wavelet transform compression applied showing that, the PSNR value decreases as the compression ratio increases.

Initially, The Adaptive White Gaussian Noise applied system has been sent through the communication channel and the reconstructed images for the given database are taken and the reconstructed images with their peak signal to noise ratio has been computed.

Table 1. A Sample image from the database with Peak Signal to Noise Ratio Results

Image Name	Compression Ratio	PSNR
Image 7	10%	66,758
Image 7	20%	58,568
Image 7	30%	52,739
Image 7	40%	48,806
Image 7	50%	45,288
Image 7	60%	42,225
Image 7	70%	39,223
Image 7	80%	36,081
Image 7	90%	32,271

Table 2. AWGN applied image database with the PSNR Values

Image	Type Of Noise	PSNR	Type Of Noise	PSNR
Image_1	Gaus. Var.=0.005	32.5161 dB	Gaus. Var.=0.025	30.5890 dB
Image_2	Gaus. Var.=0.005	33.0740 dB	Gaus. Var.=0.025	30.8569 dB
Image_3	Gaus. Var.=0.005	34.3768 dB	Gaus. Var.=0.025	31.3061 dB
Image_4	Gaus. Var.=0.005	32.6498 dB	Gaus. Var.=0.025	30.5949 dB
Image_5	Gaus. Var.=0.005	33.3082 dB	Gaus. Var.=0.025	31.3076 dB

The Gaussian variance of 0.005 and 0.025 has been considered for this application. Table 2 shows the AWGN applied image set with the Peak-Signal-to-Noise-Ratios.

Haar Wavelet Transform has been applied to the suggested database with different compression ratios in order to get the PSNR values of the images. The PSNR values of the reconstructed images are compared with the image database, before the

compression has been applied and after the compression has been applied before being sent through the communication channel. The compression ratio of more than 80 percent has been discarded because of the blurring effect of the transform on the reconstructed images.

4 Results and Conclusion

In this paper, the suggested Teledermatology System has been applied using Haar Wavelet Transform and Additive White Gaussian Noise channel model in which the image information has been applied with an additive noise given as a white noise with a spectral density. Wavelet Transform is a lossy image compression technique applied on the digital signals such as images.

Table 3 shows the results of the 5 images within the database after the compression has been applied.

The main goal for this designed system is to deliver high quality dermatology images in a short period of time to the central hospital within minimum cost of time. In this paper, an image size of 256x256 pixels has been chosen to send the image with enough information and with minimum cost of time. Increasing the sizes of the images would provide more detailed information but would take more time to send through the communication channel and decreasing the sizes of the images would not provide enough information about the dermatological images to the inspector.

Table 3. Haar Wavelet Transform applied images before AWGN with the PSNR Values

Image	Type Of Noise	PSNR	Type Of Noise	PSNR
Image_1	Gaus. Var.=0.005	32.9448 dB	Gaus. Var.=0.025	30.8937 dB
Image_2	Gaus. Var.=0.005	33.2061 dB	Gaus. Var.=0.025	31.1451 dB
Image_3	Gaus. Var.=0.005	34.5929 dB	Gaus. Var.=0.025	30.8462 dB
Image_4	Gaus. Var.=0.005	32.7471 dB	Gaus. Var.=0.025	31.2412 dB
Image_5	Gaus. Var.=0.005	33.6274 dB	Gaus. Var.=0.025	31.2123 dB

Figure 5 represents the original and reconstructed images without HWT compression and Figure 6 represents the original and reconstructed images with HWT Compression applied images.

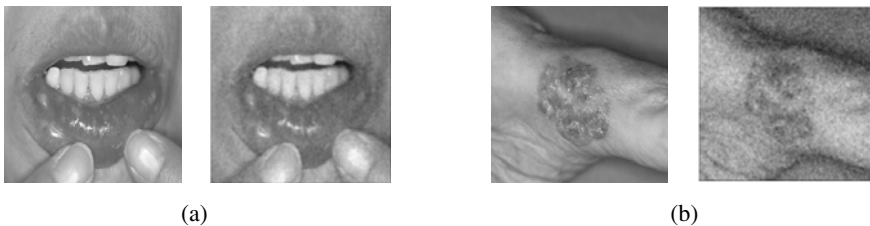


Fig. 5. Original and Reconstructed images without Haar Wavelet Transform Compression with (a)Gauss. Noise variance of 0.005 (b)Gauss. Noise variance of 0.025

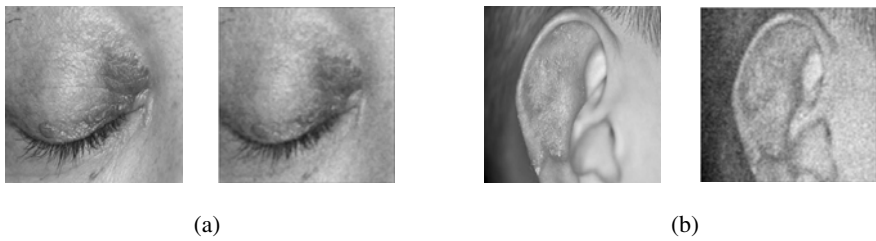


Fig. 6. Original and Reconstructed images with Haar Wavelet Transform Compression with (a) Gauss. Noise variance of 0.005 (b) Gauss. Noise variance of 0.025

Table 4 shows the results of 5 images from the database using the 0.005 and 0.025 of gaussian variance applied to the images without the application of image compression and with the application of the 80% image compression.

The results of the suggested system shows that the 73 images out of 80 (91.25%) has a higher value of PSNR. The advantage of the suggested system is that a compression of 80% provides better quality of image and less time of transmission of medical images due to the data compression. The future work will focus on the application of Biorthogonal Wavelet Transform based image compression applied on Teledermatology images and Back Propagation Neural Networks will be applied in the decision process of optimum compression ratio and optimum compression method.

Table 4. Comparison of the PSNR Values for compressed and uncompressed image set

	Gauss. Var. 0,005 without Compression	Gauss. Var. 0,005 with Compression	Gauss. Var. 0,025 without Compression	Gauss. Var. 0,025 with Compression
Image_6	33.9153 dB	33.9227 dB	30.9221 dB	31.1159 dB
Image_7	34.9135 dB	35.0613 dB	31.5637 dB	31.0352 dB
Image_8	34.1508 dB	34.2447 dB	31.1044 dB	31.0139 dB
Image_9	32.2566 dB	32.6455 dB	30.4086 dB	31.0937 dB
Image_10	32.2768 dB	32.6160 dB	30.2741 dB	31.5317 dB

References

1. Ansari, N., Fong, B., Zhang, Y.T.: Wireless Technology Advances and Challenges for Telemedicine. IEEE Communications Magazine (2006)
2. Serener, A., Kavalcioglu, C.: Teledermatology based medical images with AWGN Channel in Wireless Telemedicine System. In: Proceedings of the 1st WSEAS International Conference on Manufacturing Engineering, Quality and Production Systems, pp. 145–150. Brasov, Romania (2009)
3. Pak, H.S., Edison, K.E., Whited, J.D.: Teledermatology: A User's Guide. Cambridge University Press, Cambridge (2008)

4. Gonzalez, R., Woods, E.R., Eddins, L.S.: Digital Image Processing. Prentice - Hall, Upper Saddle River (2002)
5. Zukoski, M.J., Boulton, T., Iyriboz, T.: A novel approach to medical image compression. *Int. J. Bioinformatics Research and Appl.* 2(1), 89–103 (2006)
6. Porwik, P., Lisowska, A.: The Haar-Wavelet Transform in digital image processing: its status and achievements. *Machine Graphics & Vision* 13(1/2), 79–98 (2004)
7. Khashman, A., Dimililer, K.: Image compression using Neural Networks and Haar wavelet. *WSEAS Trans. On Signal Processing* 4(5) (2008)
8. Ashok, V., Balakumaran, T., Gowrishankar, C., Vennila, I., Kumar, N.A.: The fast haar wavelet transform for signal & image processing. *Int. Jor. Of Comp. Science and Inf. Security* 7(1) (2010)
9. Bhardwaj, A., Ali, R.: Image Compression using modified fast Haar wavelet transform. *World App. Sci. Jor.* 7(5), 647–653 (2009)
10. Khashman, A., Dimililer, K.: Intelligent System for Medical X-Rays Compression. *Trans. On Mass-Data Analysis of Images and Signals* 1(1), 3–14 (2009)
11. Khashman, A., Dimililer, K.: Comparison Criteria for Optimum Image Compression. In: *Proceeding of the IEEE International Conference on 'Computer as a Tool'*, pp. 935–938. IEEE Press, Serbia (2005)
12. Patidar, P., Gupta, M., Sriyastava, S., Nagawat, A.K.: Image De-noising by Various Filters for Different Noise. *Int. J. of Comp. App.* 9(4) (2009)

Image Processing Techniques for Glaucoma Detection

Mishra Madhusudhan, Nath Malay, S.R. Nirmala, and Dandapat Samerendra

Indian Institute of Technology Guwahati, Department of Electronics and Electrical Engineering, Guwahati, India

{m.mishra,malaya,nirmala,samaren}@iitg.ernet.in

Abstract. Glaucoma is a disease caused due to neurodegeneration of the optic nerve which leads to blindness. It can be evaluated by monitoring intra ocular pressure (IOP), visual field and the optic disc appearance (cup-to-disc ratio). Glaucoma increases the cup to disc ratio (CDR), affecting the peripheral vision loss. This paper addresses the various image processing techniques to diagnose the glaucoma based on the CDR evaluation of preprocessed fundus images. These algorithms are tested on publicly available fundus images and the results are compared. The accuracy of these algorithms is evaluated by sensitivity and specificity. The sensitivity and specificity for these algorithms are found to be very favorable.

Keywords: Glaucoma risk index, cup to disc ratio, multi-thresholding, active contours, region growing segmentation.

1 Introduction

Glaucoma is a progressive optic neuropathy leading to visual field changes due to the damage retinal ganglion cells and axons [1]. From the survey it is found that nearly 1-million Indians, age 40 and over, have glaucoma [2]. With disease progression the interconnection between the photo receptors and the visual cortex is reduced. This reduces the functional capabilities of the retina and changes the size of the cup. So cup-to-disc ratio (CDR) is used for measuring glaucoma progression. Fig.1 shows how the objects are perceived by normal vision and a patient having glaucoma. Raised intra-ocular pressure (IOP) is classified as a risk factor but is not part of the definition. However it causes typical structural changes of the optic nerve head (ONH) and the nerve fiber layer affecting the visual field of the subject [3]. The structural changes are the outcome of the slowly diminishing neuroretinal rim manifesting a degeneration of axons and astrocytes of the optic nerve resulting in expansion of cup size of the optic disc, as shown in fig. 2. However blood vessels are least affected.

The CDR expresses the proportion of the disc occupied by the cup and it is widely accepted index for the assessment of glaucoma [3]. For Normal eye it is found to be 0.3 to 0.5 [14]. The CDR value increases with the increase in neuro-retinal degeneration and the vision is lost completely at the CDR value of 0.8. Several methods of feature extraction from fundus images are reported in the literature [1, 3, 5, 6, 9, 10, and 11]. Techniques described in the literature for optic disk localization are

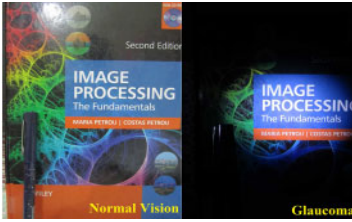


Fig. 1. Normal vision vs. patient having glaucoma where peripheral vision is lost.[7]

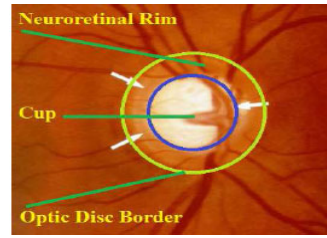


Fig. 2. Major Structures of the ONH visible in color fundus image

typically aimed at either identifying the approximate center of the optic disk or placing the disk within a specific region such as a circle or square. Lalonde et al., uses canny edge detector and Ghafar et al., uses Circular Hough-transform to detect the optic disk(OD)[5][6]. By analyzing the various local maxima in the Hough space, one can find the best fitting circular approximation of the optic disk. Walter and Klein uses color space transformation and morphological filtering for optic disk localization. Here luminance channel of hue-luminance-saturation color space is used for localizing optic disk. Then thresholding is applied to locate locus of optic disk. Precise contour is determined using green channel of RGB color space via watershed transformation. In this method slight distortion occurs due to outgoing vessels or low contrast. Youssif et al., detects OD by using the blood vessel's direction matched filter where a mask is generated [1]. The mask labels the pixels belonging to the region of interest and excludes the background for further processing. Then opening, closing and erosion are applied respectively using small kernel to obtain final ROI. These methods are based on exploiting the edge characteristics. These papers do not discuss about the glaucoma progression or classification. Bock et al., uses the concept of principal component analysis (PCA), bitsplines and fourier analysis for feature extraction and support vector machine (SVM) classifier for glaucoma predication[3]. This method attains an accuracy of 88 %. However different methods for features extraction and use of classifiers make this system more complex. Our proposed methods for glaucoma detection are simple, easy to implement and give favorable results.

In this paper three different image processing techniques namely multi-thresholding, active contour model and region based segmentation methods are proposed for detection of glaucoma. The comparative study of these proposed techniques are done with respect to their approaches and results. The rest of this paper is organized as follows. Section 2 introduces all the three methods for glaucoma detection. Experimental results are shown in Section 3 and conclusions are presented in Section 4.

2 Proposed Methods for Glaucoma Detection

In this section, we discuss our work on retinal image analysis for glaucoma detection. We have developed a scheme for automated processing and classification of the acquired images based on the usual practice in the clinic. Fig. 3 shows our proposed

system which follows a standard 3-step image analysis pipeline consisting of (i) preprocessing; (ii) segmentation of preprocessed image and (iii) classification based on evaluation of CDR. Three different techniques i.e. multi thresholding, active contours and region growing methods are applied for the segmentation of pre-processed fundus image in order to detect the disease by computing CDR.

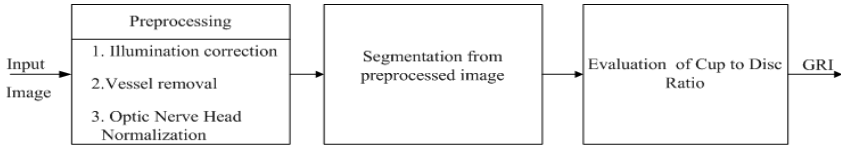


Fig. 3. Processing pipeline in detail: glaucoma risk calculation comprises of steps: (i) Preprocessing of input images for eliminating disease independent variations, (ii) Segmentation of preprocessed image and (iii) Classification based on CDR for generating Glaucoma Risk Index (GRI)

2.1 Pre Processing

The variations not related to the glaucoma disease are excluded from the images in a preprocessing step for emphasizing the desired characteristics. This includes variations due to image acquisition, such as inhomogeneous illumination and the blood vessels which are not directly linked to glaucoma. The main objective of preprocessing is to attenuate image variation by normalizing the original retinal image against a reference model or data set for subsequent viewing, processing or analysis [12]. The preprocessing retinal images may be classified in terms of the correction for non-uniform illumination, contrast enhancement and color normalization.

2.1.1 Illumination Correction

The peripheral part of the retina often appears darker than the central region because of the curved retinal surface and the geometrical configuration of the light source and camera. These interferences affect the illumination of the ONH and would have an influence to the subsequent statistical analysis, though they are not originated through glaucoma [3]. Homogeneously illuminated fundus image is obtained by subtracting the estimated retinal background from the original image. Various techniques for illumination correction like morphological operations, homomorphic filtering and median filtering have been published in the literature [13]. We implemented a correction method based on morphological operations as it has certain advantages over other techniques. The benefits of this technique over linear approaches include direct geometric interpretation, simplicity and efficiency in hardware implementation [14]. Uniformly illuminated image is obtained by subtracting the estimated background image from original image. Morphological opening is used to estimate the background illumination. Fig.4 shows the different steps needed to obtain an illumination corrected image. Original fundus image (i) is first converted into grey image (ii) and then background is estimated (iii). Image obtained from subtracting (iii) from (ii) is added with fixed dc level to get the final illumination corrected image shown in (v).



Fig. 4. Result of illumination correction using morphological operations

2.1.2 Blood Vessels Removal

Blood vessels are minimally affected by glaucoma disease; hence blood vessels need to be removed from the fundus images [15]. Blood vessel removal consists of two steps: (i) extraction of blood vessels and after that (ii) inpainting of extracted blood vessels. Image inpainting is the technique of filling in a region of an image based on the information outside the region [16]. The major blood vessel branches rising from the ONH hide large portions of the rim and their existence makes analysis of the visible parts of the ONH more difficult [3]. This can hinder the accurate segmentation of OD. Therefore, inpainting technique is applied to remove these blood vessel structures after the extraction blood vessels before further processing. The extracted blood vessels act as a mask and the region covered by the mask is inpainted. In this implementation, the vessel regions are filled iteratively layer by layer from outside inwards while the missing pixels get a weighted average of the already known neighboring values [17] [18]. Morphological operations are also used for removal of blood vessels from fundus image. Morphological closing consisting of dilation followed by erosion is applied to remove the blood vessels. Fig.5 shows the result of morphological operations for blood vessels removal.

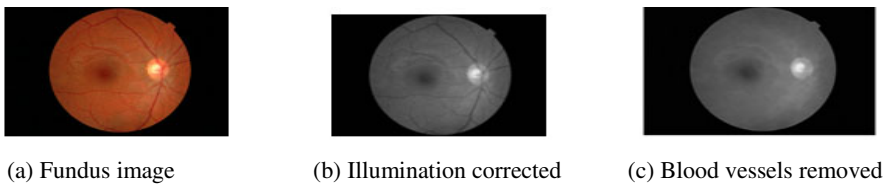


Fig. 5. Result of morphological operations for blood vessels removal

Inpainting technique gives better results in comparison to morphological operations for blood vessels removal. However the major drawback of inpainting technique lies in the fact that it requires extracted vessels as the mask and iterations process increases the computational time.

2.1.3 Normalization of the ONH Region

Papilla (ONH) is the most important structure for observing changes in order to detect glaucoma. It appears as an extremely bright, mostly circular region in fundus images. The image obtained after blood vessels removal is normalized for better analysis.

2.2 Segmentation from the Preprocessed Image

The preprocessed images are used for the segmentation of optic disc and cup which helps in classification of glaucoma. In this subsection, the three different techniques for glaucoma detection are discussed.

2.2.1 Multi-thresholding Technique

Multi-thresholding technique is one of the simplest methods and natural way to segment cup and disc of preprocessed fundus image. The preprocessed image containing the optic disc is converted to binary image and then multi-thresholding technique is applied. This technique allows the detection of cup, the brighter region of the optic disc with higher threshold value and the whole disc with lower threshold value. Fig.6 shows the detection of optic disc and cup from preprocessed image for glaucoma classification.

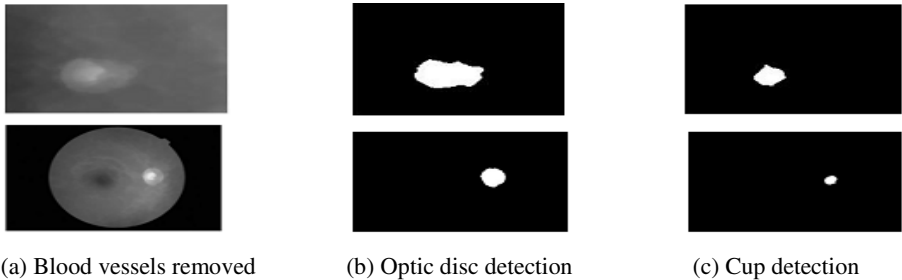


Fig. 6. Result of multi-thresholding for disc and cup detection

It has shown correct segmentation for 19 images out of 25 set of images. The accurate segmentation was not possible for rest of 6 images as the optic disc and some portions of background image have similar illumination level. The quality of original images was also one of the reasons for improper segmentation. The major demerit of the method is the manual thresholding based on the pixel intensity values. Hence other techniques like active counter method based on adaptive thresholding and region growing segmentation methods are applied for cup and disc segmentations from the preprocessed fundus image for glaucoma classification.

2.2.2 Snakes: Active Contours Method

A snake is an energy minimizing spline guided by external forces and influenced by image forces that pull it toward features such as lines and edges in order to localize the object accurately [19]. Contours stops at nearby edges detecting the desired boundary, where the external energy is the minimum. Classical snakes and active contour models, uses an edge-detector, based on the gradient of the image, to stop the evolving curve on the boundary of the desired object. However in practical situations, the discrete gradients are bounded hence the stopping function is never zero on the edges, and the curve may cross the desired boundary. Therefore, we have used a different active contour model which is not based on the gradient of the image like classical methods for the stopping process but is instead related to a particular

segmentation of the image [20]. Fig.7 shows the results of successful implementation of this technique, used in detecting cup and disc of fundus images for evaluation of CDR, a measure for glaucoma detection.

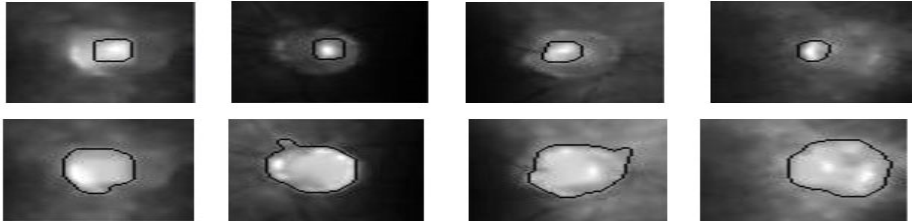


Fig. 7. Results of Active contour methods for cup and disc detection

2.2.3 Region Growing Segmentation

Region growing is a an approach to image segmentation based on selecting initial seed points and adding neighboring pixels to the region depending on certain feature such as intensity value, color similarity[8]. In our proposed method the maximum value in the green channel image serves as the seed for the region growing algorithm to be employed in the cup identification. The neighborhood pixels are iteratively connected if they fall below a certain threshold value. The centroid of the resulting region is calculated and a disk is estimated that can circumscribe the entire region. The radius of this estimated disk is noted as the radius of the cup. The area of the cup is hence known. The disk is further grown outwards till the point when a sharp intensity change is observed. This change marks the periphery of the disk and the radius at this juncture is noted as the radius for the disk. Thus we have a disk whose area is known. The ratio of area of cup to disk is evaluated and checked and if it falls below the value of 0.5 the eye can be concluded as a healthy one. If however, the value exceeds the limit, it is concluded to be defected with glaucoma. Experimental results on a 25 number of images demonstrate the effectiveness of the technique.

3 Results and Discussions of Glaucoma Detection Techniques Based on CDR Evaluation

The above discussed methods are applied to the publicly available Optic-disc databases, which consist of normal and pathological fundus images. The criteria that CDR for normal fundus images lies between 0.3-0.5 and is higher for glaucomous fundus images are used for the classification. In this section we mainly focus on the comparative approaches and results of the all three proposed systems namely multi-thresholding, active contour methods and region based segmentation techniques used for glaucoma classification.

3.1 Glaucoma Detection Based on Multi-thresholding

The processed image obtained as discussed earlier is used to determine the CDR by finding the number of ones present in the cup region to that of number of ones in disc

region of binary image. CDR of 25 fundus images downloaded from www.opticdisc.org (7 normal and 18 abnormal images) is determined in order to detect glaucoma. Efficiency of the proposed method in identifying true positive and true negative is shown the column-I of the table-1. The performance measure of this method with sensitivity (classify abnormal fundus images as abnormal) of 80% and specificity (classify normal fundus image as normal) of 60 % is summarized in column-I of table-2.

Table 1. Shows the comparative results of three methods namely (i) Multi-thresholding; (ii) Active contours and (iii) Region growing segmentation

SL. No.	Performance parameter	Efficiency (percentage)		
		Method I	Method II	Method III
1	True Positive	88.89	94.44	100
2	True Negative	42.85	57.17	85.71
3	False Positive	11.11	5.56	0
4	False Negative	57.17	42.85	14.29

Table 2. Shows the comparative accuracy of three methods namely (i) Multi-thresholding; (ii) Active contours and (iii) Region growing segmentation

SL. No.	Accuracy parameter	Efficiency (percentage)		
		Method I	Method II	Method III
1	Sensitivity	80	89.47	94.73
2	Specificity	60	83.33	100

3.2 Glaucoma Detection Based on Active Contour Method

Active contour method is applied for detecting the cup and disc from the preprocessed fundus images as discussed in the subsection 2.2.2. Their effective radii are calculated for evaluation of CDR needed for the classification. The obtained results are found to be better in-comparison to multi thresholding technique as it is based on adaptive thresholding. Efficiency of this method in identifying true positive and true negative is shown the column-II of the table-I. The accuracy measure of this second proposed method with sensitivity and specificity of 89.47 and 83.33 % respectively is summarized in column-II of table-II.

3.3 Glaucoma Detection Based on Region Growing Segmentation Technique

Region growing segmentation technique is applied to estimate radius of cup and disc from the pre processed fundus image as discussed in subsection 2.2.3. Like the other two methods, CDR evaluation was the criteria for the classification of the disease. The results obtained by this third proposed method are better compared to other two mentioned techniques as it is based on clustering of homogeneous regions. It is able to classify all eighteen abnormal images as abnormal but two images were misclassified out of seven normal images. Efficiency of this method in identifying true positive and

true negative is shown the column-III of the table-I. The performance measure of this method with sensitivity and specificity of 94.73% and 100% respectively is summarized in column-III of table-II.

4 Conclusion

In this paper, an efficient framework for early detection of glaucoma has been developed. The CDR, an important glaucoma parameter of fundus images publically available from messidor and optic data bases were evaluated using three different methods namely morphological operations based on multi-thresholding techniques, active contour models and region growing segmentation techniques. As a comparative study to these methods for glaucoma classification, we observed that region growing segmentation technique gives better result in comparisons to other two methods. The proposed methods are simple and easy to implement. The results obtained can be used as an initial investigation step in the automated diagnosis of glaucoma especially in the screening programs. These proposed methods may further be combined with some other techniques for achieving better results with large databases.

References

1. Youssif, A.A.-H.A.-R., Ghalwash, A.Z., Ghoneim, A.A.S.A.-R.: Optic disc detection from normalized digital fundus images by means of a vessels' direction matched filter. *IEEE Transactions on Medical Imaging* 27, 11–18 (2008)
2. <http://www.medindia.net/patients/patientinfo/Glaucoma.htm>
3. Bock, R., Meier, J., Nyúl, L.G., Hornegger, J., Michelson, G.: Glaucoma risk index: Automated glaucoma detection from color fundus images. *Medical Image Analysis* 14, 471–481 (2010)
4. <http://www.optic-disc.org>
5. Lalonde, M., Beaulieu, M., Gagnon, L.: Fast and robust optic disc detection using pyramidal decomposition and hausdroff based template matching. *IEEE Transactions on Medical Imaging* 11, 1193–1200 (2001)
6. Ghafar, R.A.A., Morris, T., Ritchings, T., Wood, I.: Detection and characterization of the optic disc in glaucoma and diabetic retinopathy. In: *Medical Image Understand Annual Conference*, London, UK, pp. 23–24 (September 2004)
7. <http://www.nei.nih.gov/>
8. Banerjee, B., Bhattacharjee, T., Chowdhury, N.: Color Image Segmentation Technique Using Natural Grouping of Pixels. *International Journal of Image Processing (IJIP)* 4(4), 320–328 (2010)
9. Hoover, A., Kouznetsova, V., Goldbaum, M.: Locating blood vessels in retinal images by piecewise threshold probing of a matched filter response. *IEEE Transaction on Medical Imaging* 19, 203–210 (2000)
10. Rangayyan, R.M., Zhu, X., Ells, A.L.: Detection of the optic nerve head in fundus images of the retina using the Hough transform for circles. *Journal of Digital Imaging*, 132–140 (February 2009)
11. Mendonca, A.M., Campilho, A.: Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Transaction on Medical Imaging* 25, 1200–1213 (2006)

12. Salvatelli, A., Bizai, G., Barbosa, G., Drozdowicz, B., Delrieux, C.: A comparative analysis of pre-processing techniques in color retinal images. In: 16th Argentine Bioengineering Congress and the 5th Conference of Clinical Engineering. IOP Publishing Journal of Physics: Conference series, vol. 90 (2007)
13. Kubecka, L., Jan, J., Kolar, R.: Retrospective Illumination Correction of Retinal Images. International Journal of Biomedical Imaging 2010, Article ID 780262, 10 pages (2010)
14. Vanajakshi, B., Sujatha, B., Srirama, K.: A Study on Implementation of Advanced Morphological Operations. IJCSNS International Journal of Computer Science and Network Security 10(3), 6–9 (2010)
15. http://www.optic-disc.org/tutorials/glaucoma_evaluation_basics/page13.html
16. Criminisi, A., Pérez, P., Toyama, K.: Region Filling and Object Removal by Exemplar-Based Image Inpainting. IEEE Transactions on Image Processing 13(9) (2004)
17. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH, New Orleans, USA, pp. 417–424 (2000)
18. Meier, J., Bock, R., Michelson, G., Nyúl, L.G., Hornegger, J.: Effects of Preprocessing Eye Fundus Images on Appearance Based Glaucoma Classification. In: Kropatsch, W.G., Kampel, M., Hanbury, A. (eds.) CAIP 2007. LNCS, vol. 4673, pp. 165–172. Springer, Heidelberg (2007)
19. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active contour models. International Journal of Computer Vision 1, 321–331 (1988)
20. Chan, T.F., Vese, L.A.: Active Contours Without Edges. IEEE Transaction on Image Processing 10(2) (February 2001)

Transmitter Preprocessing Assisted MIMO SDMA Systems over Frequency-Selective Channels

Shriram Swaminathan, Suraj Krishnan, and Prabagarane Nagaradjane

Department of Electronics and Communication, SSN Institutions, Chennai, India
{shriram2811, suraj.krishnan}@gmail.com
prabagaranen@ssn.edu.in

Abstract. This paper investigates the performance of multi-user transmitter preprocessing (MUTP) based on singular value decomposition (SVD) technique for multiple input multiple output (MIMO) aided space division multiple access (SDMA) system over correlated and uncorrelated frequency-selective channels with two power allocation regimes, namely ‘water-filling’ and equal power for downlink (DL) communications. Specifically, we study the effects of delay spread distribution of vehicular channel model pertaining to Long Term Evolution (LTE). In multi-user MIMO downlink (DL) communications, users conflict multi-user interference (MUI) as well as multi-stream interference (MSI) which can severely degrade the achievable performance. It is demonstrated with the aid of simulation results that MUTP can yield better achievable bit error rate (BER) by mitigating MUI and MSI and can also obviate the need for employing complex multiuser detectors (MUDs) at the mobile stations (MSs). Further, our simulation study shows that invoking the ‘water-filling’ based power allocation regime results in better achievable capacity.

Keywords: Multiple input multiple output (MIMO), preprocessing, post processing, multi-user transmitter preprocessing (MUTP), singular value decomposition (SVD), multi-user interference (MUI), multi-stream interference (MSI).

1 Introduction

Recent information theoretic results have shown that multiple input multiple output (MIMO) systems aided by antenna arrays both at the transmitter and the receiver can provide significant capacity gain through an increased spatial dimension [1]. This capacity gain comes at no expense of transmission bandwidth and power. Much of the research focus has been in the design of single user MIMO systems where only multi-stream interference (MSI) exists. Of late, space division multiple access system (SDMA) has been proposed as a possible design choice for multi-user MIMO systems. In multi-user MIMO systems, the multi-user interference (MUI) along with MSI at the mobile station (MS) in downlink (DL) communications is a critical issue that requires serious attention. To deal with these interferences, multi-user detection (MUD) can be employed [2] at the MSs, but its complexity increases excessively as the number of users grow in a system, thus making it impractical for implementation

even at the base station (BS) where complexity is acceptable. An alternative way to mitigate interference is to exploit multi-user transmitter preprocessing (MUTP) [3]. Such transmitter preprocessing techniques in the form of precoding have been widely studied in [4], [5]. In [6] and [7], it has been shown that dirty paper coding (DPC) can be applied with geometric mean decomposition (GMD) to pre-subtract the interference before transmission. Several non-linear precoders including Tomlinson-Harashima precoder (THP) and linear precoders based on transmit zero-forcing (ZF) and minimum mean square error (MMSE) have been widely studied in [7], [8]. The ZF preprocessing matrix has been derived using different methods, in [9]. In [7], the multi-user transmission schemes invoke the block diagonalization method, while in [10], SVD based MUTP is investigated in the context flat-fading channels. Furthermore, [9] obviates the need for a post processing matrix, by suppressing multi-stream per-user interference at the transmitter. A number of joint transmission schemes have been reported for DL communications and the achievable system capacity has been approached with the aid of dirty paper coding [6].

Motivation for our work is based on the recent study by W. Liu, *et al.* [10]. In their work, they have shown that MUTP can perfectly eliminate MUI, but their investigations were primarily in the context of space division multiple access (SDMA) system over flat-fading channels [10]. However, frequency-selective channels can result in significant performance degradation in terms of achievable bit error rate (BER). Also, insufficient angle spread or a lack of rich scattering environment will cause fading to be correlated thus severely affecting the performance of a MIMO system. Hence, in this treatise, in contrast to previous work, we present the performance of MUTP assisted MIMO SDMA system with two power control regimes in the context of single cell multi-user scenario over correlated and uncorrelated frequency-selective fading channels for vehicular channel model pertaining to Long Term Evolution (LTE) standard [11] along with flat-fading channel.

The rest of the paper is organized as follows: Section II describes the system model of MIMO SDMA system with MUTP. Section III addresses the power allocation regimes for the considered system. Section IV details the performance results of our analysis and in section V conclusions are drawn.

2 System Configuration

In the considered system as shown in Fig.1, we assume that K users are arbitrarily distributed in a single cell for downlink communication (DL). Each of the K users is equipped with N_r receive antennas and the base station with N_t transmit antennas. In a multi-user scenario, the users conflict MUI, as well as multi-stream interference (MSI) and there will be $(K-1)N_r$ interfering signals arriving at each of the mobile stations (MSs). Furthermore, both the BS and the MS support MUTP based on SVD technique for mitigating these interferences.

The channel model parameters of the LTE vehicular profile detailed in Table 1 [11] is invoked for evaluating the performance of SVD based MUTP for single cell multi-user MIMO space division multiple access (SDMA) system over frequency-selective

fading for downlink communications. With the aid of the parameters defined in Table 1, the channel fading coefficient connecting the i th transmit antenna and the j th receive antenna can be expressed as,

$$h_{ji}(t) = \sum_{l=1}^L h_{ji}^l \delta(t - \tau_l) \tag{1}$$

where, h_{ji}^l is a zero-mean complex Gaussian random process with variance $p(\tau_l)$ and h_{ji}^l is correlated with other paths and channels. L denotes the total number of paths between any pair of transmit and receive antennas.

Table 1. Path Delays and Relative Power Levels- Vehicular Channel Model (LTE)

Path Number(l)	1	2	3	4	5	6	7	8	9
Delay (τ_l) (ns)	0	30	150	310	370	710	1090	1730	2510
Power $p(\tau_l)$ (dB)	0	-1.5	-1.4	-3.6	-0.6	-9.1	-7	-12	-16.9

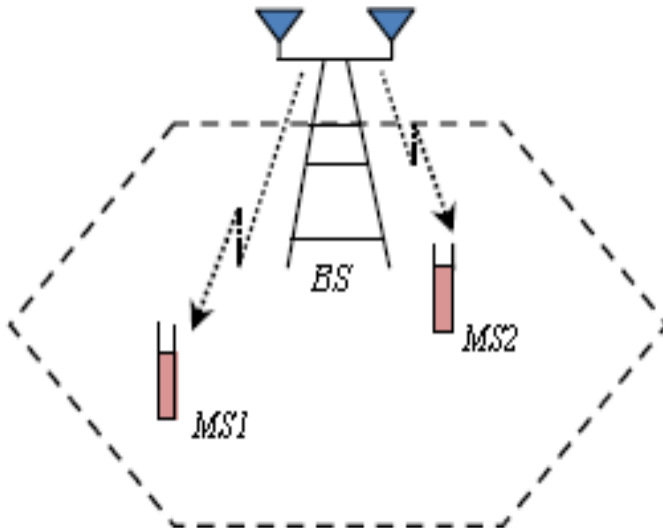


Fig. 1. MUI Elucidation

For a single cell case, let the N_r component vector transmitted to k th user be \mathbf{d}_k . It is premultiplied by $N_t \times N_r$ preprocessing matrix \mathbf{P}_k resulting in,

$$\mathbf{x}_k = \mathbf{P} \mathbf{d}_k \tag{2}$$

Here $\mathbf{P}_k = [\mathbf{V}_{ks}^H]^+$ and $\mathbf{d}_k = [\mathbf{d}_{k1} \ \mathbf{d}_{k2} \ \dots \ \mathbf{d}_{kN_r}]^T$ formulates the preprocessing and the downlink symbol vector respectively [10]. In the case of a SDMA system, after downlink preprocessing, the N_t component symbol matrix denoted by,

$$\mathbf{x} = \mathbf{P} \mathbf{d} \tag{3}$$

will be transmitted to all the K MSs, where,

$$\mathbf{d} = [\mathbf{d}_1^T, \mathbf{d}_2^T, \dots, \mathbf{d}_K^T]^T \tag{4}$$

denotes the KN_r component symbol vector and $\mathbf{P} = ([\mathbf{V}_{1s} | \mathbf{V}_{2s} | \dots | \mathbf{V}_{Ks}]^H)^+$ represents the $N_t \times KN_r$ component preprocessing matrix.

The received vector \mathbf{y}_k at the k th mobile station can be expressed as,

$$\mathbf{y}_k = \mathcal{H}_k \mathbf{x} + \mathbf{n}_k = \mathcal{H}_k \mathbf{P} \mathbf{d} + \mathbf{n}_k \tag{5}$$

$$= \underbrace{\mathcal{H}_k \mathbf{P} \mathbf{d}_k}_{\text{Desired Signal}} + \underbrace{\sum_{i=1, i \neq k}^K \mathcal{H}_k \mathbf{P} \mathbf{d}_i}_{\text{MUI}} + \mathbf{n}_k, \quad k = 1, 2, \dots, K$$

Upon carrying out the SVD on \mathcal{H}_k , we arrive at,

$$\mathcal{H}_k = \mathbf{U}_k [\mathbf{\Lambda}_k^{1/2}, 0] \mathbf{V}_k^H = \mathbf{U}_k [\mathbf{\Lambda}_k^{1/2}, 0] \begin{bmatrix} \mathbf{V}_{ks}^H \\ \mathbf{V}_{kn}^H \end{bmatrix} = \mathbf{U}_k \mathbf{\Lambda}_k^{1/2} \mathbf{V}_{ks}^H \tag{6}$$

where, \mathbf{U}_k and \mathbf{V}_k are the $(N_r \times N_r)$ and $(N_t \times N_t)$ unitary matrices, $\mathbf{\Lambda}_k$ represents $(N_r \times N_t)$ diagonal matrix containing the eigen values of $\mathcal{H}_k \mathcal{H}_k^H$, $(\cdot)^H$ denotes Hermitian transpose, $\mathbf{V}_{ks} = (N_t \times N_r)$ component matrix, constituting the eigen vectors corresponding to the nonzero eigen values of $\mathcal{H}_k^H \mathcal{H}_k$, $\mathbf{V}_{kn} = [N_t \times (N_t - N_r)]$ matrix, constituting the eigen vectors corresponding to the zero eigen values of $\mathcal{H}_k^H \mathcal{H}_k$.

Upon substituting (6) into (5), the received DL signal \mathbf{y}_k of the k th MS is given by,

$$\mathbf{y}_k = \mathbf{U}_k \mathbf{\Lambda}_k^{1/2} \mathbf{V}_{ks}^H \mathbf{P} \mathbf{d} + \mathbf{n}_k \tag{7}$$

The BS transmit preprocessing matrix \mathbf{P} is designed so as to effectively mitigate MUI. The post processing matrix that completely eliminates MUI is given by [10],

$$\mathbf{G}_k = [\mathbf{U}_k]^+ \tag{8}$$

where, $[\cdot]^+$ denotes the pseudo inverse of the matrix \mathbf{U}_k .

Applying this post processing matrix into the received vector (7), we arrive at,

$$\mathbf{y}_k = \mathbf{G}_k \mathbf{u}_k \Lambda_k^{1/2} \mathbf{d} + \mathbf{G}_k \mathbf{n}_k \quad (9)$$

Substituting the value for \mathbf{G}_k from (8) into (9), the received signal vector at the k th MS can be simplified to,

$$\mathbf{y}_k = \Lambda_k^{1/2} \mathbf{d}_k + \mathbf{G}_k \mathbf{n}_k \quad (10)$$

which results in the elimination of other user's data. Thus the SVD based MUTP completely eliminate MUI.

3 Power Allocation Policy

The power allocation regime considered in this work is based on [12], [13] and [14]. As addressed in the previous section, the preprocessing matrix is chosen in such a way that it completely eliminates MUI. Also, to realize (7), the preprocessing matrix \mathbf{P}_k of each user is formulated to satisfy the condition,

$$\mathbf{V}_{ks}^H \mathbf{P}_k = \boldsymbol{\gamma}_k \quad (11)$$

where,

$\boldsymbol{\gamma}_k = \text{diag}\{\gamma_{k1}, \gamma_{k2}, \dots, \gamma_{kN_r}\}$ designates the DL power control constraint of each user.

To satisfy (11), \mathbf{P}_k can be set to,

$$\mathbf{P}_k = [\mathbf{V}_{ks}^H]^+ \boldsymbol{\gamma}_k = \ddot{\mathbf{P}}_k \boldsymbol{\gamma}_k \quad (12)$$

The overall preprocessing matrix is given by,

$$\mathbf{P} = [\mathbf{V}_s^H]^+ \boldsymbol{\gamma} = \ddot{\mathbf{P}} \boldsymbol{\gamma} \quad (13)$$

where,

$\boldsymbol{\gamma} = \text{diag}\{\gamma_1, \gamma_2, \dots, \gamma_{KN_r}\} = \text{diag}\{\gamma_{11}, \dots, \gamma_{1N_1}, \dots, \gamma_{K1}, \dots, \gamma_{KN_r}\}$ is the overall DL power control constraint and $\ddot{\mathbf{P}} = [\ddot{\mathbf{P}}_1, \ddot{\mathbf{P}}_2, \dots, \ddot{\mathbf{P}}_K]$.

Upon substituting (13) into (7), N_r length received vector at the k th MS can be expressed as,

$$\mathbf{y}_k = \Lambda_k^{1/2} \boldsymbol{\gamma}_k \mathbf{d}_k + \mathbf{G}_k \mathbf{n}_k \quad (14)$$

In the DL communication, the power allocation is incorporated under the constraint that the total power $\mathcal{E}[\|\mathbf{P}\mathbf{d}\|^2]$ transmitted to all the MSs after preprocessing should not exceed the original transmission power $\mathcal{E}[\|\mathbf{d}\|^2]$ i.e.

$$\mathcal{E}[\|\mathbf{P}\mathbf{d}\|^2] \leq \mathcal{E}[\|\mathbf{d}\|^2] = KN_r \quad (15)$$

To be specific, power allocation for each user can be implemented under the constraint,

$$\mathcal{E}[\|\mathbf{P}_k \mathbf{d}_k\|^2] \leq \mathcal{E}[\|\mathbf{d}_k\|^2] \tag{16}$$

3.1 Joint Power Allocation Policy

This power allocation scheme maximizes the total capacity of the K DL users. Incorporating the ‘water-filling’ principle, it can be shown that the overall capacity of the K DL users can be maximized if $\{\gamma_{ii}\}$ is chosen to satisfy,

$$\gamma_{ii}^2 = \left(\mu - \frac{\sigma^2}{\lambda_{ii}} \right)^+, \quad i = 1, 2, \dots, KN_r \tag{17}$$

where $(\cdot)^+$ is $\max(\cdot, 0)$ and μ is chosen such that the power allocation constraint of (15) is satisfied. Hence, the maximum overall capacity normalized by the total number of receive antennas KN_r is given by,

$$C_{avg} = \mathcal{E} \left[\frac{1}{KN_r} \sum_{i=1}^{KN_r} \log_2 \left[1 + \frac{\lambda_{ii}}{\sigma^2} \left(\mu - \frac{\sigma^2}{\lambda_{ii}} \right)^+ \right] \right] \tag{18}$$

where $\mathcal{E}[\cdot]$ denotes Ergodic mean.

3.2 Individual Power Allocation Policy

This power allocation scheme aims at maximizing the capacity of each of the DL users. By invoking the ‘water-filling’ principle, the capacity of the k th MS can be maximized if $\{\gamma_{ki}\}$ is chosen to satisfy,

$$\gamma_{ki}^2 = \left(\mu_k - \frac{\sigma^2}{\lambda_{ki}} \right)^+ \quad i = 1, 2, \dots, N_r; \quad k = 1, 2, \dots, K \tag{19}$$

Hence, the maximum capacity of the k th MS normalized by the N_r receive antennas is given by,

$$C_{avg} = \mathcal{E} \left[\frac{1}{N_r} \sum_{i=1}^{N_r} \log_2 \left[1 + \frac{\lambda_{ki}}{\sigma^2} \left(\mu_k - \frac{\sigma^2}{\lambda_{ki}} \right)^+ \right] \right] \tag{20}$$

4 Performance Results

In this section, the simulation results for the performance of the MUTP assisted MIMO SDMA system ($N_t = 8$ and $N_r = 4$) deemed in this work for a single cell multi-user

scenario for correlated and uncorrelated fading channels are presented. The channel model considered here is based on LTE Vehicular channel specifications. Table I summarizes the channel model parameters. The modulation technique considered in this work is binary phase shift keying (BPSK). Perfect channel state information is assumed to be present both at the transmitter and receiver.

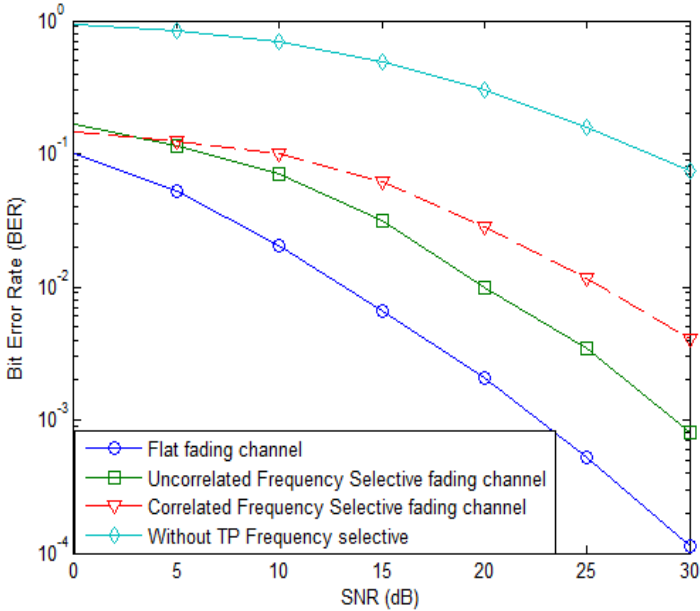


Fig. 2. BER performance of MUTP assisted MIMO SDMA system for joint power allocation policy based on the ‘water-filling’ principle with BPSK modulation

Fig. 2 illustrates the performance of MUTP assisted MIMO SDMA system for Vehicular channel model based on LTE specifications for joint power allocation policy. Here comparisons have been provided for flat-fading, uncorrelated frequency-selective fading and correlated frequency-selective fading with MUTP with correlation $\rho=0.4$, a Doppler shift of 70Hz as well as for an uncorrelated frequency-selective fading channel without employing MUTP. Monte Carlo simulation trials were carried out so as to evaluate the performance of the system and also for each SNR value, 5000 channel realizations were employed. It is discerned from the plot that MUTP based on SVD is effective in removing MUI. Also, it outperforms a system without MUTP.

Figs. 3 and 4 show the capacity per channel use performance of MUTP assisted MIMO SDMA system for Vehicular channel model based on LTE specifications for joint and individual power allocation policies respectively invoking ‘water-filling’ principle and equal power allocation. It is discerned from simulation results that a system invoking ‘water-filling’ principle results in a significant performance gain in terms of achievable capacity.

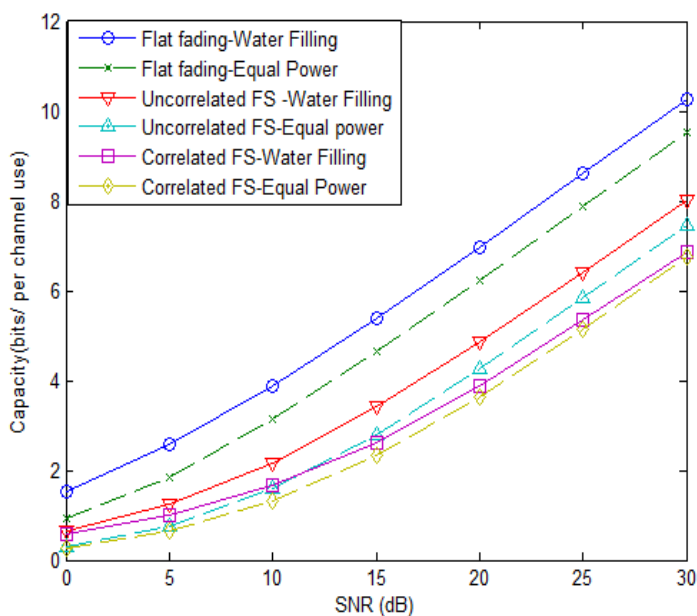


Fig. 3. Capacity per channel use of MUTP assisted MIMO SDMA system for joint power allocation policy based on ‘water-filling’ principle and equal power allocation policy

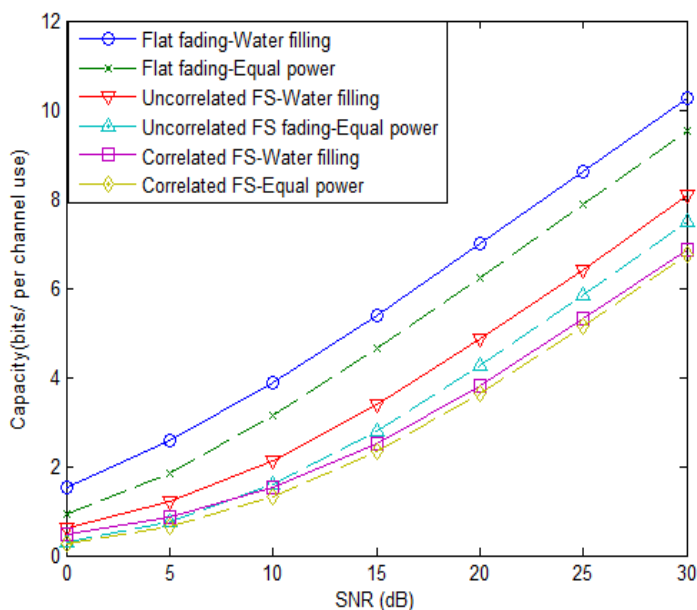


Fig. 4. Capacity per channel use of MUTP assisted MIMO SDMA system for individual power allocation policy based on ‘water filling’ principle and equal power allocation policy

5 Conclusion

In this paper, we investigated the performance of multi-user MIMO SDMA system with two power allocation regimes for DL communications with the aid of multi-user transmit preprocessing. Our studies show that a system aided by MUTP completely removes MUI resulting in significant performance improvement in terms of achievable BER and capacity. Also, the power allocation policies based on the ‘water-filling’ principle resulted in better capacity per channel use when compared to equal power allocation policy.

References

1. Foschini, G.J.: Layered space-time architecture for wireless communication in a fading environment when using multiple antennas. *Bell Lab.Tech. J.* 1(2), 41–59 (1996)
2. Verdú, S.: *Multi-user Detection*. Cambridge University Press, Cambridge (1998)
3. Vojcic, B., Jang, W.: Transmitter preprocessing in synchronous multi-user communications. *IEEE Trans. Commun.* 46, 1346–1355 (1998)
4. Fischer, H.R., Windpassinger, C., Lampe, A., Huber, J.: MIMO precoding for decentralized receivers. In: *Proc. IEEE ISIT 2002*, p. 496 (2002)
5. Shamai, S., Zaidel, B.M.: Enhancing the cellular downlink capacity via co-processing at the transmitting end. In: *IEEE VTS 53rd Vehicular Technology Conference*, Spring, Rhodes, Greece, pp. 1745–1749, 6–9 (2001)
6. Spencer, Q.H., Swindlehurst, A.L., Haardt, M.: Zero-forcing methods for downlink spatial multiplexing in multi-user MIMO channels. *IEEE Trans. on Signal Processing* 52, 461–471 (2004)
7. Kusume, K., Joham, M., Utschick, W., Bauch, G.: Efficient Tomlinson-Harashima Precoding for spatial multiplexing on flat MIMO channel. In: *Proc. IEEE ICC 2005*, vol. 3, pp. 2021–2025 (2005)
8. Joham, M., Utschick, W., Nossek, J.A.: Linear transmit processing in MIMO communications systems. *IEEE Trans. Signal Process.* 53(8), 2700–2712 (2005)
9. Yang, L.-L.: A Zero-Forcing Multi-user Transmitter Preprocessing Scheme for Downlink Communications. *IEEE Trans. Commun.* 56(6) (2008)
10. Liu, W., Yang, L.L., Hanzo, L.: SVD Assisted Joint Transmitter and Receiver Design for the Downlink of MIMO Systems. In: *IEEE Vehicular Technology Conference*, Spring (2008)
11. 3GPP (TR 36.803), Evolved universal terrestrial radio access (E-UTRA); user equipment (UE) radio transmission and reception (Release 8). Technical Specification, Sophia Antipolis, France (2007)
12. Telatar, I.E.: Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications* 10, 585–595 (1999)
13. Vucetic, B., Yuan, J.: *Space-Time Coding*. Wiley, Chichester (2003)
14. Morelli, M., Sanguinetti, L.: A novel prefiltering technique for downlink transmissions in TDD MC-CDMA systems. *IEEE Transactions on Wireless Communications* 4, 2064–2069 (2005)

A Fuzzy Neuro Clustering Based Vector Quantization for Face Recognition

Elizabeth B. Varghese and M. Wilsy

Department of Computer Science, University of Kerala
Kariavattom, Thiruvananthapuram-695581, Kerala, India
{eliza.b.varghese,wilsyphilipose}@gmail.com

Abstract. A face recognition system is a computer application for automatically identifying or verifying a person from a digital image or a video frame. In this paper, an improved codebook design method is proposed for Vector Quantization (VQ)-based face recognition which improves recognition accuracy. A codebook is created by combining a systematically organized codebook based on the classification of code patterns and another codebook created by Integrated Adaptive Fuzzy Clustering (IAFC) method. IAFC is a fuzzy neural network which incorporates a fuzzy learning rule into a neural network. The performance of proposed algorithm is demonstrated by using publicly available AT&T database and Yale database. Experimental results show face recognition using the proposed codebook is more efficient yielding a rate of 99.25% for AT & T and 98.18% for Yale which is higher than most of the existing methods

Keywords: Face Recognition, Vector Quantization, Codebook, Integrated Adaptive Fuzzy Clustering, Self Organization Map.

1 Introduction

In most situations face recognition is an effortless task for humans. Machine Recognition of faces from still and video images is emerging as an active research area spanning several disciplines such as image processing, pattern recognition, computer vision, neural networks etc [1]. Face recognition technology has numerous commercial and law enforcement applications [1]. Applications range from static matching of controlled format photographs such as passports, credit cards, photo IDs, driver's licenses to real time matching of surveillance video images [1].

A lot of algorithms have been proposed for solving face recognition problem [2]. Among these Principal Component Analysis (PCA) is the most common one. PCA [3] is used to represent a face in terms of an optimal coordinate system which contains the most significant eigenfaces where the mean square error is minimal. Fisherfaces [4] which use Linear Discriminant Analysis (LDA); Bayesian methods[5], which use a probabilistic distance metric; and SVM methods [6], which use a support vector machine as the classifier, are also present. Being able to offer potentially greater generalization through learning, neural networks have also been applied to face recognition [7]. Feature-based approach [8] uses the relationship between facial

features, such as the locations of eye, mouth and nose. Local Feature Analysis (LFA) [9], local autocorrelations and multiscale integration technique [10], etc are some of the methods.

Kotani et al. [11] have proposed a very simple yet highly reliable VQ-based face recognition method called VQ histogram method by using a systematically organized codebook for 4x4 blocks with 33 codevectors. Chen et al [12] proposed another face recognition system based on an optimized codebook which consists of a systematically organized codebook and a codebook created by Kohonen's Self Organizing Maps (SOM) [16].

VQ algorithm [13] is well known in the field of image compression. A codebook is very important since it directly affects the quality of VQ processing. In [12] an optimized codebook is created based on classification of code patterns and SOM. The Kohonen self-organizing feature map has to assume the number of clusters a priori and to initialize the cluster centroids. SOM guarantee convergence of weights by ensuring decrease in learning rates with time. Such learning rates, however, do not consider the similarity of the input pattern to the prototype of the corresponding cluster [17].

In this paper an improved codebook design method for VQ-based face recognition is proposed. At first a systematically organized codebook is created based on the distribution of code patterns [12], and then another codebook with the same size is created using Integrated Adaptive Fuzzy Clustering Method (IAFC) [17]. IAFC addresses the problems associated with SOM. In IAFC a fuzzy membership value is incorporated in the learning rule. This fuzzy membership value of the input pattern provides additional information for correct categorization of the input patterns. Moreover IAFC does not assume the number of clusters in the data set a priori, but updates it during processing of data [17].

The two codebooks are combined to form a single codebook which consists of 2x2 codevectors. By applying VQ the dimensionality of the faces are reduced. The histograms of the training images are created from the codevectors. This is considered as the personal identification information. It can represent the features of the facial images more adequately. The training set consists of 200 images from AT & T and 75 images from Yale. A recognition rate of 99.25% and 98.18% are obtained for AT & T and Yale respectively.

The rest of the paper is organized as follows: Related Work is discussed in section 2. Proposed face recognition system based on systematically organized codebook and IAFC is explained in section 3. Proposed Adaptive Codebook design is discussed in section 4. Experimental results are presented and discussed in section 5. Conclusions are given in section 6.

2 Related Works

Face recognition has become a popular area of research in computer vision and one of the most successful applications of image analysis and understanding. A general statement of the face recognition problem can be formulated as follows: Given still or video images of a scene, identify or verify one or more persons in the scene using a stored database of faces.

Kotani et al. [11] have proposed a very simple yet highly reliable VQ-based face recognition method called VQ histogram method. In VQ Histogram method [11], a special codebook was used, which is systematically organized for 4x4 blocks with 33 codevectors having monotonic intensity variation without DC component. A low-pass filtering is applied to the face image using a 2-D moving average filter. This low-pass filtering is essential for reducing high-frequency noise and extracting most effective low frequency component for recognition. The image is then divided into small 4x4 blocks. The minimum intensity in the individual block is subtracted from each pixel in the block. Only the intensity variation in the block is extracted by this process. This is very effective for minimizing the effect of overall brightness variations[11].

Vector quantization is then applied to intensity-variation blocks (vectors) by using a codebook which is prepared in advance. The most similar (matched) codevector to the input block is selected. After performing VQ for all blocks divided from a face image, matched frequencies for each codevector are counted and histogram is generated. This histogram becomes the feature vector of human face. In the registration procedure, this histogram is saved in a database as personal identification information. In the recognition procedure, the histogram made from an input facial image is compared with registered individual histograms and the best match is output as a recognition result. Manhattan distance between histograms is used as a matching measure.

Codebook which consists of typical feature patterns for representing the features of face image is very important. In this method the codebook consists of 32 codevectors and each codevector is of size 4x4[11]. This is created by changing the direction (8 different directions) and the range of intensity variation (Step values are 2, 6, 10, and 20). By adding one codevector having no intensity variation, complete codebook is organized [11].

Chen et al [12] proposed a face recognition system based on an optimized codebook which consists of a systematically organized codebook and a codebook created by Kohonen's Self Organizing Maps (SOM) [16].The optimized codebook is a combination of two codebooks. The first codebook is created by code classification and the second codebook is obtained by Kohonen's Self Organizing Map.

In [12] a second codebook is created using Kohonen's SOM [16]. The self-organizing feature map self-organizes its weights by incremental adjustments in proportion to the difference between the current weight and the input pattern. In real applications, it is often difficult to assume the number of clusters present in many real data sets. And, different initial conditions result in different results. This neural network also requires considerable time to train [17].So a better method for codebook design is needed. In the proposed method the second codebook is created using Integrated Adaptive Fuzzy Clustering (IAFC) [17] of the same size N. In IAFC a fuzzy membership value is incorporated in the learning rule. This fuzzy membership value helps in the correct categorization of the input patterns. Also IAFC does not assume the number of clusters in the data set a priori, but updates it during processing of data [17].Codebook design using IAFC is explained in section 4. The results comparing the existing method and the proposed approach are explained in section 5.

3 Proposed Face Recognition System

The proposed method starts with the pre-processing step. During pre-processing each face image in the training set is processed to get the intensity variation vectors. Pre-processing is explained in detail in section 3.1. Vector Quantization (VQ) is then applied to these vectors by using the proposed codebook which is a combination of two codebooks. The first codebook is developed by code classification [12]. The second codebook is created using IAFC [17]. During VQ the most similar codevector to each input block is selected.

After performing VQ, matched frequencies for each codevector are counted and histogram is saved in the database as Personal Identification Information. This histogram becomes the feature vector of the human face. Thus histogram is a very effective personal feature for discriminating between persons.

In the recognition procedure, the histogram made from an input test image is compared with registered individual histograms and the best match is output as the recognition result. Manhattan distance between the histograms is used as the matching measure. Figure 1 shows the block diagram of the proposed method.

Codebook which consists of typical feature patterns for representing the features of the face image is important. The proposed codebook design is explained in section 4.

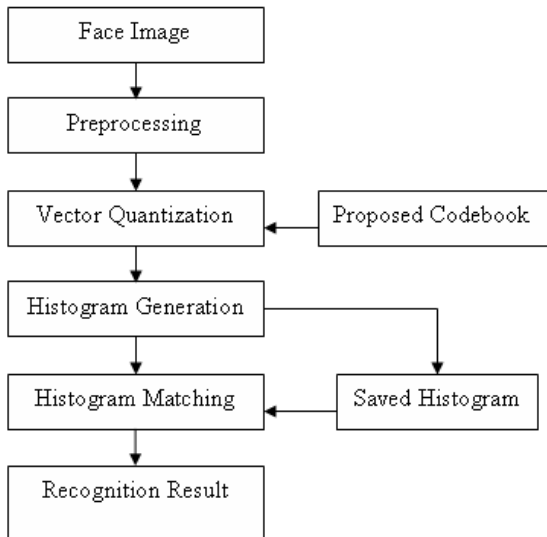


Fig. 1. Proposed Face Recognition System

3.1 Pre-processing

During preprocessing initially a low pass filtering is carried out using a simple 2D mean filter [11]. A low pass filtering is effective for eliminating the noise component.

By applying the filter, detailed facial features degrading recognition performance such as wrinkles and local hairstyle, are excluded. Only the important personal features, such as the rough shape of facial parts can be extracted.

The image is then divided into 2x2 overlapping blocks. Minimum intensity of the individual block is subtracted from each pixel in the block. Minimum intensity subtraction effectively excludes dc and vary low frequency component, such as shade variations due to small variations in lighting conditions and retains only the relevant information for distinguishing images. Figure 2 shows the steps in pre-processing.

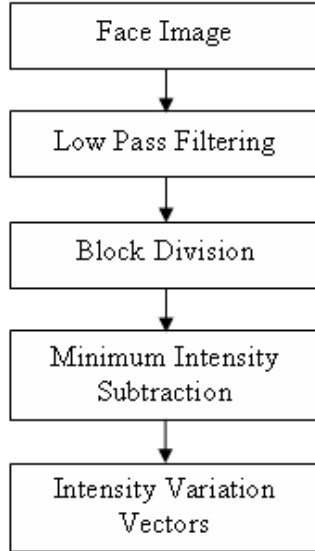


Fig. 2. Pre-processing Steps

4 The Proposed Adaptive Codebook Design

The proposed codebook for VQ is obtained from two codebooks. One codebook with size N is obtained by code classification [12] which is explained in section 4.1. This codebook is created by the variation in the intensity of the code patterns. It does not consider the intensity variations of the face images. So it cannot represent the facial features efficiently. So a second codebook is needed from the facial images to represent the facial features more efficiently.

In [12] a second codebook is created using Kohonen's SOM [16]. The self-organizing feature map self-organizes its weights by incremental adjustments in proportion to the difference between the current weight and the input pattern. In real applications, it is often difficult to assume the number of clusters present in many real data sets. And, different initial conditions result in different results. This neural network also requires considerable time to train [17]. In the proposed method the second codebook is created using IAFC [17] of the same size N . In IAFC a fuzzy membership value is incorporated in the learning rule. This fuzzy membership value

helps in the correct categorization of the input patterns. Also IAFC does not assume the number of clusters in the data set a priori, but updates it during processing of data [17].

Thus a codebook of size $2N$ is obtained. To reduce the size of the codebook from $2N$ to N , the face images in the training set is preprocessed to get the intensity variation vectors. Then VQ is applied to these intensity variation vectors, matched frequencies of each codevector are counted and histogram of each face image is generated. Then the average histogram of all images is calculated. Next, the frequencies of individual codevectors are sorted. From this sorted $2N$ codevectors, only the high frequency N codevectors are selected. Thus, the final codebook consisting of 2×2 codevectors is generated.

4.1 Codebook Generated by Code Classification

Nakayama et al. [14] have developed complete classification method for 2×2 codebook design in image compression. Figure 3 shows all categories for the 2×2 image block patterns without considering the location of pixels. In a 2×2 block, pixel intensities are marked by alphabet ‘a’, ‘b’, ‘c’, ‘d’, and $a > b > c > d$ is prescribed. In ref. [14], it was found that the number of typical patterns for all 2×2 image block is only 11. The number of varieties in pixel arrangement of each 2×2 typical pattern is also shown in figure 3. That means the total number of image patterns for 2×2 pixel blocks is theoretically only 75. By the similar consideration, Chen et al. [15] classified and analyzed the code patterns in the face images. They found that in all filter size, the number of code patterns belong to categories 7, 10, and 11 are very few. It means such code patterns are almost not used in face images. Based on this result, a new codebook for 2×2 code patterns is created, and the rules of codebook creation are as follows.

- Change the intensity difference among the blocks to from 1 to 10.
- Create very small intensity variation codes. The total number of patterns is 16
- Create code patterns of category no: 2,3,4,5,6,8 and 9
- Add one code pattern having no intensity variation

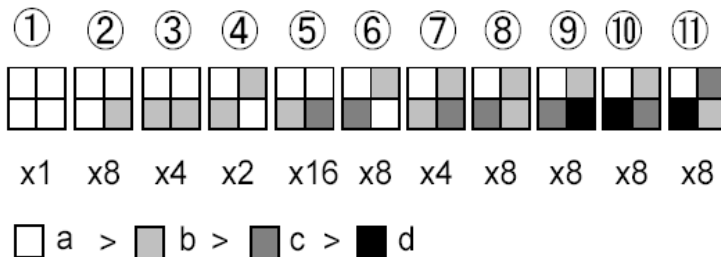


Fig. 3. Categories of 2×2 code patterns

4.2 Codebook Design Using IAFC

The IAFC model is a fuzzy neural network which incorporates a fuzzy learning rule into a neural network [17]. The learning rule, developed in IAFC, incorporates a

fuzzy membership value (μ_i), an intracluster membership value (π), and a function of the number of iterations ($f(l)$) into a Kohonen-type learning rule. The number of clusters in IAFC is updated dynamically. An intracluster membership value (π) is decided by the distance between the input pattern and the centroid of the chosen cluster. The combination of the π -function and a function of the number of iterations guarantee weights to converge. The IAFC model incorporates a similarity measure that includes a fuzzy membership value into the Euclidean distance. The similarity measure considers not only the distance between the input data point and the centroid of a winning cluster but also the relative location of the input point to the existing cluster centroids as the degree of similarity. Thus, it gives more flexibility to the shape of clusters formed [17].

IAFC consists of three major procedures: deciding a winning cluster, performing the vigilance test, and updating the centroid of a winning cluster. The input pattern X is normalized prior to presentation to the fuzzy neural network and this normalized input pattern is fed to the fuzzy neural network in parallel to the input pattern. A dot-product operation used to find the winner is shown below

$$I \bullet b_i = \frac{X \bullet V_i}{\|X\| \bullet \|V_i\|} \tag{1}$$

Where b_i is the normalized weights from the input neurons to the i^{th} output cluster, and V_i is the i^{th} cluster centroid. The output neuron that receives the largest value for the equation (1) wins the competition. In this process, the winner is decided by the angle between the input pattern and the centroids of clusters. This can cause misclassifications because a cluster of which the direction of the centroid vector has the smallest angle with the input vector wins the competition even though its centroid is located farther from the input pattern than other cluster centroids. In such a case, Euclidean distance can be used as a better similarity measure to determine a winner. However, cluster centroids cannot approach appropriate locations during the early stage of learning, thus causing poor performance of clustering algorithms. To prevent both problems, the IAFC algorithm uses a combined similarity measure to decide a winner.

After deciding a winner by the dot product, the IAFC algorithm compares the fuzzy membership value, μ_i of the input pattern in the winning cluster with the parameter σ that user can decide as a threshold of the fuzzy membership value. If the fuzzy membership value is less than the value of the parameter σ , the angle between the input pattern and the cluster centroid is the dominant similarity measure to decide a winner. On the other hand, if the parameter σ is high, the Euclidean distance between the input pattern and the cluster centroid is the dominant similarity measure to decide a winner. After selecting a winning cluster, IAFC performs the vigilance test according to the criterion:

$$e^{\gamma \mu_i} \|X - V_i\| \leq \tau \tag{2}$$

Where γ is a multiplicative factor that controls the shape of clusters, X is the input pattern, V_i is the centroid of the i^{th} winning cluster, τ is the vigilance parameter and the value of γ is normally chosen to be 1[17]. The fuzzy membership value μ_i , is calculated as follows:

$$\mu_i = \frac{\left(\frac{1}{\|X - V_i\|^2} \right)^{1/m-1}}{\sum_{j=1}^n \left(\frac{1}{\|X - V_j\|^2} \right)^{1/m-1}} \tag{3}$$

Where m is a weight exponent which is experimentally set to 2 [17] and n is the number of clusters. If a winning cluster satisfies the vigilance criterion, the centroid of a winning cluster is updated as follows:

$$V_i^{new} = V_i^{old} + \lambda_{fuzzy} (X - V_i^{old}) \tag{4}$$

Where λ_{fuzzy} is $[f(l), \pi(X; v_i^{old}, \tau), \mu_i^2]$. $f(l)$ is a function of number of iterations, l being the number of iterations, and π decides the intra-cluster membership value of the input pattern X in the i^{th} cluster as:

$$\pi(X; V_i^{old}; \tau) = \begin{cases} 1 - 2 \left(\frac{\|X - V_i^{old}\|}{\tau} \right)^2, & 0 \leq \|X - V_i^{old}\| \leq \tau/2 \\ 2 \left(1 - \frac{\|X - V_i^{old}\|}{\tau} \right)^2, & \tau/2 \leq \|X - V_i^{old}\| \leq \tau \\ 0, & \|X - V_i^{old}\| \geq \tau \end{cases} \tag{5}$$

And

$$f(l) = \frac{1}{k(l-1) + 1} \tag{6}$$

Where k is a constant

IAFC algorithm for the codebook design can be summarized by the following steps:

1. Initialize parameters τ and σ .
2. Transform the facial images in dataset to intensity variation vectors, and combine all vectors together into one training set
3. Initialize the weight vectors with the intensity variation vectors.

4. Select a new input pattern from the training set.
5. Decide a winning cluster (best matching codevector) using the combined similarity measure.
 - 5(a) Calculate the dot product between the normalized input pattern and the normalized weight vector using equation (1)
 - 5(b) Calculate the Euclidean distance between the input pattern and the weight vector.
6. Calculate the fuzzy membership value, μ_i of the input pattern in the winning cluster using equation (3).
 - 6(a) If $\mu_i < \sigma$,
 the winner neuron is selected from 5(a)
 else
 the winner neuron is selected from 5(b)
7. Perform the vigilance test using equation (2). If the criterion is satisfied then update the weights using equation (4)
8. If all the input patterns are processed go to step 9 else go to step 4
9. Stop



Fig. 4. Sample Images from AT & T Database

5 Experimental Results and Discussions

Publicly available AT & T database [18] and Yale database [19] are used for recognition experiments. The AT & T database contains 400 images in pgm format of 40 persons. There are 10 different images of each of 40 distinct 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) [18]. Figure 4 displays some images from the AT & T database. The Yale Face Database contains 165 grayscale images in GIF format of 15 individuals. There are 11 images per

subject, one per different facial expression or configuration: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink [19]. Sample images from the Yale database is shown in figure 5.



Fig. 5. Sample Images from Yale Database

Five images were selected from each person’s 10 images (in the case of AT & T) and from 11 images (in the case of Yale) for training purpose. The remaining images are used for testing. So 200 images from AT & T are used for training and the remaining 200 is used for testing. But in the case of Yale 75 images are used for training and 90 images are used for testing.

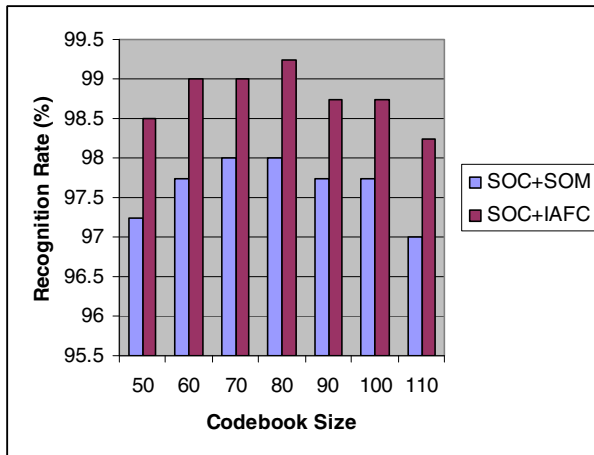


Fig. 6. Comparison of the Recognition Rate Using AT & T Database

It is necessary to choose a suitable size for the codebook. As the codebook size is large, number of codevectors increases, the resolution of histogram may become so

sensitive that noise corrupted codevectors may distort the histogram. On the contrary, if the number of codevectors is small, the histogram cannot sufficiently discriminate between faces. Recognition rate was observed for the codebook sizes 50, 60, 70, 80, 90, 100 and 110. It is clear from the figures 6 and 7 that the best performance is obtained with a codebook of size 80 for AT & T and with sizes 70 and 80 for Yale. With that size a recognition rate of 99.25% is obtained for AT & T and 98.18 % is obtained for Yale.

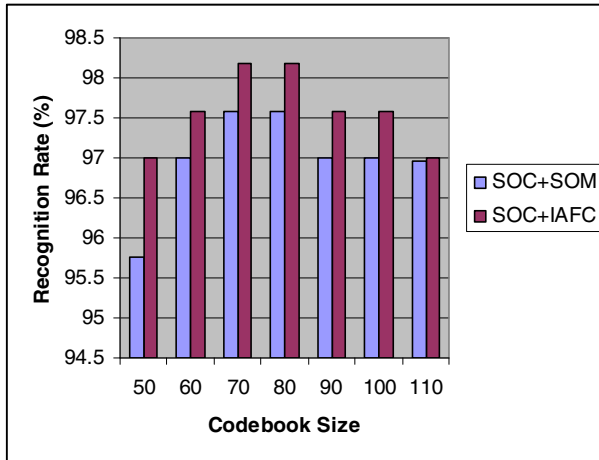


Fig. 7. Comparison of the Recognition Rate Using Yale Database

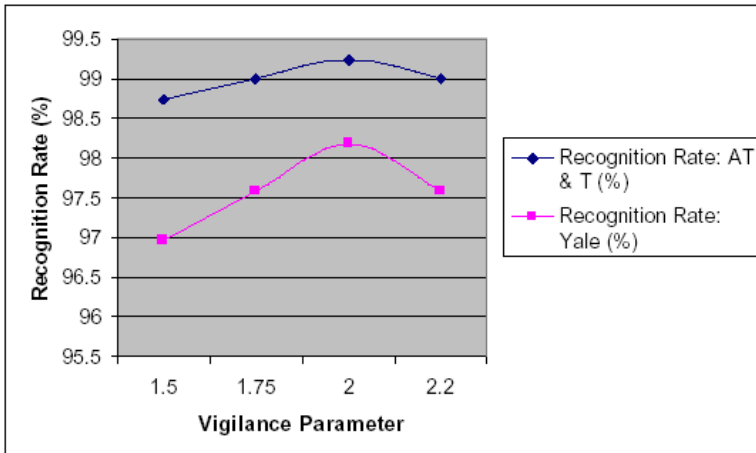


Fig. 8. Recognition Rate for different values of Vigilance Parameter (τ)

Figures 6 and 7 also show a comparison between the proposed approach and the existing method with SOM [16]. It is clear from the figures that in all the cases the

proposed method yields a better recognition result than the existing method. It can be said that the proposed codebook is more efficient in representing the facial features than the existing method using systematically organized codebook and SOM [12].

Experiments are also done by varying the values for the vigilance parameter, τ . The results are shown in figure 9. It is clear from the figure that for the value, $\tau = 2$, a higher recognition rate is achieved. In all the cases the codebook size is 80 and the value for σ is 0.5.

6 Conclusion

In this paper a new, A Fuzzy Neuro Clustering based VQ for Face Recognition, method is presented. A simple and efficient codebook design algorithm for face recognition using vector quantization is proposed. The codebook is created from two different codebooks. One codebook is created by code classification. The other codebook is created from the face images using Integrated Adaptive Fuzzy Clustering (IAFC). To create the codebook, the face images are divided into 2x2 blocks with a fixed codebook size. A good initial codebook is created from these blocks by code classification. The resultant initial codebook is combined with the codebook which is created by IAFC to become the final codebook. Utilizing such a codebook of size 80, a recognition rate of 99.25% is obtained for the AT & T database. For codebook sizes 70 and 80, a recognition rate of 98.18 % is obtained for Yale database. The results are more efficient than the existing method which consists of an optimized codebook with systematically organized codebook and Kohonen's SOM.

References

1. Chellappa, R., Wilson, C.L., Sirohey, S.: Human and machine recognition of faces: a survey. *Proc. IEEE* 83(5), 705–740 (1995)
2. Li, S.Z., Jain, A.K.: *Handbook of Face Recognition*. Springer, New York (2005)
3. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3(1), 71–86 (1991)
4. Belhumeur, P.N., Hespanh, J.P., Kriegman, D.J.: Eigenfaces vs Fisherfaces. *Recognition using class specific linear projection*. *IEEE Trans. Pattern Anal. Machine Intell.* 19, 711–720 (1997)
5. Moghaddam, B., Nastar, C., Pentland, A.: A Bayesian similarity measure for direct image matching. In: *Proceedings, International Conference on Pattern Recognition* (1996)
6. Phillips, P.J.: Support vector machines applied to face recognition. *Advanced Neural Information Processing Systems* 11, 803–809 (1998)
7. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: *Face Recognition: A Literature Survey*. *ACM Computing Surveys* 35(4), 399–458 (2003)
8. Brunelli, R., Poggio, T.: Face recognition: features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15(10), 1042–1052 (1993)
9. Penev, P.S., Atick, J.J.: Local Feature Analysis: A general statistical theory for object representation. *Network: Computation in Neural Systems* 7(3), 477–500 (1996)
10. Goudail, F., Lange, E., Iwamoto, T., Kyuma, K., Otsu, N.: Face recognition system using local autocorrelations and multiscale integration. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 18(10), 1024–1028 (1996)

11. Kotani, K., Chen, Q., Ohmi, T.: Face recognition using vector quantization histogram method. In: Proceedings of the 2002 Int. Conf. on Image Processing vol. II(III), pp. II-105-II-108 (2002)
12. Chen, Q., Kotani, K., Lee, F.F., Ohmi, T.: A VQ based fast face recognition algorithmn using optimized codebook. In: Proceedings of the 2008 Int. Conf. on Wavelet Analysis and Pattern Recognition (2008)
13. Sayood, K.: Introduction to Data Compression. Morgan Kaufmann, San Francisco (2000)
14. Nakayama, T., Konda, M., Takeuchi, K., Kotani, K., Ohmi, T.: Still image compression with adaptive resolution vector quantization technique. Int. Journal of Intelligent Automation and Soft Computing 10(2), 155-166 (2004)
15. Chen, Q., Kotani, K., Lee, F.F., Ohmi, T.: Face recognition using codebook designed by code classification. In: IEEE Int. Conf. on Signal and Image Processing, pp. 397-401 (2006)
16. Kohonen, T.: The Self-Organizing Maps. Proceedings of the IEEE 78(9) (September 1990)
17. Kim, Y.S., Mitra, S.: An adaptive integrated fuzzy clustering model for pattern recognition. Journal Fuzzy Sets and Systems (65), 297-310 (1994)
18. AT & T. The Database of Faces,
<http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>
19. Yale Face Database,
<http://cvc.yale.edu/projects/yalefaces/yalefaces.html>

3D Face Recognition Using Orientation Maps

B.H. Shekar¹, N. Harivinod¹, and M. Sharmila Kumari²

¹ Department of Computer Science, Mangalore University, Karnataka, India
{bhshekar,harivinodn}@gmail.com

² Department of Computer Science and Engineering, P.A. College of Engineering,
Mangalore, Karnataka, India
sharmilabp@gmail.com

Abstract. In this work we present a new 3D face recognition method based on orientation maps. The proposed model consists of a method for extracting distinctive features from range images of face that can be used to perform reliable matching between different poses of a face. For a 3D face scan, range image is computed and the potential interest points are identified by searching at all scales. Based on the stability of the interest point, significant points are extracted. For each significant point, we compute the significant point descriptor which consists of vector made of values from the convolved orientation maps located on concentric circles centred on the significant point, and where the amount of Gaussian smoothing is proportional to the radii of the circles. Experiments have been conducted on the standard 3D face image database. Experiments show that the newly proposed method provides higher recognition rate compared to other existing contemporary models developed for 3D face recognition.

Keywords: Range Image, Local descriptor, Orientation map, 3D face recognition.

1 Introduction

Automatic face recognition systems have wide range of applications in access control, surveillance, personal identification and pervasive computing. Although 2D intensity based face recognition is easy to acquire and process, it is vulnerable to the change of pose and illumination. Hence much of the research in the recent days focused on 3D model based face recognition. 3D model based representation of faces provides invariance to illumination and pose changes.

The 3D model based face recognition techniques are categorized based on the global information or local information of range images. In [10], Heseltine et al. applied the principle component analysis directly to the range images and used the Euclidean distance to measure similarities. In [11], Heshner et al. made statistical analysis using principle component analysis and independent component analysis, and then impose probability models on the coefficients. Achermann et al. [2] utilized the eigen face and Hidden Markov Model for recognition on range images. In [16], Moreno et al. analyze range image based face recognition on

three classes: full, upper-half and left-side facial depth map. Two-dimensional FLD based face recognition was proposed by Guru and Vikram [9].

On the other hand, we have seen the local feature based point descriptors such as *Scale Invariant Feature Transform(SIFT)* and *Speeded Up Robust Features(SURF)*, that are widely used in the area of face recognition. It is shown that these features are invariant to affine transformations and illumination changes. In [15], Lowe introduced SIFT to perform matching between different views of an object. This 128 dimensional descriptor is based on the local image gradient, transformed according to the orientation of the significant point to provide orientation invariance. Mian et al. [17] used SIFT descriptors for face recognition under illumination and expression variations by combining the 2D and 3D local feature. In [8][14], Guo et al. and Lo et al. used 2.5D SIFT Descriptor for facial feature extraction in range images. In [7], Geng and Jiangs use variant of SIFT called volume-SIFT and partial-descriptor-SIFT for face recognition on 2D faces. In [13], the SIFT descriptors are computed at specific points on a regular grid of the face image and robustness to illumination variations is achieved. Bay et al. in [3] presented SURF as a detector and descriptor. Kim and Dahyot [12] uses SURF for face components detection using support vector machines. Yunqi et al. used SURF for 2D and 3D face recognition [20][21]. In [1], An et al. used SURF for face detection and recognition with SURF for human-robot Interaction. Tola in [18] proposed DAISY, an efficient dense descriptor applied to wide-baseline stereo. This descriptor is inspired by SIFT and is computationally efficient. Velardo et al. in [19] applied this work to face recognition on 2D images.

Most of the works discussed above use the 3D data in the form of point clouds. Here, the 3D sensors, used for face capture, produces 2.5D information [6]. This data can be easily projected to a 2D image plane and is called *depth image* or *range image*. A range image is a 2D image in which each pixel represents the distance from a point of the face surface to a plane. It is observed that the range image construction normally preceded with a pose normalization phase in order to transform faces to a frontal configuration. We use *Iterative Closest Point (ICP)* algorithm for this purpose. Using interpolation methods, irregularly sampled 3D points are converted into regular (x, y) grid. Instead of comparing the whole image, we choose significant point comparison to find the match between two range images. The significant points are detected using Hessian based detector followed by the significant point descriptor computation. The significant point descriptor for all significant points are computed. The test face is matched with all images in the dataset. Two range images are said to be matched if they have maximum number of matching significant points.

The rest of the paper is organized as follows: The methodology of proposed 3D face recognition model is given in section 2. It describes the preprocessing, significant point extraction, significant point descriptor computation and matching. Database description and experimental results are provided in section 3. Conclusion is presented in section 4.

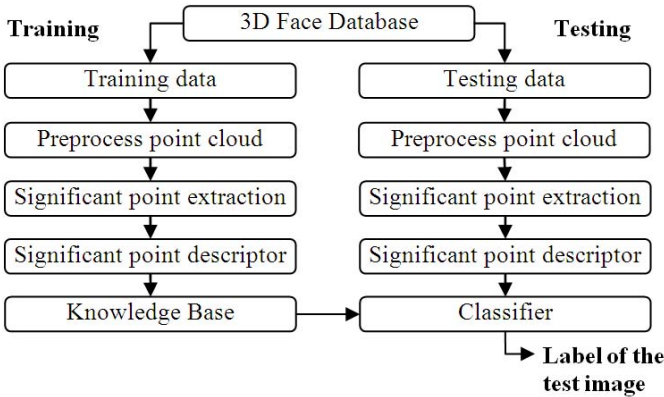


Fig. 1. Block diagram of the proposed 3D face recognition system

2 Methodology

The proposed method mainly consists of following steps: significant point extraction from the preprocessed range image, significant point descriptor computation and matching. The block diagram of the proposed face recognition system is given in Figure 1.

2.1 Preprocessing

The 3D face point clouds are obtained from a 3D face database [6]. Since the face scans of a person differ with pose, they need to be aligned. We use ICP [4] to do the registration. It performs the registration operation automatically without any human intervention. In our experiments all the scans of persons are aligned with the first frontal scan of that person.

The face scans usually do not contain corresponding data with respect to the each grid in the range image. So the scattered data is linearly interpolated. We interpolated the data using a Delaunay triangulation. Then the nose tip is identified. In our implementation, the point nearer to the scanner is identified as nose tip which posses highest depth value. Since all the scans are aligned to the frontal scan, the value with the highest depth value corresponds to the nose tip. Keeping nose tip as the center, the range image is cropped elliptically. The range images before preprocessing are shown in the left side of Figure 2 and on the right side are the range images obtained due to preprocessing.

2.2 Significant Point Extraction

The significant points are extracted using the SURF detector [3] which is based on the determinant of the Hessian matrix. We have used integral images to reduce the computational burden of the significant point extraction. Integral image helps in fast computation of sum of the rectangular region in an image.

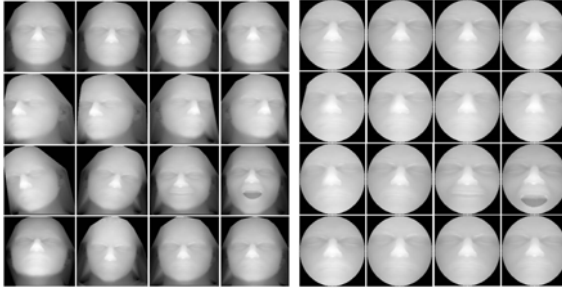


Fig. 2. All range images of a person before(left) and after(right) preprocessing of the point cloud

Due to this box type convolution filters are computed quickly. For an image I , the integral image I_Σ is defined as follows.

$$I_\Sigma(x, y) = \sum_{i=1}^{i \leq x} \sum_{j=1}^{j \leq y} I(i, j) \tag{1}$$

where x and y represent the row and column number of a pixel in image I . $I_\Sigma(x, y)$ is the sum of all $I(x, y)$ terms to the left and above the pixel (x, y) . Using integral images, the sum of the rectangular region in the image I can be calculated using three arithmetic operations [3].

Hessian based SURF detector is chosen because it is more stable and repeatable. It uses the determinant of the Hessian matrix. For a point $p = (x, y)$ in image I , the Hessian matrix $H(p, \sigma)$ is the matrix of partial derivatives of the image I , in the following form,

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(p, \sigma) & L_{xy}(p, \sigma) \\ L_{xy}(p, \sigma) & L_{yy}(p, \sigma) \end{bmatrix} \tag{2}$$

where $L_{xx}(p, \sigma)$ is the convolution of the Gaussian second order derivative with the image I in point p , and similarly for $L_{xy}(p, \sigma)$ and $L_{yy}(p, \sigma)$. Gaussians are widely used for scale-space analysis [3]. Its discrete formations are used in actual implementation. The first two diagrams in Figure 3 show discrete Gaussian second order derivative. The Hessian matrix is computed using the box filters which approximate second order Gaussian derivatives. The last two diagrams in Figure 3 illustrate the same. The 9×9 box filters in Figure 3 are approximations of a Gaussian with $\Sigma = 1.2$ and represent the lowest scale for computing the blob response maps. Suppose D_{xx} , D_{yy} and D_{xy} are the approximations to I_{xx} , I_{yy} and I_{xy} , the determinant of H_{approx} is computed as,

$$\det(H_{approx}) = D_{xx}D_{yy} - (wD_{xy})^2 \tag{3}$$

where w is the weight of the filter response used to balance the expression for the Hessian's determinant. The approximated determinant of the Hessian represents the blob response in the image at location p .

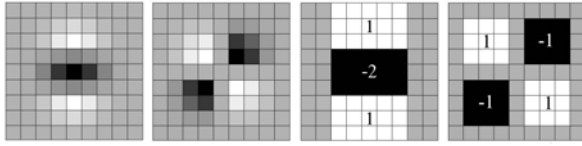


Fig. 3. The first two images represent the discretised Gaussian second order partial derivative in y (L_{yy}) and xy direction (L_{xy}) respectively. The last two images represent the approximation for the second order Gaussian partial derivative in y (D_{yy}) and xy direction (D_{xy}). The gray regions assumed to be zero. (Image Courtesy [3]).

For a given image, the potential interest points are identified by searching at all scales and based on the stability of the interest point, significant points are extracted. Lowe [15] implemented the scale spaces using the image pyramid. Each higher layer in the pyramid contains subsampled, smoothed image with Gaussian kernels. Here the image size is reduced due to subsampling. But Bay et al. [3], constructed the scale space by increasing the filter size keeping the image size fixed. They used the box filters and integral images. Box filters of any size can be applied on the original image at same speed. We employ the technique proposed in [3].

The scale space construction followed by the non-maximum suppression in the neighborhood is performed. The maxima of the determinant of the Hessian matrix are then interpolated in scale and image space with the method by Brown et al. [5]. This represents the significant point in the range image.

2.3 Significant Point Descriptor Computation

Significant point descriptor computation is similar to descriptor which was introduced by Tola et al. in [18] for matching wide-baseline image pairs. In our work, for each significant point, we compute the descriptor which consists of vector made of values from the convolved orientation maps located on concentric circles centered on the significant point, and where the amount of Gaussian smoothing is proportional to the radii of the circles.

Significant point descriptor building process is divided into three stages: computing orientation maps, convolving orientation maps with Gaussian kernels, and concatenating significant point descriptor by reading the values from convolved response maps.

For each significant point in the image I , we first compute eight orientation maps, one for each quantized direction. Orientation map $G_i(u, v)$ for the significant point (u, v) , equals the image gradient at (u, v) in the direction i , if it is greater than zero else it is equal to zero. $G_i = (\delta I / \delta i)^+$, $1 \leq i \leq H$, where H is the total orientation maps and $(.)^+$ is the operator such that $(a)^+ = \max(a, 0)$.

Each orientation map is then convolved several times with Gaussian kernels of different Σ values to obtain convolved orientation maps for different sized

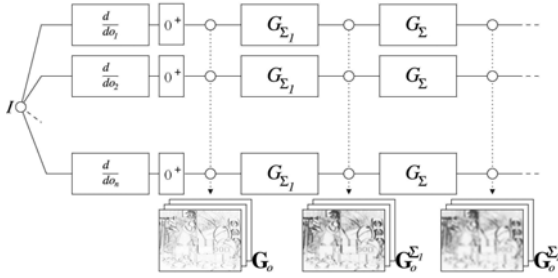


Fig. 4. Original image (I), Orientation maps (G_o) and Convolved orientation maps ($G_o^{\Sigma^i}$). (Courtesy [18]).

regions as $G_i^\Sigma = G_\Sigma * (\delta I / \delta i)^+$ with G_Σ a Gaussian Kernel. Different Σ s are used to control the size of the region. This can be done efficiently by computing these convolutions recursively. Figure 4 shows these computations.

$$G_i^{\Sigma_2} = G_{\Sigma_2} * (\delta I / \delta i)^+ = G_\Sigma * G_{\Sigma_1} * (\delta I / \delta i)^+ = G_\Sigma * G_i^{\Sigma_1} \tag{4}$$

where $\Sigma = \sqrt{\Sigma_2^2 - \Sigma_1^2}$ and $\Sigma_2 > \Sigma_1$.

The descriptor for every significant point is computed by picking the values from the convolved response maps. As depicted in Figure 5, at significant point location, say (u, v) , descriptor consists of vectors sampled in the neighborhood around it. These samples located on concentric circles and their amount of Gaussian smoothing is proportional to the radius of these circles. Let $h_\Sigma(u, v)$ be the vector made up of the values at location (u, v) in the convolved response maps.

$$h_\Sigma(u, v) = [G_1^\Sigma(u, v), G_2^\Sigma(u, v), \dots, G_8^\Sigma(u, v)]^T \tag{5}$$

where $G_1^\Sigma, G_2^\Sigma, \dots, G_8^\Sigma$ denote the Σ convolved response maps. Before concatenating these vectors to a descriptor, we normalize them to unit vector, and denote the normalized vectors by $\tilde{h}_\Sigma(u, v)$. The full descriptor $D(u, v)$ for the significant point location (u, v) can be defined as a concatenation of vectors and can be written as:

$$D(u, v) = [\tilde{h}_{\Sigma_1}^T(u, v), \tilde{h}_{\Sigma_1}^T(I_1(u, v, R_1)), \dots, \tilde{h}_{\Sigma_1}^T(I_N(u, v, R_1)), \tilde{h}_{\Sigma_2}^T(I_1(u, v, R_2)), \dots, \tilde{h}_{\Sigma_2}^T(I_N(u, v, R_2)), \dots, \tilde{h}_{\Sigma_Q}^T(I_1(u, v, R_Q)), \dots, \tilde{h}_{\Sigma_Q}^T(I_N(u, v, R_Q))]^T \tag{6}$$

where $I_j(u, v, R_1)$ is the location with distance R from (u, v) in the direction given by j when the directions are quantized into N values. Figure 5 shows the sample locations when $N = 8$, and significant point descriptor is made up of values extracted from 25 locations and 8 response maps. Therefore, descriptor length is 200 (i.e. 8×25).

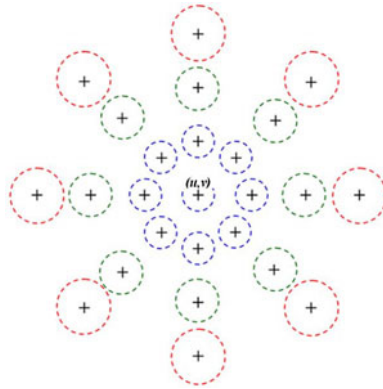


Fig. 5. Shape of the sampling locations of the descriptor. The '+' sign indicates the sampling locations. The radius of the dashed circle represents the size of the Gaussian kernel.

2.4 Similarity between Two Face Images

Once we have represented all face images as a set of interest points and their corresponding descriptions, the next step to be carried out is to find the similarity measure between two face images, in order to decide whether such images correspond to the same person or not. The comparison of two images is done by comparing the significant points. For each point of the first image, the best and second best matching points of the second image must be found. If the first match is much better than the second one, the points are said to be alike. Equation 7 shows how to apply such condition, where points B and C in range image I_2 are the best and second best matches, respectively, for point A in range image I_1 .

$$\frac{|D_{I_1}^A - D_{I_2}^B|}{|D_{I_1}^A - D_{I_2}^C|} < threshold \quad (7)$$

3 Experimental Results

3.1 Database

We conducted experiments using FRAV3D face database [6]. It contains the 3D point clouds of 106 persons with 16 scans per each person. This includes facial scans with frontal (1,2,3,4), 25 right turn in Y direction (5,6), 5 left turn in Y direction (7,8), severe right turn in Z direction (9), small right turn in Z direction (10), smiling gesture (11), open mouth gesture (12), looking up turn in X direction (13), looking down turn in X direction (14), frontal images with uncontrolled illumination (15,16). The 2D image of all 16 scan of a subject is shown in left image of the Figure 2.

Table 1. Comparison of recognition accuracy of the proposed model with subspace analysis based techniques

Test configuration	Total persons	Training samples	Testing samples	Samples tested	Recognition Rate		
					$(2d)^2$ FLD	PCA	Proposed Model
T1	90	1,2,3	4	90	96.67	95.56	98.89
T2	90	1,2,3,4	11	90	85.56	88.89	84.44
T3	90	1,2,3,4	12	90	52.22	46.67	81.11
T4	90	1,2,3,4	15.16	180	94.44	97.22	100.00
T5	90	1,2,3,4	7.8	180	94.44	87.22	90.00

Table 2. Comparison of recognition accuracy of the proposed model with local descriptor based techniques

Test configuration	Recognition Rate		
	SIFT	SURF	Proposed Model
T1	95.56	98.89	100.0
T2	80.00	84.44	90.00
T3	78.89	78.89	90.00
T4	96.67	100.0	100.0
T5	81.11	90.56	93.89

3.2 Results

We compared recognition accuracy of the proposed model with subspace analysis based techniques and local descriptor based techniques. Experiments are conducted with different test configurations.

In the subspace analysis based techniques, we used $(2D)^2FLD$ and conventional PCA directly on range image. Table 1 shows the results of various combinations of training and testing samples using these techniques. We used range images of frontal face scans in training for all test configurations. In test configuration T1, high recognition rate is observed because training and testing data contains only the frontal scan. In T2 and T3, test input contains a gesture. In T4, it is observed that illumination variation does not affect the 3D face recognition. Compared to T1, T5 results have less recognition rate due to self occlusion. On an average, it shall be observed from Table 1 that the proposed model outperform the subspace analysis based techniques.

In local descriptor based techniques we choose SIFT and SURF descriptor based algorithms proposed for 3D face recognition. Here the average number of significant points per face range image is fixed to 24. The experiments were conducted using the same test configuration as given in Table 1. Table 2 shows these results. It is observed that proposed model gives better accuracy with fewer number of significant points when compared to SIFT and SURF.

Experiments are also conducted using the leave-one-out strategy taking all 16 face scans of subjects. The results are given in Table 3. The recognition rate

Table 3. Comparison of proposed model using leave one out strategy

Total subjects	Samples tested	Recognition rate				
		$(2d)^2$ FLD	PCA	SIFT	SURF	Proposed model
10	160	91.25	93.12	91.25	96.25	98.75
20	320	87.81	91.56	92.50	96.25	95.31
30	480	86.04	90.00	91.88	95.83	94.79
40	640	84.84	87.81	90.00	94.38	94.69
90(All)	1440	81.60	86.18	87.22	91.84	92.85

decreases with the increase of persons used for experiments. It is observed that the test face with open mouth mismatches most of the times. Here also, the proposed method outperforms the other methods.

4 Conclusion

We have developed an algorithm for 3D face recognition based on the significant points described by orientation maps. Experimental results show that, the proposed model out performs the conventional holistic face recognition techniques and other local descriptor based models.

Acknowledgement

The authors would like to thank the support provided by DST-RFBR, Govt. of India vide Ref. No. INT/RFBR/P-48 dated 19.06.2009.

References

1. An, S., Ma, X., Song, R., Li, Y.: Face detection and recognition with SURF for human-robot interaction. In: IEEE International Conference on Automation and Logistics, pp. 1946–1951 (2009)
2. Achermann, B., Jiang, X., Bunke, H.: Face recognition using range images. In: International Conference on Virtual Systems and Multi Media, pp. 129–136 (1997)
3. Bay, H., Ess, A., Tuytelaars, T., Gool, L.V.: Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding* 110(3), 346–359 (2008)
4. Besl, P.J., McKay, H.D.: A method for registration of 3-D shapes. *IEEE Transactions on PAMI* 14(2), 239–256 (1992)
5. Brown, M., Lowe, D.: Invariant features from interest point groups. In: British Machine Vision Conference (2002)
6. FRAV3D Face database, <http://www.frav.es/databases/FRAV3d/>
7. Geng, C., Jiang, X.: Face recognition using sift features. In: 16th IEEE International Conference on Image Processing, pp. 3313–3316 (2009)
8. Guo, H., Zhang, K., Jia, Q.: 2.5D SIFT Descriptor for Facial Feature Extraction. In: 6th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pp. 1235–1250 (2010)

9. Guru, D.S., Vikram, T.N.: 2D Pairwise FLD: A robust methodology for face recognition. In: IEEE Workshop on Automatic Identification Advanced Technologies, pp. 99–102 (2007)
10. Heseltine, T., Pears, N., Austin, J.: Three-dimensional face recognition: An eigen-surface approach. In: ICIP, pp. 1421–1424 (2004)
11. Heshner, C., Srivastava, A., Erlebacher, G.: A novel technique for face recognition using range imaging. In: International Symposium on Signal Processing and Its Applications, vol. 2, pp. 201–204 (2003)
12. Kim, D., Dahyot, R.: Face components detection using SURF descriptor and SVMs. In: Machine Vision and Image Processing Conference, pp. 51–56 (2008)
13. Krizaj, J., Struc, V., Pavesic, N.: Adaptation of SIFT features for face recognition under varying illumination. In: MIPRO, Proceedings of the 33rd International Convention, pp. 691–694 (2010)
14. Lo, T.R., Siebert, J.P.: Local feature extraction and matching on range images: 2.5D SIFT. *Computer Vision and Image Understanding* 113(12), 1235–1250 (2009)
15. Lowe, D.G.: Distinctive Image Features from Scale Invariant Keypoints. *International Journal of Computer Vision* 20(2), 91–110 (2004)
16. Moreno, A.B., Sanchez, A., Velez, J.F.: Voxel-based 3d face representations for recognition. In: 12th International Workshop on Systems, Signals and Image Processing, pp. 285–289 (2005)
17. Mian, A., Bennamoun, M., Owens, R.: Face Recognition Using 2D and 3D Multimodal Local Features. In: Bebis, G., Boyle, R., Parvin, B., Koracin, D., Remagnino, P., Nefian, A., Meenakshisundaram, G., Pascucci, V., Zara, J., Molineros, J., Theisel, H., Malzbender, T. (eds.) ISVC 2006. LNCS, vol. 4291, pp. 860–870. Springer, Heidelberg (2006)
18. Tola, E., Lepetit, V., Fua, P.: DAISY: An Efficient Dense Descriptor Applied to Wide-Baseline Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 815–830 (2010)
19. Velardo, C., Dugelay, J.: Face recognition with DAISY descriptors. In: 12th ACM Workshop on Multimedia and Security, pp. 95–100 (2010)
20. Yunqi, L., Xutuan, J., Zhenxiang, S., Dongjie, C., Qingmin, L.: Face Recognition Method Based on SURF Feature. In: International Symposium on Computer Network and Multimedia Technology, pp. 1–4 (2009)
21. Yunqi, L., Haibin, L., Xutuan, J.: 3D face recognition by SURF operator based on depth image. In: 3rd IEEE International Conference on Computer Science and Information Technology, vol. 9, pp. 240–244 (2010)

The Optimal Wavelet for Speech Compression

Shijo M. Joseph and P. Babu Anto

School of Information Science & Technology, Kannur University
Mangattuparamba Campus, Kannur, Kerala, India 670 567

Tel.: + 91 9447447985

{shijomjose71, bantop}@gmail.com

Abstract. The main idea of speech compression is to reduce the bite rate of the speech for communication or storage without significant loss of quality. There are mainly three functional categories of speech processing methods in use. The new technique called wavelet transform is being used for speech signal analysis and synthesis. The major issues regarding the design of real time wavelet based speech coder are choosing optimal wavelets for the compression, and selecting suitable frame size etc. The performance of the different wavelets families on speech compression is evaluated and compared based on different parameters. Male and female speech is used for the comparison and analysis.

Keywords: Speech Compression, Wavelet family, Discrete Wavelet, Isolated Malayalam Spoken words.

1 Introduction

Speech is a perfect form of acoustic signal by nature and it is the most effective medium for face to face communication and telephony application. Speech is produced when the vocal tract responds to excitation signals and it is capable of conveying information with a touch of emotion [1]. Due to the rapid growth of the multimedia application, wireless communication and cellular telephony, the demand for transferring or storing digital information increased significantly. One solution to overcome this type of obstacle is to compress information by removing the redundancies present in the digital data [2]. In data compression it is desired to represent data by as small as possible number of coefficient with in an acceptable loss of visual quality. The success of the compression is based on the simplicity of the technology and algorithm used in the system. In general, compression methods can be classified in to three functional categories

- Direct Methods: the samples of the signal are directly handled to provide compression.
- Parameter Extraction methods: A preprocessor is employed to extract some features that are later used to reconstruct the signal (LPC).
- Transform Methods such as Fourier Transform (FT) Discrete Cosine Transform (DCT) and Wavelet Transform (WT).

Each method has its own advantage and disadvantage based on the application for which it is used. Fourier Transform (FT) and Discrete Cosine Transform (DCT)

techniques are commonly used as methods for speech compression and its application [3]. Wavelet is a new technique used for speech and image processing. In wavelet processing any signal can be represented by a set of scaled and translated versions of a basic function called mother wavelet. This set of wavelet function forms the wavelet coefficient at different scales and positions and results from taking the wavelet transform of the original signal. The localization feature of wavelets, along with its time frequency resolution properties makes them well suited for coding of speech signals [4]. The motivation behind speech compression involves real time processing in mobile satellite communication, cellular telephony, internet telephony, audio for video phones or video teleconferencing system. Another major application of speech compression includes storage, interactive PC software, voice memo wrist watch etc. This work is based on promising and flexible compression schemes based on different wavelets. This paper is organized as follow: Section 2 cover the wavelet family and Discrete Wavelet Transforms. Section 3 gives details about the data base used for the experiment. Section 4 discusses wavelet based speech compression. Section 5 discusses performance measurement tools. Section 6 shows the results finally section 7 gives conclusion.

2 Choice of Wavelet

The choice of the suitable mother wavelet function has prime importance in the design of high quality speech coding. Choosing a wavelet that has compact support in both time and frequency in addition to a significant number of vanishing moments is essential for an optimum wavelet speech compression [5]. Several criteria can be used for selecting an optimal wavelet function. The objective is to minimize reconstructed error (MSE) variance and maximize Peak signal to Noise Ratio (PSNR). In general optimum wavelets can be selected based on the energy conservation properties in the approximation part of the wavelet coefficient [6]. Fig.1 shows the entire Daubechies wavelets family. We cannot say one type of wavelet is better than another because every type has its own applications. The effectiveness of a wavelet family depends upon how suitable it is for a particular application. A wavelet which performs very well in one application may not be equally good in another case.

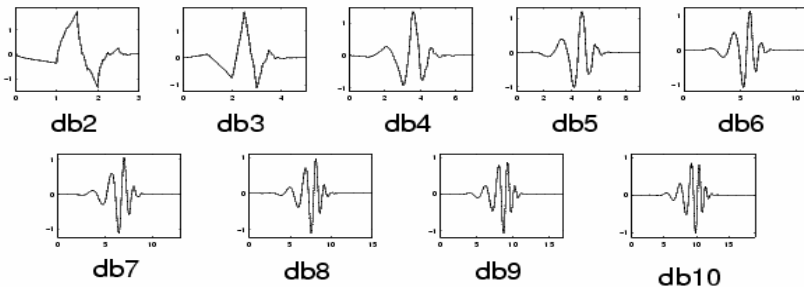


Fig. 1. Plots of Different Daubechies Orthogonal Wavelets

2.1 Discrete Wavelet Transform

Feature extraction involves the information retrieval from speech signal. Discrete wavelet transforms an alternative approach to traditional signal processing. Such as Fourier analysis for breaking a signal up into its constituent part. In wavelet transform the basic functions are compact in time. This feature allows the wavelet transform to obtain time information about a signal in addition to frequency information [7]. The original signal is successively decomposed into components of lower resolution while high frequency components are not analyzed any more.

The Discrete Wavelet Transform for one dimensional signal is defined in the following equation 1.

$$W(j, K) = \sum_j \sum_k X(k) 2^{-j/2} \Psi(2^{-j}n-k) \tag{1}$$

Where $\Psi(t)$ is the basic analyzing function called the mother wavelet

The DWT of the original signal is then obtained by concatenating all the coefficients, $a[n]$ and $d[n]$, starting from the last level of decomposition. The successive high pass and low pass filtering of the signal can be depicted by the following equations:

$$Y_{high}[k] = \sum_n x[n]g[2k-n] \tag{2}$$

$$Y_{low}[k] = \sum_n x[n]h[2k-n] \tag{3}$$

Where Y_{high} and Y_{low} are the outputs of the high pass and low pass filters obtained by sub sampling by 2 [8]. Fig 2 shows a level 3 DWT decomposition of an input signal S.

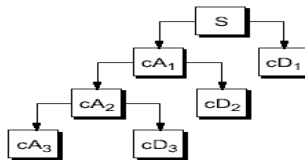


Fig. 2. Decomposition of DWT Coefficient

2.2 Signal Reconstruction

The original signal can be reconstructed or synthesized using the Inverse Discrete Wavelets Transforms (IDWT). The synthesis starts with the approximation and detail coefficient cA_j and cD_j , and then reconstructs cA_{j-1} by up sampling and filtering with the reconstruction filters. The reconstruction filters are designed in such a way as to cancel out the effect of aliasing introduced in the wavelet decomposition phase. The reconstructed filters (Lo_R and Hi_R) together with the low and high pass decomposition filters form a system known as Quadrature Mirror Filters (QMF) for multilevel analysis. The reconstructed process can itself be iterated producing successive approximation at finer resolution and finally synthesize the original signal [8].

3 Data Base used for the Experiment

The speech signals are chosen from the Malayalam (one of the south Indian languages) spoken words database. A speaker independent database has been used for the experiment. The data base consists of 10 isolated Malayalam spoken words from 10 different speakers. The speech samples were recorded from five male speakers of age above 20 years and five female speakers of age above 18. The speakers were native speakers of Malayalam and free from speech disabilities. The speech is recorded using commonly used recording microphone, at a sampling rate of 8 KHz (4 KHz band limited). The recorded speech is processed, labeled and stored in the data base. The isolated spoken words in the data base and their IPA format are given in table1.

Table 1. Speech data base and their IPA format

Words in Malayalam	Words in English	IPA format
അമ്മ	amme	//æ/m/ m/ æ//
അച്ഛൻ	acha	//æ//tʃ/ a://
മോളെ	mole	// m/ o/ / ε//
മോനെ	mone	// m/ o/ n/ ε//
എടാ	eda	// ε/ d/a://
ലെതെ	lethe	// / ε// θ/ ε//
ദേവീ	devi	// d/ ε/ v/ i//
ഞാനോ	njano	// n/ dʒ/a:/ n/ o//
കുട്ടി	kutty	//k/ʊ/tʃ/i//
മായേ	maye	// m/ a:/ j/ ε//

4 Experiment

The isolated Malayalam spoken word compression was carried out by using haar and different wavelet family. Since the speech files were of short duration, the entire signal was decomposed at once without framing. The samples in the data base are compressed using haar, db2, db4, db6, db8, db10 and db20, sym1, sym2, sym4, sym6 and sym8 wavelets. To simplify the comparison we use the most commonly spoken word amme (/æ/m/ m/ æ/) for the experiment. In each level of compression the speech signal is decomposed without losing its audibility by splitting it into high frequency and low frequency components. The compressed signal is chosen from the third level of the decomposition for the reconstruction. Further enhancement beyond the level 3 decomposition affects the intelligibility of the spoken words. The synthesized signal from the compressed one ensures perfect audibility.

5 Performance Measurement Tools

The performance of the wavelet based speech coder in terms of both reconstructed signal quality after decode and compression scores can be assessed by applying different qualitative parameters [9]. The following parameters are compared.

- Mean Square Error (MSE)
- Retained Signal Energy (RSE),
- Peak Signal to Noise Ratio (PSNR),
- Normalized Root Mean Square Error (NRMSE)
- Compression Ratios (CR).

The results obtained for the above quantities are calculated using the following formulas:

5.1 Mean Square Error

$$MSE = \frac{1}{N} \sum_{i=0}^{N-1} [S_{(i)} - S'_{(i)}]^2 \quad (4)$$

Where $S(i)$ is the original speech signal data and $S'(i)$ is the reconstructed signal

5.2 Peak Signal to Noise Ratio

$$PSNR = 10 \log_{10} \frac{NX^2}{\|x - r\|^2} \quad (5)$$

N is the length of the reconstructed signal, X is the maximum absolute square value of the signal x and $\|x - r\|^2$ is the energy of the difference between the original and reconstructed signals.

5.3 Normalized Root Mean Square Error

$$NR MSE = \sqrt{\frac{x(x(n) - r(n))^2}{(x(n) - \mu_x(n))^2}} \tag{6}$$

$x(n)$ is the speech signal, $r(n)$ is the reconstructed signal, and $\mu_x(n)$ is the mean of the speech signal.

5.4 Retained Signal Energy (RSE)

$$RSE = 100 * \frac{\|x(n)\|^2}{\|r(n)\|^2} \tag{7}$$

$\|x(n)\|$ is the norm of the original signal and $\|r(n)\|$ is the norm of the reconstructed signal. For one-dimensional orthogonal wavelets the retained energy is equal to the L^2 -norm recovery performance.

5.5 Compression Ratio

$$C = \frac{\text{Length}(X(n))}{\text{Length}(cWC)} \tag{8}$$

cWC is the length of the compressed wavelet transform vector.

Table 2. Performance of Male Speakers

Wavelet	MSE	PSNR	NRMSE	RSE	C-RATIO
haar	0.032718	81.6258	1.6426	30.9648	15.9682
db2	0.030106	82.0465	1.6059	29.5173	15.3973
db4	0.032894	81.6674	1.6377	28.9723	15.7807
db6	0.03485	81.6685	1.6511	28.9387	15.0593
db8	0.033586	81.65	1.6587	28.8379	14.7514
db10	0.032964	81.6005	1.6426	28.7903	14.4341
db20	0.032893	81.6279	1.6426	28.8014	13.0293
sym1	0.032718	81.6258	1.6059	30.9648	15.9682
sym2	0.030106	82.0465	1.6768	29.5173	15.7807
sym4	0.034115	81.4613	1.6596	28.9763	15.3973
sym6	0.033233	81.6315	1.6695	28.9021	15.0593
sym8	0.033702	81.5235	1.6527	28.7569	14.7514

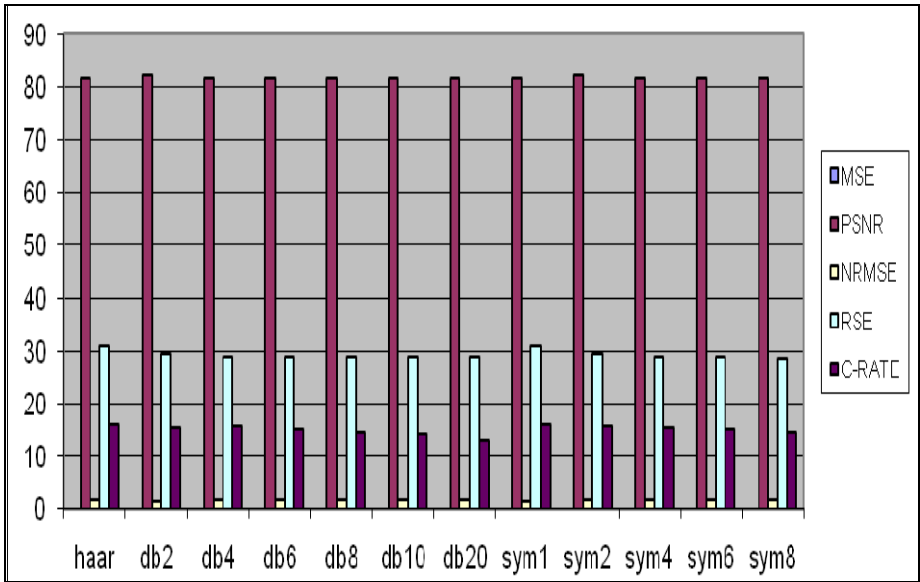


Fig. 3. Performance Analysis of Different Wavelet (Male Speakers)

Table 3. Performance of Female Speakers

Wavelet	MSE	PSNR	NRMSE	RSE	C-RATIO
haar	0.38013	62.7569	1.4554	49.1982	15.9626
db2	0.35681	63.0559	1.4253	46.0076	15.8311
db4	0.37878	62.7824	1.4423	44.5469	15.5573
db6	0.38988	62.6345	1.4703	43.5732	15.2957
db8	0.39131	62.6241	1.4684	44.042	15.0481
db10	0.38686	62.6853	1.4734	43.3773	14.8087
db20	0.38436	62.6506	1.4758	43.1865	13.6999
sym1	0.38013	62.7569	1.4554	49.1982	15.9626
sym2	0.35681	63.0559	1.4253	46.0076	15.8311
sym4	0.40392	62.477	1.4998	44.4044	15.5573
sym6	0.38853	62.614	1.4929	43.9496	15.2957
sym8	0.39703	62.5578	1.4919	43.7974	15.0481

5.6 Mean Opinion Score (MOS)

Subjective evaluation by listeners is still a method commonly used in measuring quality of reconstructed speech. MOS provides a numerical indication of the perceived quality of received media after compression and / or transmission. The MOS is expressed as a single number in the range 1 to 5. Where 1 is the highest

perceived quality and 5 is the lowest perceived quality. When taking subjective test, listeners focus on the difference between the original and reconstructed signal and rate it. The MOS is generalized by averaging the result of a set of standards, subjective tests, where a number of listeners rate reconstructed speech signals. In our experiment the average MOS rate was 1.35 in all experiment.

6 Result

The results of these performance measures are given in table 2 and table 3 and the graphical representation of the performance analysis is given in figure 2 and 3 respectively for the different wavelets we used.

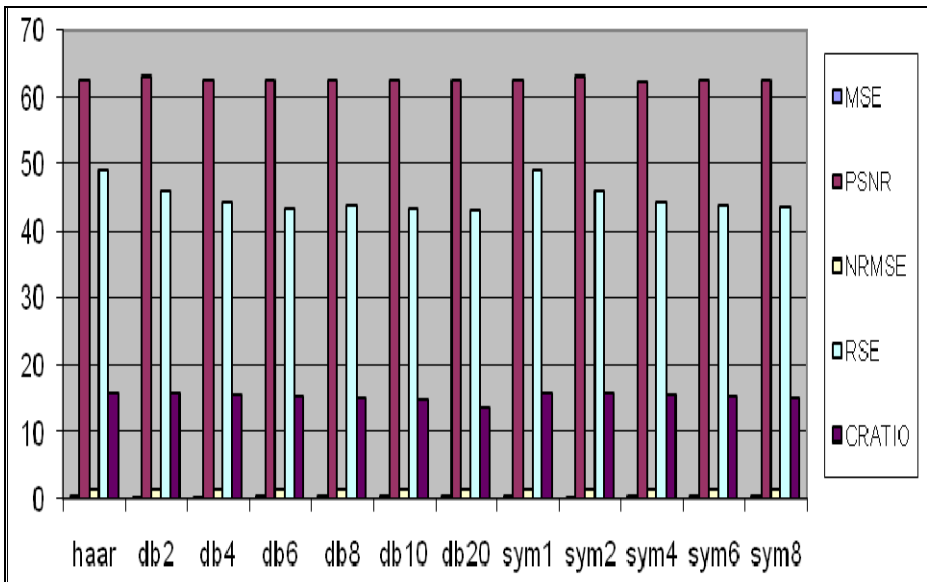


Fig. 4. Performance Analysis of Different Wavelet (Female Speakers)

7 Conclusion

This paper is a performance analysis based on the effect of different wavelet transforms on Malayalam spoken word compression. To simplify the comparison we exploit the most commonly used spoken word amme (/æ/m/ m/ æ/) for the experiment. The compressed signal is chosen from the 3rd level of the decomposition for the reconstruction. Further enhancement beyond the level 3 decomposition influence the intelligibility of the spoken words. The compressed signals are reconstructed back to their initial form with full audibility. In this work we have obtained average MSE is 0.3829, average NRMSE is 62.7209; average PSNR is 1.4647, average RSE is 45.1074 and average compression ratio is 15.3249 in the performance analysis for female speakers. In the case of male speakers average MSE

is 0.05730, Average NRMSE is 1.63482 average PSNR is 76.04585, average RSE is 30.0228 and average compression ratio is 15.1383. In general a good reconstructed signal is the one with low MSE and high PSNR. That means the signal has low error and high signal fidelity. In both male and female experiment; we have obtained low MSE and high PSNR while using db2 and sym2 wavelet for speech signal decomposition. It is also observed that both these wavelet gives highest compression ratio also. The obtained result can be improved by preprocessing the input signals. But in this work we opted raw signal for the experiment to reduce the computational complexity and to test its usefulness in real time application. The performance analysis of different wavelet is plotted in figure 2 and 3. The findings lead to the conclusion that db2 and sym2 are the optimum wavelet for spoken Malayalam speech compression and its application.

References

1. Kondoz, A.M.: Digital Speech, 2nd edn. John Wiley & Sons Ltd., West Sussex. ISBN 0-470-87007 9(HB)
2. Sayood, K.: Introduction to Data compression Book, 2nd edn. Elsevier, Amsterdam. ISBN 1 55860-558-4
3. Karam, J.: A new approach in Wavelet based Speech Compression Mathematical Methods, Computational Techniques, Non Linear Systems Intelligent Systems. ISBN 978-960-474-012-3
4. Soman, K.P., Ramachandran, K.: In Sight into Wavelets from Theory to Practice. In: PHI, 2nd edn. (2008) ISBN 978-81-203-2902-7
5. Joseph, S.M., Sebastian, R., Babu Anto, P.: The effect of different wavelets on speech compression. In: ACM Proceedings of the 2011 International Conference on Communication, Computing & Security (2011) ISBN: 978-1-4503-0464-1
6. Chan, Y.T.: Wavelet Basics. Kulwer Academic Publications, Boston (1995)
7. Daubechies: Ten Lecturers on wavelets. Society for Industrial and Applied Mathematics (1992)
8. Mallat, S.A.: Theory for Multiresolution Signal Decomposition: The Wavelet Representation. IEEE Transactions on Pattern Analysis Machine Intelligence 31, 674–693 (1989)
9. Abdul Mawala, M.A., Ram, N.A.R.: Comparing Speech Compression using Wavelets with Other Speech Compression Schemes. In: Proceedings of Student Conference on Research and Development, Putrajaya, Malaysia (2003)

Skew Angle Estimation and Correction for Noisy Document Images

M. Manomathi and S. Chitrakala

Dept. of Computer Science & Engineering,
Easwari Engineering College, Anna university, Chennai 600 089
{manomathi11, ckgops}@gmail.com

Abstract. Document skew commonly occurs during document scanning; it should be avoided because it dramatically reduces the accuracy of the OCR. Noise removal is an important procedure before on going further processing. This paper describes an approach towards noise removal, skew detection and correction for text in scanned documents. Preprocessing is a stage, comprising number of adjustments in order to obtain the noise reduced results, and then the skew angle is estimated. Instead of deriving a skew angle from the text lines, the proposed method uses various types of visual content of image skews, and HDT algorithm is used to select the useful image region dynamically. A bootstrap estimator is finally employed to combine various cues on local image blocks. Once the skew angle is being estimated it has to be rotated in the opposite direction in order to correct the skew angle.

Keywords: Bagging estimator, Visual content, Preprocessing.

1 Introduction

Document processing is affected by document skews. Document skews commonly occur during the scanning process. This has detrimental effect on document analysis, document understanding, character segmentation and character recognition. Skew estimation and correction becomes an important issue in the field of document image analysis and understanding. This paper focuses on skew detection and correction for noisy document images. Preprocessing is a stage in typical OCR system, which focuses on enhancing the acquired image to increase the ease of feature extraction and to compensate for the eventual poor quality of the scanned document. Optical character recognition (OCR) has many important applications such as: Automatic bank cheque processing, mail address recognition, and historical document recognition. If the acquired image contains noise, it is subjected to the preprocessing stage where the “de-noising” of the image takes place. Furthermore, when a document is fed into the scanner either mechanically or by a human operator, a few degrees of skew (tilt) is unavoidable. Skew correction is a process which aims at detecting the deviation of the document orientation angle from the horizontal or vertical direction.

There are different kinds of noises that can be occurred during scanning. Such as ink blobs, salt & pepper noise, stray marks, marginal noise, clutter noise [15] A kind additive noise called “Salt and Pepper Noise”, the black points and white points

sprinkled all over an image, typically looks like salt and pepper, which can be found in almost all documents. Many methods have been proposed for removing salt and pepper noise from images such as the k-Fill Algorithm, Applied k-Fill Algorithm and Median Filter [15-16]. There exist five classes of skew estimation techniques they are based on (1) projection profile, (2) Fourier method, (3) Nearest neighbor method, (4) correlation, (5) Hough transform technique respectively. [3][4]In general, Projection profile methods are limited to estimate the skew angle within ± 10 degrees. [13] Hough transform is applied to find the best detection of the straight lines in documents. Another group of skew correction methods uses k-nearest neighbor clustering on connected components, and they try to find the best overall direction of neighboring components.

The proposed method can remove these noises of any size that are smaller than the size of document objects. This method is fast and can be used effectively on binary, gray scale and color image.

2 Prior and Related Work

Amin and Fischer [1] presented a basic model for document “skew detection method using the Hough transform”. Hough transform is widely used technique for skew detection. It first transforms an image into its parameter space and then searches for the local maxima of the space to estimate the skew angle. The main disadvantage of Hough transform is it cannot be directly implemented in the real-scanned document images due to its great memory demanding and high time complexity.

Bo Yuan and Chew Lim [2] proposed a new method of “Skew Estimation for Scanned Documents from Noises”, skew estimation method that is based on the presence of a special kind of "noises" – the straight lines or edges exist in the images, which include the straight lines that separate columns or paragraphs, the inserts of photographs or drawings that have rectangular sides, the black bars around the borders of a scanned page due to the incomplete coverage of the scanning surfaces of the scanners or copiers, and any non-textual elements that purposely or accidentally, fully or partially possess straightness. The mapping Scheme is by probe-lines rather than Wallace's original feature-points based bounds intersections.

Faisal Shafait et al. [3] proposed “Response to “Projection Methods Require Black Border Removal” a novel evaluation method based on a vectorial score, and demonstrates its utility and validity by comparing it to the results obtained using Mao and Kanungo’s method.

Gaofeng Meng et al. [4] proposed a new method for “Circular Noises Removal from Scanned Document Images” in this method a fast and robust algorithm is presented to detect and correct these circular defects in scanned document images. Most of punched holes in the images can be fast detected and located after the size of the original image is appropriately reduced. To ensure that all of punched holes are correctly located, apply Hough transformation only in the roughly located regions to further confirm them. Finally, each located circular noise is removed by fitting a bi-linear blending Coons surface which interpolates along the four edges of the noisy region.

Liu et al. [5] proposed a method for “Skew detection for complex document images using robust borderlines extracted from text and non-text regions”. In general, the straight text lines based methods work well for most document images. The main disadvantage of this method is often severely affected by the presence of scan-introduced distortions and large areas of non-textual objects.

Y. Lu and C. L. Tan [6] proposed a “nearest-neighbor chain based approach to skew estimation in document images”, the nearest neighbors clustering based method first extracts connected components from a document image. For each component, the direction of its nearest neighbor is estimated and accumulated in a histogram. The angle that corresponds to the maximum peak of the histogram is taken as the dominant image skew.

Martin and Pattichis [7] has proposed the “Characterization of Scanning Noise and Quantization on Texture Feature Analysis” A quantitative approach as suggested by Halpern uses a calibrated test pattern for direct quality control with the digital media. This test pattern is designed to characterize the various scanning artifacts and noise. And it provides a mathematical model for describing the distortion of the original image due to the scanning process.

Mudit Agrawal et al. [8] presented “Clutter Noise Removal in Binary Document Images” and it is a novel approach toward clutter detection and removal for complex binary documents. And it uses an SVM classifier to detect clutter. The novelty of this approach is in its restrictive nature to remove clutter, as text attached to the clutter is neither degraded nor deleted in the process.

M. Sarfraz et al. [9] proposed a “Novel Approach for Skew Estimation of Document Images in OCR System” and this is a new Technique for skew estimation utilizing the multi- scale analysis of image.

Shen and Sun [10] proposed a new method for “Skew detection using wavelet decomposition and projection profile analysis”. Typically the method first calculates the projection profiles of a document image at various angles. Each projection profile is then evaluated by a criterion function. The function should reach its maximum when the projection profile is calculated at the correct skew angle.

The main disadvantage of the existing method is:

- Sensitive to document layouts and the presence of non-textual objects,
- Performance often deteriorates with short and sparse text lines,
- Separate mechanism required for the separation of text and non-textual objects.

3 System Description

Proposed system mainly focuses on,

- To remove the noises present in the document image using preprocessing techniques.
- Image blocks selection using the proposed HDT algorithm.
- To estimate the skew angle using bagging technique.

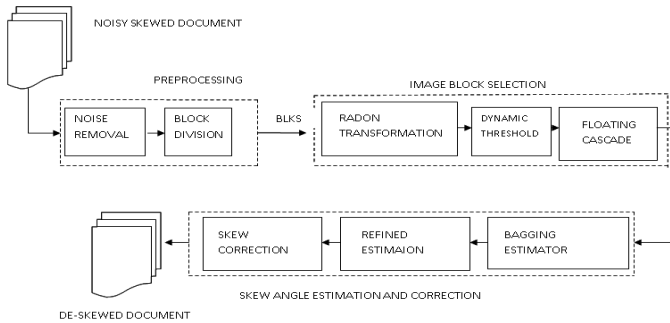


Fig. 1. Overall proposed architecture

Proposed system contains the following modules:

3.1 Preprocessing

Preprocessing is a stage in typical OCR system, which focuses on enhancing the acquired image to increase the ease of feature extraction and to compensate for the eventual poor quality of the scanned document. In the proposed system preprocessing involves removal of noise present in the document as well as image block division. After noise is being removed document image is divided into non overlapped squared blocks with equal size $L \times L$.

3.2 Noise Removal Algorithm

In this algorithm, a black pixel is defined as ON, a white pixel as OFF and k is the dynamic window size. The algorithm follows:

```

    Let k be window size, Black pixel: =ON,
    White pixel: =OFF.
    set (k-n)*(k-n) pixels as core.
    Count number of ON or OFF pixels in the core
    if number of ON or OFF pixel in the core >= half
    of all the pixel in the core
    then
    replace the ON or OFF pixel value by the
    median pixel value at the core and its
    neighborhood
    else if number of ON or OFF pixel in the core <= half
    of all the pixel in the core
    then
    replace the existing pixel value by the median
    pixel value in the core.
    else keep the pixel value as such
  
```

3.3 Image Block Selection

It is observed that some image blocks are non informative or even erroneous, especially when few visual cues of image skew are available. Only high informative

energy blocks have to be selected, all other blocks have to be rejected which is shown in Fig. 2. Eliminating such outlier blocks from the bagging estimation will help to improve its accuracy. Blocks selection can be efficiently implemented through a Cascade.

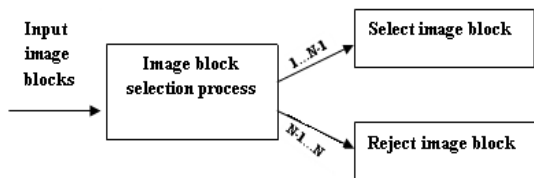


Fig. 2. Image Block Selection

Instead of using fixed threshold for image block selection dynamic thresholding method will increase the accuracy of image block selection. This is achieved by using proposed algorithm HDT (Histogram based Dynamic Thresholding).

3.4 Skew Angle Estimation and Correction

Skew angle is estimated by using bagging estimator where the angular resolution of the last stage of cascade is estimated. Symmetric peak always appears around the maximum of this symmetry of the maximum peak can be exploited to improve the estimation of skew angle. Then take three points around its maximum peak and calculate their symmetric center to refine the angle. Once the skew angle is being estimated the document has to be rotated in the opposite direction in order to correct the skew.

```

Procedure dynamic threshold( in, out, h, w)
  For i=0 to w do
    Sum ← 0
    For j=0 to h do
      Sum=sum+in[i,j]
      If i=0 then
        Int img [i,j] ← sum
      Else
        Int img[i,j] ← int img[i-1,j]+sum
      End if
    End for
  End for
  For i=0 to w do
    For j=0 to h do
      Compute histogram (each block)
      If(histogram peak==1)&&(peak gray level== very high)
        Then select threshold below the peak gray level
      Else
        Set threshold above the gray level
      End if
    End for
  End for
End for
  
```

4 Experimental Results

Original document image is taken as the input. Different levels of noises are added into the original image. Each pixel in an image has probability $p/2$ ($0 < p < 1$) to be corrupted into either a white dot (salt) or a black dot (pepper). There are different types of filter to reduce noise. Some of them are,

1. Mean Filter
2. Median Filter

4.1 Mean Filter

Mean filter is implemented by a local averaging operation where the value of each pixel is replaced by the average of all the values in the local neighborhood. The original image is taken into matrix form it contains 'n' number of columns. For each and every row Place 3×3 Window. For each and every pixel mean filter is being applied.

4.2 Median Filter

The median is estimated by first sorting all the pixel values from the surrounding neighborhood, and then replacing the neighborhood value by the median value. The original image is taken into matrix form it contains 'n' number of columns. For each and every row Place 3×3 Window.

Noise filtering algorithms are applied on images to remove the different types of noise that are either present in the image during capturing or document scanning. When document images are being scanned salt and pepper noise will occur usually. In this work, two different image filtering algorithms are compared at three different noise intensity levels.

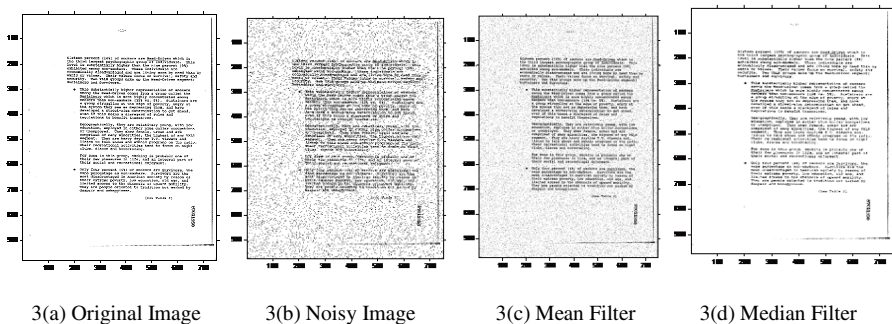


Fig. 3. Noise Removal Filters

The performances of the filters are compared using the Peak Signal to Noise Ratio (PSNR) and Image Enhancement Factor (IEF). Peak Signal to Noise Ratio (PSNR): It

is measured in decibel (dB) and for gray scale image it is defined as: It is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation.

$$PSNR = 10 \log_{10} \frac{255^2}{mse} dB \quad (1)$$

Table.1 summarizes the performance metrics of two different filters. The median type filters give better performance than the mean filter.

Table 1. Performance Evaluation for Mean and Median Filter

Different kind of text documents / images	Noise intensity level	Mean filter		Median filter		Proposed Algorithm	
		IEF	PSNR	IEF	PSNR	IEF	PSNR
Text images (.jpeg)	0.1	3.9695	6.6549	5.6151	8.7933	6.3456	9.6733
	0.2	3.8946	5.3998	7.9301	8.1704	8.7121	11.1402
	0.3	3.5515	4.4620	8.4768	7.1516	9.5677	9.7856
Text images (.png)	0.1	4.1392	8.0538	6.5038	10.103	7.0032	11.3291
	0.2	4.0680	6.0575	10.2495	9.5575	10.5432	9.9989
	0.3	3.6555	4.7809	9.8252	7.8475	9.9989	10.8746
Text Image (.bmp)	0.1	4.1044	6.3264	8.0870	9.0069	9.0870	10.8976
	0.2	3.2005	5.4264	9.1728	7.6069	10.2098	8.5634
	0.3	3.2845	4.3406	10.2530	6.8069	11.2678	8.5768



Fig. 4. Skewed Image



Fig. 5. Image block division

Skewed document image is taken as the input which is given in Fig.4. The skewed image is divided into non overlapped squared blocks with equal size $L \times L$. Estimate the height and width of each image blocks resolutions is represented in Fig.5. Each and every image block is verified and the blank ones that contain few edge points are

dropped. Instead of using fixed threshold for the image block selection dynamic thresholding method HDT is being applied to select the image blocks having more visual content. Fig. 6.(a) shows the different image blocks, Fig. 6.(b) shows the radon energy graph for different blocks, In radon energy graph x- axis represents series of angles θ in degrees, y-axis is represented as x' (radial axis). Once the radon energy graph is being obtained, it has to be plotted against set of fixed angles. Fig. 6.(c) gives the energy graph for the corresponding blocks In which energy graph is the mapping between the radon energy with serious of angles.

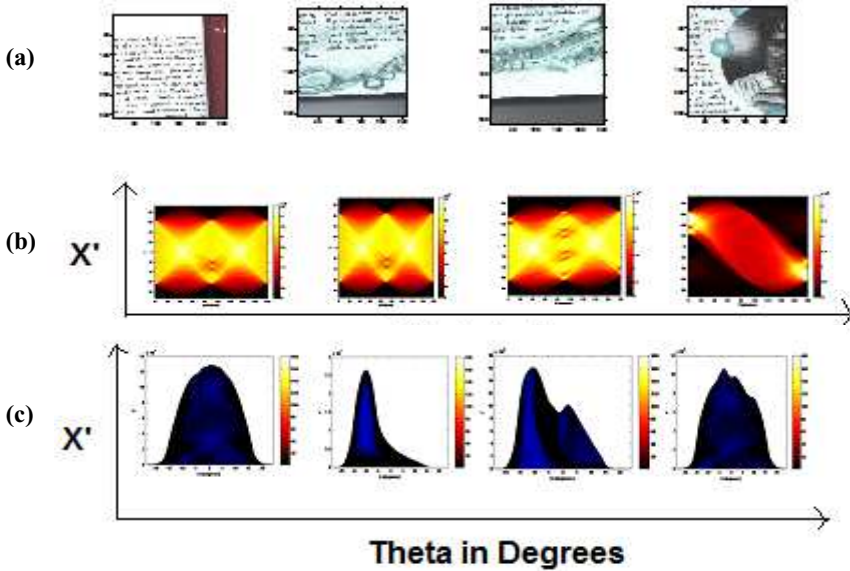


Fig. 6. (a) various types of local image blocks, (b) Radon transform energy, (c) corresponding energy functions of Radon transform

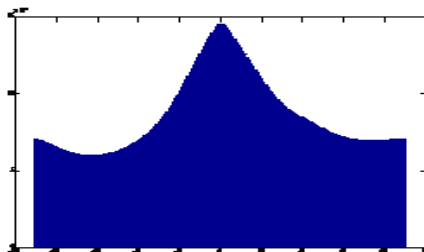


Fig. 7. Cascaded Energy Graph

Fig.7. shows the cascaded energy graph. By using bagging estimator skew angle is being estimated. Then by refined estimation the skew angle is being refined in order to get accurate results. The estimated skew angle for the image specified above is, SKEW ANGLE = 2° .

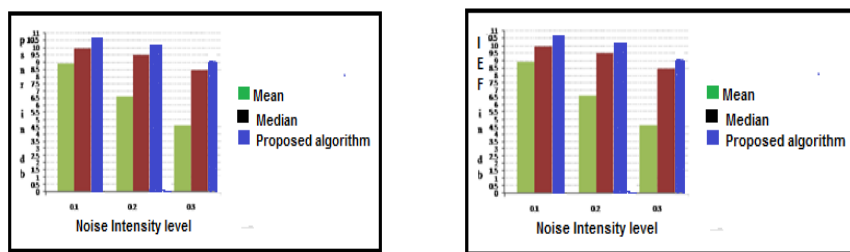


Fig. 8. Performance Graph

Fig.8. shows the performance graph with different intensity level for different document images. Different noise intensity level is taken in the x-axis, performance metrics such as IEF, PSNR is taken in the y-axis. Our proposed algorithm gives better performance than the median filter. It is clearly observed that salt & pepper noise is completely removed when using our proposed method.

5 Conclusion

This paper describes an approach towards noise removal, skew detection and correction for text in scanned documents. Preprocessing stage comprises a number of adjustments in order to obtain the noise reduced results. This proposed algorithm can remove these noises of any size that are smaller than the size of document objects. This method is fast and can be used effectively on binary, gray scale and color image. It is highly competitive in execution speed and estimation accuracy, and extremely robust to document noises, multiple different skews, short and sparse text lines, and the presence of large areas of non-textual objects of various types and quantities.

References

1. Amin, A., Fischer, S.: A document skew detection method using Hough transform. *Pattern Anal. Appl.* 3(3), 243–253 (2000)
2. Yuan, B., Lim, C.: Skew Estimation for Scanned Documents from Noises. Centre for Remote Imaging. Sensing and Processing Department of Computer Science, School of Computing National University of Singapore, *Models Image Process.* 41(6), 234–243 (2005)
3. Faisal, S.H., Daniel, K.V., Thomas, B.M.: Response to Projection Methods Require Black Border Removal. *Pattern Recognition. Lett.* 28(7), 155–162 (2009)
4. Gaofeng, M.G., Nanning, Z.A., Zhang, Y., Song, Y.: Circular Noises Removal from Scanned Document Images. Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, China (2007)
5. Liu, H., Wu, Q., Zha, H.B., Liu, X.P.: Skew detection for complex Document images using robusts border lines in both text and non-text regions. *Pattern Recognit. Lett.* 29(13), 1893–1900 (2008)
6. Lu, Tan: A nearest-neighbor chain based approach to skew Estimation in document images. *Pattern Recognit. Lett.* 24(14), 2315–2323 (2003)

7. Martin, Pattichis: Characterization of Scanning Noise and Quantization on Texture Feature Analysis. In: Computer, University of New Mexico, Albuquerque, vol. 25(7), pp. 10–22 (2004)
8. Mudit, A.L., David Dorman, D.C.: Clutter Noise Removal in Binary Document Images. In: 10th International Conference on Document Analysis and Recognition, Computer, vol. 25(7), pp. 110–212 (2009)
9. Sarfraz, M., Zidouri, A., Shahab, S.A.: Novel Approach for Skew Estimation of Document Images. In: OCR System
10. Shen, L., Sun, L.: Skew detection using wavelet decomposition And projection profile analysis. Pattern Recognition Lett. 28(5), 555–562 (2007)

Face Recognition Using ATP Feature Set under Difficult Lighting Conditions

Lincy Thomas and Komathy Karuppanan

Department of Computer Science and Engineering, Easwari Engineering College,
Anna University, Chennai, India
tlincythomas02@gmail.com, gomes1960@yahoo.com

Abstract. Recognizing face image under difficult lighting conditions is a challenging task in the Face Recognition system. Existing feature extraction using the Local Binary Pattern (LBP) fails to detect the eye position and is sensitive to noise in uniform image regions. These issues have been considered in our proposed Advanced Ternary Pattern (ATP) Feature set. The detection of eye position is based on the construction of eye model and localization of eye coordinates. Further, normalization technique attempts to solve the low sensitivity to noise in uniform image regions such as cheeks and forehead under difficult lighting. ATP feature set uses the Efficient Feature Orientation (EFO) method for effectiveness of inaccurate face normalization. The proposed scheme improves the False Acceptance Rate (FAR) up to 90% for the less dark images and about 78% in the fully dark images, thus improvising the previous results. Experimental results show that the proposed technique is promising in achieving the illumination invariant for facial images.

Keywords: Feature extraction, illumination, advanced ternary pattern, noise removal, image classification.

1 Introduction

A face recognition system (FRS) automatically identifies or verifies a person from a digital image or a video frame. Face recognition system entrenches the cutting edge technologies that can be applied to a wide variety of application domains including access control for PCs, airport surveillance, private surveillance, criminal identification and as an added security for ATM transaction. FRS also helps in identifying the blurred images of the terrorist got through any media. For example, a derived image of a terrorist or criminal under difficult illumination that is obtained from the news media has to be matched with the correct image in the data set available.

This paper primarily focuses on the issues from LBP [6] such as the difficulty in detecting eye position and the sensitivity against noise in uniform image regions. In the proposed design, the noise removal technique further eliminates the unwanted illuminations in the face image. The difficult lighting conditions of the input image are generally classified as less dark image, darker background image, shadowed image, partially darker image and fully dark image. The main processes in the face recognition system under difficult lighting condition include: equalization of

illumination, feature extraction and image classification. The equalization of illumination consists of the noise removal techniques that remove the unwanted noise in the first level. The feature extraction has the ATP and LTP feature sets for extracting the individual features from the face image. Next is the image classification which uses the *np* classification mechanism. In this classification the original face image is compared along with the noise free image, if the image matches then the face is finally recognized else it is unrecognizable.

2 Related Work

Gradient Transform Synthesis (GTS) is a new illumination formulation method which is used for estimating the too-bright and too-darker face regions. The main issues in this method include the lighting direction and correct estimation of the quality [7]. The illumination changes in a face image are an important issue in the face recognition system.

In feature extraction, the Differential Local Ternary Pattern (DLTP) is a method used for fusing the upper and lower kernels. The DLTP method is not strictly invariant to Gray-level transforms [6]. LBP, another feature extraction technique is not robust against local changes in the texture features at various view points and illumination directions [9]. The main problem in Gradient method is in reducing the dimensionality which leads to Heavy Small Sample Size (SSS) problem in Fisher's linear discriminant method (FLD). The SSS problem is greatly weakened since the dimensionality of the input feature for each FLD is much lower [5].

Yanwei Pang, et al [10], presented the Gabor method which is based on a localization of image features and a spatial constellation search over the localized features. An improved algorithm for image feature localization is used here. The Gabor filter is used in the feature extraction process that uses the Gradient Gabor (G-Gabor) filter which is used to extract multi-scale and multi-orientation features. This method is used to represent and classify a face that is defined based on the weighted Gaussian function which does not have much speed compared with Gabor Wavelet.

Face recognition system consists of several subtasks in a sequential manner, they are: face detection, face segmentation, normalization, and face verification. The main issues in the face recognition system are the illumination problem and the pose problem. The illumination problem is that the same face appears to be different in various lighting conditions. The changes induced by illumination could be larger than the differences between individual systems based on comparing images to misclassify the identity of the input image [14]. The next issue is the pose variations in the face images. Here the same face appears differently due to changes in viewing conditions. The analysis and classification of various pose problems are performed using a reflectance model. The pose problem is divided into three categories: (i) the simple case with small rotation angles; (ii) the most commonly addressed case when there are a set of training image pairs; (iii) the most difficult case when training image pairs are not available and illumination variations are present. An even more difficult case is the combined problem of pose and illumination variations.

3 Empirical Model of the Proposed System

The architecture of the proposed system is illustrated in Fig. 1. The first is the noise removal technique which consists of the four noise removal steps in which the unwanted noise in the image is removed. The second process is the feature extraction in which the image features are extracted from the pre-processed image. The third is the face recognition system which comprises image identification, classification and recognition.

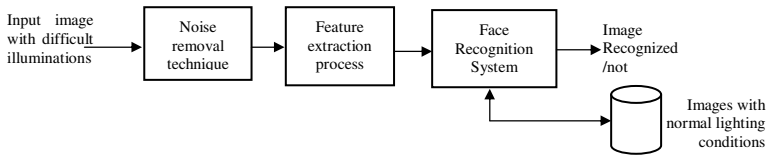


Fig. 1. Functional Architecture

3.1 Noise Removal Technique

Noise removal is a technique by which the difficult lightings are removed prior to the filtering process. During the transmission of images over the network, some random usually unwanted variation in brightness or color information may be added as noise. The noise removal technique consists of four functions that help in feature extraction [8] namely gamma correction, difference of gaussian filtering, masking, and equalization.

Gamma Correction. It is a transformation which converts the gray-level value with the user defined parameter. Gamma correction explains that the intensity of the light reflected from an object is the product of the incoming illumination L and the total surface reflectance R .

Difference of Gaussian Filtering (DoG). This filter is an easy way to achieve the resulting band pass values. Fine details remain critically important for recognition so the inner (smaller) Gaussian is typically quite narrow ($\sigma_0 \leq 1$ pixel). The outer one might have σ_1 of more than 4 pixels. The high-pass filtering removes the incidental information in the face image thus increasing the overall system performance.

Masking. Masking is the process by which the facial regions which are said to be irrelevant at any point are masked out.

Equalization. Equalization is the final step of the noise removal technique and it rescales the image intensities to normalize the overall contrast variations. The extreme values which are produced by the highlights, small dark regions near nostrils and at the image borders are compressed to a normalized form.

3.2 Feature Extraction

LTP Feature Set. The LTP feature set is a 3-valued code in which the face image is divided into 3x3 matrix format. The pixel values are -1, 0, and 1 according to the threshold t . The threshold value is calculated using the equation (1):

$$T = H_X \times W_Y / T_p^{N_p} \quad (1)$$

Here, H_X is the height of the region, W_Y is the width of the region, T_p is the type of pattern and N_p is the number of neighborhood values. The central pixel represents the threshold value of the face image. The values above the threshold T are set as +1; the values equal to threshold T is set as 0; and the values less than threshold T is -1. The LTP feature set [4] is sensitive to high noise level and it has larger bit size histogram that makes it very complex [3].

The problem with the detection of eye position in LBP is overcome by using the construction of eye model bunch, which is represented in three scales namely 256x256, 128x128, and 64x64 for locating the eye position correctly. First, Eye Model Bunch [11] of each eye is constructed from the normalized model of face images of size 256x256 followed by the second eye model bunch of size 128x128 down sampled from 256 x 256 model face images. Subsequently, the third eye model bunch about 64x64 is constructed in the same way. Next method focuses on the localization of eye coordinates that includes face detection, face normalization and multi-scale eye localization. Localized eye coordinates in 64x64 face image is based on Gabor jet similarity and it is achieved by taking the average eye coordinates [11]. Thus the issue mentioned in LBP is tackled in our proposed design and the resulting images are shown in Fig. 2. The face image in Fig. 2(a) shows the eye coordinates attained from LBP method whereas Fig. 2(b) spots the precise eye coordinates obtained from ATP method.

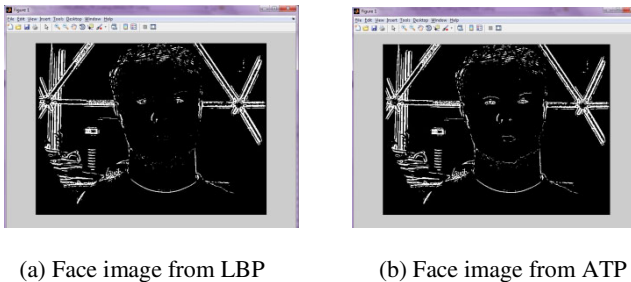


Fig. 2. Face images showing the eye coordinates

ATP Feature Set. ATP feature set presents the advanced version of the LTP feature set, which incorporates the normalization technique that helps in grouping the similar features precisely. Fig. 3 lists the pseudo code for the EFO method which is a part of ATP.

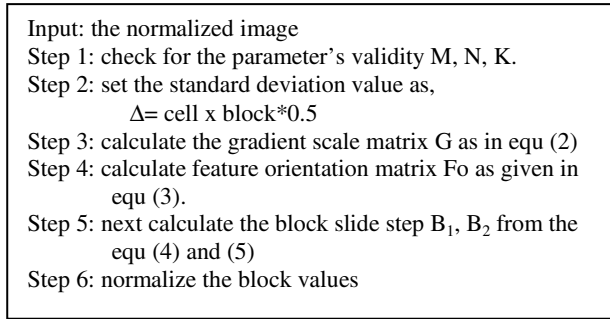


Fig. 3. EFO Algorithm

Gradient Computation. It uses many-feature extraction method and ensures normalized color and gamma values. For the gradient computation, the gradient scale matrix G is calculated from equ (2):

$$G = \text{sqrt}(\text{double}(G_x * G_x + G_y * G_y)) \quad (2)$$

where x, y are the gradient scales. The binary values from the matrix are converted as -1, 0, and 1 as done in LTP method and the values are fed into the gradient computation. This gradient filters the unwanted noise from the matrix scale x and y.

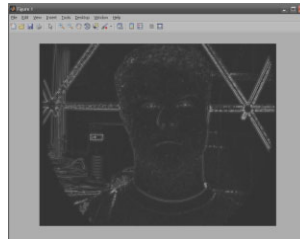


Fig. 4. Result of Feature Orientation

Feature Orientation. The result of the feature orientation process is shown in Fig. 4. Feature orientation step in EFO algorithm involves creating the cell histograms. Each pixel within the cell casts a weighted vote for an orientation-based histogram channel based on the values found in the gradient computation. The feature orientation is estimated as shown in equ (3):

$$F_o = \tan(G_y/G_x) + p_i/2 \quad (3)$$

where p_i is the pixel value of both the gradient scale x and y. Pixel contribution for the vote weight can either be the magnitude of the gradient, or some function of the magnitude. Other options for the vote weight could include the square root or square of the gradient magnitude, or some clipped version of the magnitude.

Descriptor Blocks. In order to account for the changes in illumination and contrast, the strength of gradient must be locally normalized, which requires grouping the cells

together into larger and spatially-connected blocks. Equ (4) and (5) explain the computation of blocking steps:

$$B_1 = N - \text{cell}_x * \text{block}_x + 1 . \quad (4)$$

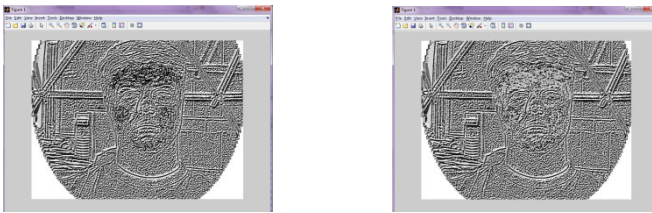
$$B_2 = M - \text{cell}_y * \text{block}_y + 1 . \quad (5)$$

Generally on a dense grid of uniformly spaced cells there presents overlapped cells and blocks. The above equations try to avoid such overlapping and to improve the accuracy of local contrast normalization. Thus the blocking feature of EFO method solves the high sensitivity in uniform image region as mentioned in LBP.

Block Normalization. Normalization follows the blocking. Let v be the non-normalized vector containing all histograms in a given block, $\|v\|_k$ be its k -norm for $k = 1, 2$ and ϵ be a small constant. Then the normalization factor can be stated as follows:

$$L_1 = \text{norm: } f = v / (\|v\|_1 + \epsilon) . \quad (6)$$

For improved accuracy, the local histogram can be normalized by contrast. The measure of intensity is calculated across a larger region of the image called a block and then this measure is used to normalize all the cells within the block. After the feature extraction is completed, the EFO normalization method controls the noise in the uniform image regions to be less sensitive as illustrated in Fig. 5(b). Dark traces in the forehead and cheeks of face image shown in Fig. 5(a) represent the noise resulting from LBP. This infers that ATP method is less sensitive to noise in uniform regions compared to LBP.



(a) Image after feature extraction in LBP (b) Image after feature extraction in ATP

Fig. 5. Face images showing the noise in uniform image regions

Face recognition and classification System. The dataset of the face images is trained first using the noise removal and feature extraction process. Features extracted from images are classified into groups according to its range of features. The normalization range is from 1-10 and the images belonging to different ranges are grouped and classified accordingly. After the classification of images, the input image is compared with the data set containing images under normal lighting conditions. If any of the image matches with the input image then the output will be the recognized image from the dataset else invalid message is displayed. For easier classification process the nearest feature space (NFS) metric is used to seek a NFS subspace to improve

the discriminating power of the subspace for classifications [1]. In this method, the feature space (FS) of each class is constructed in the manner similar to the FL constructed in NFL to model the variants of the training samples.

4 Empirical Analysis

The empirical analysis gives the intermediate results of various methods and techniques used in this paper. The noise removal technique eliminates any type of noise in the input image like gaussian, speckle, salt and pepper and in addition to this our paper also considers the noise due to lighting and illumination. ATP scheme is highlighted with the effective noise removal under fully dark lighting condition. It is observed that the fully dark image without the noise removal process produces the histogram value up to 18000, whereas the presence of noise removal process yields an improved histogram value of 30000. This implies that a lesser noise in the given image improves overall FAR rate. The histogram shown in Fig.6 (a) represents the fully dark image without equalization, and Fig. 6 (b) shows the histogram after equalization. The effectiveness of LTP and ATP on FAR with respect to equalization is illustrated in Fig. 7 and FAR reaches up to 70% and 78% respectively.

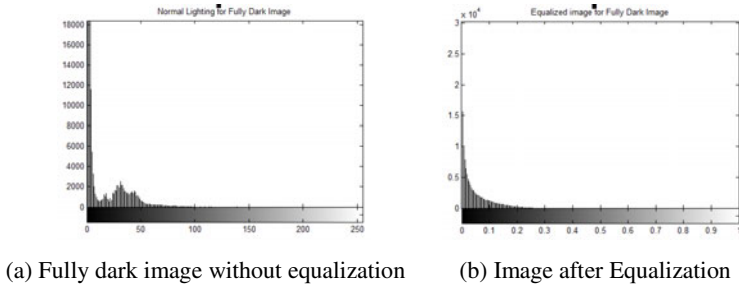


Fig. 6. Histogram comparison of FD image

The efficiency of the ATP method is illustrated by experimenting with the Extended Yale-B data base. Yale-B data set [13] comprised about ten human faces under 64 different illumination conditions has been recently upgraded as Extended Yale Face data set [12] to have 38 subjects under various 9 poses and 64 illumination conditions. The images are divided into five subsets according to the angle between the difficult lighting directions. Face recognition system is evaluated with five difficult lighting conditions and the results are given graphically in Fig.7.

From the results shown in Fig. 7, it is observed that ATP is the efficient feature set compared to other feature sets. 78% of FAR is the outcome of ATP in the case of FD image whereas LBP and LTP features produce a FAR of 65% and 70% respectively. Error, while recognizing the face, occurs when: (i) the respective image may not be available in the data set; (ii) the feature extraction process may not be efficient; (iii) the extracted image may not match with the exact normal image. The second situation is kept completely under control in ATP feature set and the error rate is restricted to

be within 0.1. In the case of less dark image (LD), ATP also shows a better FAR compared to that of LBP and LTP. The performance measure of ATP with and without EFO normalization is explained graphically in Fig. 7.

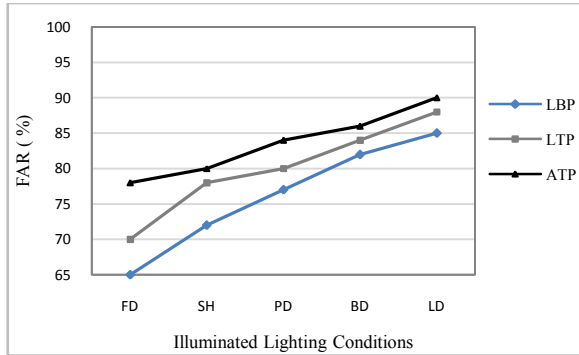


Fig. 7. FAR rate for the images under difficult lighting conditions

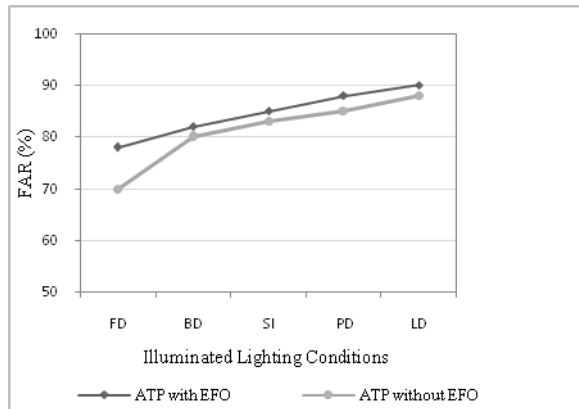


Fig. 8. Performance of ATP with and without normalization

Observations from Fig. 8 again prove that ATP feature set performs better under difficult lighting conditions. ATP with EFO normalization exhibits an improved acceptance rate under fully dark condition. EFO mainly contemplates in regularizing the features extracted from the illuminated face image. Strong illumination on a normal image may tend to change the shape and texture of the face resulting in imprecise feature extraction. EFO is preferred since it equalizes the illuminated features to a standard quality range from 1-10.

Now, ATP feature set is compared with other popular pre-processors such as histogram equalization (HE), multiscale retinex (MSR), logarithmic total variation (LTV), gross and brajovic (GB) methods. Fig. 9 shows a higher Face Recognition Rate (FRR) for ATP when compared with the above three noise removal methods.

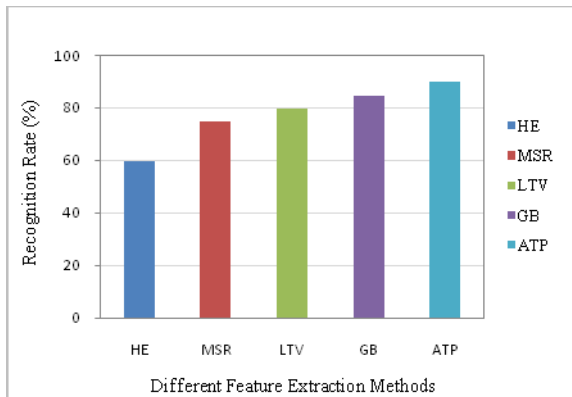


Fig. 9. Comparison of ATP with many other pre-processors

5 Conclusion

This paper has presented the Efficient Feature Orientation (EFO) algorithm used in ATP feature extraction, which is able to normalize the given image under difficult lighting condition. The major issues in LBP and LTP methods have been overcome by ATP mechanism using EFO normalization. Our proposed model of face recognition with ATP feature set can be recommended in law enforcements, such as tracing the terrorist with his/her picture under difficult lighting condition. The future work includes the exploration of various other efficient feature sets for improving the false acceptance rate under fully dark. ATP will also be enhanced to achieve 100% percentage of FAR, compared to the rest of the feature sets. The FAR rate can be further improved by using more efficient noise removal techniques.

References

1. Lu, J., Tan, Y.-P.: Nearest Feature Space Analysis for Classifications. *IEEE Signal Processing Letters* 18(1), 55–58 (2011)
2. Liao, W.-H.: Region Description Using Extended Local Ternary Patterns. In: *International Conference on Pattern Recognition*, Taiwan, pp. 1003–1006 (2010)
3. Akhlouf, M.A., Bendada, A.: Locally adaptive texture features for multispectral face recognition. In: *Computer Vision and Systems Laboratory, Laval University*, pp. 3308–3314 (2010)
4. Xie, S., Shan, S., Chen, X., Chen, J.: Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition. *IEEE Transactions on Image Processing* 19(5), 1349–1361 (2010)
5. Bendada, A., Akhloufi, M.A.: Multispectral Face Recognition in Texture Space. In: *Canadian Conference Computer and Robot Vision*, pp. 101–106 (2010)
6. Zhao, L., Chan, P.P.K., Ng, W.W.Y., Yeung, D.I.S.: A Novel Gradient Synthesis-based Illumination Formalization Method For Face Verification Under Varying Lighting Conditions. In: *Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, Qingdao*, pp. 669–674 (2010)

7. Tan, X., Triggs, B.: Enhanced Local Texture Feature Sets for Face Recognition under Difficult Lighting Conditions. *IEEE Transactions on Image Processing*, 1635–1650 (2010)
8. Cai, L.-b., Ying, Z.-l.: Face Recognition with Locally Linear Embedding on Local Binary Patterns. In: 1st International Conference on Information Science and Engineering (ICISE 2009), pp. 1246–1249 (2009)
9. Pang, Y., Yuan, Y., Li, X.: Gabor-Based Region Covariance Matrices for Face Recognition. *IEEE Transactions on Circuits and Systems for Video Technology* 18(7) (2008)
10. Kim, S., Chung, S.-T., Jung, S., Oh, D., Kim, J., Cho, S.: Multi-Scale Gabor Feature Based Eye Localization. In: *World Academy of Science, Engineering and Technology* (2007)
11. Lee, K.-C., Ho, J., Kriegman, D.J.: Acquiring Linear Subspaces for Face Recognition under Variable Lighting. *IEEE Transactions on Pattern Analysis And Machine Intelligence* 27(5) (2005)
12. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7) (1997)
13. Adini, Y., Moses, Y., Ullman, S.: Face Recognition: The Problem of Compensating for Changes in Illumination Direction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7) (1997)

Classification of Mammogram Images Using Discrete Wavelet Transformations

K.K. Rajkumar¹ and G. Raju²

¹ School of Computer Science, Mahatma Gandhi University, Kottayam, Kerala
rajatholy@yahoo.com

² School of Information Science & Technology, Kannur University, Kannur Kerala
kurupgraju@rediffmail.com

Abstract. A fractional part of biggest wavelet coefficient is enough to describe the characteristics of an image texture. Based on this concept a mammogram classification algorithm is developed. Using this classification algorithm, we classified the mammogram image into different classes as normal, benign and malignant. Ten percent of the images in each class are used for creating a class core vector. This class core vector acts as the base for the classification. The Euclidean distance is measured between the test image feature vector and the class core vector of the each class. A test image is classified into the appropriate class, which has minimum Euclidean distance measured between the test image and class core vector. Using this classification algorithm we classified 134 mammogram images into the exact class out 162 test images in the dataset. This algorithm results a detection rate of 75% for normal images, 88 % of detection rate for malignant and 100% detection rate for benign images respectively. The overall detection rate is 83%.

Keywords: Class core vector, Feature Vector, image texture, mammography, microcalcifications, Region of Interest.

1 Introduction

Breast cancer is the one of the most dangerous type of cancer found in the women all over the world. It is the second leading cause of death among the women. One among every eight women gets affected by the breast cancer at some point during their lifetime [1]. Early detection is the only provision for preventing this. However early detection of the breast cancer is not an easy task. The most accurate detection method adopted by medical experts is the biopsy. It involves some risks, patient discomfort and high cost. It also involves a high percentage of negative cases. The screen film mammography is the best suitable method accepted today for the diagnosis. It reduces the negative biopsy ratio and the cost to society by improving feature analysis and refining criteria for recommendation for biopsy

Clustered microcalcifications (MCS) are one of the mammographic hallmarks of early breast cancer. However not all MCS are the indication of malignancy, since they can occur during the course of other benign disease too [2]. X-ray mammography is currently considered as standard procedure for breast cancer diagnosis. But the

performance of X-ray mammography for breast screening is only about 75 % from the various studies performed by the different experts in this field[3]. So propose an efficient new algorithm for screening mammography is required. A digital mammogram is created when conventional mammogram is digitized through the use of specific mammogram digitizer or camera so that it can be processed by a computer [4]. Thus Digital Mammography is the best suitable diagnostic method for breast imaging today. It is widely used to detect and characterize breast cancer because of its high performance and low costs. The digital Mammography is a convenient and easy tool in classifying tumors and many applications in the literature prove its effective use in breast cancer diagnosis.

Texture information plays an important role in image analysis and detection in medical diagnosis. Texture is one of the important characteristics used in identifying an object or region of interest (ROI) in an image [5]. Since medical images are low contrast, complex anatomical structure and variability associated with their appearance, it is usually very difficult to distinguish between benign and malignant microcalcification clusters. An image region may be called cancerous (positive) or normal (negative) and a decision for detection result may be either correct (true) or incorrect (false). A decision for a detection result, therefore, will be one of four possible categories: true positive (TP), true negative (TN), false positive (FP) and false negative (FN). FN and FP represents two kinds of errors. An FN error implies that a true abnormality was not detected and a FP error occurs when a normal region was falsely identified as abnormal. A TP decision is correct judgment of an existing abnormality and a TN decision means that a normal region was correctly labeled [6].

In this paper we propose a new method for classifying mammogram images using discrete wavelet transformations. This method is an exact simulation of supervised classifier for classifying mammogram tumors into three categories, i.e. normal, benign and malignant. For the effective classification we formulated a class core vector for each class of ROI images extracted from the Mini-Mias database. Then each mammogram image in the database is tested using its feature vector against the class core vector created for each class using minimum Euclidean distance measure between them. We created class core vector using the ten percent image for each class of the image in the dataset. By using this algorithm, we classified more than fifty percent of the images in the Mini-Mias database to its appropriate classes. The ROIs' are extracted based on the abnormality center of the image in the database. This paper consists of six sections. Section II discusses the properties of mammogram images and the literature review of the recent works in mammogram classification using wavelet decomposition. The wavelet transformations and its characteristics are presented in Section III. In section IV we discussed the proposed method. The results in terms of numerical values are presented in Section V and finally Section VI concludes the work with its future scope.

2 Digital Mammograms

Mammography is a one of the most diagnostic method of analyzing breast cancer using X-rays. It is widely used to detect and characterize breast cancer because of its high performance and low costs. It is the most suited imaging technique for screening

programs. Mammogram images obtained using X-rays are gray scale images, which indicates the details inside the patient breast by means of its contrast. The details could be normal tissues, vessels, muscles, different types of masses and noise.

Each type of masses has different properties of shape, size, distribution, and brightness which act as a feature that help the radiologist to effectively diagnosis the breast tumors. This means that the experience of the radiologist and the image quality are the important factors in this manual classification. Micro calcifications are small groups of calcium deposit that has a long scale in form, size and distribution [7]. In this study we used a set of mammogram images provided by Mammographic Image Analysis Society (MIAS) [8] for the testing purpose. These images are investigated and labeled by an expert radiologist. The original mammograms are 1024 x 1024 pixels. From the selected images a regions of interest (ROI's) are extracted with size of 128 x 128 pixels. These ROI's are abnormality centered.

Cristiane Bastos Rocha Ferreira et. al. proposed and constructed a supervised classifier for classification of radial, circumscribed, micro calcifications, and normal samples. This paper also discussed the classification of benign, malign and normal mammograms using special sets of the coefficients as the features for separating each of the classes in the mammogram images using the minimum Euclidean distance measure [9].

H. Soltanian-Zadeh et. al presented an evaluation and comparison of the performance of four different texture and shape feature extraction methods for classification of benign and malignant micro calcifications in mammograms [10]. They extracted micro calcification clusters, texture and shape features using different approaches like conventional shape quantifiers, co-occurrence based method of Haralick and multi level wavelet transformations. They achieved better results on multi level wavelet transformations.

Essam A Rashed et al developed a supervised diagnosis system for digital mammograms [11]. In this model, a diagnosis process is done by transforming the data of the images into feature vector using wavelets multilevel decomposition. This vector is used as the features for separating different mammogram classes. This model classified mammogram images into tumor types and risk level. The results obtained by this method are very promising.

Essam A Rashed et al also suggested a multiresolution analysis system for interpreting digital mammograms. This system is based on using fractional amount of biggest wavelets coefficients in multilevel decomposition. They used 25% of the Mini-Mias images for creating a class core vector and the entire ROI's of mammogram images from Mini-Mias database is classified by taking the minimum Euclidean distance measure from each mammogram images to the class core vector [12].

3 Wavelet Transformations

Image texture, especially in medical images are confusing measurement that depends mainly on the scale in which the data are observed. Different types of images have different types of texture [13]. Earlier studies proved that the texture of mammograms is an irregular texture. To classify the mammogram images, we have to investigate the transformation data into different domain. Thus wavelet multiresolution domain is the best choice for automatic classification of mammogram images.

Wavelets Transform(WT) [14] were first introduced to medical imaging research in 1991 in a journal paper describing the application of wavelet transforms for noise reduction in MRI images. The wavelet transform is a decomposition of an image onto a family of functions called a wavelet family, in which all of the basic functions (called wavelets) are derived from scaling and translation of a single function that is called the mother wavelet (or analyzing function). The Wavelet transform gives a time-frequency representation of a signal that has two main advantages: (a) an optimal resolution both in the time and frequency domains; and (b) the lack of stationary nature of the signal. It is defined as the convolution between the signal $X(t)$ and the wavelet functions $\psi_{a,b}(t)$.

$$W\psi X(a,b) = \langle X(t) | \psi_{(a,b)}(t) \rangle \quad (1)$$

Where $\psi_{a,b}(t)$ are dilated (contracted) and shifted versions of a unique wavelet function $\psi(t)$

$$\Psi(a,b) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

(a, b are the scale and translation parameters, respectively). The WT gives a decomposition of $X(t)$ in different scales, tending to be maximum at those scales and time locations where the wavelet best resembles $X(t)$. Moreover, Eq. (1) can be inverted, thus giving the reconstruction of $X(t)$. The WT maps a signal of one independent variable t onto a function of two independent variables a, b . This procedure is redundant and not efficient for algorithmic implementations. In consequence, it is more practical to define the WT only at discrete scales a and discrete times b by choosing the set of parameters $\{a_j = 2^{-j}; b_{j,k} = 2^{-j} k\}$, with integers j, k .

Contracted versions of the wavelet function match the high frequency components of the original signal and on the other hand, the dilated versions match the low frequency oscillations. Then, by correlating the original signal with wavelet functions of different sizes we can obtain its details at different scales. These correlations with the different wavelet functions can be arranged in a hierarchical scheme called multi resolution decomposition. The multi resolution decomposition separates the signal into 'details' at different scales, the remaining part being a coarser representation of the signal called 'approximation'.

In this study, using discrete wavelets mammogram images are decomposed into four different levels. The decomposed image contains four scales of details and an approximation. The lower levels give the details corresponding to the high frequency components and the higher levels corresponding to the low frequencies. The decomposition coefficients produced are:

- Low frequency coefficients (A).
- Vertical high frequency coefficients (V).
- Horizontal high frequency coefficients (H).
- High frequency coefficients in both diagonal directions (D).

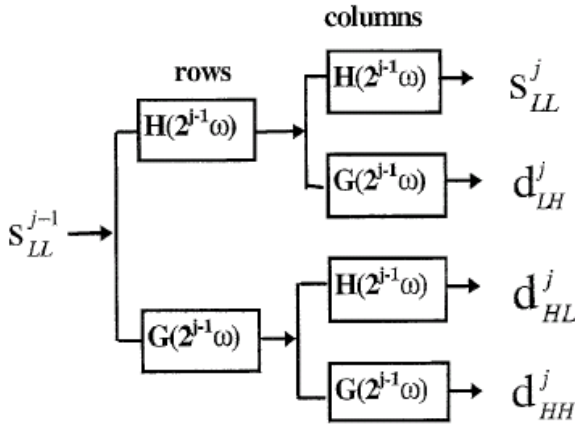


Fig. 1. Wavelet sub band decomposition

After the decomposition of the mammogram images, class core vector and feature vectors are generated by using the approximation coefficients (low frequency components of the decomposition) in each level for three different wavelet families: Daubechies, Biorthogonal spline and Haar. Both the core class vector and feature vectors are created using a fractional part of the biggest coefficients of low frequency components of the decomposition.

4 Proposed Method

The method proposed in this paper is a modification of the work suggested by Rocha Ferrerira et al and E.A Rasheed et al in their papers[9][11]. In the modification part, we computed class core vectors for each class of mammogram images and the comparison also performed with each image feature vector with the corresponding class image core vector instead of the core vector of the entire dataset by [11]. The class core vector is also created using only 10 percent of the images in each class. But the earlier study uses 25% of images from the entire dataset. The distance measure for the classification is also done for each class of the core vector instead of the entire class core vector of the earlier study. Using this way of classification, this algorithm has no misclassification of the method suggested in [9][11]. This method is analogous to the supervised classification method used in Artificial Neural Network. As a first step of the classification, we created a class core vectors for each class of images using wavelet multi level decomposition. Then the feature vectors of the individual test images are generated in the same way and classified the test image using minimum Euclidean distance measure.

The class core vectors of the image in each level of the wavelet decomposition is created randomly using 10% of the images from normal, benign and malignant classes of images in the Mini-Mias database. This is done by creating the class core vectors in four different levels of wavelet decompositions of three different wavelet families named Daubechies, Haar and Biorthogonal wavelets. In each level of the wavelet decomposition, a percentage of the biggest low frequency coefficients are used to

represent the mammogram feature vector. The selection of the percentage of low frequency coefficients is varied from small percentage to high for covering all possible values of the low frequency coefficients of the class core vector and feature vector. This may contribute promising results to the exact classification of the images in the database.

A set of 322 ROI's are extracted based on the abnormality center of the original mammogram images from the Mini-Mias database. For benign and malignant classes ROI's of size 124 x 124 are extracted based on abnormality centre. In the case normal images ROI's of same size are extracted from the center part of the image. After creating the ROI's, ten percent of the images, randomly selected from each class are used for generating class core vector. The class core vectors are created on each level of the wavelet decomposition using the following equation:

$$C_m^j = \frac{1}{N} \sum_{n=1}^N A_m^j(i) \quad (3)$$

Where C_m^j the m th class core vector at j level decomposition, N_j is the number of ROI's selected to produce the class core vector and A_m^j is the fraction of biggest wavelet coefficients of the ROI's selected from the mammogram image for the class m at decomposition level j .

The testing of individual mammogram image in the dataset is accomplished by designing a new classifier algorithm, which is based on calculating the distance between the feature vector and the class core vector on all four different levels of the wavelet decomposition. This new system automatically classifies the test image by finding the minimum Euclidean distance between feature vectors of the test image to the each of the class core vector by following equation:

$$\text{Dist}(A, C_m^j) = \frac{1}{J} \sum_1^J \sum_1^m \sqrt{(A^j(i) - C_m^j(i))^2} \quad (4)$$

Where A^m is the coefficient vector of the j^{th} decomposition level for the test image, C_m^j is the class core vector for class m at decomposition level j and m is the number of classification classes. Here m is 3 (Normal, Benign and Malignant).

5 Results

The classification algorithm described in this paper is implemented in Matlab 7.0 and tested the algorithm on images in the Mini-Mias database. In this experiment we tried to classify the risk level of images in the database. For the classification, we created a class core vector for each class using ten percent of images from all the three categories randomly. Then 162 images consists of 98 normal images, 38 benign and 26 malignant images are randomly selected from the dataset for testing purpose.

The Daubechies-4, Daubechies-8 and Daubechies-16 from Daubechies family, Haar wavelet from Harr family and Bior2.8 from Biorthogonal family are used for creating class core vector of the training class and feature vector for test image in

wavelet multilevel decomposition. In each class, four levels of decompositions are performed, and a fractional part of the biggest coefficients are used for creating class core vector. The feature vector of the test image of each class is also created. Then the Euclidean distance between the class core vector and the feature vector of the test image is calculated by taking average mean distance between the class core vector and the feature vector of the image in all four different levels of the wavelet decomposition. The tested image is then labeled to the respective class where Euclidian distance between feature vectors is minimum to that of class core vector.

Table 1. Successful classification rate of mammogram images using Daubechies wavelet decomposition

Coef. In %	Daubechies									
	Db4			Db8			Db16			
		N	B	M	N	B	M	N	B	M
25	N	72	09	17	73	08	17	72	09	17
	B	0	37	01	0	38	0	0	38	0
	M	04	03	19	04	03	19	04	03	19
	T	76	49	37	77	49	36	76	50	36
50	N	74	06	18	74	06	18	74	06	18
	B	0	38	0	0	38	0	0	38	0
	M	02	02	22	02	02	22	02	02	22
	T	76	46	40	76	46	40	76	46	40
75	N	69	06	23	69	06	23	69	06	23
	B	0	36	02	0	36	02	0	36	02
	M	02	01	23	02	01	23	02	01	23
	T	71	43	48	71	43	48	71	43	48
100	N	68	03	27	68	03	27	68	03	27
	B	0	36	02	0	36	02	0	36	02
	M	02	01	23	02	01	23	02	01	23
	T	70	40	52	70	40	52	70	40	52

N : Normal Images B:Benign Images M: Malignant Images T:Total

The classification results of each 162 mammogram images' using this classification algorithm is shown in Table 1 and Table 2. For 98 normal images 74 images are correctly classified using the 50% of the biggest wavelet coefficients in both Daubechies and Haar wavelet family decomposition, which is 75.5 %. Out of 38 benign images, all the 38 images are correctly classified in Db4, Db8, Db16 and Haar using 50% of the biggest wavelet coefficients in the decomposition, which is 100%.

The same percent of result is obtained in Db8 and Db16 of Daubechies family using only 25 % of wavelet coefficients. In 26 malignant images, 23 images are correctly classified by this system using 75% of wavelet coefficient for all the three wavelet families, which is 88%. Using Biorthogonal wavelet, irrespective of all fractional part of the wavelet coefficients, identified 23 malignant images correctly. In addition to the individual classification 134 images in the whole 162 images are correctly labeled and classified using this supervised classification algorithm, which is 83% of detection rate.

Table 2. Successful classification rate of mammogram images using Haar & Biorthogonal wavelet decomposition

Coef. In %	Haar			Biorthogonal			
		N	B	M	N	B	M
25	N	72	09	17	68	03	27
	B	0	37	01	0	36	02
	M	04	03	19	02	01	23
	T	76	49	37	70	40	52
50	N	74	06	18	68	03	27
	B	0	38	0	0	36	02
	M	02	02	22	02	01	23
	T	76	46	40	70	40	52
75	N	69	06	23	68	03	27
	B	0	36	02	0	36	02
	M	02	01	23	02	01	23
	T	71	43	48	71	40	52
100	N	68	03	27	68	03	27
	B	0	36	02	0	36	02
	M	02	01	23	02	01	23
	T	70	40	52	70	40	52

N : Normal Images B:Benign Images M: Malignant Images T:Total

6 Conclusions

The ultimate aim of this project is to formulate an effective automated mammogram classification system, which is something really should happen? The method used in this paper is the concept of using biggest wavelet coefficients as the feature vector

and the idea of multilevel decomposition of mammogram analysis. This work identifies 134 mammogram images into its respective class of images from the 162 images selected randomly out of 322 images in the Mini-mias database. The extension of this work should be the exact classification all images (322) in the database. In addition to this, the work also should be extended to the sub classification of benign and malignant images into micro calcification, speculation, ill defined, asymmetrical distortion and circular lesions. The future works should be the effective sub classification of this. Also we plan to make comparative study of this classification using stationary wavelets.

References

1. Cheng, H.D., Wang, J.: Fuzzy logic and scale space approach to microcalcification detection. In: ICASSP. IEEE, Los Alamitos (2003) ISBN No.0-7803-7663-3/03
2. Soratin, E., Schmidt, F., Mayer, H., Becker, M., Szepesvari, C., Graif, E., Winkler, P.: Computer Aided Diagnosis of Clustered Microcalcifications using Artificial Neural Nets. *Journal of Computing and Information Technology-CIT* (2), 151–160 (2002)
3. Islam, M.J., Ahmadi, M., Sid-Ahmed, M.A.: An efficient Automatic Mass Classification method in Digitized Mammogram using Artificial Neural Network. *International Journal Intelligence & Applications(IJIA)* 1(3) (2010)
4. Rajkumar, K.K., Raju, G.: Enhancement of mammograms using top-hat filtering and wavelet decomposition. In: International Conference on Mathematics and Computer Science, ICMCS (2010), ISBN 978-81-908234-2-5
5. Kim, J.K., Park, H.W.: Statistical texture features for detection of microcalcifications in Digitized Mammograms. *IEEE Transactions on Medical Imaging* 18(3) (1999)
6. Sakka, E., Prentza, A., Koutsouris, D.: Classification algorithms for microcalcifications in mammograms(review). *Oncology Reports* 15, 1049-1055 (2006)
7. Hendrik Veldkamp, W.J.: Computer aided characterization of Microcalcification clusters in Mammograms, Ph.D thesis submitted to University of Nijmegen in 2000 (2000)
8. The Mammographic Image Analysis Society: Mini Mammography Database, <http://www.wiau.man.ac.uk/service/MIAS/MIAS/mini.htm>
9. Ferreira, C.B.R., Borges, D.L.: Analysis of mammogram classification using wavelet transforms decomposition. *Pattern Recognition Letters* 24, 973–982 (2003)
10. Soltanian-Zadeh, H., Rafiee-Rad, F., Siamak Pourabdollah-Jejad, D.: Comparison of multiwavelet, wavelet Haralick and shape features for microcalcification classification in mammograms. *The Journal of the Pattern Recognition Society, Pergamon, Pattern Recognition* 37, 1973–1986 (2004)
11. Rashed, E.A., Awad, M.G.: Neural networks approach for mammography diagnosis using wavelet features. In: CSCBC First Canadian Student Conference on Biomedical Computing (2006)
12. Rashed, E.A., Ismail, I.A., Zaki, S.I.: Multiresolution mammogram analysis in multilevel decomposition. *Pattern Recognition Letters* 28, 286–292 (2007)
13. Mini, M.G., Devassia, V.P., Thomas, T.: Multiplexed Wavelet transformation technique for detection of Microcalcification in Digitized mammograms. *Journal of Digital Imaging* 17(4), 285–291 (2004)
14. Soman, K.P., Ramachandran, K.I.: Insight into wavelets theory to practice. In: PHI (2004)

Optimized Trace Transform Based Feature Extraction Architecture for CBIR

M. Meena, K. Pramod, and K. Linganagouda

Department of Electronics and Communication,
B.V. Bhoomaraddi College of Engineering and Technology, Hubli-31.

Abstract. The feature extraction and similarity measure of CBIR algorithms involve highly computation intensive and repetitive operations on a large data set yet have to satisfy the real time application requirements. One way to supplement software approaches for this purpose is to provide hardware support to the system architecture. We propose an algorithm and architecture for hardware implementation of trace transform based feature extraction for CBIR system. The proposed algorithm is focused to reduce the computational complexity in the addressing block for trace based feature extraction and is also optimized for memory consumption and speed for distance calculations in the similarity measure phase. Synthesis results show that the above measures are responsible for reducing the response time of the retrieval process by being able to process 2725 images per sec.

Keywords: Content Based Image Retrieval, Trace Transform, FPGA.

1 Introduction

The rapid growth in internet technology ,combined with increased computational processing power of hardware and software have all lead to increased ability to store and manipulate more complex data types involving images and videos .Retrieving of images from a huge data base offers challenges that demands a different approach towards the design of algorithm for retrieving from the traditional text based approach .Text based image retrieval system has its roots in the traditional information retrieval and relies on textual description of images, mostly using keywords. However, it has been shown that there are three major difficulties to this approach; subjectivity, volume, and explicability. These disadvantages of text-based image retrieval techniques call for another relatively new technique known as Content-Based Image Retrieval (CBIR). CBIR Technology boils down to two intrinsic problems i) how to mathematically describe an image ii) how to assess the similarity between a pair of images based on their abstracted descriptions. The mathematical description of the image for retrieval purpose is referred to as the image signature. The image signature is computed directly from visual features of the image such as colors, shapes, contrast, etc. These features are used to describe the entire image and provide a dramatic reduction of data, which is required to perform online retrieval. The approach towards similarity measure depends on the types of signatures. Euclidean

and geodesic distance measures have been adopted for feature vector based signature whereas weighted sum of vector distances is used for region based signature.

The feature extraction and similarity measure of CBIR algorithms involve highly computation intensive and repetitive operations on a large data set yet have to satisfy the real time application requirements. One way to supplement software approaches for this purpose is to provide hardware support to the system architecture [1]. Unfortunately, very little has been explored in this direction, partly due to the lack of agreed-upon indexing and retrieval methods. There are a notable few hardware implementations reported in the literature [2, 3, 4, 5]. The authors of [2] propose an FPGA implementation for sub image retrieval within an image database for binary image. Color histogram based method has been proposed by [3]. The proposal implementation in hardware proved very efficient, since it requires elements such as filters and shifters, which perform very well on hardware in terms of speed and silicon area. The authors of [4] have proposed a parallel architecture for distance calculations in the similarity measure stage of the architecture proposed in is based on the unique combination of reconfigurable resources combined to Flash memory, and allows for a speed-up of 45 as compared to existing software solutions. In [5], a hardware implementation of a fuzzy modified ant colony processor that is suitable for image retrievals is presented for the first time. The main characteristic of the proposed digital circuit is the use of pipelined data supply and parallel processing. examples of micro-architectures related to the dominant colors descriptor and the compact color descriptor.

We propose an algorithm and architecture for hardware implementation of trace transform based feature extraction for CBIR system. The proposed algorithm is focussed to reduce the computational complexity in the addressing block for trace based feature extraction and is also optimized for memory consumption and speed for distance calculations in the similarity measure phase. The above measures are responsible for reducing the response time of a retrieval process.

2 Proposed Algorithm and Architecture for Trace Transform Based Feature Extraction

Transform based feature extraction for CBIR algorithms have been implemented using wavelet, Radon, Hough and Trace transform. Wavelets [6] have often been considered for their locality and their compression efficiency. Wavelet transforms have been applied to texture representation [7, 8] sometimes in conjunction with Markovian analysis. The authors of [9] propose two rotation and scale invariant features extracted from the Hough transform domain to guide a CBIR system in the search of relevant building images. [10] have used local Radon Transform by aligning the centre of the Radon Transform with the centroids of the region of interest to provide a robust CBIR feature. [12] Propose a radon transform and 1-D Discrete Cosine Transform based hashing algorithm which is more secure and robust.

Trace transform [13] which is an alternative image representation and from which we can construct the so called triple features. In [14] it was shown how one can

construct such features invariant to rotation, translation and scaling, hence object signatures invariant to affine transforms can be constructed. The trace transform is a generalization of the Radon transform in such a way that instead of computing only the integral along the lines tracing image many functionals [Table 1] can be computed of the image function to obtain the trace matrix or the trace image in Fig 1. Applying two others functionals (Circus and Diametric) to the each parameter of tracing line, yields to a Triple feature, which characterize the original image in number format. To compute invariant Triple features a sophisticated combination of three functionals (Trace, Circus and Diametric) is needed.

We propose to use trace transform for obtaining the image descriptor as the transform is invariant to translation, rotation, scaling, local texture distortion, and local geometrical deformation. We have selected trace functionals among the multitude of well defined trace functionals that are suitable for hardware implementations. The selected functionals involve integers and not floating point operations and hence are appropriate for FPGA environment. The identifier so obtained has comparable performance in terms retrieval accuracy at the same time performing better in terms of retrieval speed with respect to the reported CBIR algorithms.

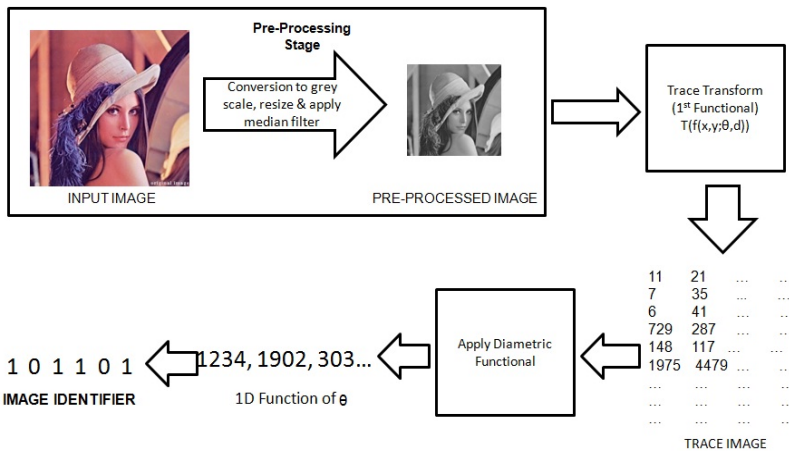


Fig. 1. Algorithm flow

2.1 Proposed Algorithm

I. PRE-PROCESSING

- i. The input colored image is converted to grey scale.
- ii. Image is downscaled to 64 X 64 size.
- iii. Apply median filter to remove the noise.

II. FEATURE EXTRACTION

- i. Trace Transform is applied to obtain the trace transform matrix.
- ii. A diametric function is applied to obtain -1d.

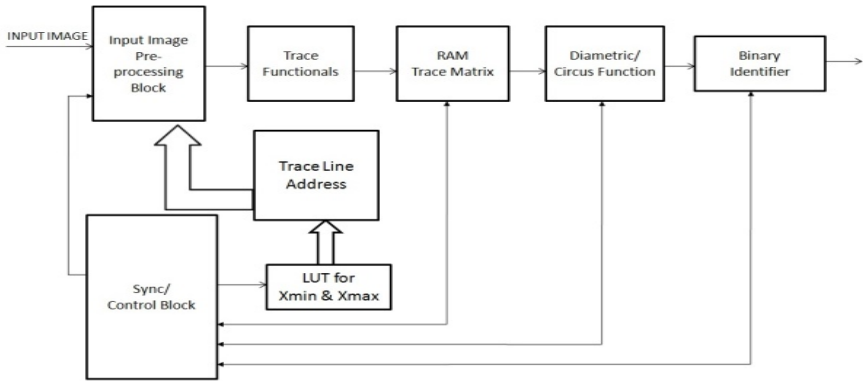


Fig. 2. Functional diagram of the proposed architecture

III. EXTRACTION OF BINARY BIT STRING TO BE STORED IN DATA BASE.

IV. STEPS 1, 2 AND 3 ARE REPEATED FOR QUERY IMAGE.

V. SIMILARITY MEASURE IS DONE USING XOR OPERATION.

2.2 Proposed Architecture

2.2.1 Pre-processing

Preprocessing involves conversion of images to grey scale, filtering and down scaling the images. Grey scale conversion contributes to the reduction in the volume of the input data hence also the memory size. Filtering is achieved using median filters. The choice of median filters is that they belong to a class of spatial window based filters and involve integer operations and hence are suitable for implementation on FPGA platform. Architecture for implementing median filter has been proposed in [15]. Median filters show optimum performance in the removal of impulsive noise and impulsive noise effects the results of trace transform functional, especially those that involve the minimum and maximum intensity values along the trace lines when compared to Gaussian noise because impulsive noise changes the pixel values drastically. Median filters effectively remove these outliers thus have an averaging effect on the pixel values which contributes to the robustness of the identifier against transformations like rotation. Downscaling is utilized to decrease the run-time of the feature extraction process complexity and the storage space requirement. Downscaling does not affect retrieval accuracy of the trace feature based retrieval algorithm.

2.2.2 Trace Transform Matrix

We have used functional 1 of Tab.1.to obtain the 2D trace transform matrix. Trace transform involves mapping of the pixel value from spatial domain to the trace domain Fig.3. The pixels in the trace domain are defined by the trace line t which is a function of angle θ and distance d from the origin. The pixels along the trace lines need to be addressed for applying the desired trace functionals. Though the

computation of the trace functionals is simple the extraction of the pixels is a complicated task. For an image of size $N \times N$, the maximum number of trace lines n_t for each angle of rotation θ , are $\sqrt{2}N$. Let the number of rotation angles be n_θ , then the number of pixels n_p along the trace line will be $\sqrt{2}N$, the total number of addresses to be generated is

$$n_{addr} = n_t n_\theta n_p.$$

$$n_{addr} = 2N^2 n_\theta$$

Each address is obtained by

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

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

The cost of addressing in terms of operations for one degree resolution is

$$C_{tot} = 1400N^2 [C_{trig} + C_{mul} + 2C_{add}]$$

Table 1. Trace Functionals

Trace Transform	Functionals
1	$T(f(x, y)) = \int_0^\infty f(t) dt$
2	$T(f(x, y)) = \int_0^\infty (f(t))^2 dt$
3	$median_{r>0}\{f(r), (f(r))^{1/2}\}$ where $r = t - c$

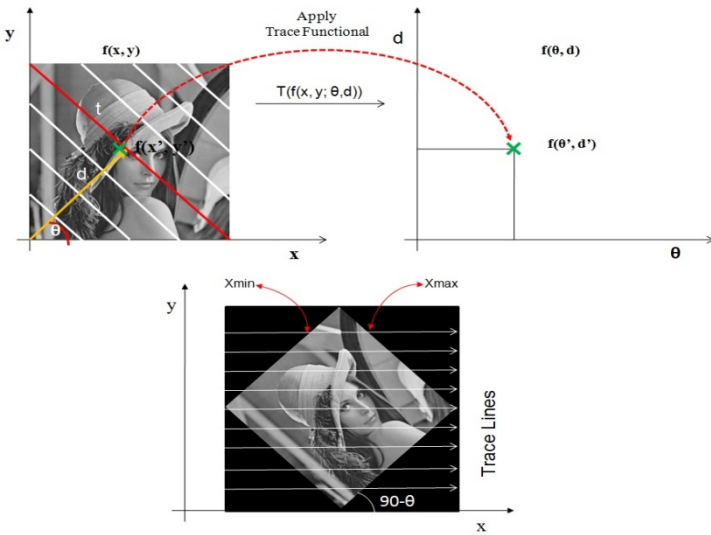


Fig. 3. Trace Transform

We have applied two diametric functions sum along the columns of the trace matrix, and maximum slope along each column to obtain the identifier bit string. We have reduced this complexity by rotating the image by $(90 - \theta)$, so that the trace lines are horizontal and the pixels can be addressed in scan mode Fig 3. One of the factors to be looked into while using this approach was that size of the rotated image changes due to which we need to consider only those pixels along the trace line within the image hence we have calculated the start and end points for the trace lines X_{min} and X_{max} for each angle of rotation θ . These values are stored in the LUT. Addresses for each trace line are generated by simply incrementing from X_{min} until X_{max} is reached. Table 2 gives the comparison of the costs with [16] where the reduction in the addressing complexity is achieved by using LUT for $\sin \theta$ and $\cos \theta$.

Table 2. Comparison of costs

Sl. No.	Approach	Memory Size	Mathematical Complexity
1	LUT for $\sin \theta$ and $\cos \theta$ [16]	6Kbytes	$C_{tot} = 1400N^2[C_{mul} + C_{add}]$ Floating point multiplications and additions
2	LUT for all the required pixel addresses	84 Mbytes(256X256)	Nil.
3	LUT X_{min} and X_{max}	3Kbytes	$n_{\theta} * n_t$ comparator operations

3 Implementation and Results

The proposed architecture for trace transform based feature extraction has been captured using VHDL and designs instantiated using Spartan 3. Xilinx ISE 12.2 with XST synthesizer was used in the synthesis of this design. ModelSim-XE III 6.5c has been used for functional simulation and testing. Table 3 summarizes the implementation results of the proposed architecture. Fig 4 gives the RTL schematic. Fig 5 gives the functional simulation of the proposed architecture for downscaled image of size 64x64 and angle resolution of 5 degrees. Fig 6 gives the functional of the proposed architecture when parallelism is used by operating on four rotated versions simultaneously. Synthesis results are given in Table 3.

Table 3. Synthesis Results

One Image	Parallel Approach(4 Images)
No. of Slices: 1 out of 768	No of Slices:4 out of 768
No. of 4 i/p LUTs: 1 out of 1536	No of 4 i/p LUTs: 1 out of 1536
No of IOs: 51	No of IOs: 51
No of bonded IOBs: 51 out of 124	No of bonded IOBs:51 out of 124
No of BRAMs:1 out of 4	No of BRAMs:4 out of 4
No of GCLKs:1 out of 8	No of GCLKs:1 out of 8
Frequency: 681.4 Images/Sec	Frequency: 2725.58 Images/Sec

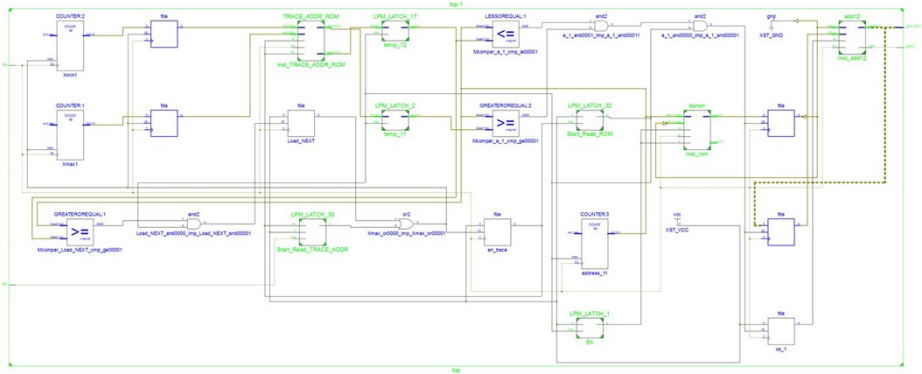


Fig. 4. RTL Schematic for Trace Functional

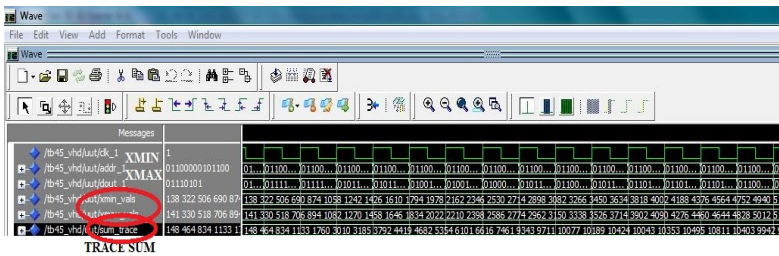


Fig. 5. Snapshot of Trace Sum

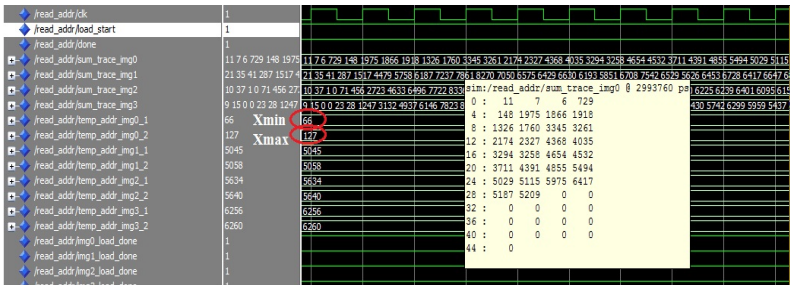


Fig. 6. Snapshot of Trace Sum (Four Images parallel)

4 Conclusion

We have proposed architecture for trace based feature extraction which has reduced addressing complexity by using a scan mode of addressing to retrieve the pixels along the trace line in scan mode at the cost of extra memory (10 KB) for storing the start and end addresses of the trace lines. The proposed algorithm reduced the computational complexity of trace functional based on integer operations. The architecture was further optimized for speed by introducing parallelism to calculate

the functional on four rotated images simultaneously at the cost of incremental increase in the memory. Implementation results show optimization in speed and device utilization when compared to reported architectures [3,16].

We propose to optimize the architecture for similarity measure and memory utilization for the identifier database.

References

1. Datta, R., Joshi, D., Li, J., Wang, J.Z.: Image Retrieval: Ideas, Influences, and Trends of the new age. *ACM Computing Surveys* 40(2), Article 5 (April 2008), doi:10.1145/1348246.1348248
2. Nakano, K., Takamichi, E.: An image retrieval system using FPGAs. In: *Asia and South Pacific Design Automation Conference (ASP-DAC 2003)*, pp. 370–373. ACM, New York (2003)
3. Kotoulas, L., Andreadis, I.: Colour histogram content-based image retrieval and hardware implementation. *IEEE Proc.-Circuits Devices Syst.* 150(5), 387–393 (2003)
4. Chikhi, R., Derrien, S., Noumsi, A., Quinton, P.: Combining Flash Memory and FPGAs to Efficiently Implement a Massively Parallel Algorithm for Content-Based Image Retrieval. In: Diniz, P.C., Marques, E., Bertels, K., Fernandes, M.M., Cardoso, J.M.P. (eds.) *ARCS 2007*. LNCS, vol. 4419, pp. 247–258. Springer, Heidelberg (2007)
5. Konstantinos, Georgios, Ioannis: Design and implementation of a fuzzy-modified ant colony hardware structure for image retrieval. *Trans. Sys. Man Cyber. Part C* 39(5), 520–533 (2009)
6. Daubechies: *Ten Lectures on Wavelets*. SIAM, Philadelphia (1992)
7. Choi, H., Baraniuk, R.: Multiscale Texture Segmentation Using Wavelet-Domain Hidden Markov Models. In: *Proc. 32nd Asilomar Conf. Signals, Systems, and Computers*, vol. 2, pp. 1692–1697 (1998)
8. Laine, A., Fan, J.: Texture Classification by Wavelet Packet Signature. *IEEE Trans. Pattern Analysis and Machine Intelligence* 15(11), 1,186–1,191 (1993)
9. Xiang, L.: CBIR approach to building image retrieval based on invariant characteristics in Hough domain. In: *IEEE ICASSP*, pp. 1209–1212 (2008)
10. Xiong, W., Ong, S.H., Lee, W., Foong, K.: Local Radon Transform and Earth Mover's Distances for Content-Based Image Retrieval. In: Satoh, S., Nack, F., Etoh, M. (eds.) *MMM 2008*. LNCS, vol. 4903, pp. 436–445. Springer, Heidelberg (2008)
11. Vieux, R., Benois-Pineau, J., Domenger, J.-P., Braquelaire, A.: ESPI Image Indexing and Similarity Search in Radon Transform Domain. In: *IEEE Seventh International Workshop on Content Based Multimedia Indexing*, pp. 231–236 (2009)
12. Ou, Y., Rhee, K.H.: A Key-Dependent Secure Image hashing Scheme by Using Radon Transform. In: *IEEE International Symposium on Intelligent Signal Processing Communication Systems(ISPACS)*, pp. 595–598 (2009)
13. Kadyrov, A., Petrou, M.: The Trace Transform and Its Applications. *IEEE Trans. PAMI* 23(8), 811–828 (2001)
14. Kadyrov, A., Petrou, M.: Object Signatures Invariant to Affine Distortion Derived from Trace Transform. *Image and Vision Computing* 21, 1135–1143 (2003)
15. Meena, S.M., Linganagouda, K.: Implementaion and Analysis of optimized architectures for rank order filter. *Journal of Real-Time Image Proc.* 3, 33–41 (2008)
16. Fahmy, S.A., Bouniggins, C.-S., Cheung, P.Y.K., Wayne Luk, J.: Real-Time Hardware Accelration of The Trace Transform. *Journal of Real-Time Image Proc.* 2, 235–248 (2007)

Multi-algorithm Fusion for Speech Emotion Recognition

Gyanendra K. Verma, U.S. Tiwary, and Shaishav Agrawal

Indian Institute of Information Technology, Allahabad, India
Allahabad, India - 211012
{gyanendra,ust}@iitaa.ac.in,
shaishav.engr@gmail.com

Abstract. In this paper, we have proposed a speech emotion recognition system based on multi-algorithm fusion. Mel Frequency Cepstral Coefficients (MFCC) and Discrete Wavelet Transform (DWT), the two prominent algorithms for speech analysis, have been used to extract emotion information from speech signal. MFCC, a representation of the short-term power spectrum of a sound is a classical approach to analyze speech signal whilst the DWT, a multiresolution approach mainly approximate the frequency information along with time information. Feature level fusion of algorithms has been performed after extraction of features by acoustic analysis of speech emotion signal. The final emotion state was determined by classification using Support Vector Machine. Popular Berlin emotion database is used for evaluation of the proposed system. The results achieved are very promising as the proposed fusion algorithm performed well compared to individual algorithms.

Keywords: Multi-algorithm Fusion, MFCC, DWT, Speech Emotion Recognition.

1 Introduction

Natural human-human interactions are made in two ways: verbal which includes speaking, singing and tone of voice, and Non-verbal which involves facial expression, body language, sign language, touch, eye contact etc. Emotion can be present in both types of communication i.e. verbal and non-verbal. Emotion helps to improve the level of interaction. Emotion is expressed in several ways: facial expressions, body gesture and speech are few of them. Recognizing emotion from human speech is an interesting and challenging problem for researchers working in the field of human computer interaction. Emotion recognition is one of the prime factors of Human-computer Interaction (HCI).

Fundamental frequency is an important voice feature for emotion recognition from speech signal [1] and can be easily extracted. Vocal expressions are conveyed by prosodic features, which include the fundamental frequency, intensity and rhythm of the voice [2]. Other important attributes that can be used for emotion recognition are pitch intensity and spectral measures. [3]. Some attributes are very proficient in particular emotion state, such as pitch is good to capture anger and fear whereas intensity is good for happiness. The success of the emotion recognition system depends upon the features incorporating different emotion states of the speaker.

We have already proposed [3] a novel approach for emotion recognition based on relative amplitude of speech signal. Many researchers have used wavelet based approach [4, 5] for emotion recognition [6, 7] used MFCC features for emotion recognition. Liqin Fu et. al. [8] utilized temporal features i.e. pitch, amplitude energy, energy frequency value, and zero cross ratio and formant frequency.

We propose a multi-algorithm approach for emotion recognition from speech signal. The MFCC and Discrete Wavelet Transform based algorithms have been successfully used to extract emotional information from speech signal. MFCC is a classical approach to analyze speech signal. It represents the short-term power spectrum of a sound, based on a linear cosine transform of a log power spectrum on a nonlinear Mel scale of frequency [9]. In the other approach approximation and detail coefficients were calculated by decomposing the input speech signal using DWT. The wavelet features for each input speech signal are obtained from 4th and 6th level decomposition using db4 wavelets. The similarity between the extracted features and a set of reference features is calculated by K-NN and SVM classifiers. We have used Berlin emotion database to evaluate the proposed system. The results obtained from fusion are better as compared to the separate performance reported in the literature [4, 5, 6, 7].

The rest of the paper is organized as follows: MFCC and wavelet transform algorithms are described in section 2. Multi-algorithm fusion is explained in section 3. Finally the experiment and the result analysis are given in section 5 and concluding remarks are given in section 6.

2 Our Approach

We have been used wavelet transform and Mel Frequency Cepstral Coefficient (MFCC) algorithms to recognize emotion from speech signal. The emotion recognition system based on multi-algorithm fusion is given in figure 1.

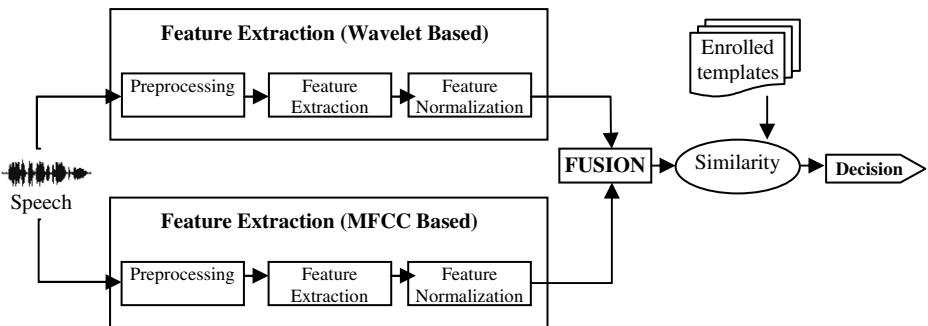


Fig. 1. Multi-algorithm fusion emotion recognition system

2.1 Wavelet Transform

Wavelet transform provides a compact representation that depicts the energy distribution of the signal in time and frequency domain [10]. We have used discrete wavelet transform to decomposes the signal into multilevel successive frequency bands utilizing two sets of function called scaling function ϕ and wavelet functions (ψ) associated with low pass and high pass filters respectively [11]. Information captured by wavelet transform depends on properties of wavelet function family like Daubechis, Symlet, Biorthogonal, Coiflet etc. and properties (waveform) of the target signal. Information extracted by wavelet transforms using different family of wavelet function need not be same. It is required to choose or evaluate the wavelet function that provides more useful information for particular application.

In Discrete wavelet decomposition of signal, the output of high pass filter and low pass filter can be represented mathematically by equation 1 and 2.

$$Y_{high} [k] = \sum X [n]g[2k - 1] \tag{1}$$

$$Y_{low} [k] = \sum X [n]h[2k - 1] \tag{2}$$

Where Y_{high} and Y_{low} are the outputs of the high band pass and low band pass filters respectively.

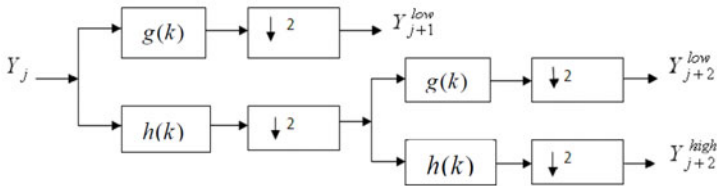


Fig. 2. Schematic of Discrete Wavelet decomposition of a speech signal

2.2 Mel Frequency Cepstral Coefficients (MFCC) Algorithm

MFCC's are based on the known variation of the human ear's critical bandwidths with frequency. The technique is based on two types of filters, namely linearly spaced filter and logarithmically spaced filters. The phonetically important characteristics of speech can be captured by representing the signal at the Mel frequency scale. This scale has a linear frequency spacing below 1000 Hz and a logarithmic spacing above 1000 Hz. Normal speech waveform may vary from time to time depending on the physical condition of speakers' vocal cord. MFCCs are less susceptible to these variations [12].

The Algorithm

The process of calculating MFCC consists of the following steps:

- Framing: In this step the speech signal segmented into N samples with 25% overlapping frames.
- Windowing: this step is applied for spectral analysis of the speech signal. We have used Hamming window, given in equation 3.

$$W(n) = 0.54 - 0.46 \times \cos\left(2 \times \frac{n}{N-1}\right) \quad (3)$$

Where $0 \leq n \leq N-1$

- Fast Fourier Transform (FFT): FFT was applied on each frame to obtain the spectral information from the time domain signal.
- Mel-frequency: For a given frequency, the Mel scale is being calculated by equation 4.

$$Mel(f) = 2595 \times \log_{10}\left(1 + \frac{f}{700}\right) \quad (4)$$

- Cepstrum: Finally the discrete cosine transform (DCT) was applied to the signal in order to obtain MFCC coefficients.

2.3 Fusion of Algorithms

Information fusion from different streams is a major challenge, specially an adaptive and optimal fusion of information. Information fusion can be at three levels: (i) signal level (ii) feature level (iii) decision level. In this study we performed feature level information fusion of both algorithms. The fusion of algorithms is illustrated in fig 3.

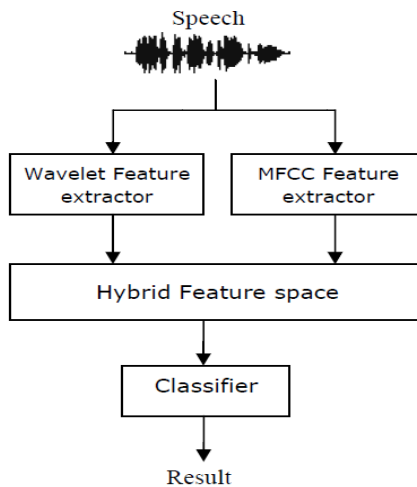


Fig. 3. Feature level fusion of algorithms

We get two feature vectors $F_{mfcc} = (f_1, f_2, \dots, f_m)$ and $F_{wav} = (f_1, f_2, \dots, f_n)$ by applying mfcc algorithm and wavelet transform respectively. The feature vectors having vector length n and m . Let $Fi = (f_1, f_2, \dots, f_{m+n})$ be the fused feature vector of feature vectors F_{mfcc} and F_{wav} . Fi is obtained by augmenting the normalized feature vectors F_{mfcc} and F_{wav} , then performing the feature selection on the concatenated vector formed.

3 Feature Extraction

3.1 Feature Extraction from Wavelet Transform

The speech signal is passed into successive high pass and low pass filter in order to extract wavelet coefficients. Selection of suitable wavelet coefficients and the number of levels of decomposition is important. Daubechis wavelet family provides good results for non-stationary one dimensional signal analysis [13]. Therefore we have used Daubechis wavelet in this study. The feature vectors obtained at six level wavelet decomposition provides compact representation of the signal. The coefficients occur in the whole bandwidth from low to high. The original signal can be represented by the sum of coefficients in every sub band, which is cD6, cD5, cD4, cD3, cD2, cD1. Feature vectors are obtained from the detailed coefficients applying common statistics i.e. standard deviation, mean etc.

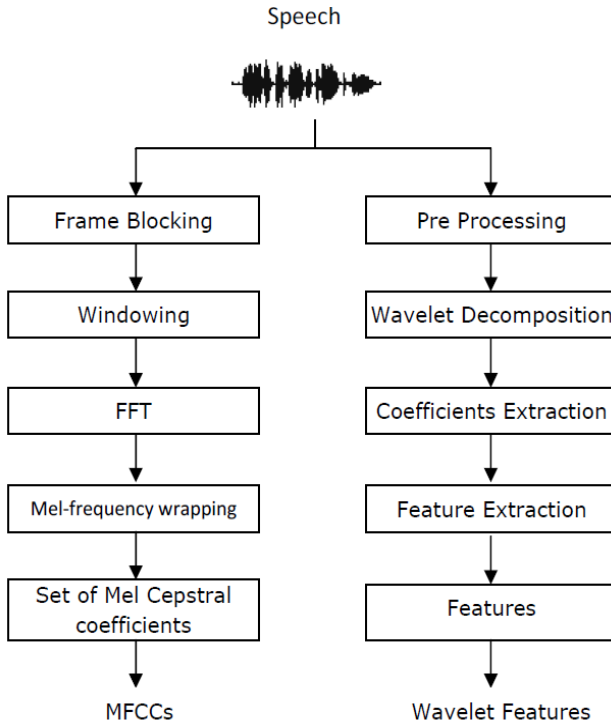


Fig. 4. Feature extraction process using MFCC and DWT

3.2 Feature Extraction from MFCC

For the short period of time the characteristics of the speech signal are fairly stationary, therefore the short-time spectral analysis is the most common way to characterize the speech signal [14]. The input emotion speech signal is segmented into a number of frames. Windowing operation is performed to capture the static property of the signal. Hamming window with 20 ms size and 25% overlapping has been used here. Then Fast Fourier transform is applied to produce the spectral characteristic of the speech signal. For the given frequency the mel frequency was calculated by equation 4. Finally the log Mel spectrum is converted back to time domain in order to get mel frequency cepstral coefficients (MFCC).

4 Feature Matching

K-NN algorithm has been used for feature matching between feature set (reference features) and query feature. The training set s contains l points $\{f_1, \dots, f_l\}$, $f_{xl} \in R^n$ and their corresponding class labels $\{y_1, \dots, y_N\}$, $y_i \in c$, $c = \{1, \dots, N_c\}$ where N_c is the number of different classes.

The distance between two vectors X and Y having length n can be calculated by Euclidean distance, Manhattan distance etc.

Euclidean distance is defined as follow:

$$d(X, Y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (5)$$

5 Experiments and Results

The Berlin emotional dataset, recorded by speech workgroup headed by Prof. Dr. W. Sendlmeieris was used for experimental purpose [15]. The database accommodates spoken sentences of 10 actors (5 males and 5 females) in seven states of emotional references: Anger, Boredom, Disgust, Fear, Joy, Sad and Neutral. The actors have spoken 10 predefined sentences with each emotion. They are sampled at 16 KHz sampling frequency. We have rearranged the database by renaming the folder of each emotion state from one to seven for seven emotion classes. Then we also renamed the emotion files from one to thirty. All the experiments were performed on MatLab 7.0 platform.

The steps involved are Pre-processing, segmentation, coefficient extractions, feature vector generation, features normalization and classification. The feature vectors obtained from multi-algorithms has been normalized using min-max normalization for classification purpose. The normalization was applied into both feature vectors obtained from MFCC and wavelet transform. We have seven emotion classes and each class contains thirty emotion files. Twenty five emotion files were used for training and rest files were used for testing. The experiments were also

performed individually by taking wavelet and mfcc features. A feature level fusion was also performed as discussed in section 3. Each emotion samples of same emotion state was assigned same class i.e. twenty-five emotion samples of same emotion states were assigned the same class. In this way the whole dataset is grouped into seven classes for seven emotion states. Euclidean distance was used to calculate the distances among feature vectors.

The wavelet coefficients were formed by 6th level wavelet decomposition of emotion samples. General statistics was applied on wavelet coefficients in order to form feature vectors. Feature vectors are generated by applying common statistics described in section 3. The performance results of different algorithms are shown in table1 and corresponding graph is illustrated in figure 5.

Table 1. Classification Results

Sl.	Emotion	Wavelet	MFCC	Fusion
1	Anger Vs Boredom	0.10	0.15	0.10
2	Boredom Vs Disgust	0.30	0.25	0.20
3	Disgust Vs Fear	0.25	0.35	0.25
4	Fear Vs Happy	0.20	0.25	0.15
5	Happy Vs Neutral	0.25	0.34	0.20
6	Neutral Vs Sad	0.30	0.25	0.20

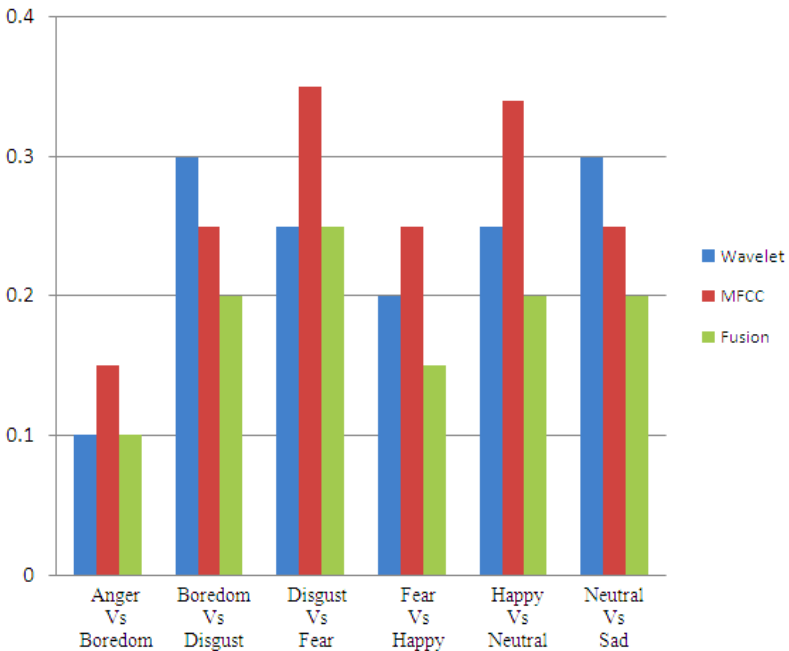


Fig. 5. Performance graph of different algorithms with fusion

6 Conclusion

In this study, we have proposed speech emotion recognition system based on MFCC and Wavelet Transform. Features were extracted using the proposed algorithm and evaluated using Berlin emotion speech database. A Comparative study is also performed here. The experimental results are very promising. A feature level fusion of algorithms was performed in this study.

References

1. Cohn, J.F., Katz, G.S.: Bimodal expressions of emotion by face and voice. In: Workshop on Face/Gesture Recognition and their Applications, the Sixth ACM International Multimedia Conference, Bristol, England (1998)
2. Fasel, B., Luetttin, J.: Automatic facial expression analysis: A survey. *Pattern Recognition* 36, 259–275 (2003)
3. Kudiri, K.M., Verma, G.K., Gohel, B.: Relative Amplitude based Features for Emotion Detection from Speech. In: 3rd IEEE Int. Conf. on Signal and Image Processing, pp. 301–304 (2010)
4. Rizon, M.: Discrete Wavelet Transform Based Classification of Human Emotions Using Electroencephalogram Signals. *American Journal of Applied Sciences* 7(7), 865–872 (2010)
5. Shah, F., et al.: Discrete Wavelet Transforms and Artificial Neural Networks for Speech Emotion Recognition. *International Journal of Computer Theory and Engineering* 2(3), 1793–8201 (2010)
6. Kwon, O.-W.: Emotion Recognition by Speech Signals. In: EUROSPEECH-2003, Geneva (2003)
7. Mao, X.: Speech Emotion Recognition based on a Hybrid of HMM/ANN. In: Proceedings of the 7th WSEAS International Conference on Applied Informatics and Communications, Athens, Greece, August 24-26 (2007)
8. Liqin, F., et al.: Relative Speech Emotion Recognition Based Artificial Neural Network. In: IEEE Pacific-Asia Workshop on Computational Intelligence and Industrial Application (2008)
9. http://en.wikipedia.org/wiki/Mel-frequency_cepstrum
10. Dutta, T.: Dynamic Time Warping Based Approach to Text Dependent Speaker Identification Using Spectrograms. In: Congress on Image and Signal Processing, vol. 2, pp. 354–360 (2008)
11. Tzanetakis, G., Essl, G., Cook, P.: Audio Analysis using the Discrete Wavelet Transform. In: Proc. Conf. in Acoustics and Music Theory Applications, Skiathos, Greece (2001)
12. Lindasalwa, M., Begam, M., Elamvazuthi, I.: Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. *Jour. of Computing* 2(3), 138–143 (2010)
13. Toh, A.M., Togneri, R., Northolt, S.: Spectral entropy as speech features for speech recognition. In: The Proceedings of PEECS, Perth, pp. 22–25 (2005)
14. Kan, P.L.E., Allen, T., Quigley, F.: A GMM-Based Speaker Identification System on FPGA. In: 6th International Symposium on Reconfigurable Computing: Architectures, Tools and Applications. LNCS. Bangkok, Thailand (March 2010)
15. Burkhardt, F., Paeschke, A.: A database of German emotional speech. In: Interspeech, Lisbon, Portugal, pp. 1517–1520 (2005)

Combining Chain-Code and Fourier Descriptors for Fingerprint Matching

C.Z. Geevar and P. Sojan Lal

Department of Computer Application, MES College of Engineering
Kuttippuram, India

School of Computer Science, Mahatma Gandhi University
Kottayam, India

{geevarcz,padikkakudy}@gmail.com

Abstract. The performance of an fingerprint recognition system is measured by its accuracy in recognition. For a feature-based fingerprint recognition system, the accuracy is heavily depend on the chosen feature set. A fingerprint image may suffer from problems like translation, rotation, scaling and elastic distortion due to different imaging conditions. A fingerprint recognition algorithm should address these problems before building the feature set. We present a novel method of representing the fingerprint ridge shape as the feature set by combining chain code and fourier descriptor for fingerprint recognition. Experimental results shows that our proposed algorithm is reliable for fingerprint recognition.

Keywords: fingerprint matching, chain code, fourier descriptors.

1 Introduction

Fingerprint is one of the reliable and widely accepted biometric systems for automatic personal identification. Also, due to their uniqueness and unchangeability, it is a trusted biometric system for many security purposes. An automatic fingerprint identification/verification system either verifies an individual for a one-to-one template match or identifies an individual for a one-to-many template match. The performance of an fingerprint recognition system can be measured using its matching accuracy and matching speed and these two criterions heavily depends mainly on methods of feature extraction and matching algorithm used.

A fingerprint matching algorithm compares two fingerprints, the one which stored in the database and the other which is queried, and returns either a similarity score (score between 0 and 1) or a binary decision (matched/non-matched). One of the most popular method is to utilize the feature set extracted from the fingerprint image as an input for the fingerprint matching. Fingerprint feature extraction and matching algorithm are quite similar for both fingerprint verification and identification problem [1]. Due to different imaging conditions, feature extraction from a fingerprint image may suffer from problems like translation, rotation, scaling and elastic distortion and thus also same for fingerprint

matching. Fingerprint matching can be broadly classified into minutiae-based, correlation-based and feature-based.

The most popular and widely used methods are minutiae-based [6,8]. These use a feature vector extracted from fingerprints as sets of points in a multi-dimensional plane. Minutiae-based matching methods consider minutiae points, special points of fingerprint impressions representing ends and bifurcations of the fingerprint ridge structure, that has maximum number of pairing.

Minutiae-based fingerprint recognition often fails because minutiae extraction is extremely difficult in low-quality fingerprint images. The feature-based recognition methods use features such as local orientation and frequency, ridge shape, and texture information for its matching [11]. The features for these types may be extracted more reliably than those of minutiae.

In correlation-based matching two fingerprint images are superimposed and the scores from the intensities of the corresponding pixels are computed [11].

In an automatic fingerprint recognition system, feature-based matching would be better because the deterioration is gradual in a noisy image. This paper proposes a novel method of representing fingerprint ridge shape as the feature set by combining fourier descriptor with freeman chain code [3] for fingerprint recognition. Chain code is extensively used in [9,10,15,4] representing the contour of an object. The advantage of using chain code to represent the shape of an object is that, it gives a lossless representation of the contour. But chain code processing is computationally costly and does not has much discriminating power for a match. On the other hand, fourier descriptor method can overcome the effect of noise and boundary variations on shape feature extraction by analyzing shape in spectral domain. In addition, shape representations based on fourier descriptor method are compact and computationally light. Furthermore, they are easy to normalize and their matching is very simple process [7].

2 Shape Representation

In general, a shape representation is any 1-D function representing 2-D area or boundaries. The following sections represent two popular methods for representing the shape.

2.1 Chain Code

Chain codes describe an object by a connected sequence of straight-line segments of length and direction [9]. This representation is based on 4-connectivity or 8-connectivity of the segments [5]. The direction of each segment is coded by using a numbering scheme as shown in (Fig. 1) below. Chain code representation of object contour is extensively used in document analysis [10].

The contour can be represented using 8-connectivity chain encoding [3] method. When 8-way connectivity is used, we will get 8 different representation of directions, where each direction varying in 45° angle from the previous one. A chain code can be generated by moving along a boundary of an object in

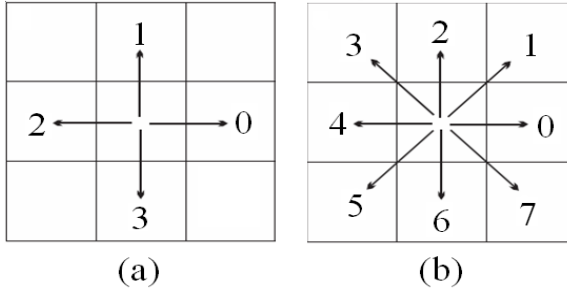


Fig. 1. Chain code Direction: (a) chain code in four directions (4-connectivity), (b) chain code in eight directions (8-connectivity)

an anti-clockwise direction and assigning a direction to the segments connecting each pair of pixels. This direction along the object boundary varies from 0,1,2,....,7 in anti-clockwise direction. It is compact and translation invariant in representing an object, however, it depends on rotation and scaling transformations. Processing the resulting chain code, also, poses difficulties, since they tend to be quite long.

3 Fourier Descriptors

The Discrete Fourier Transformation (DFT) is generally used to describe the shape features from the image boundary and so it is a popular method for shape analysis [13,16]. The fourier transformation utilizes the coefficients form the fourier descriptors of the object shape. The advantage of using fourier descriptors is that, it represent the object shape in frequency domain. The lower frequency of the fourier descriptor contain information about the global features of the shape, and the higher frequency contain information about local details of the shape. The high frequency information that specifies the local details of the shape does not contain the discriminating power for shape classification since it contains noises, and therefore, they can be ignored. Since only a subset of the coefficients generated from the transform would suffice in capturing the overall features of the shape, the dimensions of the fourier descriptors get significantly reduced. The Discrete Fourier Transformation method is explained below.

Consider that the boundary of a particular object shape has K pixel points numbered from 0 to K-1. Starting from an arbitrary point the coordinate pairs of these pixel points can be $(x_0, y_0), (x_1, y_1), \dots, (x_{k-1}, y_{k-1})$ by traversing the boundary in some order. Therefore, we can describe the boundary as:

$$\begin{aligned} x(k) &= x_k \\ y(k) &= y_k \end{aligned}$$

Here the coordinate pair of the point in an image plane can be converted as a complex number by writing

$$z(k) = x(k) + iy(k)$$

The complex-valued boundary function using the Discrete Fourier Transform (DFT) to form the descriptors [5]. The discrete fourier transform of $z(k)$ is

$$a_k = \frac{1}{k} \sum_{k=0}^{K-1} z(k)e^{-j2\pi uk/K}, \quad u = 0, 1, \dots, K - 1$$

The complex coefficients a_k are called the *Fourier Descriptors* of the boundary.

4 Proposed Method

In the proposed method of fingerprint recognition, we combine two shape signatures, Freeman chain code [3] and Fourier descriptor.

4.1 Pre-processing

The first step in preprocessing is to binarize the fingerprint to make the image pixel values to range between 0 and 1. We used Otsu's algorithm [12] for the purpose of selecting the threshold and binarizing the gray scale images, so that the binarized image has 0 (black) as background pixels and 1 (white) as foreground pixels. The noises in the image is removed using morphological open and close operations. Then we performe labeling using connected component labeling method. Connected Component Labeling scans a fingerprint image and projects out fingerprint ridges based on their pixel connectivity.

4.2 Chain Code Processing

Fingerprint ridge contour is chain encoded by using 8-way chain encoding [3] method. We have chosen contour instead of a thinned image as thinning introduces some loss of information and takes high processing time. Our chain code extraction method is obtained by the following method:

- Step 1:** Scanning the image from top to bottom and right to left
- Step 2:** Detect the transition from black (background) to white (foreground)
- Step 3:** Trace the contour anti-clockwise by analyzing the pixel label
- Step 4:** Mark the chain code number and express it as an array elements
- Step 5:** Repeat the Step 1 to 4 for every fingerprint ridges

But before extracting the contour using chain code, image has to be normalized to fixed sample of points. This is important because the length of the chain code of an object shape tends to be too long. In addition, shape normalization is also important due to the fact that stored object shape and querying object shape can have different size. The sampling process not only eliminates the noises along the shape boundary but also corrects the small details along the shape boundary as well. In order to facilitate the use of the fast fourier transform (FFT), the number of sampled points is chosen to be power-of-two integer. There are generally three methods of normalization (i) equal points sampling (ii) equal arc sampling and (iii) equal arc length sampling. Among the above three sampling methods, we have used the equal arelength sampling method mentioned in [17].

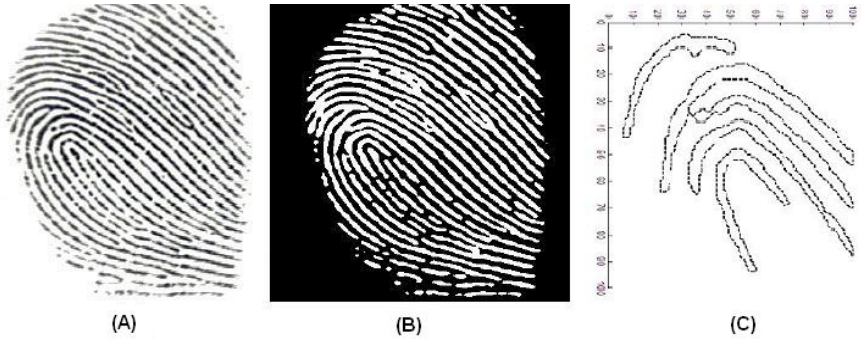


Fig. 2. (a) Original Image (b) Image after Binarization and Labeling (c) Plot of direction of chain encoded ridge contour(For viewability only part of the ridge is plotted)

4.3 Fourier Descriptor

For a given shape described above, $z(k), k = 0, 1, \dots, L$ assuming it is normalized to N points in the sampling stage, the discrete fourier transform of $z(k)$ is given by

$$z_n = \frac{1}{N} \sum_{i=0}^{N-1} z(k)e^{-j2\pi nt/N}, \quad n = 0, 1, \dots, N - 1 \tag{1}$$

The coefficients $z_n, n = 0, 1, \dots, N - 1$, are usually called Fourier Descriptor (FD) of the shape denoted as $FD_n, n = 0, 1, \dots, N - 1$

Our method of computing fourier descriptor follows the following steps:

- Step 1:** Represent the chain code numbers of the fingerprint ridges in the complex plane using the lookup table (Table I) as described in [14]
- Step 2:** Apply the fourier transform to these set of numbers using equation(1)
- Step 3:** Obtain the fourier descriptor coefficients
- Step 4:** Repeat Step 1 to 3 for every array elements

Table 1. Lookup Table [14]

Chain code	0	1	2	3	4	5	6	7
Complex number	1	1+i	i	-1+i	-1	-1-i	-i	1-i

4.4 Feature Extraction

Feature extraction is a critical phase in any automatic fingerprint matching system. The performance of a matching algorithm is directly related with how good the features extracted from the image. As in the case of fingerprint image, its the global shape feature determines the different fingerprint classes. A valid feature set extracted should be able to capture this global information effectively. Due

to different imaging conditions, a fingerprint image may suffer from problems like translation, rotation, scaling and elastic distortion. In order to make a valid match between stored object shape and queried object shapes, the feature set extracted from the object shape must be invariant to the above transformation. Shape invariance is difficult to achieve under spatial domain. Most invariance techniques in spatial domain, especially rotation invariance techniques, involve large amount of computation. However, shape invariance is easy to achieve in frequency domain using FDs.

The FD's are invariant under translation by its nature. Rotation invariance of the FD's are achieved by ignoring the phase information and by taking only the magnitude values of the FD's. Scale invariance is achieved by dividing the magnitude values of all the other descriptors by the magnitude value of the second descriptor [17].

For the object shape representations using complex coordinates, all the descriptors are needed to describe the shape. The DC component depends only on the contour position of the shape, it is not useful in describing shape and thus is ignored [7,17].

The invariant feature vector used to describe the shape [17] is then given by

$$f = [\frac{|FD_2|}{|FD_1|}, \frac{|FD_3|}{|FD_1|}, \dots, \frac{|FD_{N-1}|}{|FD_1|}]$$

4.5 Matching Similarity Measures

In this proposed method Euclidian distance in the equation(2) is used as the measure of similarity for matching the input image. During the matching process the feature extracted are compared against the stored feature data set. The difference score between the queried feature set (f_q) and stored feature set (f_s) will be calculated. Features of the image having less score are more similar.

$$score = \left(\sum_{i=1}^N |f_q^i - f_s^i|^2 \right)^{\frac{1}{2}} \tag{2}$$

5 Test Results

The proposed reference point estimation method has been tested with a sample set of fingerprint images taken form Neurotechnologija web site [2] named Verifinger Match DB. From the database we have taken 400 fingerprint images (8 fingerprint images of 50 persons)

To evaluate the performance of the proposed matching method, False Acceptance Ratio(FAR) and False Rejection Ratio(FRR) are used. A Receiver Operating Characteristic (ROC) curve is used to plot the Genuine Acceptance Rate (1-FRR) against False Acceptance Rate.

To compute the False Acceptance Rate (FAR)

$$FAR = \frac{\text{Number of accepted imposter claims}}{\text{Total number of imposter acceses}} * 100 \tag{3}$$

and the False Rejection Rate (FRR)

$$FAR = \frac{\text{Number of rejected genuine claims}}{\text{Total number of genuine accesses}} * 100 \tag{4}$$

In ideal case these two values would be zero. But in real scenario, in ROC curve these two values will meet at an operating point. An Equal Error Rate (ERR) is the point where these two values meet in an ROC curve.

To compute FAR and FRR, the genuine match score and imposter match score need to be calculated. To compute genuine match score for the entire database, one fingerprint sample of each person is matched against other fingerprint samples of the same person. To compute imposter match, one fingerprint sample is matched against the remaining sample of the other persons.

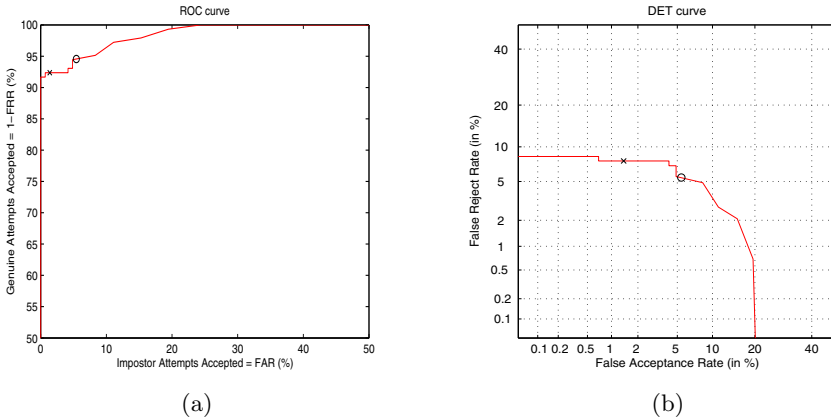


Fig. 3. Performance Measure Plots (a) ROC curve from similarity measures with the Neurotechnologija Verifinger Match DB database (b) DET curve. EER is shown with a label 'O'

The results has been plotted using Biosecure Tool, a program for performance evaluation of a biometric verification system [1]. Fig 3 shows the performance measure plots on the Neurotechnologija Verifinger Match DB database.

Table 2. Confusion matrix of the fingerprint image matching

	R	L	W	A	T
R	30	0	0	0	0
L	0	26	3	0	0
W	0	3	21	0	0
A	0	0	0	6	0
T	0	0	0	1	15

For a classification problem, we manually classified the fingerprint images of the database into five different class such as Whorl, Arch, Tented Arch, Right Loop, and Left Loop. Out of 105 sample test images, the matching algorithm produced a classification accuracy of 93.3%. The confusion matrix is shown in the (Table 2). By classifying into five different classes, no fingerprint images were rejected during the classification stage. More misclassification occurred in the Whorl class due to its small interclass variation.

6 Conclusion

The performance of an fingerprint recognition system can be measured using its matching accuracy and matching speed and these two criterions heavily depends mainly on methods of feature extraction and matching algorithm used. We present a novel method of representing the fingerprint ridge shape as the feature set by combining chain code and fourier descriptor for fingerprint recognition. Fingerprint ridge shape is a good feature set for fingerprint matching because fingerprint is formed by ridges and valleys. By using fourier descriptor a translation invariant feature set can be extracted from the chain encoded fingerprint image. The experiment results shows that this proposed method can be used reliably for fingerprint matching.

References

1. Biosecure Tool:Performance Evaluation of a Biometric Verification System, http://svnext.it-sudparis.eu/svnview2-eph/ref_syst/Tools/PerformanceEvaluation/
2. Neurotechnology, <http://www.neurotechnologija.com/download.html>
3. Freeman, H., Davis, L.S.: A corner finding algorithm for chain coded curves. *IEEE Transaction in Computing* 26, 297–303 (1977)
4. Geevar, C.Z., Sojan Lal, P.: Reference point estimation in fingerprint image. In: *IEEE International Conference on Computational Intelligence and Computing Research* (2010)
5. Gonzale, R.C., Wintz, P.: *Digital Image Processing*. Addison-Wesley, Reading (1987)
6. Jain, A.K., Prabhakara, S., Hong, L., Pankanti, S.: An identity-authentication system using fingerprints. *IEEE Trans. Systems Man Cybernet.* 85(9), 1365–1388 (1997)
7. Kauppinen, H., Seppanen, T., Pietikainen, M.: An experimental comparison of autoregressive and fourier based descriptors in 2d shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(2), 201–207 (1995)
8. Liu, J., Huang, Z., Chan, K.: Direct minutiae extraction from gray-level fingerprint image by relationship examination. In: *International Conference on Image Processing*, pp. 427–430 (2000)
9. Liu, Y.K., Wei, W., Wang, P.J., Zalik, B.: Compressed vertex chain codes. *Pattern Recognition* 40(11), 2908–2913 (2007)
10. Madhvanath, S., Kim, G., Govindaraju, V.: Chain code processing for hand-written word recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21, 928–932 (1997)

11. Maltoni, D., Maio, D., Jain, A.K., Prabhakar, S.: Handbook of Fingerprint Recognition. Springer, Heidelberg (1991)
12. Otsu, N.: A threshold selection method from gray-level histogram. *IEEE Trans. Systems Man Cybernet.* 9, 62–66 (1979)
13. Persoon, E., Fu, K.: Shape discrimination using fourier descriptors. *IEEE Trans. On Systems, Man and Cybernetics SMC-7(3/7)*, 170–179 (1977)
14. Rajput, G.G., Horakeria, R., Chandrakant, S.: Printed and handwritten mixed kanada numerals recognition using svm. *International Journal on Computer Science and Engineering* 2(5), 1622–1626 (2010)
15. Siddiqi, I., Vincent, N.: A set of chain code based features for writer recognition. In: *Proceedings of the 10th International Conference on Document Analysis and Recognition* (2009)
16. Smach, F., Lematre, C., Gauthier, J.-P., Miteran, J., Atri, M.: Generalized fourier descriptors with applications to objects recognition in svm context. *Journal of Mathematical Vision and Imaging* 30(1), 43–71 (2006)
17. Zhang, D., Lu, G.: A comparative study of fourier descriptors for shape representation and retrieval. In: *The 5th Asian Conference on Computer Vision*, pp. 23–25 (2002)

Facial Emotion Recognition Using Different Multi-resolution Transforms

Gyanendra K. Verma¹, U.S. Tiwary¹, and Mahendra K. Rai²

¹ Indian Institute of Informaiton Technology, Allahabad
Allahabad, India – 211012

² Gyanganga Institute of Technology and Science, Jabalpur
Jabalpur, India – 482003

{gyanendra, ust}@iiita.ac.in,
mahendrakumar.raai@gmail.com

Abstract. The present work investigates the performance of different multi-resolution transforms in the application of emotion recognition from facial images. Multi-resolution analysis of image provides frequency information along with time information in different scale, orientation and locations. The emotion information from facial images was being captured by different multiresolution algorithm such as Wavelet Transform, Curvelet Transform and Contourlet Transform. Wavelet transform mainly approximate frequency information along with time whereas curvelet transform is best to capture edges information with very few coefficients. Various statistical features obtained from different algorithms have been used to build reference model. The classification part was done using support vector machine (SVM) and K-Nearest Neighbor (KNN) classifier with JAFFE, a Japanese facial emotion database. The individual as well as comparative study of different algorithms was done successfully.

Keywords: Multi-resolution transforms, Emotion Recognition, Curvelet Transform, Wavelet Transform, Contourlet Transform.

1 Introduction

In Human Computer Interaction (HCI) domain, nowadays researchers focus on to make a better interactive system that should be capable to perceive, interpret and express emotions as like human being. From the machine perspective, recognizing the user's emotional state is one of the main requirements for computers to successfully interact with humans [1]. Multi-media and man-machine communication systems could promote more efficient performance if machine understanding of emotion is improved [2].

Natural human interaction occurs through either two ways: verbal which involving speaking, tone of voice and Non-verbal involves facial expression, sign language etc. Emotion can be present in both type of communication i.e. verbal and non-verbal. It helps to better human-human interactions. Emotion can be expressed by several ways; facial expressions, body gesture and speech are few of them. Recognizing human facial expression and emotion by computer is an interesting and challenging problem

before today researchers emphasis on better human computer interaction. Emotion recognition is one of the prime factor in the Human-Computer Interaction.

Multiresolution methods provide frequency information in different scale, orientation and locations. The multiresolution transforms can be successfully used to analyze the image. In this study we used three different multiresolution transforms namely Wavelet, Curvelet and Contourlet Transform. Wavelet transform well approximate the frequency information along with time however some properties like directionality and anisotropy are not supported by wavelet in image processing. So we have to explore new transform like curvelet and contourlet for image representation. The main characteristic of curvelet transform is that edge information can be better approximated by curvelet and few curvelet coefficients are required to capture edge information. Most facial expression recognition systems are of the unimodal type, as they focus only on facial expressions when determining mental activities. However, the evaluations of multiple communication channels may robustness as well as improve correct interpretation of facial expressions in ambiguous situations. At present, most attempts of channel fusion are of the bimodal type and integrate voice in addition to facial expressions [3].

We propose a framework for human emotion recognition using different multiresolution transforms. The Curvelet Transform has problem of redundancy. Contourlet has all the desirable properties of curvelets including directionality and anisotropy and at the same time reduces the redundancy of the curvelets [4]. The similarity between the extracted features and a set of reference features is calculated by means of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) classifier. JAFFE database has been used in experiments.

The organization of the paper is as follows: Related works are discussed about in section 2. Section 3 described different multiresolution algorithms i.e. wavelet, curvelet and contourlet algorithms. Experimental setup is explained in section 4 and concluding remarks are given in section 5.

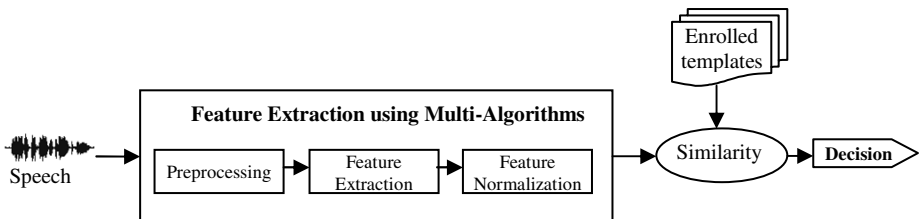


Fig. 1. Typical model of Emotion Recognition System

2 Related Work

Multi-resolution approaches are relatively new for emotion recognition, hence few research work has been reported in this area. Wavelet transform based study is discussed in [5] and [6]. Many researcher emphasis that the emotion recognition system should be multimodal as single modality is not sufficient to recognize emotion completely. Muharram Mansoorizadeh [7] has proposed a multimodal information

fusion framework for human emotion recognition. Curvelet based emotion recognition work has been presented by [8, 9]. Chien-Cheng Lee et al. [10] used contourlet features and a regularized discriminate analysis for facial expression analysis.

3 Multi-resolution Image Analysis

Multi-resolution image analysis provides both spatial as well as frequency information of image. It yields good time resolution but poor frequency resolution at higher frequencies, on the other hand good frequency resolution and poor time resolution at lower frequencies [11]. Multi-resolution techniques approximate the frequency information along with time information, eventually useful in image analysis. In this study, we investigated different multi-resolution algorithms i.e. wavelet, curvelet and contourlet transform with their pros and cons.

3.1 Wavelet Transform

The two-dimensional wavelet can also be constructed from the tensor product of one dimensional ϕ and ψ by setting: [12]

$$\begin{aligned}\phi(x, y) &= \phi(x)\phi(y) , \quad \psi^H(x, y) = \phi(x)\psi(y) , \\ \psi^V(x, y) &= \psi(x)\phi(y) , \quad \psi^D(x, y) = \psi(x)\psi(y)\end{aligned}$$

where $\psi^H(x, y)$, $\psi^V(x, y)$, $\psi^D(x, y)$ are wavelet functions.

Wavelet transform analyzes the image at different scales or resolutions. It decomposes the image into subbands that are localized in both space and frequency domains. A wavelet coefficient is large only if the singularities are present in the support of a wavelet basis function. The magnitudes of coefficients tend to decay exponentially across scale. Maximal energy of images concentrates on these large coefficients. Wavelet coefficients of images tend approximately de-correlated because of the ortho-normal property of wavelet basis functions.

Wavelet transform makes it possible to be multi-resolution by providing the means to analyze an image signature at several levels of resolution. DWT decomposes a signal in multilevel successive frequency band utilizing two sets of functions called scaling function ϕ and wavelet functions (ψ) associated with low pass and high pass filters respectively [13]. Information captured by wavelet transform depends on properties of wavelet function family like Daubechies, symlet, biorthogonal, coiflet etc and features (waveform) of target signal.

3.2 Curvelet Transform

The Curvelet transform is a higher dimensional generalization of the Wavelet transform designed to represent images at different scales and different angles [14]. The benefit of using Curvelet transform is that curved singularities can be well approximated with very few coefficients and in a non-adaptive manner [14]. The limitation of wavelet transform is that it fails to represents objects containing

randomly oriented edges and curves as it is not good at representing line singularities [15]. The curvelet transform is able to catch the edge singularities efficiently. Fig.1 shows the curvelet construction.

There are two variants of discrete curvelet transform namely Unequispaced FFT transform and Wrapping transform. In unequispaced FFT, the curvelet coefficients are formed by irregularly sampling the Fourier coefficients of the image. In wrapping algorithm, curvelet coefficients are formed by using a series of translations and a wraparound technique. The wrapping based algorithm performs faster in computation and more robust as compare to USFFT however both algorithms give the same output. Therefore we have used wrapping based algorithm to find out curvelet coefficients. Figure 2 shows curvelets in frequency as well as spatial domain. The shaded area is one of the wedge. The wedges are formed by dividing the frequency plane into different partitions and the spatial plane is divided in respect to θ (angular division). The angular division partitions each subband image into different angles.

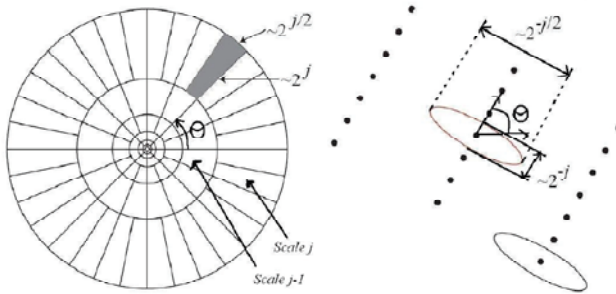


Fig. 2. Curvelets in frequency domain (left) and spatial domain (Right) [16]

The Algorithm (USFFTs)

Algorithm for finding the curvelet coefficients [22]:

STEP1: Take the Fourier transform of the signal

STEP2: Divide the frequency plain into polar wedges (as shown in fig.2)

STEP3: Find the curvelet coefficients at a particular scale (j) and angle (θ) by taking the inverse FFT of each wedge at scale j and oriented at angle θ .

Discrete curvelet coefficient can be given by the equation 1 [16]

$$C^D(j, l, k_1, k_2) = \sum_{\substack{0 \leq m < M \\ 0 \leq n < N}} f[m, n] \varphi_{j, l, k_1, k_2}^D[m, n]. \tag{1}$$

Here $\varphi_{j, l, k_1, k_2}^D[m, n]$ is digital curvelet transform. j is scale, l is orientation and (k1, k2) are location parameters.

3.3 Contourlet Transform

The Contourlet Transform is a directional transform, which is capable of capturing contours and fine details in images. The contourlet expansion is composed of basis function oriented at various directions in multiple scales, with flexible aspect ratios [17].

The contourlet transform is implemented by pyramidal directional filter band proposed by Minhdo and Vetterli [18]. It computes the multi-scale decomposition into octave bands where low pass channel is sub-sampled. A directional filter bank (DFB) was applied to each high pass channel. As DFB is designed to capture high frequency components (representing directionality), the LP part of the PDFB permits subband decomposition to avoid “leaking” of low frequencies into several directional subbands, thus directional information can be captured efficiently. Fig. 3 depicts the directional filter bank decomposition. for more reading please refer to [19].

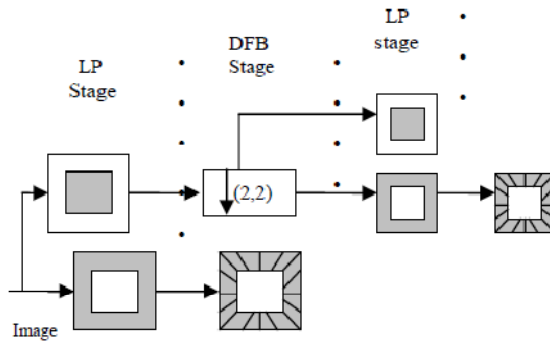


Fig. 3. Double filter bank decomposition [17]

4 Experiments and Results

4.1 Emotion Corpus

JAFFE database [20]: The experiments were performed on JAFFE database. The JAFFE database contains 213 images of 7 facial expressions (6 basic facial expressions and one neutral) posed by 10 Japanese female models. As shown in Fig. 5 (a), the images were taken against a homogeneous background with extreme expression variation, and the image size is of 256×256 .

4.2 Experimental Setup

The performance of the face recognition system depends upon the features extracted from images. The features should have high intraclass similarity and low interclass similarity. The experiments comprise two modules: training and testing. Different Transforms were applied to decompose each image in order to extract the coefficients

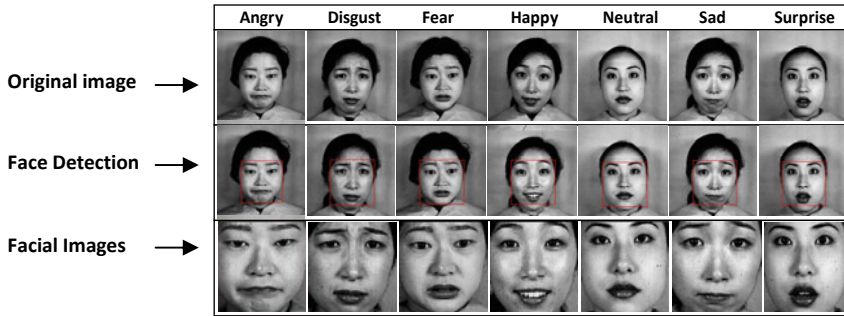


Fig. 4. Face detection and crop operation on JAFFE Database

as discussed in section 2. To extract the curvelet coefficients we applied fast discrete curvelet transform to images with 4th and 6th level decomposition using Curvelab-2.1.2 [21]. We used only basic software of Curvelab and further modification was done by us. Some common statistical methods applied on each coefficient obtained from different transforms at each subband level to get feature vector. The curvelet subbands at 4th level decomposition are {1x1 cell} {1x16 cell} {1x32 cell} {1x1 cell} and 6th level decomposition {1x1 cell} {1x16 cell} {1x32 cell} {1x32 cell} {1x64 cell} {1x1 cell}. Min-max algorithm has been used for feature set normalization in order to improve the recognition accuracy before the classification for large dataset. All the experiments are performed on Mat Lab 7.0

Table 1. Classification Results with SVM

Sl.	Emotion	Wavelet	Curvelet	Contourlet
1	Anger Vs Disgust	35	40	35
2	Disgust Vs Fear	25	60	50
3	Fear Vs Happy	50	45	40
4	Happy Vs Neutral	20	50	50
5	Neutral Vs Sad	55	45	40
6	Sad Vs Surprise	30	50	45

Table 2. Classification Results with KNN

Sl.	Emotion	Wavelet	Curvelet	Contourlet
1	Anger Vs Disgust	56.6	36.6	35.5
2	Disgust Vs Fear	53.3	73.3	70.2
3	Fear Vs Happy	53.3	46.6	40.5
4	Happy Vs Neutral	60	66.6	55.6
5	Neutral Vs Sad	50	50	45.2
6	Sad Vs Surprise	36.6	36.6	36.0

For classification purpose image of same emotion is assigned same class. In this way the whole database is grouped in class. Manhattans' distance is used to calculate the distances among feature vectors using KNN algorithm. We have also used SVM for classification purpose. The performance results are shown in table 1 and table 2.

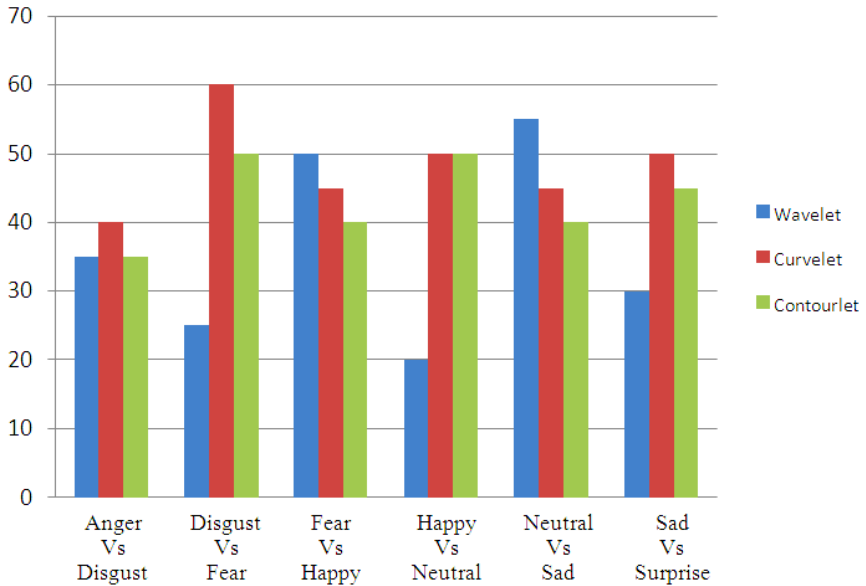


Fig. 6. Performance with SVM classifier

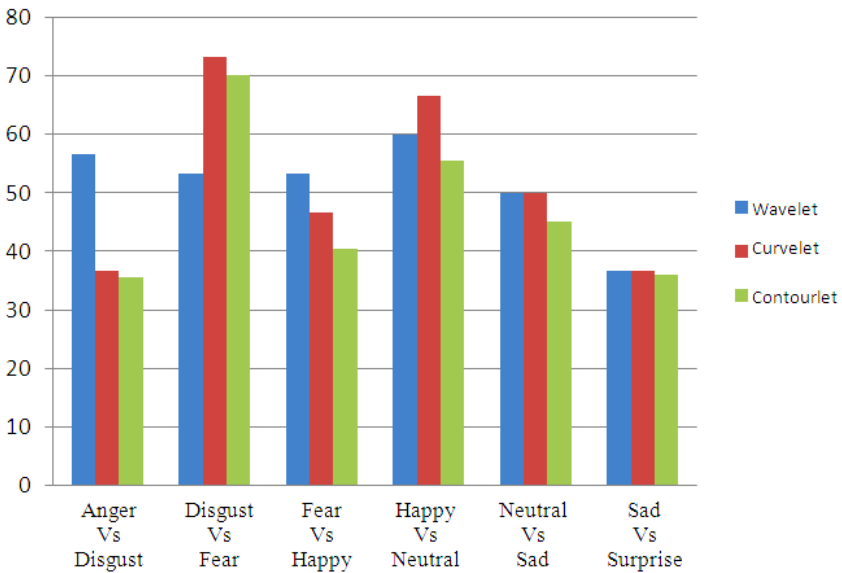


Fig. 7. Performance with KNN classifier

5 Conclusion

A multi-resolution based facial emotion recognition approach has been proposed in this paper. The wavelet, curvelet and contourlet transform was used to extract the coefficients from input images. KNN and SVM classifier have been used for pattern matching. The experiments were performed on JAFFE database. The experimental result shows comparative performance among different multi-resolution algorithms.

References

1. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in human computer interaction. *IEEE Signal Process. Magazine* 20, 569–571 (2001)
2. Galateia, I.: Emotional Facial Expressions recognition & classification. MS thesis, Delft University of Technology, Delft, Netherland
3. Fasel, B., Luetin, J.: Automatic facial expression analysis: A survey. *Pattern Recognition* 36, 259–275 (2003)
4. Delac, K., Grqic, M., Bartlett, M.S.: Recent advances in face recognition. In-Tech Publication, Crosia (2008)
5. Chuang, Y., Yuning, H., Zhao, K.: The Method of Human Facial Expression Recognition Based on Wavelet Transformation Reducing the Dimension and Improved Fisher Discrimination. In: 3rd International Conference on Intelligent Networks and Intelligent Systems (ICINIS), pp. 43–47 (2010)
6. Zhi, R., Ruan, Q.: Robust Facial Expression Recognition Using Selected Wavelet Moment Invariants. In: WRI Global Congress on Intelligent Systems, GCIS 2009, pp. 508–512 (2009)
7. Muharram, M., Charkari, Moghaddam, N.: Multimodal information fusion application to human emotion recognition from face and speech. In: *Multimedia Tools and Applications*. LNCS, vol. 49(2), pp. 277–297. Springer, Heidelberg (1977)
8. Saha, A., Jonathan, Q.M.: Facial Expression Recognition using Curvelet based local binary patterns. In: *IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp. 2470–2473 (2010)
9. Wu, X., Zhao, J.: Curvelet feature extraction for face recognition and facial expression recognition. In: *Sixth International Conference on Natural Computation (ICNC)*, pp. 1212–1216 (2010)
10. Lee, C.-C., Shih, C.-Y.: Facial expression recognition using contourlets and regularized discriminant analysis-based boosting algorithm. In: *International Computer Symposium (ICS)*, pp. 1–5 (2010)
11. Shen, Y., Li, X., Ma, N.-W., Krishnan, S.: Parametric Time-Frequency Analysis and Its Applications in Music Classification. *EURASIP Journal on Advances in Signal Processing* 2010, Article ID 380349, 9 pages (2010)
12. Dai, D.-Q., Yan, H.: *Wavelets and Face Recognition*. I-Tech, Austria (2007)
13. Tzanetakis, G., Essl, G., Cook, P.: Audio Analysis using the Discrete Wavelet Transform. In: *Proc. Conf. in Acoustics and Music Theory Applications, Skiathos, Greece* (2001)
14. Curvelet Literature, <http://www.curvelet.org>
15. Lajevardi, S.M., Hussain, Z.M.: Contourlet structural similarity for facial expression recognition. In: *IEEE International Conference on Acoustics Speech and Signal Processing*, pp. 1118–1121 (2010)

16. Sumana, I., Islam, M., Zhang, D.S., Lu, G.: Content Based Image Retrieval Using Curvelet Transform. In: Proc. of IEEE International Workshop on Multimedia Signal Processing, Cairns, Queensland, Australia, pp. 11–16 (2008)
17. Esakkirajan, S., Veerakumar, T., Murugan, V.S., Sudhakar, R.: Fingerprint Compression Using Contourlet Transform and Multistage Vector Quantization. *International Journal of Biological and Life Sciences* 1, 2 (2005)
18. Do, M.N., Vetterli, M.: Pyramidal directional filter banks and curvelets. In: Proc. of IEEE Int. Conf. on Image Processing, Thessaloniki, Greece, vol. 3, pp. 158–161 (2001)
19. Do, M.N., Vetterli, M.: Contourlet Transform: An Efficient Directional Multiresolution Image Representation. *IEEE Trans. on Image Processing* (2001)
20. Lyons, M.J., Akamatsu, S., Kamachi, M., Goba, J.: Coding facial expressions with gabor wavelets. In: IEEE International Conference on Automatic Face and Gesture Recognition (1998)
21. <http://www.curvelet.org/software.html>
22. Verma, G.K., Prasad, S., Bakul, G.: Robust Face Recognition using Curvelet Transform. In: International Conference on Communication, Computing & Security. ACM, Rourkela (2011)

Robust Watermarking through Spatially Disjoint Transformations

Reena Gunjan, Saurabh Maheshwari, Vijay Laxmi, and M.S. Gaur

Department of Computer Engineering, Malaviya National Institute of Technology,
Jaipur, Rajasthan, India
{reenagunjan, ersammah, vlgaur, gaurms}@gmail.com

Abstract. An image watermarking scheme is introduced with an aim to satisfy the imperceptibility and robustness requirements. The image is divided into spatially disjoint areas. Watermarking is done by interleaving DCT (Discrete Cosine Transform), DWT (Discrete Wavelet Transform) and DFT (Discrete Fourier Transform). Multiple watermarks are simultaneously embedded in a single image in random order. Each area is selected sequentially and watermarked independently. The main advantage of this concurrent watermarking method is that it makes the watermark invariant to all the available frequency based attacks and geometric attacks. The shortcomings of one transform based watermarking scheme are overcome by the other. DFT based watermarking is invariant to geometric attacks and watermark detection attacks but it is not best suited for frequency based attacks like filtering and noise addition. Its shortcomings are overcome by applying the DCT and DWT based method in other areas which are resilient to frequency based attacks but fail against geometric attacks. Hence if any method fails against an attack the watermark can be efficiently extracted by the other implemented methods. Simulation results show that proposed method is able to withstand many image processing attacks.

Keywords: Digital Watermarking, Discrete Cosine Transform, Discrete Fourier Transform, Discrete Wavelet Transform.

1 Introduction

With the growth of the internet, exponential growth rate has been observed in the use of digital media for data transfer. And this increased use has led to piracy of data in transit. It is due to this reason that the protection of intellectual property rights has gained so much of importance. Digital watermarking [3, 10] is one such technique for protection of the digital data by inserting some visible or invisible classified information into it such that the ownership of the media can be claimed. Digital watermarking has been divided into three broad categories depending upon the media in which the watermark is embedded as image, audio and video watermarking.

The paper implements a strategy for image watermarking. Image watermarking can again be deployed in two domains as: spatial and frequency. In spatial domain the watermark is embedded by modifying the pixel values of the image as in contrast to the frequency domain techniques in which the watermark is inserted by modifying the

transform coefficients applied on the image. A large variety of transforms are available that exploit various characteristics of the image to make the embedded watermark resistant towards various attacks. The technique implemented here for image watermarking uses the transforms viz. DFT, DWT and DCT. Watermark is embedded in different areas of the image by exploiting the advantages of each transform.

The organization of the paper is as follows. Section 2 reviews related works in this area, Section 3, 4, 5 describe the theory of DCT, DWT, DFT watermarking respectively. Section 6 elaborates upon the implemented strategy. Section 7 shows simulation results and section 8 puts forward the conclusions.

2 Related Works

Several image watermarking algorithms, with the application of multiple transforms, have been proposed to meet the requirements of robustness and security. An algorithm based on joint DWT and DCT transform is proposed [7] in which a binary image is embedded in three levels DWT transform of host image. For each DWT sub-band, the DCT transform is computed and the binary bits of the watermark are inserted in the middle frequency coefficients.

Another watermark embedding scheme by Kang et al. [6] deals with DWT and DFT together. In DWT domain a spread spectrum based watermark is embedded in the coefficients of the LL sub-band while in DFT domain a template is embedded in the middle frequency components.

Approach on combined DWT-DCT transform by Ali Al-Haj [1] is proposed. In this method the DWT non-overlapping sub-bands are calculated. Second level DWT is applied on horizontal resolution (HL) sub-band. Now the block based DCT is applied in the second level DWT HL sub-band.

Liu et al. [9] describes a blind watermarking algorithm in mixed transform domain. A chaotic sequence is used for locating the watermark. The DCT is applied on the low frequency coefficients obtained through the DWT and chaotic sequence is used to locate the position of the watermark information in DCT domain.

A blind watermarking algorithm [14] is based on combination of DCT and DWT where the DWT of image is performed and the DCT of LL sub-bands is computed. The watermark is scrambled using Arnold and the pseudo random sequence bits are embedded into the middle frequency components of the blocks of DCT.

3 Theory of DCT Watermarking

Discrete Cosine Transform [4] watermarking uses the block based approach. DCT-based method transform 8×8 sized block image into the transform coefficients with the same size and watermark is embedded by modifying them [2]. For extraction the image is divided into similar sized blocks and DCT is applied to them. The watermark is extracted by employing the same approach as in embedding process. The complete schema of watermarking procedure is shown in Figure 1 (a).

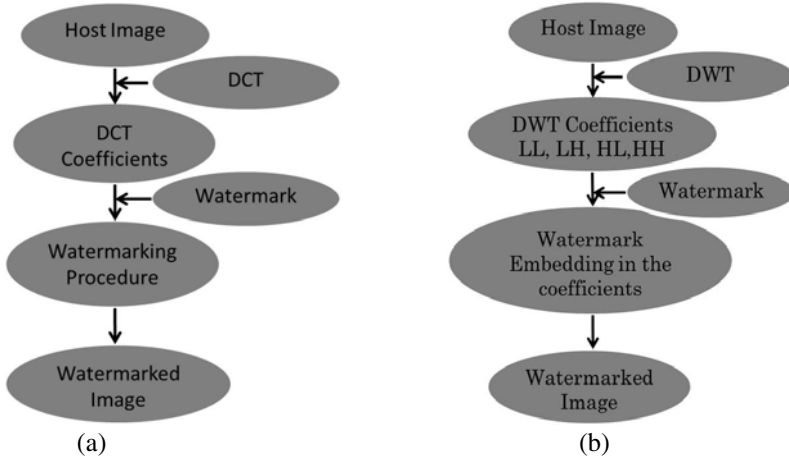


Fig. 1. (a) Schema of DCT based watermarking. (b) Schema of DWT based watermarking.

4 Theory of DWT Watermarking

The DWT based watermarking method utilizes the characteristics of decomposition of image using DWT [15]. DWT is performed on the image to decompose it into four non – overlapping multi-resolution coefficient sets LL1, HL1, LH1 and HH1. HL1 coefficients are further decomposed by applying DWT and then watermark is embedded in second level coefficients, which is performed using ‘haar’ wavelet and the coefficient mean E_j of each wavelet block S_D is computed [6].

$$E_j = \frac{1}{M} \sum_{(x,y) \in S_D} C_j^{(l,s)}(x,y)$$

$C_j^{(l,s)}(x,y)$ is the wavelet coefficient of the j^{th} sub-block, l is the level of wavelet decomposition, s refers to the three sub bands LH, HL and HH, (x,y) is the coordinate of the wavelet coefficient in the wavelet-block S_D , M is the total number of wavelet coefficient in S_D .

If $wm(j) = 1$, we modify the whole wavelet coefficients in S_D , using to the given equation :

$$C_j'^{(l,s)}(x,y) = C_j^{(l,s)}(x,y) - E_j$$

If $wm(j) = 0$, we use following equation :

$$C_j'^{(l,s)}(x,y) = C_j^{(l,s)}(x,y) - E_j + P$$

Here P is defined as $P = A + T(j) * B$ and $T(j)$ is defined as

$$T(j) = texture(j)^\beta = \left(\sqrt{\sum_{(x,y) \in S_D} (C_j^{(l,s)}(x,y) - E_j)^2} \right)^\beta$$

Texture (j) is the local variance of the j^{th} wavelet sub-block and β is a constant. The complete schema of DWT watermarking is shown in Figure 1 (b). A 2-level DWT decomposition of each block is done to extract the embedded bits from a sub-block, Then we compute the coefficient sum (Sum^*) of the wavelet-block S_D , using equation:

$$Sum^* = \sum_{(x,y) \in S_D} C_j^{(l,s)}(x,y)$$

Sum^* is compared with the predefined constant threshold T to decide whether the coefficient carries a bit '0' or '1' using the formula:

$$wm(j) = \begin{cases} 0 & Sum^*(j) \geq T \\ 1 & otherwise \end{cases}$$

Where $wm(j)$ is an extracted watermark bit.

5 Theory of DFT Watermarking

Discrete Fourier Transform (DFT) of a 2 dimensional matrix (image) is calculated as follows.

$$I(k_1, k_2) = \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} i(n_1, n_2) e^{-j2\pi n_1 k_1 / N_1 - j2\pi n_2 k_2 / N_2}$$

Where I is the matrix storing the DFT coefficients, N_1, N_2 are the dimensions of the image. These coefficients are complex numbers. The watermark used with the DFT watermarking is circularly symmetric [12].

Watermark in DFT domain is embedded in the magnitude part of the coefficients. The magnitude of DFT coefficient is resistant to the RST attacks, which makes DFT transform based watermarking scheme RST attacks invariant [5]. The magnitude is given by the absolute value of the DFT coefficients.

$$mag(k_1, k_2) = \sqrt{(real(I(k_1, k_2)))^2 + (imaginary(I(k_1, k_2)))^2}$$

This phase information is required to rebuild the image from the DFT coefficients and calculated as follows.

$$\theta(k_1, k_2) = arctan(imaginary(I(k_1, k_2)) / (real(I(k_1, k_2))))$$

To insert the watermark into the magnitude, additive watermark is applied after multiplying the watermark by watermark strength factor α [11].

$$mag'(k_1, k_2) = mag(k_1, k_2) + f(mag(k_1, k_2), W(k_1, k_2), \alpha)$$

The function f is defined as:

$$f = \alpha * mag(k_1, k_2) * W(k_1, k_2)$$

To rebuild the image we use

$$W(k_1, k_2) = mag(k_1, k_2)cos\theta + i mag(k_1, k_2)sin\theta$$

As a result of the above equation the watermarked magnitude again transformed into the complex values representing the watermarked DFT coefficients. Now inverse DFT is applied to these coefficients so that these are converted into the image intensities. The correlation between the watermarked image DFT magnitude and the watermark W is used for watermark detection purpose. This correlation is performed by following equation.

$$W' = \sum_{n_1=1}^N \sum_{n_2=1}^N W(n_1, n_2)M'(n_1, n_2)$$

So W' is the finally extracted watermark from the image. Figure 2(a) gives the schematic diagram of the DFT based watermarking.

$$W' = IDFT(W)$$

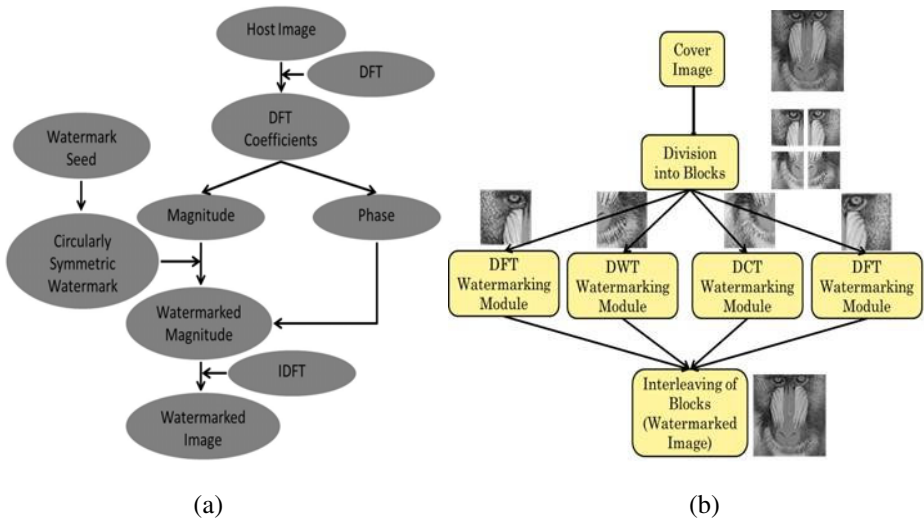


Fig. 2. (a) Schema of DFT based watermarking. (b) Embedding schema using multiple transforms.

6 Proposed Spatially Disjoint Watermarking Procedure

In this paper, a unique image watermarking method is introduced for interleaving the same or different watermarks using various watermarking strategies. The interleaving of watermarks is done by dividing the image into several equal sized areas. The image is divided into four equal areas. Each of the areas is then watermarked with a different strategy. For every area a different watermarking procedure is applied treating that particular area as a reduced sized image. All the four areas are watermarked

independently with same watermark or using the same watermark as a seed to generate the watermark. After the watermarking procedures, all the four areas are joined at their original spatial location to get the final watermarked image as depicted in Figure 2(b). To extract the watermark the watermarked image has to be broken into the same number of areas of same size as it was done during watermark embedding. Each area is then fed to various watermark embedding modules which return the extracted watermark.

7 Experimental Results

The images shown here for the simulation are Baboon and Lena of size 512 x 512. A circularly symmetric watermark is constructed by using watermark as a seed. The watermark for embedding in the DCT and DWT based methods is also generated by using the same watermark. Both the watermarks are shown in Figure 3.

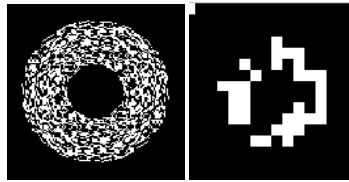


Fig. 3. Watermarks used in DFT, DCT and DWT based schemes

The scheme embeds these four watermarks alternatively to the four areas of Lena image as shown in Figure 4. Figure 5 shows the original and watermarked image of Baboon.

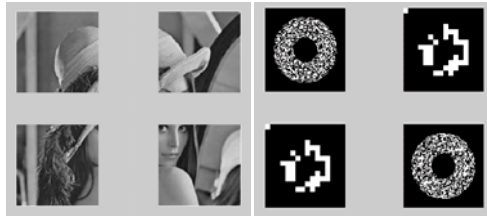


Fig. 4. Image Lena divided into four equal sized partitions and four watermarks embedded into the image

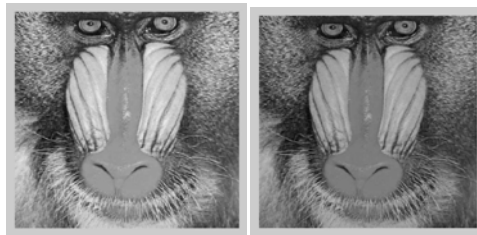


Fig. 5. Original and watermarked image Baboon

7.1 Result Analysis for PSNR

This scheme is tested on gray scale images Lena and Baboon. PSNR is the universally chosen metric for watermark information added to the original image. As a measure of the quality of a watermarked image, the peak signal to noise ratio (PSNR) is typically used. For image with 255 gray levels, the PSNR is defined as:

$$PSNR = 10 \log_{10} \left(\frac{(2^n - 1)^2}{MSE} \right) (dB) \quad MSE = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N (x_{ij} - x'_{ij})^2$$

M and N represent image size. Here the host images are taken as Lena, Baboon and Peppers. n is the number of bits used for storing the pixel intensities. For gray scale images $n=8$. MSE is the mean standard error. x and x' denote the intensity values of pixels of the original and the watermarked images respectively.

7.2 Result Analysis for Similarity Factor

The similarity factor (SM) is a metric which determines the similarity between the watermark inserted and the extracted watermark. The watermark can be extracted from watermarked images with SM 1 and 0.9991 respectively. SM is the similarity factor ranging from 0 to 1. SM=1, when the embedded watermark and the extracted watermark are same. Any other value indicates the difference between them. Generally value of SM >.75 is acceptable. SM can be calculated using the following equation [4]:

$$SM = \frac{\sum_{i=1}^M \sum_{j=1}^N w(i, j) \cdot w^*(i, j)}{\sqrt{\sum_{i=1}^M \sum_{j=1}^N w(i, j)^2 \cdot \sum_{i=1}^M \sum_{j=1}^N w^*(i, j)^2}}$$

The expression has w and w^* as the original and the detected watermark.

7.3 Results for PSNR and SM against Various Signal Processing Attacks

The PSNR value of the watermarked image having hidden watermarks is 39.61 dB for Baboon image and 40.06 dB for Lena image. These values are quite high and provide good quality of the watermarked image even with the presence of four watermarks. The attacks applied to the watermarked image are implemented from Stirmark Benchmark tool [11]. The table 1 shows the performance of the watermarking algorithm against various signal processing attacks. Results consist of SM values detected during the experiments of extraction of watermark from Lena and Baboon images. The high SM values prove that at least one of the watermarks is extractable from any of the areas even after application of attacks. Also the attacks where DFT watermark is not invariant, the DCT based watermark can be detected with high SM values. It depicts the usefulness of the proposed watermarking scheme. Higher the PSNR, better the watermarked image quality.

It is concluded from the results that DCT and DWT performs well for frequency based attacks. Figure 6 shows the watermarked image Baboon which is filtered with disk filter with filtering coefficient 0.8 and watermarks extracted from the filtered watermarked image. The results also show that there is no effect of addition of various types of noise to the image in the detection of the watermarks.

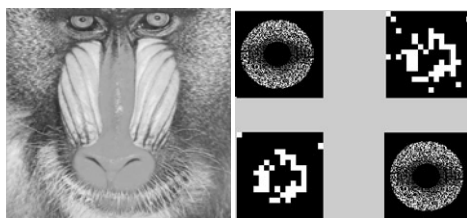


Fig. 6. Watermarked image Baboon filtered with Gaussian filter and watermarks extracted from the filtered watermarked image

It is also seen that the DCT and DWT based watermarks are not robust against RST (Rotation, Scaling and Translation) attacks. This problem is counterfeited by DFT based watermark. The watermark can be detected completely even after rotating the image through 30° or scaling up to 1.5 times as shown in figure 7.

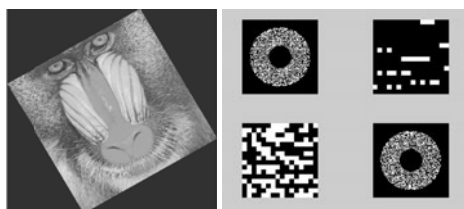


Fig. 7. Watermarked image Baboon rotated through 30° and watermarks extracted from the rotated watermarked image

Table 1. Results of Watermark Extraction for Baboon and Lena image SM values

Attack Category	Watermarked Image(Baboon)			Watermarked Image(Lena)		
	DFT	DCT	DWT	DFT	DCT	DWT
Average filter (2x2)	1.0000	0.7971	0.6764	1.0000	0.6926	0.6602
Median filter (5x5)	0.9984	0.4917	0.6421	0.9865	0.5102	0.6160
Gaussian filter (5x5)	0.9695	0.9854	0.9853	0.9382	0.9732	0.9756
Gaussian noise 0.6	0.9123	0.5207	0.7282	0.9217	0.2936	0.7100
Salt and Pepper noise	0.9832	0.4476	0.7282	0.9764	0.2496	0.7406
Blurring 0.6	1.0000	0.9483	0.8109	1.0000	0.9111	0.8083
Sharpening 0.6	1.0000	0.9342	0.8203	1.0000	0.9210	0.8002
Rotation 30°	1.0000	0.3243	0.6971	1.0000	0.3200	0.6463
Scaling 1.5	0.9894	0.5018	0.6055	0.9887	0.5016	0.6012
JPEG compression 20%	1.0000	0.3615	0.4023	1.0000	0.3378	0.4112
JPEG compression 40%	1.0000	0.5006	0.6123	1.0000	0.5100	0.6111
Shape Distortion						
Affine transform	1.0000	0.4919	0.7282	1.0000	0.4432	0.6743
Random Shape	1.0000	0.5436	0.6136	1.0000	0.5623	0.6159
Inverting Pixels						
Side by side	1.0000	0.5051	0.6239	1.0000	0.4432	0.4532
Diagonally	1.0000	0.5533	0.5952	1.0000	0.5114	0.5589

The results of JPEG compression for Lena and Baboon images are also shown in Table 1. JPEG 20 means the JPEG image is compressed by 20%. Even after this compression the SM between the original and the detected watermark is very high due to invariance of DFT watermark against JPEG compression.

8 Conclusion

In the proposed strategy, the watermark embedding is done using three transforms, DCT, DWT and DFT simultaneously in different spatially disjoint areas of the image making the method robust against different type of signal processing and geometric attacks. Experimental results demonstrate that after JPEG compression the PSNR is high due to invariance of DFT watermark against JPEG compression. The high watermark detection ratio (SM) values prove that the watermark is extractable even after the application of attacks. The scheme suffers from the disadvantage that the transformations have not been implemented by taking into account the statistical properties of the image. The transformations can be applied depending upon the high and low frequency regions of the image. The size of the areas can be further reduced which may increase the processing time but at the same time making the algorithm more robust.

References

1. Al-Haj, A.: Combined DWT-DCT Digital Image Watermarking. *Journal of Computer Science*, 740–746 (2007)
2. Chen, L., Li, M.: An Effective Blind Watermark Algorithm Based on DCT. In: *Proceedings of the 7th World Congress on Intelligent Control and Automation*, pp. 6822–6825 (June 2008)
3. Gonzales, F.P., Hernandez, F.R.: A Tutorial on Digital Watermarking. In: *IEEE 33rd Annual International Carnahan Conference on Security Technology*, pp. 286–292 (October 1999)
4. Gunjan, R., Maheshwari, S., Gaur, M.S., Laxmi, V.: Permuted Image DCT Watermarking. In: *Herrero, Á., Corchado, E., Redondo, C., Alonso, Á. (eds.) Computational Intelligence in Security for Information Systems 2010. Advances in Intelligent and Soft Computing*, vol. 85, pp. 163–171. Springer, Heidelberg (2010)
5. Hernandez, M.C., Miyatake, M.N., Meana, H.M.P.: Analysis of a DFT-based watermarking algorithm. In: *2nd International Conference on Electrical and Electronics Engineering*, pp. 44–47 (September 2005)
6. Kang, X., Huang, J., Shi, Y.Q., Lin, Y.: A DWT-DFT composite watermarking scheme robust to both affine transform and JPEG compression. *IEEE Transactions on Circuits and Systems for Video Technology* 13(8), 776–786 (2003)
7. Kasmani, S.A., Nilchi, A.N.: A New Robust Watermarking Technique based on Joint DWT-DCT Transformation. In: *International Conference on Convergence and Hybrid Information Technology*, pp. 539–544 (July 2008)
8. Lin, P.Y., Lee, J.S., Chang, C.C.: Dual Digital Watermarking for Internet Media Based on Hybrid Strategies. *IEEE Transactions on Circuits and Systems for Video Technology* 19(8), 1169–1177 (2009)
9. Liu, J., Gao, F., Zhang, H.: A Blind Watermark Algorithm in Mixed Transform Domain Based on Chaotic Sequence Locating. In: *The 3rd International Conference on Innovative Computing Information and Control (ICICIC 2008)*, p. 23 (June 2008)

10. Macq, B., Dittmann, J., Delp, E.: Benchmarking of Image Watermarking Algorithms for Digital Rights Management. *Proceedings of the IEEE* 92(6), 971–984 (2004)
11. Podilchuk, C.I., Delp, E.J.: Digital Watermarking: Algorithms and Applications. *IEEE Signal Processing Magazine*, 33–45 (July 2001)
12. Solachidis, V., Pitas, I.: Circularly Symmetric Watermark Embedding in 2D-DFT Domain. *IEEE Transactions on Image Processing* 10(11), 1741–1753 (2001)
13. Suhail, M.A., Obaidat, M.S.: Digital Watermarking-Based DCT and JPEG Model. *IEEE Transactions on Instrumentation and Measurement* 52(5), 1640–1647 (2003)
14. Liu, P.F., Liang, B.Z., Peng, C.: A DWT-DCT Based Blind Watermarking Algorithm for Copyright Protection. In: *3rd IEEE International Conference on Computer Science and Information Technology (ICCSIT)*, vol. 7, pp. 455–458 (2010)
15. Zhang, G., Wang, S., Nian, G.: A blind watermarking algorithm based on DWT for color image. In: *The 2004 Joint Conference of the 10th Asia-Pacific Conference on 5th International Symposium on Multi-Dimensional Mobile Communications Proceedings*, August 29-September 1, vol. 2, pp. 634–638 (2004)

An ANN Based Approach to Calculate Robotic Fingers Positions

Ankit Chaudhary^{1,*}, J.L. Raheja², Kunal Singal³, and Shekhar Raheja⁴

¹ Computer Vision Research Group
Dept. of Computer Science, BITS, Pilani, India
ankitc.bitspilani@gmail.com

² Machine Vision Lab, Digital Systems Group
CEERI, Pilani, India

jagdish@ceeri.ernet.in

³ PEC, Chandigarh, India

Kunal5596@gmail.com

⁴ BKBIET, Pilani, India

shekhar.raheja@rediffmail.com

Abstract. In the certain applications, a human hand like robotic hand is needed to do the operations alike human can do. The fingers in the human hand have the capability to bend on different angles and they can apply the force or can hold an object. Current focus of our research is on controlling the robot fingers using vision based techniques, which have the joints in finger like human hand. This paper describes our approach of robotic fingers positions calculation using supervised Artificial Neural Network. User has to show a gesture to the system without any limitation or restriction. Hand cropping gives the region of interest and made the algorithm faster by processing smaller image. The gesture would be extracted from the input image and after detecting fingertips in the region of interest, fingers bending angles would be calculated using ANN.

Keywords: Human Computer Interface, Intelligent Systems, Artificial Neural Network, Fingertip Detection, Action Recognition, Digital Image Processing.

1 Introduction

There are applications or situations like bomb detection and diffusion, landmine removal, and others where more degree of freedom is required than a static machine shaft, the shaft should have more functionality. People have an urgent need of a robotic hand, which can work same as the human hand. The robotic hand should be able to bend fingers like human and it should be easily controllable. Also it should be able to work in user interactive situations with a near to real time response time.

* Author is currently working toward his PhD at Machine Vision Lab, Central Electronics Engineering Research Institute, Pilani INDIA (www.ceeri.res.in), affiliated to the Birla Institute of Technology & Science, Pilani, INDIA (www.bits-pilani.ac.in).

This paper describes a novel method to calculate positions of robot fingers using Artificial Neural Network. In the past researchers have done significant work in this area of hand gesture recognition but they used a wired glove in which sensors were planted or they used colors on fingers to recognize the gesture clearly in the image. This work requires no use of sensors or no use of colors or no paper clips are required to detect fingers in the image. The detection of moving fingers in real time needs a fast and robust implementation. Processing time is the very critical factor in real time applications which depends on several factors. Our vision based system that can be used to control a remotely located robotic hand. it would be able to perform same operation as the human hand. User have to bend his fingers to hold object (virtual object), and the robotic hand will do the same to hold actual object on its location. Our system capture images in 2D, which will be preprocessed and skin filter, segmentation will be applied on the images. This will result in region of interest (ROI) in hand gesture even if there are skin colors like objects in the gesture background. It detects fingertips and center of palm in ROI and calculates the distances between each fingertip and centre of palm and classifies it with ANN. The block diagram of proposed system is given in the figure 1.

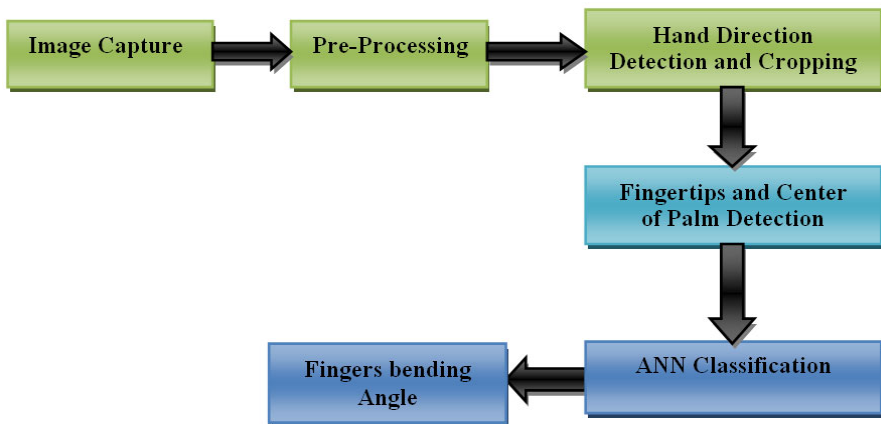


Fig. 1. Block Diagram flow of the system in different phases

2 Related Work

The survey and current state of our system can be found in [1-3]. Nolker [4] focuses on large number of 3D hand postures in her system called GREFIT. She used finger tips in hands as natural determinant of hand posture to reconstruct the image. It takes 192x144 size gray scale image to process. Nolker used ANN based layer approach to detect fingertips. After obtaining fingertips vectors, it is transformed into finger joint angles to an articulated hand model. For each finger separate network were trained on

same feature vectors, having input space 35 dimensional while output dimensional as only 2. Lee [5] used Hidden Markov Model (HMM) for gesture recognition using shape feature. Gesture state is determined after stabilizing the image component as open fingers in consecutive frames. He also used maxima and minima approach like Raheja [6] for construction the hand image and FSM. Wang [7] proposed an optical flow based powerful approach for human action recognition using learning models. It labels hidden parts in image also. This mas-margin based algorithm can be applied to gesture recognition. Kim [8] in his system used learning model for dynamic gestures recognition. Many other applications could be found in literature of real time control in human computer interaction, Computer games control [9], human robot interaction [10], and sign language recognition [11].

3 System Description

Our system use real time input from the user and calculate the fingers positions in different gestures. There are no restrictions on user to use any kind of sensor or wired device, even it is not mandatory to wear full sleeve shirt. User is free to show hand gesture in any direction to the camera. Hand can be on any angle in any direction from the camera. System first has to find out the direction of hand to extract ROI. The captured image would be pre-processed and then skin filter and segmentation would be applied on the image. The pre-processing details of captured image sequences are discussed in our previous work [2]. In gesture recognition colour based methods are applicable because of their characteristic of colour footprint of human skin. The colour footprint is usually more distinctive and less sensitive than in the standard (camera capture) RGB colour space. Then system detects fingertips and centre of palm in the extracted gesture image. Finally it measures distance between centre of palm and fingertips in pixels and calculated angle for each finger with the help of ANN.

The image was captured using a simple camera which was connected to a windows XP® running PC. There can be different lightning conditions, in which user is giving input to system. An HSV color space based skin filter was applied on the RGB format image to reduce lighting effect. The resultant image was segmented to get a binary image of hand. To remove errors in the resultant image, the BLOB was applied to the noisy images, and then we find out the wrist in the image to know hand direction. ROI was extracted by image cropping by using formula shown below.

$$imcrop = \left\{ \begin{array}{l} origin_{image}, \text{ for } Xmin < X < Xmax \\ \quad \quad \quad Ymin < Y < Ymax \\ 0, \quad \quad \quad \text{elsewhere} \end{array} \right\}$$

Where imcrop represents the cropped image, Xmin, Ymin, Xmax, Ymax represent the boundary of the hand in the image.

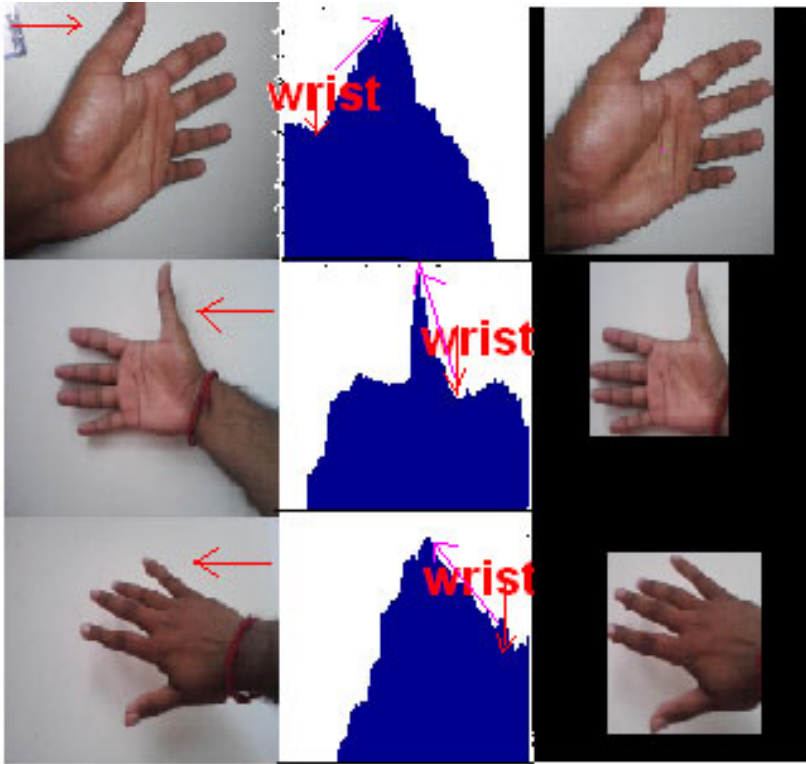


Fig. 2. Results of Hand cropping process [2]

Some results of image cropping for different inputs are shown in figure 2 where the arrows in the main frames indicate the direction of scan which is found from hand direction detection step. In all the histograms in figure 2 it is clear that at the wrist point, a steeping inclination starts in the scanning direction.

Fingertips have been used in different systems for different purposes. In our approach the hand directions are already known, so fingertips would be found out using the following function:-

$$Finger_{edge}(x,y) = \begin{cases} 1 & \text{if } modified_{image}(x,y) = 255 \\ 0 & \text{otherwise} \end{cases}$$

Here $Finger_{edge}$ would give the fingertips. Automatic centre of palm (COP) would be detected by applying a mask of dimension 30x30 to the cropped image and counting the number of on pixels lying within the mask. This process was made faster using summed area table of the cropped binary image for calculating the masked values. The fingertips and COP detection result is shown in the figure 3.



Fig. 3. Fingertips and Center of palm detected in the extracted image

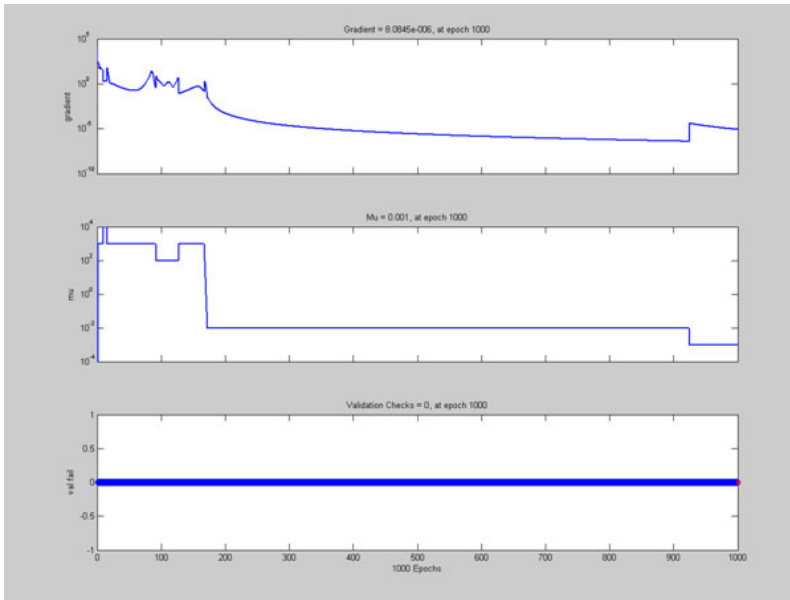


Fig. 4. Training State using 1000 Iterations

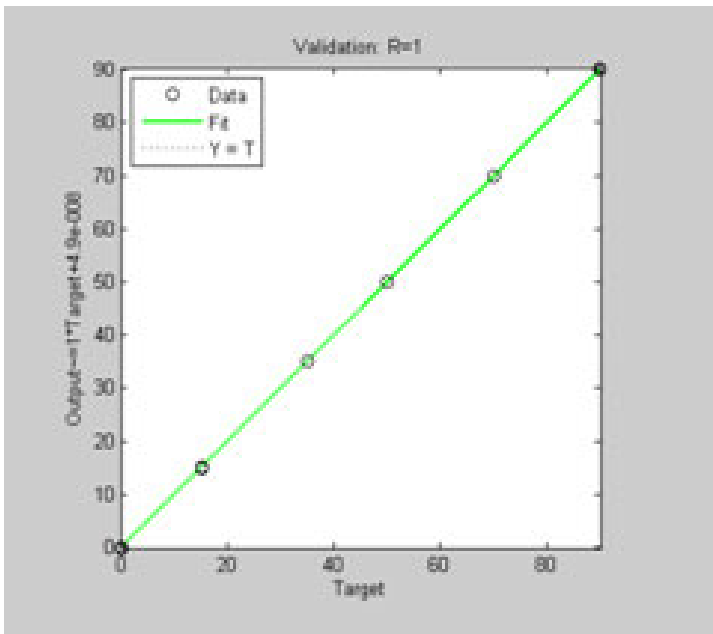


Fig. 5. Data Validation State Graph

4 ANN Implementation

A supervised ANN was implemented using the 8000 sample data for all fingers in different positions and ANN was trained for 1000 iterations. The training state during iterations and data validation is shown in figure 4 and figure 5 respectively. ANN architecture has 5 inputs for all fingers positions and 5 outputs for corresponding angles. The design includes 2 hidden layer and 23 Neurons for processing, which we found out for better results after many iterations for different designs. The input data include the distance from COP to fingertips for all fingers at different angles. The Mean squared error in ANN is of the order of 10^{-12} as shown in figure 6.

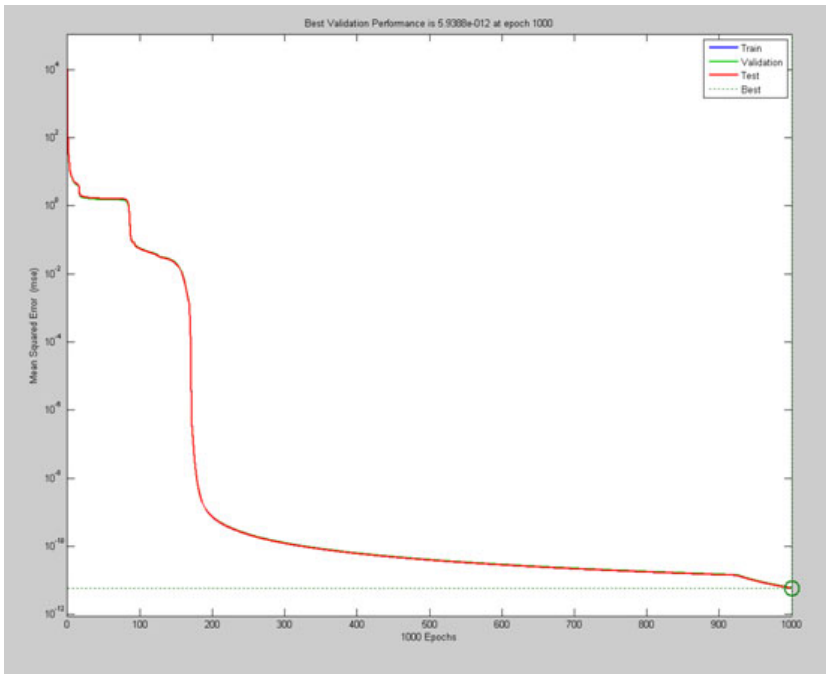


Fig. 6. Mean Squared Error in the ANN

5 Results

The system takes 160x120 frame resolution in RGB format, to which it process internally. It takes the reference distance (RD) from the first frame of hand, which user has to show as straight hand without bending any finger. This RD will work throughout the session of system. If hand disappears from the camera view, the system will take the new RD in the same session, so different users can work in the same session. Now if the user bends the fingers, the new distance (ND) will be calculated from the COP. The ratio of ND to RD would be sent to ANN to match with manually collected data and ANN will results in the corresponding angles for all

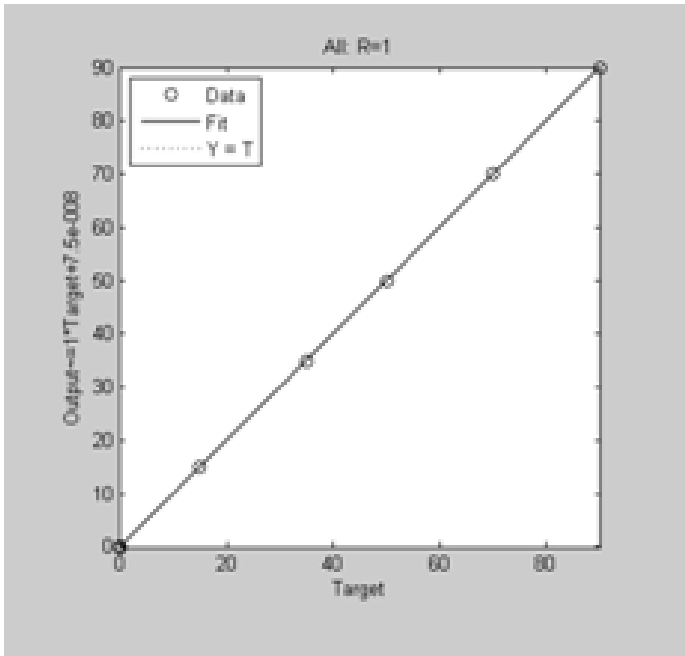


Fig. 7. Output Vs Training Graph

the five fingers. The output vs training graph for ANN is shown in figure 7. The response time of the system to display angles for one gesture is 0.001 seconds, that is very satisfactory and near to real time.

6 Conclusion

This paper gives a new research direction in gesture recognition technologies where trained ANN is used to calculating fingers positions online. The system is able to detect the fingertip in the real time movements of hand. The gesture was extracted even from the complex background and cropping of image made the algorithm faster. User can show any hand to camera in any direction, and fingertips and centre of palm would be detected in the captured image sequences. The bending angles of fingers were calculated using the ANN by performing gesture in front of camera without wearing any gloves or markers. The results are satisfactory and this technology can be used in a many real life applications to save human life. We are still working to refine the results and fastening the response time.

Acknowledgement

This research is being carried out at Central Electronics Engineering Research Institute (CEERI), Pilani, India as part of our project 'Controlling the robotic hand using hand gesture'. Authors would like to thank Director, CEERI for his active encouragement and support.

References

1. Chaudhary, A., Raheja, J.L.: ABHIVYAKTI: A Vision Based Intelligent Pervasive System for Elder and Sick Persons. In: 3rd IEEE International Conference on Machine Vision, Hong Kong, December 28-30, pp. 361–364 (2010)
2. Raheja, J.L., Das, K., Chaudhary, A.: An Efficient Real Time Method of Fingertip Detection. In: Proceedings of 7th International Conference on Trends in Industrial Measurements and Automation (TIMA 2011), January 6-8, pp. 447–450. CSIR Complex, Chennai (2011)
3. Chaudhary, A., Raheja, J.L., Das, K., Raheja, S.: A Survey on Hand Gesture Recognition in Context of Soft Computing. In: Meghanathan, N., Kaushik, B.K., Nagamalai, D. (eds.) CCSIT 2011, Part III. Communications in Computer and Information Science, vol. 133, pp. 46–55. Springer, Heidelberg (2011)
4. Nolker, C., Ritter, H.: Visual Recognition of Continuous Hand Postures. *IEEE Transactions on Neural Networks* 13(4), 983–994 (2002)
5. Lee, D., Park, Y.: Vision-Based Remote Control System by Motion Detection and Open Finger Counting. *IEEE Transactions on Consumer Electronics* 55(4), 2308–2313 (2009)
6. Raheja, J.L., Shyam, R., Kumar, U., Prasad, P.B.: Real-Time Robotic Hand Control using Hand Gesture. In: 2nd International Conference on Machine Learning and Computing, Bangalore, India, February 9-11, pp. 12–16 (2010)
7. Wang, Y., Mori, G.: Max-Margin Hidden conditional random fields for human action recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, Miami, Florida, USA, June 20-25, pp. 872–879 (2009)
8. Kim, H., Fellner, D.W.: Interaction with hand gesture for a back-projection wall. In: Proceedings of Computer Graphics International, pp. 395–402 (June 19, 2004)
9. Freeman, W.: Computer vision for television and games. In: Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, p. 118 (1999)
10. Triesch, J., Malsburg, C.V.D.: Mechanical gesture recognition. In: Gesture Workshop, pp. 233–244 (1997)
11. Starner, T., Pentland, A.: Real-time American Sign Language recognition from video using hidden markov models. In: SCV 1995, pp. 265–270 (1995)

Word Classification Using Neural Network

A. Muthamizh Selvan^{1,*,**} and R. Rajesh^{2,***}

¹ Research Scholar, School of Computer Science and Engineering
Bharathiar University, Coimbatore - 641 046, India

muthamizh@ieee.org

² Senior Research Associate, Network System and Technologies (NeST)
Technopark, Thiruvananthapuram, India

kollamrajeshr@ieee.org

Abstract. Classification of words plays a primary vital role to develop a robust automatic speech recognition (ASR) applications due to the diversity in the vocal tract of speakers. This paper presents Neural Network based word classification using the combination of features like, MFCC, Zero Crossing, Zero-Crossing Rate (ZCR) and Formants. The results of word classification are promising.

Keywords: MFCCs, Zero-Crossing Rate (ZCR), Formants, Speech Classification.

1 Introduction

Automatic speech recognition (ASR) is one of the hottest topics for the last several decades and it has lot of applications in man-machine communication systems, mobile phones and security applications [8]. Recently many approaches have been developed to find relevant features from the speech [3].

The most widely used features are cepstral coefficients (LPCCs) obtained through Linear Predictive Coding (LPC) [6] [18], Perceptual Linear Prediction (PLP) Coefficients [14] [17] [24], Mel-Filter Bank (MFB) [6], Mel-Frequency Cepstral Coefficients (MFCCs) [2] [4] [5] [10] [15] [19] [20] [21] [24], Auto-Correlation based MFCCs (AMFCCs) [18], Fractional Fourier Transform based MFCCs (FrFT-MFCC) [13], Minimum Mean Square Error (MMSE) [10], ZCR [1] [2] [23] [26]. The RASTA-PLP, Normalization, Principal Component Analysis (PCA), Independent Component Analysis (ICA) [11], Linear Discriminant Analysis (LDA) are the other methods to achieve the relevant features [15]. Each of these methods and features has advantages.

In this paper, we make use of MFCCs, Zero Crossing Rate (ZCR) and Formants for word classification using Neural Network. The results of classification are promising.

* Corresponding author.

** This work is partially supported by NTS-India, CIIL.

*** Dr. R. Rajesh, Assistant Professor (on leave), Bharathiar University is currently working as SRA at NeST.

This paper is organized as follows. Section 2 gives a short description of the features used. Section 3 presents the results and discussion. Section 4 concludes the paper.

2 Features for Classification

2.1 Mel-Frequency Cepstral Coefficients (MFCCs)

The Mel-Frequency Cepstral Coefficients (MFCCs) are the dominant features widely used for speech processing applications such as identification, classification and recognition [22]. Mel-Frequency Cepstrum (MFC) is the short-term power spectrum of a sound based on the linear cosine transform of the log power spectrum on the Mel-Scale frequency. Mel-Frequency Cepstral Coefficients (MFCCs) collectively make up a MFC and are derived from the cepstral representation of the sound files. The difference between the Cepstrum and the Mel-Frequency Cepstrum is that the frequency bands are equally spaced on the Mel-Scale, which approximates the human auditory system's response more closely than the linearly-spaced frequency bands used in the normal cepstrum. The block diagram for extracting MFCC is shown in the below fig.1

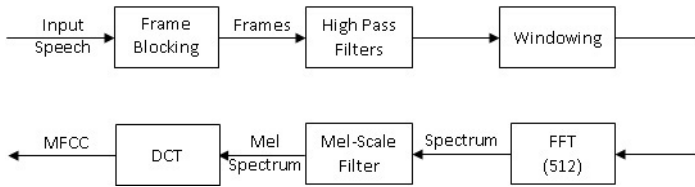


Fig. 1. Block Diagram of extracting MFCCs

In this paper MFCCs on the short-time frames are extracted as described below.

1. 20 Short time frames are extracted with 50% overlapping for each of the words (wav files). As the result of overlapping total number of frames is 39.
2. High pass pre-emphasis filter applied with the coefficients denominator $a = 1$ and numerator $b = [1, -0.97]$.
3. Multiply by the Hamming window into each frame.
4. Fast Fourier Transformation (FFT) is applied to each frames with size of FFT is 512.
5. The amplitudes of the mel-scale filter bank are calculated with number of filters as 24.
6. The log spectrum of each filter is calculated.
7. Discrete Cosine Transformation (DCT) is applied to produce the Mel-filtered cepstral coefficients.
8. Mel-Frequency Cepstral Coefficients (MFCCs) are produced in 24 dimensions in each frame.

2.2 Formants (F1, F2, F3 and F4)

The peaks of energy in the spectrum at particular frequencies of a wav file which contribute to the perceived quality of that wave are called Formants or Harmonics frequencies. Formants are powerful frequency features to describe sounds produced by humans, but also by instruments and are caused by the acoustic resonance phenomenon. The formants with lower frequencies are dominant in distinguishing the vowel patterns [9] [13]. Here we used the first four formants F1, F2, F3 and F4, calculated by an Auto-Regressive Model which is linear prediction model that gives an output $y(n)$ as a linear combination from the preceding outputs $[y(n-1), y(n-2), y(n-3), \dots, y(n-p)]$.

The formants are extracted as given below.

1. Auto-Regressive Model applied to the input sound.
2. Transfer function and frequency responses are computed.
3. Peaks locations of the frequency response have been detected.
4. Resonant frequencies of all peaks are extracted and limited on the lowest four frequencies.

2.3 Zero Crossing Rate

The number of zero crossings (ZC) of the input sound is divided by the duration of the frame to get the rate of zero crossings (ZCR) [1]. Zero-Crossing Rate (ZCR) is associated to the mean frequency for a given input signal. The number of zero crossings is null in the silent [23]. Occurrence of null zero-crossings is a relevant feature for identifying speech activity.

3 Results and Discussions

The speech material have been prepared too narrow and recorded in the closed room by the use of normal microphone with the sampling rate 16 KHz. The structure of the utterances or words is with the 12 initial vowels "short A, long A, short I, long I, short U, long U, short E, long E, diphthong AI, short O, long O and diphthong AU" available in Tamil language. The form of producing the utterance in each speaker is <Amma> <Aadu> <Ilai> <Iitty> <Ural> <Uudha> <ERumbu> <EeNi> <Aindhu> <Ottakam> <Oodam> <Auvai>. Totally 504 word utterances have been recorded from 42 speakers with 30 female and 12 male. MFCCs, Formants (F1, F2, F3 and F4), Zero Crossings and Zero Crossing Rate (ZCR) are extracted for all the 504 utterances and got the feature vector with the size of 504 X 942 for the classification. The Feature Matrix shown in Table [1].

Neural Network is an optimal classifier to provide pattern classification as a result of the ability to train and recognize the patterns. In this work we used pattern recognition network which is a feed-forward network with tan-sigmoid

Table 1. Feature Matrix

Feature	No. of Speakers	No. of Utterances	Size of Features
MFCC	42	12	504 X (39 X 24)
F1,F2,F3 and F4	42	12	504 X 4
ZeroCrossings (ZC)	42	12	504 X 1
ZCR	42	12	504 X 1
Total Features	42	12	504 X 942

transfer functions in both the hidden layer and the output layer. We used 100 neurons in one hidden layer for the function-fitting. The Scaled Conjugate Gradient algorithm has been used for training with the 50% of the feature vector extracted from the 15 female and 6 male speakers. The remaining 50% of the feature vector has been used to testing process. The performance goal has been evaluated by Mean Squared Error (MSE).

Four combinations of the features have been used to find the best feature sets and the results are deployed. The results show that the proposed method with combination of MFCCs and Formants (F1,F2,F3 and F4) gives better classification accuracy.

Two times cross validation has been done by interchanging the training and testing features. Table 2 shows the average word classification rate for two cross validation with input runs of each validation.

Table 2. Word Classification Rate

No.	Method	Classification
1	MFCC + NN	85.01 %
2	MFCC + Formants (F1,F2,F3 and F4) + NN	90.77 %
3	MFCC + ZeroCrossing + NN	85.95 %
4	MFCC + Formants + ZeroCrossing + NN	87.47 %

The classification results are comparable with the work done by Ching-Tang Hsieh et.al. [7], Jaehwan Kang et. al. [16], Hisham Al-Mubaid [12] and Tong Zhang et.al. [26]. The comparison results are shown in Table 3.

Table 3. Comparison of Classification Rate

Author	Classification
Ching-Tang Hsieh et.al. [7]	88.86 %
Jaehwan Kang et. al. [16]	71.50 %
Hisham Al-Mubaid [12]	86.57 %
Tong Zhang et.al. [26]	86.00 %

4 Conclusion

This paper has proposed a word classification method by use of pattern recognition neural network classifier. The feature matrix consists MFCC, Formants (F1,F2,F3 and F4), Zero Crossings and Zero Crossing Rate is computed and applied to neural net for classifying the words. The classification results are promising.

Acknowledgement

Dr. R. Rajesh is thankful to Bharathiar University and NeST. The first author is gratified to National Testing Service (NTS) - India, Central Institute of Indian Languages (CIIL), Ministry of HRD, Govt. of India for the valuable fellowship and thankful to all the staff members and research scholars of the Department of Computer Applications, Bharathiar University for their valuable support.

References

1. Pirkakis, A., Giannakopoulos, T., Theodoridis, S.: A Speech/Music Discriminator of Radio Recordings Based on Dynamic Programming and Bayesian Networks. *IEEE Transactions on Multimedia* 10(5), 846–857 (2008)
2. Al-Haddad, S.A.R., Samad, S.A., Hussain, A., Ishak, K.A.: Isolated Malay Digit Recognition Using Pattern Recognition Fusion of Dynamic Time Warping and Hidden Markov Models. *American Journal of Applied Sciences* 5(6), 714–720 (2008)
3. Anusuya, M.A., Katti, S.K.: Speech Recognition by Machine: A Review. *International Journal of Computer Science and Information Security* 6(3), 181–205 (2009)
4. Maier, A., Haderlein, T., Stelzle, F., Noth, E., Nkenke, E., Rosanowski, F., Schutzenberger, A., Schuster, M.: Automatic Speech Recognition Systems for the Evaluation of Voice and Speech Disorders in Head and Neck Cancer. *EURASIP Journal on Audio, Speech, and Music Processing* 2010 (2010)
5. Lee, C.-H., Hanand, C.-C., Chuang, C.-C.: Automatic Classification of Bird Species From Their Sounds Using Two-Dimensional Cepstral Coefficients. *IEEE Transactions on Audio, Speech, and Language Processing* 16(8), 1541–1550 (2008)
6. Jankowski Jr., C.R., Vo, H.-D.H., Lippmann, R.P.: A Comparison of Signal Processing Front Ends for Automatic Word Recognition. *IEEE Transactions on Speech and Audio Processing* 3(4), 286–293 (1995)
7. Hsieh, C.-T., Hsu, C.-H.: Speech Classification Based on Fuzzy Adaptive Resonance Theory. In: *Proceedings of the 2006 Joint Conference on Information Sciences*. Atlantis Press, Taiwan (2006)
8. Levy, C., Linares, G., Bonastre, J.-F.: Compact Acoustic Models for Embedded Speech Recognition. *EURASIP Journal on Audio, Speech, and Music Processing* 2009 (2009)
9. Gläser, C., Heckmann, M., Joubin, F., Goerick, C.: Combining Auditory Preprocessing and Bayesian Estimation for Robust Formant Tracking. *IEEE Transactions on Audio, Speech, and Language Processing* 18(2), 224–236 (2010)
10. Yu, D., Deng, L., Droppo, J., Wu, J., Gong, Y., Acero, A.: A Minimum-Mean-Square-Error Noise Reduction Algorithm on Mel-Frequency Cepstra for Robust Speech Recognition. In: *IEEE: ICASSP*, pp. 4041–4044 (2008)

11. Kolossa, D., Astudillo, R.F., Hoffmann, E., Orglmeister, R.: Independent Component Analysis and Time-Frequency Masking for Speech Recognition in Multitalker Conditions. *EURASIP Journal on Audio, Speech, and Music Processing* 2010 (2010)
12. Al-Mubaid, H.: A Learning - Classification based Appro Word Prediction. *The International Arab Journal of Information Technology* 4(3), 264–271 (2007)
13. Nadeu, H.Y.C., Hohmann, V.: Pitch and Formant Based Order Adaptation of the Fractional Fourier Transform and Its Application to Speech Recognition. *EURASIP Journal on Audio, Speech, and Music Processing* 2009 (2009)
14. Boril, H., Hansen, J.H.L.: Unsupervised Equalization of Lombard Effect for Speech Recognition in Noisy Adverse Environments. *IEEE Transactions on Audio, Speech, and Language Processing* 18(6), 1379–1393 (2010)
15. Park, H., Takiguchi, T., Ariki, Y.: Integrated Phoneme Subspace Method for Speech Feature Extraction. *EURASIP Journal on Audio, Speech, and Music Processing* 2009 (2009)
16. Kang, J., Lee, H.: Automatic Voice Classification System Based on Traditional Korean Medicine. *World Academy of Science, Engineering and Technology* 56, 35–38 (2009)
17. Frankel, J., King, S.: Speech Recognition Using Linear Dynamic Models. *IEEE Transactions on Audio, Speech, and Language Processing* 15(1), 246–256 (2007)
18. Morales-Cordovilla, J.A., Peinado, A.M., Snchez, V., Gonzlez, J.A.: Feature Extraction Based on Pitch-Synchronous Averaging for Robust Speech Recognition. *IEEE Transactions on Audio, Speech, and Language Processing* 19(3), 640–651 (2011)
19. Muda, L., Begam, M., Elamvazuthi, I.: Voice Recognition Algorithms using Mel Frequency Cepstral Coefficient (MFCC) and Dynamic Time Warping (DTW) Techniques. *Journal of Computing* 2(3), 138–143 (2010)
20. Korba, M.C.A., Messadeg, D., Djemili, R., Bourouba, H.: Robust Speech Recognition Using Perceptual Wavelet Denoising and Mel-frequency Product Spectrum Cepstral Coefficient Features. *Informatica* 32, 283–288 (2008)
21. Morales, N., Toledano, D.T., Hansen, J.H.L., Garrido, J.: Feature Compensation Techniques for ASR on Band-Limited Speech. *IEEE Transactions on Audio, Speech, and Language Processing* 17(4), 758–774 (2009)
22. Wang, N., Ching, P.C., Zheng, N., Lee, T.: Robust Speaker Recognition Using Denoised Vocal Source and Vocal Tract Features. *IEEE Transactions on Audio, Speech, and Language Processing* 19(1), 196–205 (2011)
23. Panagiotakis, C., Tziritas, G.: A speech/music discriminator based on RMS and zero-crossings. *IEEE Transactions on Multimedia* 7(1), 155–166 (2005)
24. Dharanipragada, S., Yapanel, U.H., Rao, B.D.: Robust Feature Extraction for Continuous Speech Recognition Using the MVDR Spectrum Estimation Method. *IEEE Transactions on Audio, Speech, and Language Processing* 15(1), 224–234 (2007)
25. Scheirer, E., Slaney, M.: Construction and Evaluation of a Robust Multifeature Speech/Music Discriminator. In: *Proc. Int. Conference on Acoustics, Speech and Signal Processing (ICASSP)*, vol. 2, pp. 1331–1334 (1997)
26. Zhang, T., Jay Kuo, C.C.: Audio content analysis for online audiovisual data segmentation and classification. *IEEE Transactions on Speech and Audio Processing* 9(4), 441–457 (2001)

Very Short Term Wind Power Forecasting Using PSO-Neural Network Hybrid System

E. Pratheepraj*, Anuj Abraham, S.N. Deepa, and V. Yuvaraj

Dept. of Electrical and Electronics Engineering
Anna University of Technology
Coimbatore, India

{pratheep.be, deepapsg}@gmail.com,
{aunj1986aei, yuva520}@yahoo.co.in

Abstract. In recent years, the small-scale wind power generation increases rapidly worldwide. The wind power depends on the wind speed, which is a random variable and is irregular. For efficient operation of wind power plants accurate short-term forecasts are essential. The knowledge of future power generation from wind turbines is useful for schedulers, transmission operators and energy traders. In this paper, wind speed and the power generation are predicted using a particle swarm optimization (PSO) algorithm -Neural hybrid system. The neural network is used in many prediction systems and it gives very successful results compared to other forecasting techniques like persistence, mean methods etc. At present many prediction systems are built with ANFIS but it is a more complex structure and not suitable for small scale wind power plants. The result of this hybrid system shows that it is more accurate and reliable for short term wind power forecasting. In this paper we have compared the results of hybrid model with the earlier available techniques like standard neural network and Genetic Neural Network for verification purpose.

Keywords: Short term wind power forecasting, Artificial neural networks, particle swarm optimization, Wind power generation, Artificial intelligence.

1 Introduction

It is seen from the literature survey that the trend towards the sustainable energy and green power sources such as wind energy was largely increasing [1]. Wind energy is one of the economic renewable sources and a valuable supplement to conventional energy sources. The important problems in wide usage of wind power are difficulties in accuracy of wind power forecast and it is not able to store. Fluctuations in wind power production, also makes it difficult for owners of wind power plants to compete in electricity markets. Economical efficient schedulers are required to meet an accurate forecast for grid operators with the demand of electrical customers [2]. In the operation of wind power forecasting, up to 48 hours ahead is recognized from wind

* Corresponding author.

farm operators, transmission system operators, and other end-users, which is a major contribution for a reliable addition of wind generation in a power system. Enhancement in the position of wind energy as a prediction tool can be seen in the open market environment with other forms of power generation.

The Back Propagation Neural Networks (BPNN) is one the most popular method for wind speed prediction due to its simple ANN structure. But the training process of a BPNN is slower and could fall into a local minimum easily because of the unrestricted nonlinear optimization process of the BPNN [3, 4]. So here we use the particle swarm optimization algorithm in combination with neural network to avoid the convergence towards nearer local minima and to obtain accurate results.

The key factor of the wind power generation forecasting is prediction of the wind speed, because there is a huge relationship between the wind speed and the wind power generation of a wind turbine. Many study works have been done in the area of the short-term wind speed prediction based on the wind speed measurement data. So the most significant factor influencing wind power generation is the local wind speed [7].

This paper consists of two models; the first one is PSO-neural network which is used to predict the wind speed and second one is back propagation neural network which is used to predict the future power generation of a plant. In PSO-neural hybrid system, PSO algorithm is used to update the weights and bias of the neural network. Due to this combination of PSO-neural network the very short term wind power prediction will be more accurate compared to other systems [5]. In other cases there is a possibility to calculate the power mathematically using standard power curves of the wind turbine.

2 PSO-Neural Network

The PSO-neural network system is a cascaded model that uses particle swarm optimization algorithm to find optimal solution for a fitness function, which will be given as an input to the first neural network and another neural network is used to predict the future power generation from the predicted values of first neural network. The weights and bias are optimized using PSO and given to neural network depending upon the error equation i.e. fitness function.

2.1 Artificial Neural Network

Artificial Neural Networks (ANNs) are non-linear mapping architectures based on the neurons structure in the human brain. ANN is powerful tools for prediction, especially when the data relationships are unknown. ANNs can learn correlated patterns from input data sets and target values. After training, ANNs can be used to predict the outcome of new independent input data. ANN mostly deals with data which are vague and noisy. Thus they are ideally suited for the modeling of wind power data prediction which are known to be complex and often non-linear.

A neural network can perform pattern matching task that has a large number of highly interconnected processing elements (nodes). The interconnection of neurons gives the ability to learn and generalize the training patterns. A strong learning capability is the major advantage of neural network in prediction system [10].

Neural network architecture consists of number of layers, number of neurons in each layer and activation functions etc.

Number of layers. A three layers ANN structure is used in the prediction system. These are input layer, hidden layer and output layer respectively. The hidden layer is used to give the ability to generalize.

Number of neurons in each layer. The number of neurons in the input layer is as same as the number of inputs to the network. The number of neurons used in the hidden layer is greater than the number of neurons in the input layer to give better convergence and it can be calculated as a square of multiplication of number of neurons in input layer and number of neurons in output layer. The number of neurons in the output layer is as same as the number of output parameters of the network.

Activation function. The Tan-sigmoid transfer function is used in the both hidden layer and the output layer of neural network due to their nonlinearity and constantly differentiable properties. This activation function gives output in the range of 0 to 1 as the neuron's net input goes from negative to positive values.

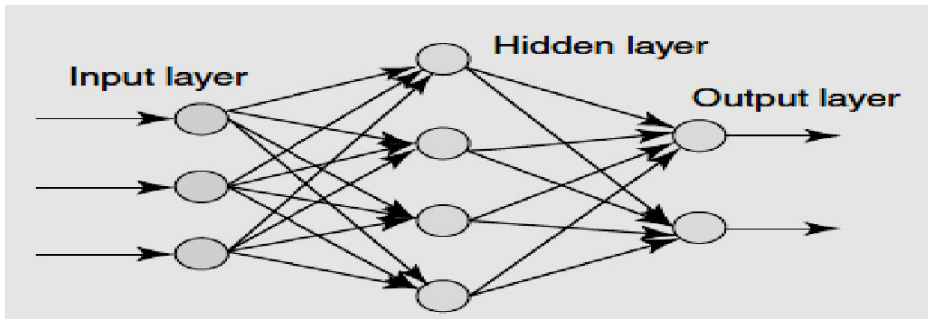


Fig. 1. General block diagram of artificial neural network model

The error measure (performance index) is usually defined by the sum of the squared difference between actual and desired outputs.

2.2 Particle Swarm Optimization

The PSO algorithm is an adaptive algorithm based on the food searching pattern of birds in the sky.

The PSO algorithm basically has two parameters for each particle. These are Velocity and Position components. During each generation, every particle is accelerated towards the previous best position and the global best position of the particles. The next step is updating of new velocity values using its past velocity, the distance from its previous best position, and distance from the global best position. The new velocity value is then used to calculate the next position of the particle in the search space.

This process is repeated until a particular number of iterations or a minimum error is achieved. Furthermore, the space in which the particles move is heterogeneous with

respect to fitness, which is some regions are better than others. A number of particles can be evaluated and there is presumed to be some kind of preference or attraction for better regions of the search space.

The fitness function for a particle swarm optimization algorithm is given as:

$$f(x) = \frac{1}{n} [\sum_{i=1}^n |O(i) - T(i)|^2]. \quad (1)$$

$$Fitness = \frac{1}{f(x)}. \quad (2)$$

Here $O(i)$ is the output and $T(i)$ is the target supplied to the neural network, and their difference is the error value and n is the number of iterations. The solution with minimum error will give maximum fitness and vice versa.

3 Short Term Wind Power Forecasting

The wind power forecasting process consists of two separate phases, the first one for predicting the wind speed and second for predicting the wind power. By the two cascaded prediction system, the total power (i.e.) predicted wind power is most accurate for very short term wind power forecasting process compared to other single prediction models.

3.1 Wind Speed Prediction

In this first step, the future wind speed is predicted from a past wind speed data using a PSO-neural network by training it to get the most accurate result. A past wind speed data series collected from a real wind farm is given as an input and target data to this network. The input data values may be used in the range of 6 to 18 hours and the target data can be 1 to 6 hours for better result. The training data for the network is taken consists of values of wind speed at a regular interval about 10 minutes. This step makes a suitable training data for the network that gives very less RMS error and the network is now ready to forecast the wind speed of the future. For this, another set of measured data is given as an input and the system is simulated. Now the RMS error is calculated as a difference between the actual measured output data and the simulated output data.

This predicted wind speed is used as an input for the next step of the work. By this separate prediction of wind speed and power generation, the operators can easily get both the predicted values of wind speed and power generation instead directly predicting the wind power from wind speed.

3.2 Wind Power Prediction

The second step of this work is to predict the wind power generation from the predicted wind speed using another neural network. This network may have different learning functions and number of inputs and outputs from the first one because each system handles different patterns. So this network is designed to give minimum root mean square (RMS) error with respect to its pattern of inputs given to it.

For this power forecasting network the training data are: predicted wind speed, hub height and power generation. The power generation is the target and other two are the inputs for the training process. The relation between the wind speed and the power generated is similar as in the manufactures power curve, but with varying height of the mast the output power is also varied proportionally, so that is major necessity of this second stage of prediction. From this two-step prediction system the operators can get both the predicted wind speed and wind power, but if we used single stage prediction system the predicted wind speed can't able to get. The key factor to use second neural network for this stage is variation of hub height from sea level. So this network is trained with a various values of hub height and from that it gives the exact wind power value in simulation period. This system has equal number of inputs and outputs i.e. for every input of wind speed a respective output power value is available. After training process the predicted wind speed is given to this network and is simulated to get a power generation value of the turbine. The relationship between wind speed and the respective wind power generation value is predicted using a manufacturer power curve.

4 Results

The two models are trained with a real wind farm data from the Western North Carolina Renewable Energy Initiative research lab located near the Pinnacle Inn Beech Mountain and turkey ridge.

Here we used 500 samples of real wind speed data taken at interval of 10 minutes and the first 400 samples are used in training process and the remaining 100 is used for the validation or checking process of the system.

The RMS error of the system during training period is shown in the figure 2. From that we can notice that the RMS error is gradually reduced and at 150 iterations it reaches a minimum value of less than 0.31 in the training period of the first PSO-neural network. The simulated PSO-neural network predicted output and measured wind speed is shown in figure 3. The second neural network is trained with a standard power curve. After training process the standard power curve and predicted power curves are plotted as shown in figure 4. The second neural network is now simulated with an input of forecasted wind speed and it predicts the respective wind power generation output of a system as shown in figure 5. It shows that these two systems are performed well and gives reduced RMS error compare to other methods of prediction like standard back propagation neural network and genetic algorithm with neural network.

This simulation is done in a MATLAB environment and the results are given in Table 1. It includes the power generation prediction errors of neural network, genetic neural network and PSO-neural network system. From Table 1, we can know that the minimum Root Mean Square Error (RMSE) of proposed system was perceptibly less compared to genetic neural network and standard neural network. So the two stages of neural network model give most accurate results to predict wind speed and wind power generation of a wind turbine.

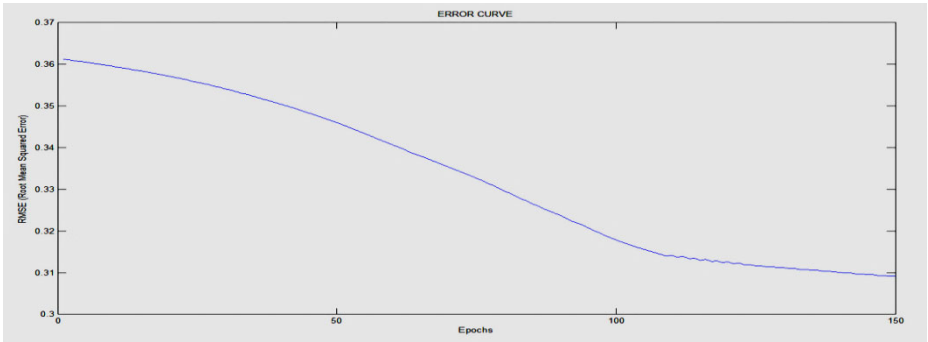


Fig. 2. RMS error during the training period

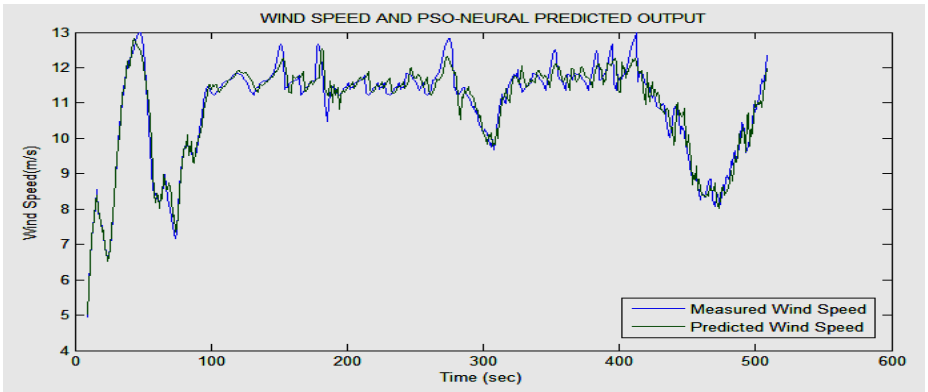


Fig. 3. Measured and Forecasted wind speed outputs.

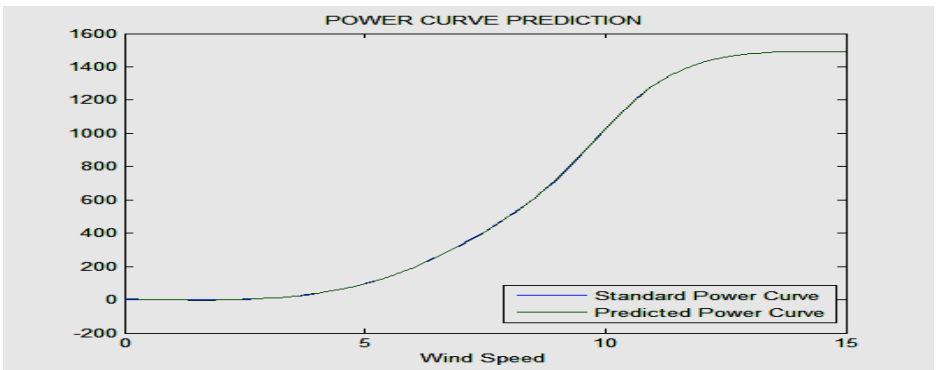


Fig. 4. Standard and predicted power curves

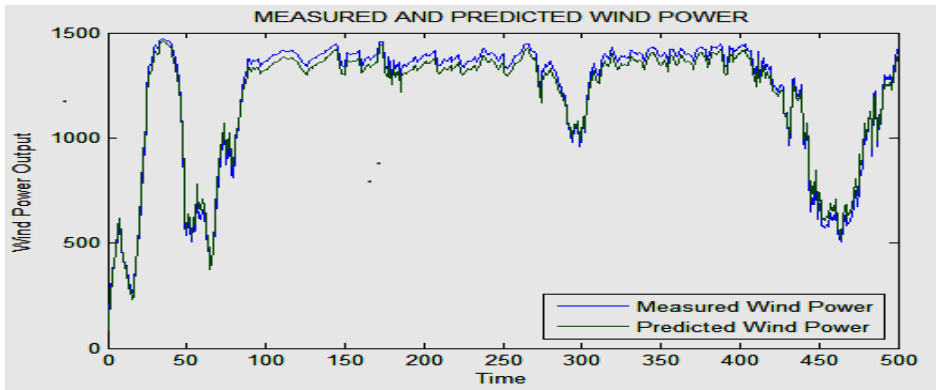


Fig. 5. Measured and Forecasted wind power outputs.

Table 1. Prediction Errors of Wind Speed for Various Methods

METHODS	RMSE
Standard neural network with BP	0.384
Genetic neural network	0.298
PSO neural network	0.162

5 Conclusion

This paper presented is an overview of the use of modern forecasting techniques for predicting wind power generation and wind speed by combination of particle swarm optimization technique and neural network models. We considered the areas of application of neural based methods [5], genetic neural network [3] and PSO-neural network, which offers a promising approach to building very short-term wind power prediction models. By combining the two systems, we can obtain a powerful tool for precise prediction of short term wind power generation of wind turbines. But this paper has not covered the factors like unit commitment and scheduling problems, temperature, humidity, density, etc., which will also affect the prediction results quietly, so in further research, these factors will be considered to give enhanced results. However it will take a lot of time to build and train the network.

Acknowledgment. The author would like to thank The Western North Carolina Renewable Energy Initiative research lab-Raleigh, USA, for their provision of time series data for the author’s research on wind power prediction system. Specific thanks go to Dr. S N Deepa, M.E, PhD, Head of the Department, Anna University of Technology, Coimbatore, for their personal assistance in this research.

References

1. World Wind Energy Report 2008, from World Wind Energy Association (2008)
2. Ernst, B., Oakleaf, B., Ahlstrom, M.L.: Predicting the Wind. *IEEE Power & Energy Magazine* (2007)
3. Xin, W., Liu, Y., Li, X.: Short-Term Forecasting of Wind Turbine Power Generation based on Genetic Neural Network. In: 8th World Congress on Intelligent Control and Automation, July 6-9. North China Electric Power University, China (2010)
4. Zack, J.: Overview of Wind Energy Generation Forecasting. True Wind Solutions. LLC and AWS Scientific, Inc. (December 17, 2003)
5. Moghaddas-Tafreshi, S.M., Panahi, D.: One-Hour-ahead Forecasting of Wind Turbine power Generation using Artificial Neural Networks.: Energy Management Research Laboratory. K.N.Toosi University of Technology, Tehran (2007)
6. Abraham, A., Guo, H.: Swarm Intelligence: Foundations, Perspectives and Applications. In: Nedjah, N., Mourelle, L. (eds.) *Swarm Intelligent Systems. Studies in Computational Intelligence*, vol. ch.1, pp. 3–25. Springer, Heidelberg (2006)
7. Abraham, A., Liu, H., Zhang, W.: Scheduling Jobs on Computational Grids Using Fuzzy Particle Swarm Algorithm. In: Gabrys, B., Howlett, R.J., Jain, L.C. (eds.) *KES 2006. LNCS (LNAI)*, vol. 4252, pp. 500–507. Springer, Heidelberg (2006)
8. Sierra, M.R., Coello, C.A.C.: Multiobjective particle swarm optimizers: A survey of the state-of-the-art. *Int. J. Comput. Intell. Res.* 2(3) (2006)
9. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: *Proc. IEEE Int. Conf. Neural Networks*, Perth, Australia, pp. 1942–1948 (1995)
10. Pinson, P., Siebert, N., Kariniotakis, G.: Forecasting of Regional Wind Generation by a Dynamic Fuzzy-Neural Networks based Up-scaling Approach. In: *European Wind Energy Conference & Exhibition (EWEC)*, Spain (2003)
11. Negnevitsky, M., Potter, C.W.: Innovative Short-Term Wind Generation Prediction Techniques. In: *IEEE Conference. University of Tasmania, Australia* (2006)
12. Potter, C.W., Negnevitsky, M.: Very Short-Term Wind Forecasting for Tasmanian Power Generation. *IEEE Transactions On Power Systems* 21 (May 2006)
13. Methaprayoon, K., Yingvivananpong, C., Lee, W.-J., Liao, J.R.: An Integration of ANN Wind Power Estimation Into Unit Commitment Considering the Forecasting Uncertainty. *IEEE Transactions On Industry Applications* 43(6)(November/December 2007)
14. Mishra, A.K., Ramesh, L.: Application of Neural Networks in Wind Power (Generation) Prediction. Dr. MGR University, Chennai
15. Hong, L.Y., Jiang, D., Huang, Q., Ding, Y.: Wind Speed Forecasting Using Fully Recurrent Neural Network in Wind Power Plants. Tsinghua University, Beijing (2010)
16. Li, S., Wunsch, D.C., O’Hair, E.A., Giesselmann, M.G.: Using Neural Networks to Estimate Wind Turbine Power Generation. *IEEE Transactions On Energy Conversion* 16 (September 2001)
17. Bracale, G.C., Mangoni, M., Proto, D.: Wind Power Forecast Methods and Very Short-term Steady-State Analysis of an Electrical Distribution System. In: *Electrical Engineering Research Report*, vol.1. University of Naples Parthenope (2009)
18. Hashimoto, A., Hattori, Y., Kadokura, S.: The forecasting system for wind power generation combined use of numerical and statistical models. Central Research Institute of Electric Power Industry, Japan (2008)
19. Borg, R., Rothkrantz, L.J.M.: Short-Term Wind Power Prediction with Radial Basis Networks. Delft University of Technology, the Netherlands (2008)

20. Ernst, B., Wan, Y.-H., Kirby, B.: Short-Term Power Fluctuation of Wind Turbines. Looking at Data from the German 250 mw Measurement Program from the Ancillary Services Viewpoint, National Renewable Energy Laboratory (2000)
21. Fitzwater, L.M., Winterstein, S.R.: Predicting Design Wind Turbine Loads from Limited Data: comparing Random Process and Random Peak Models. The American Institute of Aeronautics and Astronautics Inc. and the American Society of Mechanical Engineers (2001)
22. Cali, U., Lange, B., Dobschinski, J., Kurt, M., Moehrlen, C., Ernst, B.: Artificial neural Network based wind power forecasting using a multi-model approach, Germany (2009)
23. Sreelakshmi, K., Ramakanthkumar, P.: Neural Networks for Short Term Wind Speed Prediction. World Academy of Science, Engineering and Technology (2008)
24. Stoutenburg, E.D., Jenkins, N., Jacobson, M.Z.: Power output variations of co-located offshore wind turbines and wave energy converters in California. Institute of Energy, Cardiff University, UK (2010)

A Class of Recurrent Neural Network (RNN) Architectures with SOM for Estimating MIMO Channels

Kandarpa Kumar Sarma and Abhijit Mitra

Dept. of Electronics and Electrical Engineering (EEE)
Indian Institute of Technology Guwahati, Guwahati-781039, Assam, India
{s.kandarpa,a.mitra}@iitg.ernet.in

Abstract. Artificial Neural Network (ANN)s can handle Multi Input Multi Output (MIMO) channel prediction and estimation. With suitable modification, ANNs also tackle time - varying properties of the wireless links. But these turn out to be cumbersome to configure and train for which alternative ANN architectures are required for such applications. The immediate option that emerges is the Recurrent Neural Network (RNN) which has the capacity to deal with time - dependent inputs. But a problem is observed with respect to the approach in which RNNs are trained to deal with signals with real and imaginary components. Signals with bifurcated real and complex components help the RNN to learn better. But for tightly coupled transmissions which exists most of the times in wireless channels, the performance of such RNNs suffer. The present work attempts to reduce this tradeoff and adopts a split - activation RNN training approach with exclusive blocks for in-phase and quadrature components. The responses of such blocks, are combined and optimized with Self Organizing Map (SOM). The results show better performance and ease of implementation than MLPs with temporal characteristics.

1 Introduction

Artificial Neural Network (ANN)s like the Multi Layer Perceptron (MLP)s are able to deal with Multi Input Multi Output (MIMO) channel prediction and estimation [1] [2] with certain complexities while tackling time - varying nature of the MIMO - channels. As a result temporal characteristics are required to be incorporated into the MLP using ARMA synaptic weight links [1]. The resulting configuration turns out be cumbersome to configure and train though the performance levels are satisfactory. Hence, alternate and innovative ANN architectures are required for such applications. The immediate option that emerges is the Recurrent Neural Network (RNN) which has the capacity to deal with time varying inputs due to the presence of atleast one feed-backward loop [3]. The feedback in a RNN in singular or multiple delayed forms from the output to the hidden layer and then to the input layer, store information for one step at a time. This process continues and learning cumulates over the given time cycles.

Moreover, the RNN architecture, by using the state vector relates the present and the past context and transitions in-between the layers to describe the future response of the system thereby dynamically learning temporal behaviour of input samples [3].

This work is related to the formulation of a RNN based architecture for efficient MIMO channel estimation. Signals encountered in wireless communication demonstrate variations in terms of their complex valued nature. One way to ensure that ANNs learn magnitude and phase components separately is to use split - complex activation functions. The practice is to provide real and imaginary components in splitted form to RNN blocks and allow them to carry on the training as mutually exclusive events [4]. However for cases where real and imaginary components are tightly coupled the split - complex activation approach yields poor performance. Therefore channel estimation set-ups with splitted RNNs trained to handle real and complex components separately shall demonstrate lower performance [4]. Two contrasting pictures therefore emerge. This trade off between the learning functions of RNNs requiring real and imaginary components separately and the necessity of better performance for signals with tightly coupled in-phase and quadrature components needs to be sorted out. An attempt is made here to formulate a an RNN structure designed to reduce this trade off. Split - complex activation is adopted to allow the RNNs to learn the real and complex parts separately and the output of these exclusive blocks are optimized with Self Organizing Map (SOM)s. By allowing the RNN blocks to train with real and imaginary components as exclusive events, the error surfaces are aligned appropriately to form specific classification boundaries. It helps in segregation of inputs from real and imaginary segments and their placement as per nearest classification decision. This at times though is undesirable but reduces computational complexity. The output from the real and imaginary blocks are combined and optimized using the SOM which finds the best outcome using competitive learning and a Euclidean distance norm. Some of the relevant works are [2] to [10].

2 Basic Considerations of RNN

RNNs are ANNs with one or more feedback loops. The feedback can be of a local or a global kind. The RNN maybe considered to be an MLP having a local or global feedback in a variety of forms [3]. The architecture of a generic RNN that follows naturally from a MLP is shown in figure 1. The present value of the model input is denoted by $u(n)$, and the corresponding value of the model output is denoted by $y(n + 1)$; that is, the output is ahead of the input by one time unit [3]. The dynamic behavior of the the RNN model is described by

$$y(n + 1) = F(y(n), \dots, y(n - q + 1), u(n), \dots, u(n - q + 1)) \quad (1)$$

where F is a nonlinear function of its arguments. The basic RNN training algorithms can be summarized as below:

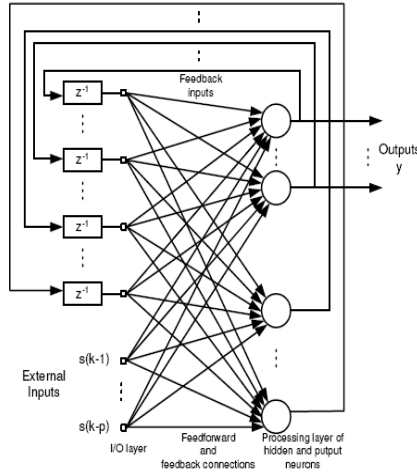


Fig. 1. Time Delay Fully connected RNN (TDFRNN)

1. **Back Propagation Through Time:** The Back Propagation Through Time (BPTT) algorithm for training of a RNN is an extension of the standard back-propagation algorithm. It may be derived by unfolding the temporal aspect of the structure into a layered feedforward network, the topology of which grows by one layer at every time step [3].
2. **Real-Time Recurrent Learning:** Another learning method referred to as *Real Time Recurrent Learning* (RTRL), derives its name from the fact that adjustments are made to the synaptic weights of a fully connected RNN in real time. The RTRL algorithm can be summarized as follows [3]:

Parameters:

m = dimensionality of input space, q = dimensionality of state space, p = dimensionality of output space and w_j = synaptic weight vector of neuron.

Initialization:

- (a) Set the synaptic weights of the algorithm to small values selected from a uniform distribution.
- (b) Set the initial value of the state vector $x(0) = 0$.
- (c) Let $A_j(n)$ be a $q - by - (q + m + 1)$ matrix defined as the partial derivative of the state vector $x(n)$ with respect to the weight vector w_j . Set $A_j(0) = 0$ for $j = 1, 2, \dots, q$.

Computations: Compute for $n = 0, 1, 2, \dots$

$$\begin{aligned}
 A_j(n + 1) &= \Phi(n)[W_a(n)A_j(n) + U_j(n)] \\
 e(n) &= d(n) - Cx(n) \\
 \Delta w_j(n) &= \eta C A_j(n) e(n)
 \end{aligned}$$

3. **Decoupled Extended Kalman Filtering (DEKF) Algorithm and RNN Learning:** Despite its relative simplicity (as compared to RTBP), it has been demonstrated that one of the difficulties of using direct gradient descent algorithms for training RNNs is the problem of vanishing gradient.

The solution to this problem has been derived by considering the training of RNN to be an optimum filtering problem similar to that shown by the Kalman filter such that each updated estimate of the state is computed from the previous estimate and the data currently available. This algorithm is derived from the Extended Kalman Filter (EKF) algorithm which is a state estimation technique for nonlinear systems [3], [7]. EKF is a second-order gradient descent algorithm, in that it uses curvature information of the (squared) error surface. By configuring the network with decoupled sub-networks, results in considerable reduction in computations as proved by Feldkamp et al. (1998) which gives rise to the Decoupled Extended Kalman Filter (DEKF) algorithm [3] [7].

4. **Complex - Valued RTRL Algorithm:** Figure 1 shows a TDFRNN, which consists of N neurons with p external inputs. The network has two distinct layers consisting of the external inputs. The network has two distinct layers consisting of the external input-feedback layer and a layer of processing elements. Let $y_l(k)$ denote the complex valued output of each neuron, $l = 1, \dots, N$ at time index k and $s(k)$ the $(1 \times p)$ external complex-valued input vector. The overall input to the network $I(k)$ represents the concatenation of vectors $y(k)$, $s(k)$ and the bias input $(1 + j)$. A complex valued weight matrix of the network is denoted by W , where for the i^{th} neuron, its weights form a $(p + F + 1) \times 1$ dimensional weight vector $W_i = [w_{i,1}, \dots, w_{i,p+F+1}]^T$ where F is the number of feedback connections. The feedback connections represent the delayed output signals of the TDFRNN. The output of each neuron can be expressed as

$$y_l(k) = \phi(\text{net}_l(k)), l = 1, \dots, N. \tag{2}$$

where

$$\text{net}_l(k) = \sum_{n=1}^{p+N+1} w_{l,n}(k)I_n(k) \tag{3}$$

is the net input to l^{th} node at time index k . For simplicity, we state that

$$y_l(k) = \phi^r(\text{net}_l(k)) + j\phi^i(\text{net}_l(k)) = u_l(k) + jv_l(k) \tag{4}$$

$$\text{net}_l(k) = \sigma_l(k) + j\tau_l(k) \tag{5}$$

where ϕ is a complex nonlinear activation function [8]. The output error consists of its real and imaginary parts and is defined as

$$e_l(k) = d(k) - y_l(k) = e_l^r(k) + je_l^i(k) \tag{6}$$

$$e_l^r(k) = d^r(k) - u_l(k), e_l^i(k) = d^i(k) - v_l(k), \tag{7}$$

where $d(k)$ is the reference signal [8].

3 System Model

Statistical and MLP based methods of modeling and estimation of MIMO channels are useful and proved their worth. But with increase in transmission rates and system complexity, the possibility always exists to explore other means for such applications specially to configure ANNs with better temporal and optimization capacity so that they are able to adapt to a given environment in a suitable manner [6]. Moreover, ANNs for their capacity to learn can exploit transmitter side information (TSI), channel side information (CSI) and receiver side information (RSI) better than the statical techniques [6] and specially configured RNN architectures can exploit such information for making reception better in a wireless medium. This advantage can be adopted to simply reception and receiver design despite using higher data rate techniques like MIMO - OFDM. The probable advantages of such schemes are

- Application of the complete learning ability of the ANN which will help in controlling precision.
- No pilot symbol bits are required in the MIMO - OFDM transmission thereby preserving bandwidth.
- Combined channel estimation and symbol recovery in MIMO-OFDM transmission and
- Exploiting the robustness of the ANN to time - dependent and frequency - dependent variations of the channel to improve BER rates in high data rate and wide-band applications.

3.1 Complex Time Delay Fully Recurrent Neural Network (CTDFRNN) with Self Organizing Map (SOM) Optimization

Here split - complex activation is adopted to allow the TDFRNNs to learn the real and imaginary parts separately (Figure 2). In this form the TDFRNN is called a Complex TDFRNN (CTDFRNN). As a result the weight update process of the RNNs obtain the opportunity to learn in-phase and quadrature signal components separately and adapt as per requirement. But since for cases where complex parts are tightly coupled, the output of these exclusive blocks are combined and optimized with Self Organizing Map (SOM)s.

Let $u_{Ri}[n]$ be a sequence of real components of the input which is a MIMO-OFDM signal. Also let $x[n]$ be the response of the output layer at time n to be fed-back as state vectors to the input layer. The expression of the input to the hidden layer is given as

$$y_{R1}[n] = \sum_j f_j(u_{Ri}[n]w_{1i}[n] + x[n-1]w_{01} + x[n-2]w_{02}) \quad (8)$$

where $f(\cdot)$ is the activation function and $w[\cdot]$ are synaptic links. The expression for the output of the hidden layer is given as

$$y_{R2}[n] = \sum_j f_j(y_{1i}[n]w_{2i}[n]). \quad (9)$$

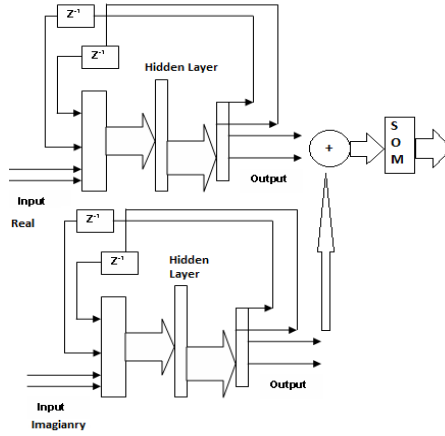


Fig. 2. CTDFRNN Estimator with SOM Optimization

The expression for the response of the output layer is given as

$$y_{R0}[n] = \sum_k f_k(y_{R2k}[n]w_{3k}[n]). \tag{10}$$

A similar expression for the quadrature component can be obtained. The expression for the output of the hidden layer is given as

$$y_{Q0}[n] = \sum_k f_k(y_{Q2k}[n]w_{3k}[n]). \tag{11}$$

Let d_r and d_q be the desired outputs for the RNN modules respectively. The error signals can be obtained as

$$e_R[n] = \frac{1}{2} \sum_{k=1}^N (y_{R0k}[n] - d_{rk}[n])^2 \tag{12}$$

$$e_Q[n] = \frac{1}{2} \sum_{k=1}^N (y_{Q0k}[n] - d_{qk}[n])^2 \tag{13}$$

The minimization of the error signals is carried out by four different training methods as explained in Section 3.2. The SOM block is used to carry out an optimization of the outputs generated by the RNN modules in a given time of say n seconds. Let the input to the SOM block be a combined signal $y_0[n]$ formed by $y_{R0}[n]$ and $y_{Q0}[n]$. The output of the SOM block is given by

$$y_j = y_{0j}^T W_j \quad j = 1, \dots, m \tag{14}$$

where W_j is the random weights of the competitive layer of the SOM. The ‘winners take all’ competition starts in this layer such that the winning neuron index, J , satisfies

$$y_J = \max_j \{y_{0j}^T, W_j\}. \tag{15}$$

3.2 Competitive Learning of the SOM Optimization Block

Competitive learning requires that the weight vector of the winning neuron be made to correlate with the input vector [3]. This is done by variation of only the winning weight vector $W_J = (w_{1J}, \dots, w_{nJ})^T$ to approach the input vector. A scalar form of this learning law is presented below:

$$w_{iJ}^{k+1} = w_{iJ}^k + \eta x_i^k \quad i = 1, \dots, n. \quad (16)$$

where η is a eigen vector in a neighbourhood of weight w. From the above, the following standard competitive form in discrete time is obtained:

$$w_{ij}^{k+1} = w_{ij}^k + \eta s_j^k (x_i^k - w_{ij}^k) \quad i = 1, \dots, n \quad j = 1, \dots, m \quad (17)$$

where $s_j^k = 1$ only for $j=J$ and is zero otherwise, for a hard competitive field.

Table 1. Average MSE convergence after 1000 epochs of training for the RNN- architecture trained with four different methods

Architecture	BPTT	RTRL	DEKF	CRTRL
CTDFRNN	6×10^{-5}	0.5×10^{-5}	0.35×10^{-5}	0.35×10^{-5}

4 Experimental Results and Discussion

The sample size considered for training includes two different forms of Clarke-Gans channel model each generated using three different AWGN values viz.-3dB, 1dB and 3dB for the generic, Rayleigh and Rician faded channels. This way several sample sets with different SNRs are obtained. Orthogonal Frequency Division Multiplexing (OFDM) signals are included with the MIMO to verify whether the system is able to perform symbol recovery. The samples are accumulated for a 4 x 4 MIMO - OFDM set-up for each of the channel types considered. The testing includes a range of signal conditions with SNR values ranging from -10 to 10 dB. The testing carried out with inputs from the receiver side calculates channel coefficients and compares them to the theoretically generated values for a frequency range of 0.2 Ghz to 8 GHz. The number of training epochs are limited to a few thousand sessions only to minimize the chances of overtraining. But the process is repeated with several sets of samples. The performance achieved during these epochs are noted. The ANN training considers the Mean Square Error (MSE) convergence and precision generated in channel estimation and the Bit Error Rate (BER) calculation. If the MSE has converged to the fixed target value, the precision levels and the associated BER values are calculated. If the values fall within the desired levels, the training is extended to include more number of samples which represent varying channel conditions. The iterations within a few thousands provide a minimum MSE convergence of 0.5×10^{-6} which provides improved estimation of channel coefficients and lower BER values.

Table 2. Training time convergence parameters for LS, MMSE, MLP, temporal-MLP and RNN architecture

Sl Num	Method	Epochs	Time (S)	Average Precision (%)
1	LS	251	19.2	91
	MMSE	289	18.9	92
2	MLP	326	16.7	94
	ARMA-MLP	225	13.4	94
	LAPC	295	10.1	95.6
3	CTDFRNN	218	10.2	96.2

Table 3. Configuration of the SOM Optimizer

Sl Num	Parameter	Specification
1	Input Size	10 x 20
2	Channel length	20
3	Multipaths	10
4	Output grid	1 x 20
5	Topology	<i>Gridtop</i> and <i>Hextop</i>
6	Output grid	1 x 20

Table 1 shows average MSE convergence after 1000 epochs of training for each of the RNN- architectures trained with four different methods. Among the four different methods DEKF and CRTRL algorithms are found to be most suitable. The DEKF algorithm is better suited for RNN training for this application due to the fact that it is easy to implement than CRTRL. Hence, DEKF algorithm is adopted as the training method for the RNN-architecture. Table 2 provides a comparative depiction of the average training time performance of the CTDFRNN architecture with respect to LS, MMSE, MLP and temporal-MLP configurations [1]. In terms of number of epochs required, the RNN blocks show significant improvement. There are marked rise of precision levels as well. The CTDFRNN architecture generates the best performance in terms of time required to reach an MSE value while generating an average success rate of around 96.2 % which is note-worthy in comparison to LS, MMSE, MLP and temporal-MLP techniques. This improvement is also reflected in the BER values generated by taking SNR values between -10 and 10 dB. The composite plot is depicted in Figure 3. The improvement in BER values as shown by the RNN architectures reflect their capacity to deal with time varying signals. Thus these are found to be suitable for MIMO channel estimation and OFDM symbol recovery. The role played by the SOM blocks in CTDFRNN is paramount due to the fact that the result generated is an optimized one. For a given window of N-sec.s several sets of output are generated. The optimization process carried out by following the competitive learning algorithm selects the best set of result at the end of about 50 iterations. After 50 iterations the result is found to generate the best approximation using which the BER values are generated. Initially all the neurons in the

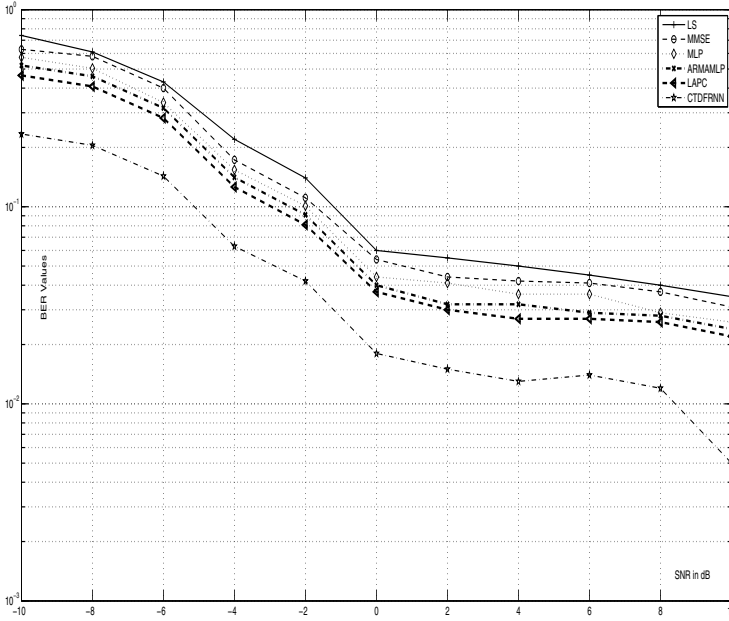


Fig. 3. Comparative BER plot generated by LS, MMSE, MLP, ARMA-MLP, LAPC, CTDFRNN and CTDFRNN-SOM

SOM block are given some coordinates and the connecting wights some random values. As the training continues, the updating process of the weights continue upto, on an average, 50 sessions after which the generated value is taken to be the result of the RNN- estimator. For a 20 tap channel filter, a training session of the SOM consists of 50 filter coefficient vectors from which the most suitable set is extracted at the end of around 55 to 60 sessions of training of the SOM. The SOM optimizer is constituted by following the considerations as provided in Table 3.

5 Conclusion

The inability of the conventional MLP to deal with time-varying characteristics of the MIMO-channel can be rectified to a large extent by incorporating temporal features to the MLP which provides encouraging result at the cost of increase in implementation complexity. The RNN is a configuration immediately available for such application due to its excellent ability to deal with time-varying signals. The basic RNN model however needs to be modified to be able to deal time varying signals with in-phase and quadrature components. This work provided some insight towards that direction. It also showed how RNN in split activation configuration can be used for MIMO channel estimation with optimization carried out by a SOM. There are four methods for RNN training out of which

the DEKF algorithm is found to be the most suitable for RNN configuration implemented in this work. The work can be further extended to explore the possibility of incorporating heterogeneous elements the presence of which can make the learning of the RNNs better. Another possibility is to consider fuzzy - neural hybrid architecture as a handy tool to make the precision levels even better and lower BER values even further.

References

1. Sarma, K.K., Mitra, A.: MIMO Channel Modeling using Temporal Artificial Neural Network (ANN) Architectures. In: Proceedings of the 1st ACM IITM, IIIT Allahabad, India, pp. 37–44 (December 2010)
2. Ling, Z., Xianda, Z.: MIMO Channel Estimation and Equalization Using Three-Layer Neural Networks with Feedback. *Tsinghua Science and Technology* 12(6), 658–661 (2007)
3. Haykin, S.: *Neural Networks A Comprehensive Foundation*, 2nd edn. Pearson Education, New Delhi (2003)
4. Kim, T., Adali, T.: Fully complex backpropagation for constant envelope signal processing. In: Proceedings of the 2000 IEEE Signal Processing Society Workshop, USA, vol. 1, pp. 231–240 (December 2000)
5. Ozcelik, H., Czink, N., Bonek, E.: What Makes a Good MIMO Channel Model? In: Institut für Nachrichtentechnik und Hochfrequenztechnik, Technische Universität Wien, Vienna, Austria, VTC
6. Jiang, M., Hanzo, L.: Multiuser MIMO-OFDM for Next-Generation Wireless Systems. *Proceedings of the IEEE* 95(7), 1430–1469 (2007)
7. Jaeger, H.: A tutorial on training recurrent neural networks, covering BPPT, RTRL, EKF and the “echo state network” approach. German National Research Center for Information Technology (March 2005)
8. Goh, S.L., Mandic, D.P.: A Complex-Valued RTRL Algorithm for Recurrent Neural Networks. *Neural Computation* 16, 2699–2713 (2004)
9. Potter, C.G.: Multiple-input multiple-output wireless communications with imperfect channel knowledge. Missouri University of Science and Technology, USA (2008)
10. Potter, C.G., Venayagamoorthy, G.K., Kosbar, K.: RNN based MIMO channel prediction. *Elsevier Journal of Signal Processing* 90(2), 440–450 (2010)

An Efficient Framework for Prediction in Healthcare Data Using Soft Computing Techniques

Veena H. Bhat^{1,2}, Prasanth G. Rao³, S. Krishna¹, P. Deepa Shenoy¹,
K.R. Venugopal¹, and L.M. Patnaik⁴

¹Department of Computer Science and Engineering,
University Visvesvaraya College of Engineering, Bangalore, India

²IBS-Bangalore, Bangalore, India

³Research Scholar, Bangalore, India

⁴Vice Chancellor, Defense Institute of Advanced Technology, Pune, India
{veena.h.bhat, prasanthgrao, krishna.somandepalli}@gmail.com,
shenoypd@yahoo.com

Abstract. Healthcare organizations aim at deriving valuable insights employing data mining and soft computing techniques on the vast data stores that have been accumulated over the years. This data however, might consist of missing, incorrect and most of the time, incomplete instances that can have a detrimental effect on the predictive analytics of the healthcare data. Preprocessing of this data, specifically the imputation of missing values offers a challenge for reliable modeling. This work presents a novel preprocessing phase with missing value imputation for both numerical and categorical data. A hybrid combination of Classification and Regression Trees (CART) and Genetic Algorithms to impute missing continuous values and Self Organizing Feature Maps (SOFM) to impute categorical values is adapted in this work. Further, Artificial Neural Networks (ANN) is used to validate the improved accuracy of prediction after imputation. To evaluate this model, we use PIMA Indians Diabetes Data set (PIDD), and Mammographic Mass Data (MMD). The accuracy of the proposed model that emphasizes on a preprocessing phase is shown to be superior over the existing techniques. This approach is simple, easy to implement and practically reliable.

Keywords: Imputation, soft computing, categorical data, continuous data.

1 Introduction

Quality of the data under study plays a pivotal role in the accuracy of modeling using data mining techniques and predictive analytics. The objective of data mining is to find patterns in the data that gives accurate outcomes with newer, unseen cases. In healthcare databases, a huge number of recorded samples are used to analyze characteristic hidden patterns; the validity of the model is verified by comparing the outcome of the model for an unseen case and its corresponding medical expert opinion.

In the real world, data has a cost attached to it, especially in the field of medical diagnostics. The medical experiments conducted to record data are expensive, often

repetitive and invasive in nature, which sometimes might increase the threat on the patient's life, hence not repeatable. Therefore it is of utmost importance to have reliable and cost efficient methods to reuse data and give an opinion to the patient without much medical re-experimentation but using simple software that holds the simulated statistical model.

Cabena et al., [1] in their work estimate that about 20% of the efforts are spent on problem identification, about 60% on data preparation and data preprocessing while the rest about 20% for data mining and knowledge discovery; which definitely upholds the importance of data preprocessing. Missing values or missing data in any database, especially in a healthcare data set is often an indelible hurdle for efficient analyses. Such missing values can occur due to non response or a simple missed entry. This reduces the representativeness of the sample thus distorting the inferences drawn. The presence of missing values at rates less than 1% are generally considered trivial, 1-5% manageable, 5-15% require sophisticated methods to handle the missing values and more than 15% severely affect any kind of interpretation [2].

Soft computing techniques aim at improving quality of data by reducing imprecision and uncertainty employing approximate reasoning and logic, in order to achieve tractability, robustness and low cost solutions. Popular methods include neural networks, swarm intelligence, rough sets, fuzzy logic and genetic algorithms.

This work proposes a novel modeling methodology, specifically for healthcare data which is chosen so as to contain continuous, categorical, numeric alongside textual data as variables in the several attributes considered. The data is chosen so as include instances of missing values which are imputed using separate approaches for continuous and categorical data. The proposed methodology for imputation is validated for prediction accuracy, using Artificial Neural Networks (ANN) as the classifier. Results show considerable improvement in the accuracy of the proposed model as against other techniques in vogue.

Section 2 outlines the related work. The motivation to carry out this work is explained in section 3. To make this paper self-contained we have described the data set details in section 4. The framework proposed for data preprocessing is explained in detail in section 5. Section 6 describes the implementation of the entire model – the preprocessing framework along with the prediction model. Results and performance analyses are described in section 7. Conclusions and future work is given in section 8.

2 Literature Survey

Reference [3] details the three rules any effective imputation model should follow – retaining the data distribution, the relationship between the attributes and cost-time efficiency. Some of the predominant methods for imputation include Mean Imputation (MI), Regression Imputation, C4.5 and k-Nearest Neighbor (k-NN) [1, 4, 5, 6, 14]. The PIMA data set has been explored by researchers, applying many data mining techniques. In the previous works adopting neural networks [7,8,9] and other classifiers [12, 13, 15] in the prediction modeling stage, the accuracy in predicting the diabetes status ranges from 66% to 82.29%. Works related with Mammographic Mass Data (MMD) have reported prediction accuracy upto 80.9% [10]. Most of the research works either work on complete instances or ignore the tuples with attribute

values set to zero [7, 8] while [9] considers data instances of PIDD (to predict continuous values) with an equal mix of patients who had diabetes onset and patients who were non-diabetic, but with attributes that contained missing/incorrect values.

In this work, the data sets selected have missing value range (incorrect value) of 4% to 48%. The proposed framework is aimed at achieving a cost effective imputation method to assess the accuracy of the imputed values at each iteration. A prediction model is then constructed which works on the corrected data, which now neither has missing values or incorrect values. Further the accuracy of the model is established and analyzed using various parameters as explained in section 7.

3 Motivation

Today, healthcare information systems amass a large volume of digital data in various forms, but the analyses of such data for preventive diagnosis; prediction of disease transmission and epidemic outbreaks is not frequent. Analyzing healthcare data for an efficient patient-care and tracking the patients' records that include critical parameters such as medicine intake or dosage may give a peek into the long term effects that could be caused by a particular diagnostic tool or treatment. An intelligent, robust and reliable prediction model could help in the identification of this kind of analyses and predicting long-term probable health issues. This could also help the doctors or other health coordinators who oversee the treatment of a certain patient to give quality diagnosis and help identify the changes in the lifestyle that may be necessary in the patient's routine to prevent short and long term health complexities caused by an illness. Though several models for such prediction and analyses are in vogue, a robust and reliable model that predicts with a good accuracy even when the attributes include missing and erroneous values is a challenge. In this work we propose a solution to such analyses by employing various soft computing techniques in preprocessing the data to impute both categorical and numerical missing values that further improves accuracy of the prediction.

4 Data Set Details

The data sets selected for building the predictive model are PIMA Indian Diabetes Data (PIDD) set and the Mammographic Mass Data (MMD) from the Machine Learning Database Repository at the University of California, Irvine. These data sets may be downloaded from [11].

As per previous studies, Pima Indians may be genetically predisposed to diabetes and it was noted that their diabetic rate was 19 times that of any typical town. The National Institute of Diabetes and Digestive and Kidney Diseases of the NIH originally owned the PIDD and was received by UC-Irvine Machine Learning Repository in 1990. The data set contains 768 patient data records and each record is described by 8 attributes, while the 9th attribute associates the class attribute that is a label indicating the onset of diabetes within 5 years. A '0' indicates the onset and '1' indicates non-occurrence of diabetes. This data constitutes of female patients of age 21 years and above belonging to the Pima Indian heritage. This population lives near Phoenix, Arizona, USA.

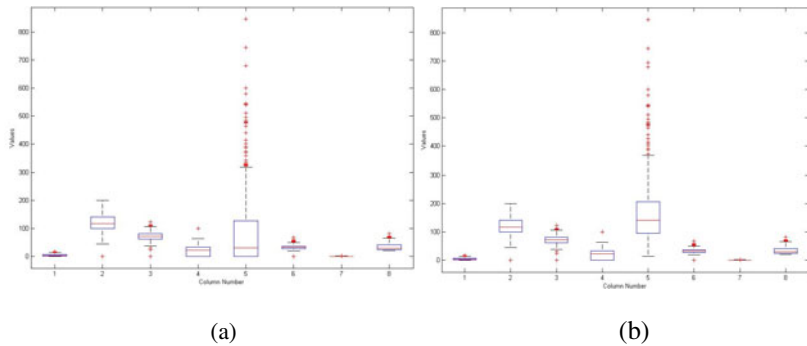


Fig. 1. Boxplots of the PIDD attribute values before and after soft computing preprocessing

The Mammographic Mass Data (MMD) was received by UC-Irvine Machine Learning Repository in 2007. This data set contains the Breast Imaging–Reporting and Data Systems (BI-RADS) assessment and the patient’s age related data for 916 instances that have been identified on full field digital mammograms collected at the Institute of Radiology of the University Erlangen-Nuremberg between 2003 and 2006. The data is categorical in nature with BI-RADS assessment ranging from 1 (definitely benign) to 5 (highly suggestive malignancy) assigned in a double-review process by physicians. The data has a total of 5 attributes with the 6th attribute being the severity (benign = 0, malignant = 1). The age of the patient is an integer (maximum value being 96 and minimum 18), the shape of the mass takes values ranging from 1 to 4 (where round=1 oval=2 lobular=3 irregular=4), the mass margin takes values from 1 to 5 (where circumscribed=1 microlobulated=2 obscured=3 ill-defined=4 spiculated=5) and mass density being ordinal, takes values from 1 to 4 (high=1 iso=2 low=3 fat-containing=4).

Data sets are usually explored either graphically or analytically. The quality of data being explored is an important aspect considering presence of missing and/or outlier values. These affect the pattern recognition and predictive analyses of the data being mined. Several methods are used to handle the noisy and missing data like replacing the incorrect values with ‘mean’ of the variables in the attribute, its ‘median’ or sometimes user defined values. PIDD is explored using boxplot as the visualization technique and is illustrated in Fig. 1(a). This boxplot indicates the quartiles, median besides the outlier values for each attribute.

The Pima Indian Diabetes Data (PIDD) set when examined manually, is observed to contain few incorrect values for specific attributes, as listed in Table 1. PIDD includes 500 non-diabetic patients (class = negative; ‘0’) and 268 diabetic patients (class = positive; ‘1’) giving an incidence rate of 34.9%. Thus if a random classifier is used for prediction for all instances as belonging to negative class (that is all instances being predicted negative) then accuracy of such a model would be 65.1% (or an error rate of 34.9%). Using this simple strategy of selecting tuples that contain complete and correct data values, we statistically represent the data selected, using the boxplot to aid analysis of the data.

Table 1. Attribute-value study outcome, indicating attributes with incorrectly present values for PIDD set

Sl. No.	Attribute Name	No. of 'Zero' value instances for attributes listed, out of 768 tuples.
1	Plasma glucose concentration, in 2 hours, in an oral glucose tolerance test	5
2	Diastolic blood pressure (mm Hg)	35
3	Triceps skin fold thickness (mm)	227
4	2-hour serum insulin (μ U/ml)	374
5	Body mass index (weight in kg/(height in m) ²)	11

Mammographic Mass Data (MMD), consists of 916 instances, that include 516 instances of benign (class = negative; '0') and 445 instances of malignant masses (class = positive; '1'), with an incidence rate of 48.58%. In this particular data set, the attribute that requires imputation is 'density' as this attribute shows the highest number of missing or erroneous value shown in Table 2.

Table 2. Attribute-value study outcome, indicating attributes with incorrectly present values for MMD

Sl.No.	Attribute Name	No. of '-1' value instances for attributes listed, out of 961 tuples.
1.	BI-RADS assessment	2
2.	Age	5
3.	Shape	31
4.	Margin	48
5.	Density	76

5 Proposed Data Preprocessing Framework

The data imputation framework provides a reliable solution to both continuous and categorical missing values. This makes the proposed model unique and self-contained. PIDD data is used to demonstrate the imputing of continuous values where as MMD set is identified to impute categorical values.

5.1 Proposed Approach to Impute Continuous Missing Values

A regression model, using Classification and Regression Trees (CART), is developed, using tuples that have complete and correct values for all its attributes. These tuples constitute the training set for the model. Further, the instances that would need values

to be imputed are input to this model. The set of outliers of the original data is compared with the set of generated outlier values.

A second order quadratic error function is generated. Genetic Algorithms is used to optimize the error function in any given domain. Using a Genetic Algorithm function call, we find a local unconstrained minimum, x , to the objective function, $fitnessfcn$.

$$x = ga(fitnessfcn, nvars) \tag{1}$$

$nvars$ is the dimension (number of design variables) of $fitnessfcn$. The objective function, $fitnessfcn$, accepts a vector x of size 1 -by- $nvars$, and returns a scalar evaluated at x . The predicted value for the incorrect or missing value is corrected using this optimization through genetic algorithms. The proposed module to impute categorical values is illustrated in Fig. 2.

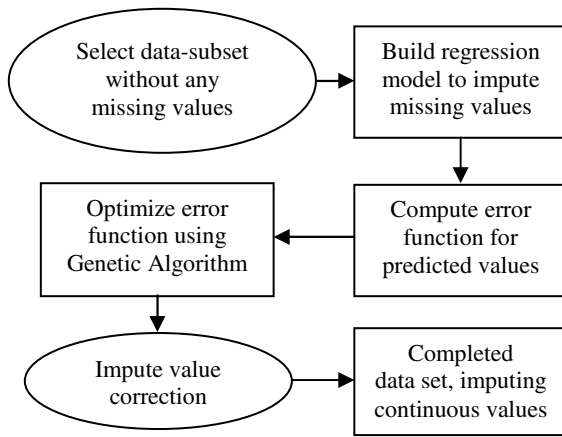


Fig. 2. Module to impute continuous values

PIDD set contains missing/incorrect value ranging from 4% (taking only the records or instance containing zero values to the attribute –‘diastolic blood pressure’) to 48% (taking only the records or instances containing zero values to the attribute – ‘2 hour serum insulin presence’). This work implements the proposed data preprocessing framework, to impute continuous values, for the attribute ‘2-hour serum insulin’ (with 48% incorrect values). A box-plot is once again employed to inspect the mean and the outliers in the data, for any further improvement shown as shown in Fig 1(b). The attribute ‘2-hour serum insulin’ is represented by column five in the boxplots which indicates the differences before and after the imputation process.

The now ‘complete’ data set (with respect to the attribute ‘2 hour serum insulin presence’) with an optimized error is ready to be used for predictive analytics.

5.2 Proposed Approach to Impute Categorical Missing Values

With the exception of the attribute ‘age’ (continuous), the nature of all other attributes of MMD is categorical. Hence a model that identifies and detects the regularities and correlations between these input functions is one of the crucial steps of this framework.

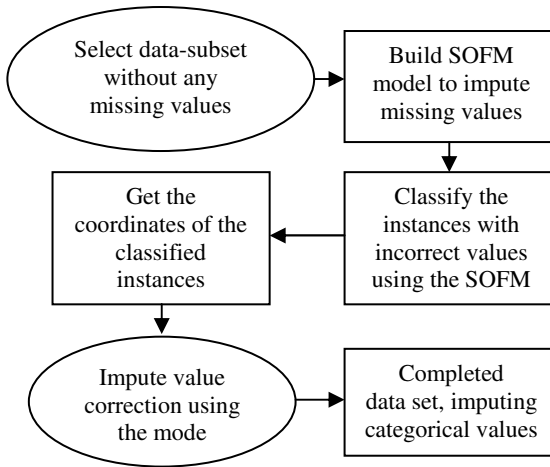


Fig. 3. Module to impute categorical values

The technique of Self Organizing Feature Maps (SOFM) is selected to model the complete instances of the MMD, as the neurons of the competitive network learn to recognize the groups of similar vector inputs. SOFM is trained using the complete tuples. Learning Vector Quantization (LVQ) method is used for supervised training of the competitive layers. The instances that include erroneous values are input to the thus simulated model, for classification. The co-ordinates of the classification, deduced from the simulated SOFM, is used to map the value to be imputed, using the mode of the values. The proposed module to impute categorical values is illustrated in Fig. 3.

The imputation of categorical data was experimented using MMD, wherein the data set contains missing/erroneous value ranging from 0.2% to 7.9%. The attribute selected from imputation is ‘density of the mass’, which had the highest error percentage in the data of 7.9%. The MMD data is input to the proposed framework to impute categorical values, using the SOFM concept.

The completed data set where the data was imputed for the attribute ‘density of the mass’ alone, is now ready to be tested on the prediction model, where multi-layer perceptron neural network is selected as the classifier.

6 Prediction Model

The architecture of the proposed model is depicted in Fig. 4. Artificial Neural Networks (ANN), a supervised learning data mining approach, is selected as the classifier as it is sensitive to non-linear input values and high on prediction efficiency. The preprocessed data, worked on: as per the proposed preprocessing framework is input to the prediction model. The output statistics from the proposed model is discussed in section 7.

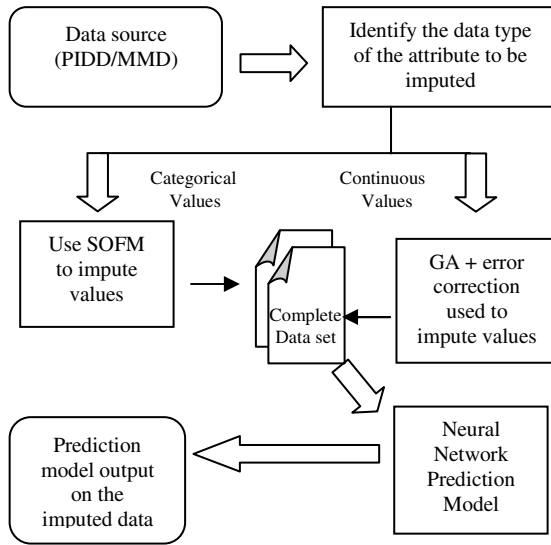


Fig. 4. Overview of the Prediction Model with Preprocessing Implemented

7 Performance Analysis

The parameters used to evaluate the performance of the prediction model are the accuracy, sensitivity and specificity, Receiver Operating Characteristic (ROC) and Area under ROC (AUR). Accuracy measures the proportion of true results in a test. Sensitivity or true positive rate, measures the proportion of actual positives which are correctly identified while specificity, the true negative rate, measures the proportion of negatives which are correctly identified. Receiver Operating Characteristic (ROC) curves have been used to compare the performances of different predictive models based on the input data set. Comparison in terms of the AUR, gives an insight to the performance of the classifier over an entire range of values and is independent of the prevalence of the condition unlike the accuracy, which weights sensitivity and specificity in proportion to their prevalence.

The PIDD set, when used without any preprocessing, as input to a neural networks prediction model exhibits an accuracy of 75.82%.

When PIDD is input to the prediction model, suggested in this work, the preprocessing framework works on the imputation of the ‘2-hour serum’ attribute. The now-complete PIDD was partitioned for training and testing the model using a random split of 80-20 split where 80% (615 records) constitute the training set and 20% (153 records) that form the test set. This splitting of data facilitates easy validation of the model in order to improve its reliability. The performance for 4 runs are as indicated in Table 3. The accuracy, sensitivity (the true positive rate) and specificity (the true negative rate) calculated from the confusion matrix reflect the performance of the prediction model.

The accuracy of the proposed model is also compared with the work presented by Kayaer K., Yildirim T. [12] where the classifier selected is the general regression

Table 3. PIMA prediction accuracy details

	Imputation	Accuracy	Sensitivity	Specificity
1	No	75.8171%	81.73%	63.27%
2	Yes	82.3529%	95.79%	73%
3	Yes	81.6993%	86%	73.59%
4	Yes	84.3137%	86.54%	79.59%
5	Yes	83.0065%	85.45%	76.74%

Table 4. Comparison of Proposed Model with works [12, 13] on PIDD data set

	Classifier	Accuracy
Source data	ANN	75.82
Kayaer K., Yildirim T. [12]	General Regression Neural Networks	82.29
Aslam. M. W., Nandi. A. K [13]	GP	80.7
Data after Imputation	ANN	82.84

neural network. Aslam. M. W., Nandi. A. K [13] have adopted genetic programming with comparative partner selection (GP CPS). The results are detailed out in table 4.

MMD data, when used without any preprocessing, exhibited a prediction accuracy of 81.3%. In this work, MMD data was preprocessed as per our proposed framework for categorical data and data was imputed for the attribute ‘density’.

The now-complete MMD data was partitioned randomly to give a training and test data set as per the 80-20 split. The prediction model’s accuracy is found to be 90.76%.

Accuracy, sensitivity and specificity of the model with preprocessed data are compared with the results in [10], detailed out in table 5.

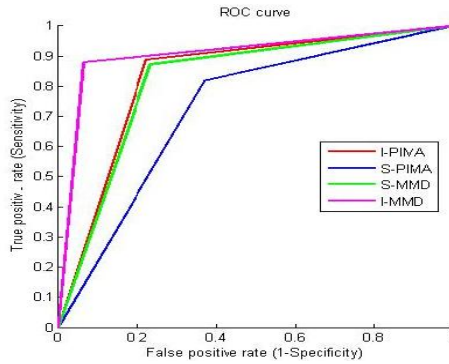
Table 5. Comparison of Proposed Model with work [10] on MMD data set

	Accuracy	Sensitivity	Specificity
Source data	81.34715	87.2093	76.63551
Elsayad [10]	80.9	85.29	76.97
Data after Imputation	90.76358	87.87879	93.61702

The ROCs for the prediction model are illustrated in Fig. 5. Imputed PIDD set is indicated as I-PIMA and S-PIMA refers to the source/original PIMA data set, I-MMD represents the imputed MMD as data set and S-MMD the source/original MMD data set. The area under the ROC (AUR) is given in Table 6. An AUR of 1 represents a perfect test; an area of 0.5 represents a worthless test. It can be observed that Neural Network Prediction Model gives the best response on applying the soft computing imputation method.

Table 6. Area under the ROC for the prediction models

Data	Model	Area Under ROC	
		Before Imputation	After Imputation
PIDD	ANN	0.7249	0.833
MMD	MLP	0.8192	0.907

**Fig. 5.** ROC curves of prediction before and after imputation

8 Conclusions

This work proposes a new approach for preprocessing real-time data, used for predictive analyses and data mining in the health sector using two data sets of diabetes and cancer prediction. The preprocessing involves two unique and distinct processes, based on the data set considered for imputation. In case of continuous values, CART along with optimization of the error function using a Genetic Algorithm, maps the imputed values in a valid domain of values for each attribute. The proposed model when tested with Pima Indian Diabetes Data set (PIDD) shows an average accuracy of 82.84%, as against the accuracy of the prediction model when built on data without preprocessing which is 75.816%. This improved prediction accuracy fares superior to some of the other models in vogue. In case of categorical values, Self Organizing Feature Maps is used to impute the selected attribute's incorrect values. The thus completed data is used to build the multi-layer-perceptron model for prediction. The accuracy of the proposed model, when tested with the Mass Mammographic Data (MMD) increased from 81.35% to 90.76%.

This framework has to be tested on similar health sector data, which tracks ailments and the factors that influence the occurrence or recurrence of a particular disease. The preprocessing steps that have been discussed here can be used on any prediction model development framework, to impute missing, incorrect and incomplete values. This approach being simpler to implement and results that are easy to comprehend provides a reliable solution to real-time predictive analyses or other problems that require imputation that could further contribute to the improvement of performance of the model.

References

1. Cabena, P., Hadjinian, P., Stadler, R., Verhees, J., Zanasi, A.: *Discovering Data Mining: from Concepts to Implementation*. Prentice Hall, Englewood Cliffs (1998)
2. Acuna, E., Rodriguez, C.: The Treatment of Missing Values and its Effect in the Classifier Accuracy. In: *Multiscale Methods in Science and Engineering*. LNCS, pp. 639–647. Springer, Heidelberg (2004)
3. Peng, L., Lei, L.: A Review of Missing Data Treatment Methods. *Intelligent Information Management Systems and Technologies* 1(3), 412–419 (2005)
4. Bhat, V.H., Rao, P.G., Shenoy, P.D., Venugopal, K.R., Patnaik, L.M.: An Efficient Prediction Model for Diabetic Database Using Soft Computing Techniques. In: Sakai, H., Chakraborty, M.K., Hassanien, A.E., Ślęzak, D., Zhu, W. (eds.) *RSFDGrC 2009*. LNCS, vol. 5908, pp. 328–335. Springer, Heidelberg (2009)
5. Mehala, B., Ranjit, J.T.P., Vivekanandan, K.: Selecting Scalable Algorithms to Deal with Missing Values. *International Journal of Recent Trends in Engineering* 1(2) (2009)
6. Batista, G.E.A.P.A., Monard, M.C.: K-Nearest Neighbour as Imputation Method. *Experimental Results*. Tech. Report 186, ICMC-USP (2002)
7. Breault, J.L.: Data Mining Diabetic Databases: Are Rough Sets a Useful Addition? *Artificial Intelligence in Medicine* 27, 227–236 (2003)
8. King, M.A., Elder IV, J.F., et al.: Evaluation of Fourteen Desktop Data Mining Tools. In: *Proc. of IEEE International Conference on Systems, Man and Cybernetics*, San Diego, CA (1998)
9. Khan, A.H.: Multiplier-free Feedforward Networks. In: *Proc. of the IEEE International Joint Conference on Neural Networks (IJCNN)*, Honolulu, Hawaii, vol. 3, pp. 2698–2703 (2002)
10. Elsayad, A.M.: Predicting the Severity of Breast Masses with Ensemble of Bayesian Classifiers. *Journal of Computer Science* 6(5), 576–584 (2010)
11. Machine Learning Database Repository at the University of California, Irvine, <http://www.ics.uci.edu/mllearn/MLRepository>
12. Kayaer, K., Yildirim, T.: Medical Diagnosis on Pima Indian Diabetes using General Regression Neural Networks. In: *Proc. of the International Conference on Artificial Neural Networks/International Conference on Neural Information Processing*, Istanbul, Turkey, pp. 181–184 (2003)
13. Aslam, M.W., Nandi, A.K.: Detection of Diabetes using Genetic Programming. In: *18th European Signal Processing Conference*, Denmark, pp. 1184–1188 (August 2010)
14. Magnani, M.: *Techniques for Dealing with Missing Data in Knowledge Discovery Tasks*. By Department of Computer Science, University of Bologna (2004)
15. Estébanez, C., Aler, R., José, M.: Method Based on Genetic Programming for Improving the Quality of Data Sets in Classification Problems. *International Journal of Computer Science and Applications* 4(1), 69–80 (2007)

Process Oriented Guided Inquiry Learning for Soft Computing

Clifton Kussmaul

Muhlenberg College, 2400 Chew St, Allentown, PA 18104-5586 USA
kussmaul@muhlenberg.edu

Abstract. This paper describes several process oriented guided inquiry learning (POGIL) activities focused on soft computing (genetic algorithms, neural networks, and fuzzy systems). The paper provides background on POGIL, illustrates several activities, summarizes student feedback, and discusses lessons learned and possible future directions.

Keywords: active learning, fuzzy logic, genetic algorithms, inquiry learning, neural networks, POGIL.

1 Introduction

To improve learning quality and retention, particularly in science, technology, engineering, and mathematics (STEM) disciplines, educators have developed a wide variety of approaches to engage students, enhance learning, and emphasize attitudes and skills rather than only knowledge (often rote). These approaches are generally referred to as *active learning*, and approaches where students work together are referred to as *collaborative learning*. Process oriented guided inquiry learning (POGIL) is a particular approach to collaborative learning which combines a set of effective practices, and has been developed and validated over many years.

This paper provides some general background on POGIL. It then describes activities focused on soft computing (genetic algorithms, neural networks, and fuzzy systems) that have been used successfully and are being revised and generalized. This paper is novel in that it describes: a) POGIL in computer science in general and soft computing in particular; b) POGIL outside the USA, and specifically in India; and c) POGIL activities using presentation slides rather than paper handouts.

2 Process Oriented Guided Inquiry Learning

POGIL has three distinguishing features [11]. First, teams of learners (typically 3-5) follow processes with specific roles, steps, and reports that help students develop process skills and encourage individual responsibility and meta-cognition. Second, teams work on scripted inquiry activities and investigations designed to help them construct their own knowledge, often by modeling the original processes of discovery and research. POGIL activities and processes are designed to achieve specific learning objectives; typically an activity is designed to focus on 1-2 (disciplinary)

concepts and 1-2 process skills. Third, the POGIL instructor serves as a facilitator, not a lecturer. Typically this means that the instructor circulates among the student teams, checks on progress, and helps teams to resolve problems. Periodically the instructor may ask teams to share answers or insights with the rest of the class.

POGIL is based on learning science (e.g. [14]), and shares characteristics (e.g. teams, processes, inquiry) with other forms of active, discovery, and inquiry-based learning (e.g. [5]). POGIL is distinctive in the way it combines a set of characteristics that support each other and strengthen learning outcomes.

POGIL activities are generally designed to follow a *learning cycle* with 3 phases [1, 11]. In *exploration*, students look for trends or patterns in data they have collected or that is provided, and generate and test hypotheses to help understand or explain the data. In *concept invention*, the trends, patterns, or hypotheses are used to define a new concept or term; importantly, students have constructed understanding before the concept is introduced. In *application*, the new concept is applied in other situations or contexts to help students generalize its meaning and applicability. Thus, the scripted activity provides information and asks questions to guide students through the learning cycle and help them develop process and learning skills. (For various reasons, some POGIL activities introduce a concept, and follow it with application and exploration.)

Designing effective POGIL activities can be time-consuming, but supporting resource are available [7]. Generally, the author first identifies learning objectives and the focus of the activity. For examples, teams could analyze data, derive equations, or explore the behavior of a physical system. Next, the author creates a sequence of *key questions* that guide teams through the inquiry process. Finally, the author identifies and develops supporting information, such as prerequisites, glossary of terms, references, handouts, and subsequent assignments or projects.

Activities involve three types of key questions. *Directed questions* have definite answers, are based on material available to students, and provide a foundation for later parts of the activity. *Convergent questions* may have multiple answers, and require teams to analyze and synthesize information to reach non-obvious conclusions. *Divergent questions* are open-ended, do not have right or wrong answers, and may lead teams and individual students in different directions.

POGIL has been developed and validated over the last 15 years. Multiple studies have found that POGIL significantly improves student performance, particularly for average and below-average students. (e.g. [6], [9], [10]). POGIL has been used extensively in chemistry, and also in materials science and engineering [4], and computer science (CS) [3]. Although active learning and discovery learning are increasingly popular in CS (e.g. [2], [13]), POGIL is not yet well known in CS education.

3 Context and Examples

The author developed and used POGIL activities and other active learning activities for a course on soft computing, including genetic algorithms, neural networks, and fuzzy systems [12]. The course was taken by roughly 20 first term graduate students in computer science at the University of Kerala, Karyavattom, where the author was a Visiting Fulbright-Nehru Scholar.

Although most POGIL classrooms give each team or student a set of paper handouts for each activity, in this course the activities were presented as a sequence of PowerPoint slides which were posted online after class for students to review. The initial motivation was to reduce paper use. It also provided the instructor with more flexibility to tweak activities by adding steps or providing more information, and to manage the class's pace since students couldn't race forward to finish an activity or look for clues. This could make the class less student-centered, but it could also be more familiar to students who are accustomed to lectures.

The following subsections briefly describe slides and other parts of several activities from the course. These particular examples were selected because they illustrate some (though not all) aspects of POGIL, and require relatively little context or background.

Slides presenting concepts were covered in a few minutes of lecture and questions. Slides with questions improve students' understanding and enables the instructor to help resolve any questions or confusion before moving on to new topics. Early questions are simpler, and later questions may be more challenging or require more reflection. After the topic is covered in class, students apply it in a programming assignment, usually working with a partner.

Given time (for development, and in the classroom) these activities could be improved or extended in various ways, some of which are discussed below. For these activities to be more readily adopted for other courses, it would be useful to: define learning objectives, prerequisites, resources, and vocabulary; provide more complete background information; and package them in a standard format. This work is ongoing (and perhaps unending).

3.1 Genetic Algorithms

The slides in Figure 1 are from an activity that introduces schemata (patterns of solutions), some issues in schemata counting, and the first steps in deriving the schema theorem. Before starting the activity, the instructor ensures that students are organized into teams, and might spend a few minutes introducing the activity. Slide *Schemata* defines terms and provides some examples. The instructor could reinforce these ideas verbally, or have the teams study them. Slide *Schemata Count: Questions* begins with a directed question about counting solutions, and uses convergent questions to develop formulas. The instructor could ask teams to share their answers to see if there is agreement. Slide *Schemata Selection: Questions (1)* defines some more terms, and uses more convergent questions to lead students toward an equation showing how more fit solutions become more common over time. Slide *Schemata Selection Questions (2)* extends this to schemata, as the first of three major steps in defining the schema theorem. These might prompt a brief class discussion of how the equation could be used or improved. Together, the slides lead teams to explore schemata terminology and invent concepts which will be applied in later slides or subsequent assignments, thereby completing the learning cycle and lead students to deeper understanding.

To improve this activity and make it more widely applicable, we might:

- Use exploration tasks to motivate schemata rather than starting with a definition.
- Motivate the terms defined in the slides.
- Provide more intermediate steps, particularly for students who are less comfortable with mathematical derivations & proofs.

<p style="text-align: center;">Schemata</p> <p>In analyzing & discussing GAs & solutions, we often want to describe solution patterns, called schemata. In each schema, some values are required, and others are not (represented by *)</p> <p>Examples $AB^{**}AB^{**}$ $0010^{*}0011$ Schemata present counting problems...</p>	<p style="text-align: center;">Schemata Selection: Questions (1)</p> <p>$A(t)$ is the population at time t $f(a)$ is the fitness of solution $a \in A(t)$ $m(a,t)$ is the count of a at t</p> <ol style="list-style-type: none"> 1. What is the total fitness of $A(t)$? 2. What is the average fitness? 3. What is $m(a,t+1)$, the count of solution a at $(t+1)$? 4. Does #3 make sense? How could it be used?
<p style="text-align: center;">Schemata Counting: Questions</p> <p>Assume there are s symbols in the alphabet, and each solution has L symbols.</p> <ol style="list-style-type: none"> 1. Count the solutions for each schema: $1^{*}0^{*}$ $1^{**}0^{**}0^{*}$ $AA^{**}C$ 2. How many solutions are there for a schema with w wildcard values? 3. How many schemata match a given solution? 4. How many schemata are possible, total? 	<p style="text-align: center;">Schemata Selection: Questions (2)</p> <p>H is a schema</p> <ul style="list-style-type: none"> • pattern does not depend on t, but members do <p>$m(H,t)$ is the count of H at t</p> <ol style="list-style-type: none"> 1. What is $m(H,t+1)$? Hint: sum over solution counts 2. Substitute into #1 your earlier expression for $m(h,t+1)$. 3. Explain the equality: $\sum f(h_i) = m(H,t)f(H)$ 4. Substitute #3 into #2. 5. Does #4 make sense? How could it be used?

Fig. 1. Sample slides from activity on genetic algorithms

3.2 Neural Networks

The slides in Figure 2 are from an activity that explores the capabilities of single and multi layer neural networks. Previously, students investigated the form and behavior of a single neuron, and the role and possible shapes for activation functions. Before starting the activity, the instructor might briefly summarize some of these concepts, or give a brief quiz to see how well students understand them. Slide *ANN Design Questions (1)* asks convergent questions about choosing weights for neurons that perform specific functions, to help the teams better understand how neurons compute results. Depending on the class, the instructor might provide a handout showing truth tables for each operation, or a spreadsheet or software widget where students could enter weights and see the output for different inputs. Slide *ANN Design: Questions (2)* leads students to plot contours (for specific values of y) in order to see how a neuron combines its inputs, and some of the effects of activation functions. The instructor might provide blank graphs. Slide *ANN Design: Questions (3)* introduces the XOR operation, which cannot be computed with a single neuron but can be computed with a sequence of neurons. This might prompt a brief class discussion of single vs. multi-layer networks. Slide *ANN Design: Questions (4)* asks students to design networks to

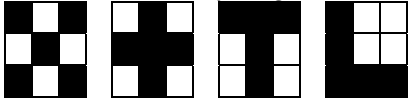
<p>ANN Design: Questions (1) Consider a neuron with $F=-1$, $T=+1$ and a step activation function.</p> <ol style="list-style-type: none"> 1. Try to choose weights to compute <ol style="list-style-type: none"> a. logical NOT (1 input) b. logical OR (2+ inputs) c. logical AND (2+ inputs) 2. Can one activation function serve all 3 functions? 	<p>ANN Design: Questions (3)</p> <ol style="list-style-type: none"> 1. Try to choose weights to compute XOR with a single node. What new issues arise? 2. What general statements (pros & cons) can you make about one-layer networks? 3. Try to choose weights to compute XOR with a multi-layer network. 4. What general statements (pros & cons) can you make about multi-layer networks?
<p>ANN Design: Questions (2)</p> <ol style="list-style-type: none"> 1. On an (x_1, x_2) graph, draw lines for $y = x_1w_1 + x_2w_2$ for w_1 & w_2, $y = \{-1, 0, +1\}$ 2. How would $f()$ affect the lines? 3. Extend & compare with a 3+ input network. 	<p>ANN Design: Questions (4) Consider a 3x3 binary image:</p>  <ol style="list-style-type: none"> 1. Design a network where each pattern above has a unique output. 2. How could you use less neurons? 3. How could you allow noisy data (1-2 incorrect pixels)?

Fig. 2. Sample slides from activity on neural networks

identify simple patterns, and prompts them to begin considering issues of efficiency (number of neurons) and robustness when data is noisy or incomplete. Here again, the instructor might provide a handout or a software widget. Teams that progress quickly could spend more time exploring ways to reduce the number of neurons or allow noisy data.

To improve this activity and make it more applicable, we might:

- Use more exploratory examples in slides a) and d), perhaps including some with continuous rather than binary inputs, and examples with multiple outputs.
- Use a series of questions to better motivate the graphs in slide b).

3.3 Fuzzy Sets and Relations

The slides in Figures 3 and 4 are from an activity that introduces fuzzy sets and relations. This activity focuses more on application than on concept invention, in part because the students were already familiar and comfortable with these concepts in regular (crisp) sets. In a setting where students were less familiar with these concepts, the activity might start more gradually and review these concepts.

In Figure 3, the first three slides provide context and define some terms and relations, and slide *Fuzzy Set Relations: Questions* prompts students to apply these ideas. The instructor could ask teams to share their answers to see if there is

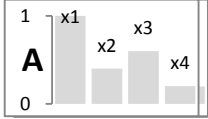
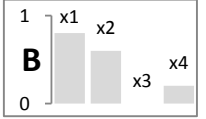
<p style="text-align: center;">Fuzzy Sets</p> <p>fuzzy set generalizes regular (crisp) set with degree or grade [0.0 ... 1.0]</p> <ul style="list-style-type: none"> crisp set elements have degree 0 (non-member) or 1 (member) <p>membership function $m(x)$ maps each element $x \in$ Universe to its degree. (Universe is a crisp set)</p> <ul style="list-style-type: none"> finite/infin, discrete/continuous 	<p style="text-align: center;">Fuzzy Set Relations: Questions</p> <p>Given fuzzy set $C =$ $\{bla, whi, yel, mag\}$ with $m_R(C) = \{ bla/0, whi/1, yel/1, mag/1 \}$ $m_G(C) = \{ bla/0, whi/1, yel/1, mag/0 \}$ $m_B(C) = \{ bla/0, whi/1, yel/0, mag/1 \}$</p> <ol style="list-style-type: none"> Identify all subset relations among R, G, & B. Identify support of R, G, & B.
<p style="text-align: center;">Fuzzy Sets: Examples</p> <p>materials $A = \{wax, steel, diamond\}$ hardness $m_H(A) = \{w/.1, s/.7, d/1\}$ ages $A = \{0\dots\}$ (non-neg ints) oldness $m_O(x) = \{0.0 \text{ for } x \leq 20$ $(x-20)/60 \text{ for } 20 \leq x \leq 80$ $1.0 \text{ for } 80 \leq x\}$</p>	<p style="text-align: center;">Fuzzy Set Operations: Familiar Union & Intersection</p> <p>$A \cup B \{x/\max(m_A(x), m_B(x)) x \in U\}$ $A \cap B \{x/\min(m_A(x), m_B(x)) x \in U\}$</p> <p style="text-align: center;">Complement</p> <p>$A' \{x/(1-m_A(x)) x \in U\}$</p>
<p style="text-align: center;">Fuzzy Set Relations</p> <p>Equality $A=B \equiv m_A(x)=m_B(x) \forall x \in U$ Subset $A \subseteq B \equiv m_A(x) \leq m_B(x) \forall x \in U$ $A=B \equiv A \subseteq B \text{ and } B \subseteq A$ Member $x \in A$ undefined for fuzzy but U is universe (crisp set) Support $\{x m_A(x) > 0, x \in U\}$ Fuzzy Singleton set with $\text{support} = 1$</p>	<p style="text-align: center;">Fuzzy Set Operations: Questions</p> <div style="display: flex; justify-content: space-around;">   </div> <ol style="list-style-type: none"> Given sets A & B, draw: $A', B', A \cup B, A \cap B, A' \cap B'$ How are 1c & 1e related? Are crisp and fuzzy definitions for $\{\cup, \cap, '\}$ consistent?

Fig. 3. Sample slides from activity on fuzzy sets (first part)

agreement. Slide *Fuzzy Set Operations: Familiar* lists operations that students have encountered previously, and slide *Fuzzy Set Operations: Questions* uses convergent questions to prompt teams to apply these operations to fuzzy sets displayed as graphs, and to consider relationships among the new concepts. This might lead to a brief classroom discussion. The instructor might provide handouts for drawing the results of the 5 operations, and post each team’s answers so other teams can study them.

Similarly, in Figure 4, slide *Fuzzy Set Properties: Familiar* defines concepts and slide *Fuzzy Set Properties: Unfamiliar* prompts students to consider the definitions before applying them; some teams may notice some subtleties before applying the concepts in slide *Fuzzy Set Operations: Questions (1)* and slide *Fuzzy Set Operations: Questions (2)*. Again, the instructor might provide handouts and post answers for review and discussion.


<p>Fuzzy Set Properties: Familiar</p> <p>Commu. $A \cup B = B \cup A$ $A \cap B = B \cap A$</p> <p>Assoc. $(A \cup B) \cup C = A \cup (B \cup C)$ $(A \cap B) \cap C = A \cap (B \cap C)$</p> <p>Distrib. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$</p> <p>DeMorgan $(A \cup B)' = A' \cap B'$ $(A \cap B)' = A' \cup B'$</p> <p>Involution $(A')' = A$</p> <p>Idempotent $A \cup A = A$</p> <p>Identity $A \cup \emptyset = A$ $A \cap \emptyset = \emptyset$ $A \cup U = U$ $A \cap U = A$</p>	<p>Fuzzy Set Operations: Questions (1)</p>  <p>1. Given fuzzy sets A and B, each with 4 elements, draw: CON(A), DIL(B), NORM(A), NORM(B)</p>
<p>Fuzzy Set Properties: Unfamiliar</p> <ul style="list-style-type: none"> Concentration & Dilation $CON(A) \{x/m_A(x)^2 x \in U\}$ $DIL(A) \{x/m_A(x)^{1/2} x \in U\}$ Normalization $NORM(A) \{x/m_A(x)/Max x \in U\}$ CON decreases degree, DIL increases degree. Do these names make sense? 	<p>Fuzzy Set Operations: Questions (2)</p> <p>Given previous definitions</p> <ul style="list-style-type: none"> ages $A = \{0 \dots\}$ (non-negative ints) oldness $m_o(x) = \{0.0 \text{ for } x \leq 20$ $(x-20)/60 \text{ for } 20 \leq x \leq 80$ $1.0 \text{ for } 80 \leq x\}$ <p>1. For $x = 0..100$, draw: $m_o(x)$, $CON(m_o(x))$, $DIL(m_o(x))$</p>

Fig. 4. Sample slides from activity on fuzzy sets (third part)

To improve this activity and make it more widely applicable, we might:

- Review (crisp) set relations, operations, and properties to assess understanding and resolve gaps or misconceptions.
- Use questions to explore and invent concepts such as membership degree and membership function. For example, the activity could start by considering different ways to describe someone’s age (e.g. “under 21”, “over 65”, “middle aged”, “young”, “old”).
- Use examples to explore and invent ways to define membership functions for finite and infinite sets.
- Use questions to explore and invent ways to extend set relations and operations to fuzzy sets and the concept of support.

4 Results

At the end of the course, students were given a written survey and asked to briefly describe things that worked well and things that worked less well. Two examples follow:

Group discussions & exchanging ideas with other groups is a better one. This makes us think about it in a better way. As you come to each group to talk, we can clarify our doubts by discussing it with you. And you encourage us to think more about the topic. Assignments also help us a lot to understand the subject in detail. Exams like this is new to us. This exams also make us think more about the things that we prepared for exams.

Your way of teaching is entirely different from other classes. It helps me to think more and understand more things. When you are giving an assignment to us it is very different from the assignments that we are already done, at first was a strange experience but now I can realize that this is one of the best ways to think for students. In the case of exams also I feel the same. The group activities that you gave us is very helpful to know more about how helpful when working in a team rather than single.

Other responses were consistent, though not always as eloquent. In sum, the responses included several recurring themes:

- Excitement about having a visiting faculty member from the US.
- Initial uncertainty about an unfamiliar teaching and learning style, followed by the realization that they were understanding concepts better and acquiring useful process skills.
- Some difficulty understanding the instructor's (American) English. However, within a team students could (and did) converse in English, Hindi, and Malayalam, so that students who had difficulty understanding the instructor could seek clarification from other team members.

5 Conclusions

POGIL is based on learning science and has a proven track record in other disciplines. POGIL classrooms are very different from lecture-based classrooms, but POGIL shares characteristics with other forms of active, discovery, and inquiry-based learning, so that faculty familiar with such approaches should not have difficulty adapting to POGIL. There are a variety of materials to help faculty develop and improve POGIL materials (e.g. [7], [8], [11]).

There are several benefits to using POGIL, particularly in computer science and software engineering. Particularly in settings where students have varied academic and linguistic backgrounds, POGIL encourages students to collaborate and learn from each other rather than focusing attention on a single instructor. Software development is primarily a problem-solving activity where most background information is readily available, and POGIL helps students to develop their problem-solving abilities. Teams are an important part of most IT organizations, and POGIL helps students to develop important team process skills.

There are also some challenges to using POGIL. Developing and revising effective activities can be time-consuming. Students and faculty who are not used to POGIL or other active learning approaches may be reluctant to try it or have difficulty at first.

This paper is novel in that it describes: a) POGIL in computer science in general and soft computing in particular; b) POGIL outside the USA, and specifically in India; and c) POGIL activities using presentation slides rather than paper handouts.

The author continues to develop, use, and revise POGIL activities for other topics in computer science, including Java programming, project management, and data structures and algorithms. One future direction is to extend and adapt the soft computing activities described here for courses and faculty at other institutions. Another direction is to explore the potential of supporting infrastructure, such as personal computers, learning management systems, or classroom response systems. Finally, the author hopes develop an international community of faculty using and improving POGIL activities in computer science.

Acknowledgements

The author thanks the US-India Educational Foundation and the University of Kerala for a Fulbright-Nehru teaching award that provided the opportunities and other support for this work. The author also thanks Muhlenberg College and the US National Science Foundation for support to develop and refine POGIL activities for computer science courses. Finally, the author thanks the National POGIL Project (<http://www.pogil.org>) for ongoing encouragement and support.

References

1. Abraham, M.R.: Inquiry and the learning cycle approach. In: Pienta, N.J., Cooper, M.M., Greenbowe, T.J. (eds.) *Chemists' Guide to Effective Teaching*, pp. 41–52. Prentice Hall, Upper Saddle River (2005)
2. Baldwin, D.: Discovery learning in computer science. In: *Proceedings of the 27th SIGCSE Technical Symposium on Computer Science Education (SIGCSE)*, pp. 222–226. ACM, Philadelphia (1996)
3. De Palma, A.: Engaging students through the guided-inquiry cycle. In: *Proceedings of Redesigning Pedagogy: Research, Policy, Practice*. National Institute of Education, Nanyang Technological University, Singapore (2005)
4. Douglas, E.P., Chiu, C.: Use of guided inquiry as an active learning technique in engineering. In: *Proceedings of the 2009 Research in Engineering Education Symposium*, Palm Cove, Queensland, Australia (2009)
5. Eberlein, T., Kampmeier, J., Minderhout, V., et al.: Pedagogies of engagement in science. *Biochemistry and Molecular Biology Education* 36(4), 262–273 (2008)
6. Farrell, J.J., Moog, R.S., Spencer, J.N.: A guided-inquiry general chemistry course. *Journal of Chemical Education* 76(4), 570 (1999)
7. Hanson, D.M.: Designing process-oriented guided-inquiry activities. In: Beyerlein, S.W., Apple, D.K. (eds.) *Faculty Guidebook - A Comprehensive Tool for Improving Faculty Performance*, 2nd edn. Pacific Crest, Lisle (2005)
8. Hanson, D.M.: *Instructor's Guide to Process-Oriented Guided-Inquiry Learning*. Pacific Crest, Lisle (2006)
9. Hanson, D.M., Wolfskill, T.: Process workshops - A new model for instruction. *Journal of Chemical Education* 77(1), 120 (2000)

10. Lewis, S.E., Lewis, J.E.: Departing from lectures: An evaluation of a peer-led guided inquiry alternative. *Journal of Chemical Education* 82(1), 135 (2005)
11. Moog, R.S., Spencer, J.N. (eds.): *Process-Oriented Guided Inquiry Learning (POGIL)*. Oxford University Press, USA (2008)
12. Munakata, T.: *Fundamentals of the New Artificial Intelligence: Neural, Evolutionary, Fuzzy and More*, 2nd edn. Springer, Heidelberg (2008)
13. Trono, J.A.: A discovery-based capstone experience. *Journal of Computing Sciences in Colleges* 23(4), 195–200 (2008)
14. Zull, J.: *The Art of Changing the Brain: Enriching the Practice of Teaching by Exploring the Biology of Learning*, 1st edn. Stylus Publishing (2002)

A Modified and Efficient Shuffled Frog Leaping Algorithm (MSFLA) for Unsupervised Data Clustering

Suresh Chittineni¹, Dinesh Godavarthi¹, ANS Pradeep¹,
Suresh Chandra Satapathy¹, and PVGD Prasad Reddy²

¹Anil Neerukonda Institute of Technology and Sciences
Sangivalasa, Visakhapatnam, Andhra Pradesh, India
{sureshchittineni, dinesh.coolguy222,
pradeep.6174, sureshsatapathy}@gmail.com

²Andhra University Engineering College
Visakhapatnam, Andhra Pradesh, India
prasadreddy.vizag@gmail.com

Abstract. Shuffled frog leaping Algorithm (SFLA) is a new memetic, population based, meta-heuristic algorithm, has emerged as one of the fast, robust with efficient global search capability. In order to enhance the algorithm's stability and the ability to search the global optimum, the conventional SFL Algorithm has been modified in our work by using the local best value of each memplex instead of generating a new frog, to enhance the effectiveness of the SFLA. This paper implements the application of Modified SFLA in Partitional clustering of the unlabelled data. This algorithm is applied on various classification problems and the simulated results demonstrate that, this modified SFLA has outperformed the conventional SFL Algorithm.

Keywords: Modified Shuffled Frog Leaping Algorithm (MSFLA), Shuffled Frog Leaping Algorithm (SFLA), Memetic Algorithm (MA), Partitional Clustering.

1 Introduction

Memetic algorithms (MA) represent one of the recent growing areas of research in evolutionary computation. The term MA is now widely used as a synergy of evolutionary or any population-based approach with separate individual learning or local improvement procedures for problem search [1]. There are various subtypes and algorithms exist in the memetic algorithms. Shuffled frog-leaping algorithm (SFLA) is a new memetic meta-heuristic algorithm with efficient mathematical function and global search capability [2]. This was developed by Eusuff and Lansey [3]. SFLA is a population based, cooperative Search metaphor inspired by behavior of frogs in nature searching for food. It is originated from the research of food hunting behaviors of frog. Researchers found that, in theory at least, individual members of the school can profile from the discoveries and previous experience of all other members of the school during the search for food. The advantage can become decisive, outweighing the disadvantages of competition for food items, whenever the resource is unpredictably distributed in patches. Their behaviors are unpredictable but always

consistent as a whole, with individuals keeping the most suitable distance. The key idea is to create a population of candidate solutions to an optimization problem, which is iteratively refined by alteration and selection of good solutions for the next iteration [4]. Candidate solutions are selected according to a fitness function, which evaluates their quality with respect to the optimization problem. To enhance the algorithm’s performance and the ability to search the global optimum, the conventional SFL Algorithm has been modified by replacing the worst frog with the local best frog of each memplex instead of generating a new random frog. This paper applied our concepts of Modified Shuffled Frog Leaping Algorithm (MSFLA) in the process of clustering, and this algorithm is tested to the various real-world classification datasets. The rest of the paper is organized as follows: Scientific background of clustering is explained in Section 2 followed by detailed description of SFLA in Section 3. The clustering algorithm using MSFLA is proposed in section 4 and the results are displayed in the section 5 followed by conclusion in the last section.

2 Scientific Backgrounds

All the objects can be represented in a physical or abstract structure called *patterns*. These can be distinguished from each other by a collective set of attributes called *features*, which together represent a pattern [5]. Let $P = \{P_1, P_2, \dots, P_k\}$ be a set of k patterns or data points, each having m features. These patterns can be represented as $X_{k \times m}$ with k m -dimensional row vectors. The \vec{X}_i characterizes the i^{th} object from the set P , and each element $X_{i,j}$ in \vec{X}_i corresponds to the j^{th} real-value feature ($j = 1, 2, \dots, m$) of the i^{th} pattern ($i = 1, 2, \dots, k$). The clustering algorithm tries to cluster the given n $X_{k \times m}$ matrix in to n clusters $C = \{C_1, C_2, \dots, C_n\}$ using a Partitional clustering algorithm such that the similarity of the patterns in the same cluster is maximum and patterns from different clusters differ as far as possible. The partitions should maintain three properties.

- 1) At least one pattern should be assigned to each cluster i.e., $C_i \neq \emptyset \forall i \in \{1, 2, \dots, N\}$.
- 2) No pattern should be common to two different clusters C_i, C_j i.e. $C_i \cap C_j = \emptyset \forall i \neq j$ and $i, j \in \{1, 2, \dots, K\}$
- 3) Each pattern should definitely be attached to a cluster i.e., $\cup_{i=1}^n C_i = P$.

Clustering process may also be defined as the optimization problem and it can be solved by using evolutionary search heuristics [6]. Since the given data set can be partitioned in a number of ways, maintaining all of the fore mentioned properties and a fitness function (some measure of the adequacy of the partitioning) must be defined [7]. The problem then turns out to be one of finding a partition C^* of optimal or near-optimal adequacy, as compared to all other feasible solutions $C = \{C^1, C^2, \dots, C^{N(k,N)}\}$, where

$$N(k, N) = \frac{1}{N!} \sum_{i=1}^N (-1)^i \binom{N}{i} (N - i)^i \tag{1}$$

is the number of feasible partitions. This is the same as

$$\text{Optimize}_f(X_{k \times m}, C) \tag{2}$$

Where C is a single partition from the set \mathbf{C} , and f is a statistical–mathematical function that quantifies the goodness of a partition on the basis of the distance measure of the patterns [8].

3 Shuffled Frog Leaping Algorithm (SFLA)

Shuffled frog leaping (SFL) is a population based, cooperative search metaphor inspired by natural memetics. The combination of EAs with local search was named memetic algorithms (MAs) in [10]. Its ability of adapting to dynamic environment makes SFL become one of the most important memetic algorithms. The frogs can communicate with each other, and can improve their memes by infecting (passing information) each other [11]. Improvement of memes results in changing an individual frog’s position by adjusting its leaping step size.

3.1 Steps in SFL Algorithm

The following are the steps involved [12] in the Shuffled Frog Leaping Algorithm (SFLA):

Step 1: Generate Random population. Initially generate the random population of k frogs within the feasible region.

Step 2: Rank Frogs. Calculate the fitness value of each frog according to the given problem and sort them according to their fitness values. Record the best frog’s position X_g in the entire population.

Step 3: Divide in to memplexes. Divide the frogs into p memplexes each holding q frogs such that $k = p \times q$. The division is done with the first frog going to the first memplex, second one going to the second memplex, the p^{th} frog to the p^{th} memplex and the $(p+1)^{th}$ frog back to the first memplex.

Step 4: Memetic evolutions. Identify X_b and X_w in each memplex which represents the Best and Worst frog respectively. Also the frog with the global best fitness X_g is identified. Then the position of the worst frog X_w for each memplex can be improved as follows:

$$B_i = \text{rand}(\cdot) \times (X_b - X_w) \tag{3}$$

$$\text{New } X_w = \text{Old } X_w + B_i \quad -B_{\max} \leq B_i \leq B_{\max} \tag{4}$$

Where, $\text{rand}(\cdot)$ is a random number between 1 and 0, and B_{\max} is the maximum allowed change in the frogs position. If the evolutions produce a better frog (solution), it replaces the older frog. Otherwise, X_b is replaced by X_g in (3) and the process is repeated. If non improvement becomes possible in this case a random frog is generated which replaces the old frog.

Step 5: Local Exploration. Perform the process from step 3 for a specific number of iterations.

Step 6: Shuffling. After the defined number of evolutionary steps in step 5, perform the process of shuffling by combining all the frogs in each memplex in to a single group.

Step 7: Check convergence. If the convergence criteria are satisfied, stop. Otherwise, return to step 2.

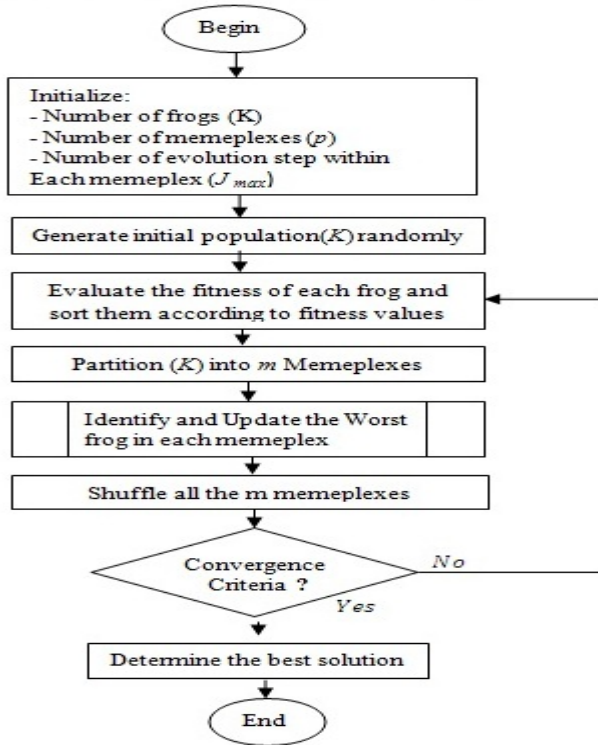


Fig. 1. SFL Algorithmic Flow Chart

4 MSFLA to Data Clustering

4.1 Modified SFLA (MSFLA)

The SFLA is one of the most Efficient Algorithm in reaching the global optimal. But, this Algorithm consumes more time because of the redundant steps present in it. This problem can be modified to increase its capability of reaching the global optimal faster. In Step (4) of SFL Algorithm (as mentioned in above section), if the fitness obtained from the new frog (*New X_w*) is not greater than the fitness of the old frog (*Old X_w*), a random value or a new frog is generating to replace the old frog (*Old X_w*). This generation of a new frog can affect the performance of the Algorithm. By generating a new frog again, the best value that has been achieved so far for that frog is not remained. The old frog which is converged up to that extent is going to change its value with a new value i.e. the frog is again going to rebirth by leaving all the work that has been done till now. Thus the new frog should start from beginning and it should take all the steps to converge again. This leads to low convergence rate and requires more time consumption to attain global optimal thus resulting the poor performance.

This problem should be avoided in order to improve the performance. This can be avoided by utilizing the services of best frog rather than generating a new frog. From, the Knowledge of Clonal selection [13], (it specifies replacement of the worst with the best value) we can avoid this problem. In the step (5) of the basic SFLA, if the fitness of the new frog $New X_w$ is not better, then it can be replaced by the value of the best frog X_b found in that Memplex. This replacement helps to start the next step of convergence from the best value rather than generating a new frog randomly. This may reduces the time consumption of the Algorithm since the frog is not generated randomly again. This modification in the Algorithm helps to attain solution faster with more convergence rate, and thus helps to attain the global optimal. The Pseudo code of the Modified step can be described as follows:

$$B_i = \tau \times (X_b - X_w) \tag{5}$$

$$New X_w = Old X_w + B_i \tag{6}$$

$$(-B_{max} \leq B_i \leq B_{max})$$

If (Fitness is not better) Replace X_b with X_g in equation (5)

Else (fitness is not better) Replace with best solution of that memplex

4.2 Application to Data Clustering

The Euclidian distance between the centroid vector of a class to the each data vector can be calculated as follows:

$$d(Z_p, m_j) = \sqrt{\sum_{k=1}^{N_d} (Z_{pk} - m_{jk})^2} \tag{7}$$

Where, Z_p denotes p^{th} data vector, m_j denotes the centroid vector of cluster j , N_d denotes the input dimension, i.e. the number of parameters of each data vector and k subscripts the dimension

The MSFLA based clustering is described as follows:

Step 1: Initialize the number of frogs N_{pop} , Number of memplexes N_m , Number of frogs in each memplex N_f , such that $N_{pop} = N_m \times N_f$

Step 2: Generate N_{pop} sample frogs $F(1), F(2), \dots, F(N_{pop})$ in the feasible space. The i^{th} frog can be represented as a vector with $F(i) = (F_i^1, F_i^2, \dots, F_i^{K^d})$ that is a candidate solution containing K cluster centers.

Step 3: Calculate the fitness value of each frog $F(i)$, fitness can be calculated in terms of Quantization Error (QE) of the Frog. The QE can be calculated by using the formula as follows:

$$q_e = \frac{\sum_{j=1}^{N_c} \sum_{z_p \in C_{ij}} d(z_p, m_j) / |C_{ij}|}{N_c} \tag{8}$$

Where N_c represents the number of clusters and $|C_{ij}|$ represents the number of data vectors that belong to the i^{th} particle and j^{th} cluster and $d(z_p, m_j)$ represents the Euclidian distance which can be calculated using Equation (7).

Step 4: Sort all the frogs in ascending order according to the fitness (q_e). Divide all the N_f Frogs in to N_m memeplexes such that, 1st frog goes to 1st memeplex and the second frog goes to second memeplex e.t.c.,. So, that($N_m + 1$) frog goes to 1st memeplex again.

Step 5: Identify the global best frog (F_g) and local best (F_{lb}), local worst (F_{lw}) frog in each memeplex. Change the value of worst frog of each memeplex by using Equation (5) and (6). If no improvement is seen in this case, the old frog can be replaced with the best frog in that memeplex.

Step 6: Shuffle the frogs in all the memeplexes i.e., combine all the Frogs in to a single array or in a single group.

Step 7: Check for the convergence criteria, if it is not satisfied, then go to step 3 again. Typically, the decision on when to stop is made by a pre-specified number of consecutive time loops or at least one Frog carries the “best memetic pattern” without change or a maximum total number of function evaluations can be defined.

5 Experimental Studies

In this section, we evaluate and compare the performances of the SFLA with MSFLA on Four real-world data sets (*Iris*, *Ripley's glass*, *wine* and *Wisconsin Breast cancer*) which are described as follows:

5.1 Experimental Settings

For all the results reported in this paper, the best values are obtained by choosing the following for both SFLA and MSFLA:

The initial Number of Frogs (N_f) = 20
 Number of Memeplexes (N_m) = 5
 Number of Frogs in each memeplex=4
 Number of iterations=300
 Number of iterations within each memeplex=30

5.2 Results and Discussion

MSFLA clustering is applied on four datasets where quality is measured according to the following two criteria:

- The quantization error (q_e) as defined in equation (8).
- The Standard Deviation of the Quantization Error: (std (q_e)).

The following Table 1 summarizes the results obtained from SFLA and MSFLA clustering algorithm for four Classification problems stated above. The results are tabulated by performing 30 independent runs and the Standard deviation (std) is calculated by taking the average in all iterations.

Table 1. Simulation Results of SFLA and MSFLA for Four Data Sets

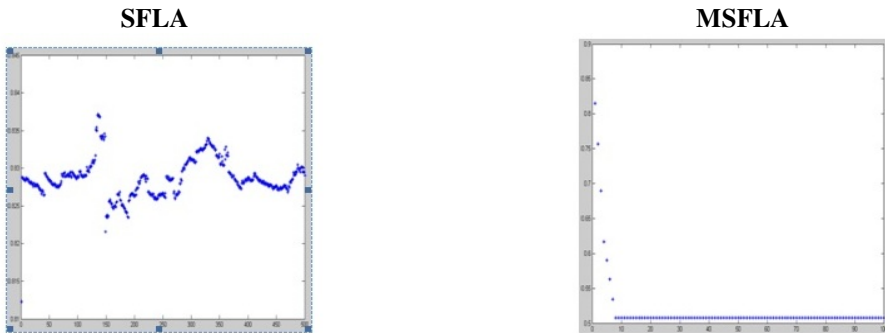
DATA SET	PARAMETERS	SFLA			MSFLA		
		BEST	AVG	WORST	BEST	AVG	WORST
IRIS	Mean q_e	0.6956	0.7132	0.8262	0.5024	0.5078	0.5162
	std(q_e)	0.0094	0.0137	0.0195	0.0011	0.0057	0.0125
GLASS	Mean q_e	1.2574	1.2907	1.3289	1.1134	1.1934	1.2171
	std(q_e)	0.0490	0.0426	0.0359	0.0214	0.0266	0.031
WINE	Mean q_e	101.7571	101.7923	101.8243	98.9923	99.7923	101.7643
	std(q_e)	0.0089	0.0354	0.0521	0.0072	0.0254	0.0292
BREAST CANCER	Mean q_e	6.6946	6.75422	6.8012	6.1946	6.2345	6.2467
	std(q_e)	0.0758	0.0783	0.0801	0.0385	0.0568	0.0790

5.3 Simulation Results

First, consider the fitness of solutions, i.e. the quantization error. For all Classification problems, the MSFL Algorithm has the smallest average Quantization error as compared to SFLA. The standard deviation (std) found to be very small, thereby indicating the better convergence of the algorithm.

In [14] by Naveen et al., has submitted the best Quantization error for the IRIS, GLASS data sets, which are found to be: 0.60702 and 1.38083 respectively. But, by using this MSFLA, the best results are obtained as: 0.5024 and 1.1134 respectively.

Therefore, this MSFLA produces better results in all Datasets. For the BREAST CANCER Dataset, the result obtained is closer to the best value. With reference to these criteria, the MSFLA approaches succeeded most in finding clusters with larger separation with minimum Quantization Error which yields the more compact clusters.

**Fig. 2.** IRIS Data Set

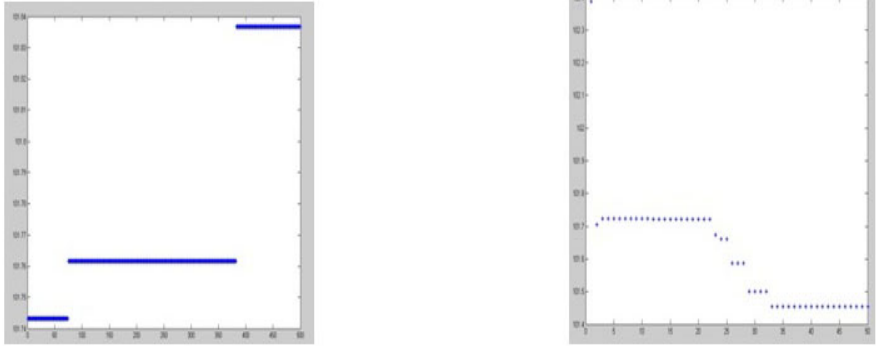


Fig. 3. WINE Data Set

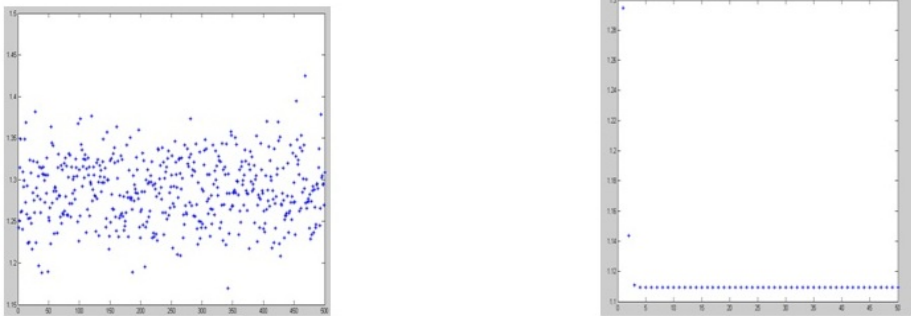


Fig. 4. RIPLEY'S GLASS Data Set

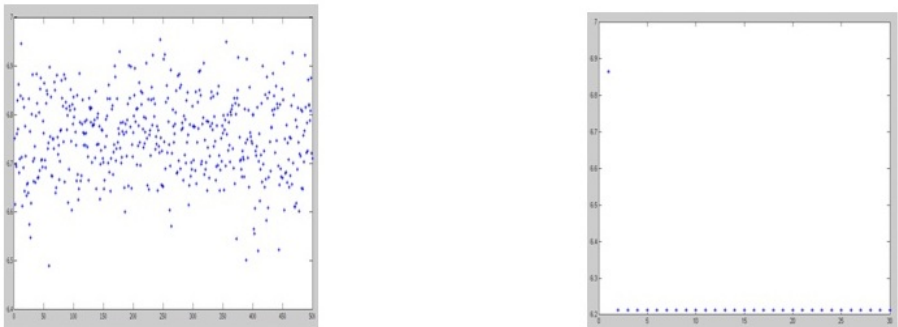


Fig. 5. BREAST CANCER Data Set

Fig. 2 to Fig. 5 illustrates the convergence behavior of the MSFL Algorithm for the four classification Problems. It summarizes the effect of varying the number of Iterations to the Quantization Error. The MSFL Algorithm exhibited a faster convergence to a smaller quantization error.

6 Conclusions

This paper proposed a new Algorithm, A Modified Shuffled Frog Leaping Algorithm (MSFLA) by overcoming the problems in the conventional SFLA. It also investigated the performance of MSFLA to cluster data vectors. Four Datasets were tested, namely Iris, Ripley's glass, wine and Wisconsin Breast cancer. This MSFLA approach has a better convergence to lower quantization errors and outperformed SFLA. Future studies will involve more elaborate tests on higher dimensional problems and large number of patterns. The MSFLA clustering algorithms can also be extended to dynamically determine the optimal number of clusters.

References

1. Moscato, P.: On evolution, search, optimization, GAs and martial arts: toward memetic algorithms, California Inst. Technol., Pasadena, CA, Tech. Rep. Caltech Concurrent Comput. Prog. Rep. 826 (1989)
2. Elbeltagi, E., Hegazy, T., Grierson, D.: Comparison among five evolutionary-based optimization algorithms. *Advanced Engineering Informatics* (19), 43–53 (2005)
3. Eusuff, M.M., Lansey, K.E., Pasha, F.: Shuffled frog-leaping algorithm: a memetic meta-heuristic for discrete optimization. *Engineering Optimization* 38(2), 129–154 (2006)
4. Liong, S.-Y., Atiquzzaman, M.: Optimal design of water distribution network using shuffled complex evolution. *J. Inst. Eng.* 44(1), 93–107 (2004)
5. Rao, M.R.: Cluster analysis and mathematical programming. *Amer. Stat. Assoc., J.* 66(335), 622–626 (1971)
6. Abraham, A., Das, S., Roy, S.: *Swarm Intelligence Algorithms for Data Clustering*
7. Jain, A.K., Murty, M.N., Flynn, P.J.: *Data clustering: A review*
8. Rao, M.R.: Cluster analysis and mathematical programming. *Amer. Stat. Assoc., J.* 66(335), 622–626 (1971)
9. Dasgupta, D.: *An overview of artificial immune systems and their applications. Artificial Immune System and Their applications* (1999b)
10. Zhen, Z., Wang, D., Liu, Y.: Improved Shuffled Frog Leaping Algorithm for Continuous Optimization Problem. In: *2009 IEEE Congress on Evolutionary Computation, CEC 2009* (2009)
11. Xue-Hui, L., Ye, Y., Xia, L.: Solving TSP with Shuffled Frog-Leaping Algorithm. In: *Eighth International Conference on Intelligent Systems Design and Applications*
12. Zhang, X., Hu, X., Cui, G., Wang, Y., Niu, Y.: An Improved Shuffled Frog Leaping Algorithm with Cognitive Behavior. In: *Proceedings of the 7th World Congress on Intelligent Control and Automation, Chongqing, China, June 25-27* (2008)
13. De Castro, L.N., Von Zuben, F.J.: Learning and Optimization using the clonal selection principle. *J. IEEE Transactions on Evolutionary Computation* 6(3), 239–251 (2002)
14. Naveen, I.G., Hali, Dessouki, E., Nahed, Mervat, A.N., Bakrawi, L.: Exponential Particle Swarm Optimization Approach for Improving Data Clustering. *International Journal of Electrical and Electronics Engineering*

Neighborhood Search Assisted Particle Swarm Optimization (NPSO) Algorithm for Partitional Data Clustering Problems

R. Karthi¹, C. Rajendran², and K. Rameshkumar³

¹Department of Computer Science and Engineering, Amrita Vishwa Vidyapeetham
Coimbatore – 641 105, India

²Department of Management Studies, IITM Chennai, Chennai – 600036, India

³Department of Mechanical Engineering, Amrita Vishwa Vidyapeetham
Coimbatore – 641 105, India

{r_karthi, k_rameshkumar}@cb.amrita.edu,
craj@iitm.ac.in

Abstract. New variant of PSO algorithm called Neighborhood search assisted Particle Swarm Optimization (NPSO) algorithm for data clustering problems has been proposed in this paper. We have proposed two neighborhood search schemes and a centroid updating scheme to improve the performance of the PSO algorithm. NPSO algorithm has been applied to solve the data clustering problems by considering three performance metrics, such as TRace Within criteria (TRW), Variance Ratio Criteria (VRC) and Marriott Criteria (MC). The results obtained by the proposed algorithm have been compared with the published results of basic PSO algorithm, Combinatorial Particle Swarm Optimization (CPSO) algorithm, Genetic Algorithm (GA) and Differential Evolution (DE) algorithm. The performance analysis demonstrates the effectiveness of the proposed algorithm in solving the partitional data clustering problems.

Keywords: Data Clustering, Particle Swarm Optimization, Trace Within criteria, Variance Ratio Criteria, Marriott Criteria.

1 Introduction

Clustering is a technique that attempts to group unlabeled data objects into clusters of similar objects. Each group called a cluster consists of objects that are similar to each other than objects belonging to other groups or clusters. A review of the clustering methods can be found in Xu and Wunsch [1]. Clustering techniques based on Evolutionary Computing and Swarm Intelligence algorithms have outperformed many classical methods of clustering. Two recent approaches, the particle swarm optimization and differential evolution has been used for solving the clustering problems. PSO introduced by Kennedy and Eberhart [2] is one of the most recent metaheuristics introduced to optimize various continuous nonlinear functions. Many variants of PSO algorithms were developed over the years and applied to solve the various optimization problems.

Literature review reveals that only few attempts have been made to solve the clustering problem using PSO algorithms, compared to other metaheuristics such as Genetic algorithm. The performance of the proposed Neighborhood search assisted Particle Swarm Optimization (NPSO) algorithms are studied with the consideration of three clustering metrics such as TRace Within criteria (TRW), Variance Ratio Criteria (VRC) and Marriott Criteria (MC) using real world data sets. The results are compared with the published results of the basic PSO, CPSO, GA and DE for all the clustering metrics.

The remaining part of the paper is organized as follows: Section 2 defines the formal clustering problem. In section 3, PSO algorithm applied to clustering problems are discussed. Section 4 describes the general structure of the proposed NPSO algorithm for data clustering. Section 5 presents the benchmark data sets and parameter settings used for experimentation. In Section 6 the results obtained by NPSO algorithm for different performance metrics are presented. In Section 7, conclusions from this study were made.

2 Data Clustering Problem Formulation

Notations Used:

N the number of data objects to be clustered.

D the dimension of each of the data objects.

K the number of clusters.

O set of N data objects to be clustered, $O = \{\vec{O}_1, \vec{O}_2, \dots, \vec{O}_N\}$.

n_k number of data objects allocated to cluster k .

Each data object is represented as: $\vec{o}_i = \{o_{i1}, o_{i2}, \dots, o_{id}\}$, where o_{id} represents value of data i at dimension d .

c set of K partitions with data objects assigned to each partition $C = \{C_1, \dots, C_K\}$.

z cluster centers to which data objects are assigned, $Z = \{\vec{Z}_1, \vec{Z}_2, \dots, \vec{Z}_k\}$.

Each cluster center is represented as:

$\vec{Z}_i = \{z_{i1}, z_{i2}, \dots, z_{id}\}$, where z_{id} represents value of cluster i at dimension d .

To solve the clustering problem, it is necessary to define a measure of similarity or dissimilarity between the data objects and centroids (see Jain et al., [3]). Different statistical criteria or fitness measures have been proposed. We have considered the fitness measures considered by Paterlini and Krink [4] for comparing partitions generated by different clustering algorithms.

3 PSO Algorithms in Clustering

PSO is a population based, cooperative search heuristic introduced by Kennedy and Eberhart [2] to find optimal or near solutions to an unconstrained optimization

problem. The ideas that underlie PSO are inspired by the social behavior of bird flocking and fish schooling. PSO is an iterative method that is based on the search behavior of the swarm in a multidimensional space. The algorithm iterates by updating the velocities and positions of the particles until the stopping criteria is met. Several variations [5-12] of this PSO scheme have been proposed in the literature for solving continuous optimization problems.

PSO based clustering algorithm was first proposed by Van der Merwe et al. [13]. Xiao et al. [14] proposed a hybrid approach based on Self Organization Maps (SOM) and PSO to cluster the gene data. Chen and Ye [15] employed a representation in which each particle corresponds to the centroids of the clusters. Orman et al. [16] proposed a dynamic clustering system based binary PSO and K-means algorithm. The algorithm automatically identifies the number of clusters in the data and employs a validity index to access the clusters obtained. Using the concept of socio-cognition, Cohen et al., [17] proposed a modified PSO algorithm called Particle Swarm Clustering (PSC) algorithm. Paterlini and Krink [4] compared the performance of K-means, Differential evolution (DE), Random Search (RS), PSO and GA for partitional clustering problems. The empirical results show that PSO and DE perform better compared to GA and K-means algorithms.

Jarboui et al. [18] compared the performance of CPSO with GA for the TRW and VRC measures. Jarboui et al. [18] proposed an approach based on Combinatorial Particle Swarm Optimization (CPSO) algorithm. Each particle is represented as a string of length n (where n is the number of data points) and each element of the string represents the cluster number assigned to the data point. CPSO algorithm was compared with GA by solving data clustering problems. A Discrete Particle Swarm Optimization algorithm using the principle of probability to construct the velocity of particle followed by a search scheme to construct the clustering solution was proposed by Karthi et al. [19]. A detailed comparative evaluation of PSO algorithms for partitional data clustering were reported by Karthi et al. [20].

4 General Structure of Proposed NPSO Algorithm

Notations used:

- t iteration counter.
- T maximum number of iterations.
- S swarm size.
- D maximum numbers of dimensions in each data object.
- K maximum number of clusters.
- N number of data objects to be clustered.

The data objects to be clustered are represented as a set: $O = \{\bar{o}_1, \bar{o}_2, \dots, \bar{o}_N\}$

Each data object is represented as: $\bar{o}_i = \{o_{i1}, o_{i2}, \dots, o_{id}\}$, where o_{id} represents value of data i in dimension d .

x_n^t position of Current particle n (Current) at iteration t .

$f(x_n^t)$ value of objective function for particle n (Current) at iteration t .

Each particle is represented as $x_n = \{x_{n11}, x_{n12}, \dots, x_{n1d}, x_{n21}, x_{n22}, \dots, x_{n2d}, \dots, x_{nk1}, x_{nk2}, \dots, x_{nk d}\}$, where $x_{nk d}$ represents value of cluster k in dimension d for particle n .

v_n^t velocity of particle n at iteration t .

Each velocity is represented as $v_n = \{v_{n11}, v_{n12}, \dots, v_{n1d}, \dots, v_{nk1}, v_{nk2}, \dots, v_{nk d}\}$, where $v_{nk d}$ represents value of cluster k in dimension d for particle n .

p_n Best position of particle n till iteration t .

$f(p_n)$ value of objective function for p_n .

g G-Best position of the whole swarm.

$f(g)$ value of objective function for g .

The general structure of the NPSO algorithm is given below:

Step 1: Generate $2s + 1$ initial particle x_n^t according to the swarm size s .

Step 2: Apply the centroid updating scheme to each of the generated $2s + 1$ initial particles x_n^t . Evaluate the objective value of the particles $f(x_n^t)$ based on fitness measures.

Step 3: For each of the x_n^t initial particles apply the neighborhood search schemes.

Step 4: Initialize x_n^t (Current particle), p_n (Particle Best) and g (Global Best).

Step 5: Initialize v_n^t , the velocity vector of the particles.

Step 6: $t = t + 1$

Step 7: For each particle $n = 1 \dots s$

Step 7.1: Update the position and velocity vectors of the Current particle x_n^t .

Step 7.2: Apply the centroid updating scheme to x_n^t . Evaluate the objective value of the particles $f(x_n^t)$ based on the considered fitness measure.

Step 7.3: For each x_n^t generated in Step 7.2 apply the neighborhood search schemes.

Step 7.4: Update the Best (p_n) and G-Best (g) particle positions.

Step 8: If $t < T$ go back to Step 6
else Step 9.

Step 9: Return G-Best (g) particle.

4.1 Generation of Initial Particles

In the proposed NPSO algorithm, we generate $2s + 1$ initial particles. The particles are initialized within the maximum and minimum limit in each dimension. Cluster centers of the particles are randomly initialized using the data points taken from the corresponding datasets. This method of randomly selected data point initialization ensures diversity in the population to obtain high quality results.

4.2 Centroid-Updating Scheme

The main task of the centroid updating scheme is to find best possible set of κ centroids, one for each cluster. The scheme works as follows: The clustering is done by assigning each data object to the nearest centroid encoded in the particle. The centroids in the particle are updated based on mean of the data objects in the cluster. It is to be noted that in the conventional PSO algorithm, the centroid encoded in the particle is not updated after every iteration. The proposed centroid updating scheme for the data clustering problems is given below:

Algorithm: Centroid Updating Scheme

/* Input: A Centroid set $Z^t = \{z_1, z_2, \dots, z_K\}$ $Z_k = \{z_1, z_2, \dots, z_d\}$ where d is the number of dimensions in the data object */

Step 1: Assign each data object in the dataset to a centroid in Z^t .

Step 2: Set $t = t + 1$

Step 3: Update Z^t such that it is the mean vector of all the data objects in that

$$\text{cluster. } z_{kd}^t = \frac{\sum_{i=1}^N (O_{id} \times \delta_{ik})}{n_k}, \text{ where } \delta_{ik} = \begin{cases} 1 & \text{data } i \text{ belongs to cluster } k \\ 0 & \text{other wise} \end{cases}. n_k \text{ is the numbers of data objects in cluster } k.$$

Step 4: If $Z^t = Z^{t-1}$, then Z^t is a converged centroid, Stop
Else go to step 2.

4.3 Neighborhood Search Schemes

The neighborhood search schemes proposed in this paper search for better solution in the proximity of the particle considered. The solution space is explored through a sequence of positions that are close to the particle centroids. Neighborhood search may be viewed as a procedure that proposes the direction in which the centroids should move for greater improvement in partitioning. These schemes accelerate the convergence and possibly lead to superior solution by escaping local optima. The two Neighborhood Search Schemes proposed in this paper are discussed below:

4.3.1 Neighborhood Search Scheme 1

Neighborhood search scheme 1 is characterized by the following strategy. Each of the clustering centroid solution is a point in the solution space. An optimal solution is assumed to be in the proximity of the current centroid. The solution space is explored through construction of centroid solutions that are within the limits of the current cluster diameter. Thus centroids move to better positions within its neighborhood.

Algorithm

Z_{kd}^{\min} represents minimum centroid set in each dimension d based on the data objects assigned to cluster k .

Z_{kd}^{\max} represents maximum centroid set in each dimension d based on the data objects assigned to cluster k .

O_{id}^k represents data object i assigned to centroid k , d is the dimension of the data object.

U_1 uniformly distributed random number between 0 and 1.

Step 1: Assign each data object O_i in dataset to a centroid in Z .

Step 2: For each centroid $Z_k \in Z, k = 1 \dots K$

For each dimension $d \in Z_k, d = 1 \dots D$

Set $Z_{kd}^{\min} \in \min\{O_{id}^k\}, i = 1 \dots N$; Set $Z_{kd}^{\max} \in \max\{O_{id}^k\}, i = 1 \dots N$

Step 3: For each centroid $Z_k \in Z$

For each dimension $d \in Z_k$

$Z'_{kd} = Z_{kd} + m \times U_1 \times (Z_{kd}^{\max} - Z_{kd}^{\min})$, where $m \in \{1, -1\}$

Step 4: Apply the centroid updating scheme to Z'_{kd} . Evaluate $f(Z^l_{kd})$.

return Z' .

4.3.2 Neighborhood Search Scheme 2

The centroids generated are within the neighborhood of the initial centroid. The neighborhood search schemes allow exploration to new areas within the neighborhood to uncover the near optimal value.

Algorithm

Step 1: Assign each data object O_i in dataset to a centroid in Z .

Step 2: Update the centroid using the following heuristic

For each centroid $Z_k \in Z$

For each dimension d in Z_{kd}

$$Y_{kd} = \left(\frac{\sum_{i=1}^N O_{nd} \times \delta_{nk}}{\sum_{i=1}^N \delta_{nk}} \right), \text{ where } \delta_{nk} = \begin{cases} 1 & \text{data } n \text{ belongs to cluster } k \\ 0 & \text{other wise} \end{cases}$$

Δ_{nk} - the Euclidean distance between the data objects n and cluster k

Step 3: Generate new neighborhood centroids from Y_{kd} using the following heuristics.

For each centroid $Z_k \in Z, k = 1 \dots K$

For each dimension d in $Z_k, d = 1 \dots D$

$$Y'_{kd} = U_1 \times Y_{kd} + Z_{kd} * (1 - U_1).$$

For each centroid $Z_k \in Z, k = 1 \dots K$

For each dimension d in Z_{kd} , $d = 1 \dots D$

$$Y_{kd}'' = Z_{kd} + (-1 + 3 \times U_1) \times (Y_{kd} - Z_{kd}).$$

Step 4: Apply the centroid updating scheme to Y' and Y'' . Evaluate $f(Y')$ and $f(Y'')$

Step 5: If $f(Y') < f(Y'')$ return Y' Else return Y''

5 Performance Analysis of NPSO Algorithm

5.1 Experimental Setup

To evaluate the performance of the proposed NPSO algorithm the results were compared with the published results of GA, DE, Basic PSO, CPSO and CK-PSO algorithm. We have implemented Constriction type PSO algorithm called CK-PSO algorithm, proposed by Clerc and Kennedy [8] for solving data clustering problems. We have considered the data sets used by Paterlini and Krink [4] available at UCI repository of machine learning database (http://www.daimi.au.dk/~krink/medoid-evolution/me_benchmark_data) to evaluate the performance of the NPSO algorithm.

Several experiments were conducted to determine parameter setting that yields the best performance with respect to all the datasets. In this paper we have considered population size as 50 and maximum functional evaluations as 10000 (termination criterion) to report the results. All the algorithms considered in this paper were repeated 30 times for each of the benchmark instances. CK-PSO and NPSO algorithms were implemented in C++. All experiments with CK-PSO and NPSO were run in windows XP on a HP PC, with Intel P4, 1.8 GHz processor.

6 Results and Discussions

Performance of the NPSO algorithm is measured based on mean best fitness values, mean relative percentage increase in objective function values and the best, average and worst fitness value of TRW, VRC and MC measures considered in this paper. The mean best fitness values of the CK-PSO, NPSO algorithms are compared with the published results of GA, DE and BPSO algorithms. The results are reported in the Table 1. Mean relative percentage increase in the objective function values are also calculated to relatively evaluate the heuristics considered in this paper. The results are shown in Table 2. The results shown in the Table 1 and Table 2 indicates the better performance of the NPSO considered in this study for most of the benchmark instances compared with B-PSO, CK-PSO, GA and DE algorithm. The best, average and worst fitness values for the NPSO, CPSO and CK-PSO algorithms are reported in the Table 3. For all the data sets, comparison of the results reveals that NPSO algorithm finds better quality solution for TRW measure. For the VRC measure CK-PSO and NPSO algorithms performs better than the CPSO algorithm.

Table 1. Comparison of mean best fitness of NPSO, CK-PSO with GA, DE and Basic PSO algorithm

Dataset	Criteria	Mean best fitness [#]				
		GA	DE	B-PSO	CK-PSO	NPSO
Iris	MC	0.1984	0.1984	0.1984	0.0877	0.0653
	TRW	7885.14	7885.14	7885.14	7885.28	7885.14
	VRC	561.63	561.63	561.63	628.68	613.49
Cancer	MC	0.3565	0.3546	0.3527	0.2006	0.1679
	TRW	19323	19323	19324	19323	19323
	VRC	1026.26	1026.26	1026.26	1621.19	1644.38
Glass	MC	0.02661	0.01984	0.03176	0.0146	0.0056
	TRW	341.09	336.06	339.04	345.69	336.06
	VRC	121.94	124.62	122.74	144.54	141.52
Vowel	MC	0.3199	0.2906	0.3032	0.2502	0.1618
	TRW	30943106	30690785	30734068	31128244	30686225
	VRC	1450.45	1465.55	1463.33	1578.82	1534.59

[#]Mean results were reported to 4 decimal places accuracy over 30 simulation runs.

Table 2. Mean relative percentage increase in objective value of heuristics

Dataset	Criteria	Mean relative percentage increase in objective function value				
		GA	DE	B-PSO	CK-PSO	NPSO
Iris	MC	203.87	203.87	203.26	34.35	0
	TRW	0	0	0	0	0
	VRC	10.66	10.66	10.66	0	2.42
Cancer	MC	112.33	111.20	110.07	19.46	0
	TRW	0	0	0.01	0	0
	VRC	37.59	37.59	37.59	1.41	0
Glass	MC	372.35	252.17	463.76	159.62	0
	TRW	1.50	0	0.89	2.86	0
	VRC	15.64	13.78	15.08	0	2.09
Vowel	MC	97.76	79.65	87.44	54.69	0
	TRW	0.84	0.01	0.16	1.44	0
	VRC	8.13	7.17	7.31	0	2.80

It is to be noted that number of functional evaluations required for reporting the results of CK-PSO and NPSO algorithms are relatively less compared to the CPSO algorithm. We run the CK-PSO and NPSO algorithms for a maximum of 10000 functional evaluations, where as CPSO algorithm has been allowed to run for a maximum of 250000 functional evaluations to report the results. The improved performance of the NPSO algorithm over the CK-PSO algorithm is due to the implementation of the neighborhood search scheme and the centroid updating scheme. Both the CK-PSO and NPSO algorithms outperform the CPSO algorithm in terms of quality of the solution and convergence speed.

Table 3. Worst, best and average fitness of NPSO, CPSO and CK-PSO algorithms

Data set	Criteria	CPSO			CK-PSO			NPSO		
		Worst	Average	Best	Worst	Average	Best	Worst	Average	Best
Iris	TRW	7885.14	7885.14	7885.14	7885.56	7885.28	7885.14	7885.14	7885.14	7885.14
	VRC	561.63	561.63	561.63	628.13	628.68	629.14	602.94	613.49	622.32
Cancer	TRW	19323	19323	19323	19323	19323	19323	19323	19323	19323
	VRC	1026.26	1026.26	1026.26	1621.10	1621.19	1621.21	1533.66	1644.38	1672.34
Glass	TRW	375.79	336.17	336.06	383.45	345.69	336.06	336.06	336.06	336.06
	VRC	123.62	124.62	124.44	119.78	144.54	159.45	137.80	141.52	147.41
Vowel	TRW	30712302	30691704	30689508	34999368	31128244	30686212	30686228	30686225	30686222
	VRC	1465.47	1465.65	1465.88	1468.67	1578.82	1610.54	1516.60	1534.59	1558.25

7 Conclusion

In this paper we have investigated the problem of data clustering by using particle swarm optimization algorithms. A new variant of PSO algorithm is developed for data clustering. The effectiveness of the proposed algorithm is evaluated by comparing NPSO with the basic PSO algorithm, CK-PSO, CPSO, GA and DE algorithms. The proposed NPSO perform better for most of the benchmark datasets for VRC, TRW and MC criterion. NPSO improve the best-known solution available in the literature for the TRW, VRC and MC measures for all the benchmark datasets. The better performance of the NPSO algorithm is due to the implementation of novel neighborhood search scheme and the centroid updating scheme.

References

1. Xu, R., Wunsch, D.: Survey of Clustering Algorithms. *IEEE Transactions on Neural Network* 16(3), 645–678 (2005)
2. Kennedy, J., Eberhart, R.C.: Particle Swarm Optimization. In: *Proceedings of the 1995 IEEE International Conference on Neural Networks*, vol. 4, pp. 1942–1948. IEEE Press, Piscataway (1995)
3. Jain, A.K., Murty, M.N., Flynn, P.J.: Data Clustering: A Review. *ACM Computing Survey* 31(3), 264–323 (1999)
4. Paterlini, S., Krink, T.: Differential evolution and particle swarm optimization in partitionial clustering. *Computational Statistics & Data Analysis* 50(5), 1220–1247 (2006)
5. Shi, Y., Eberhart, R.: A modified particle swarm optimizer. In: *Proceeding of the 1998 IEEE World Congress on Computational Intelligence*, pp. 69–73. IEEE Press, Piscataway (1998)
6. Shi, Y., Eberhart, R.: Empirical study of particle swarm optimization. In: *Proceeding of the 1999 IEEE World Congress on Evolutionary Computing*, pp. 1945–1950. IEEE Press, Piscataway (1999)

7. Eberhart, R., Shi, Y.: Tracking and optimizing dynamic systems with particle swarms. In: Proceeding of 2001 IEEE World Congress on Evolutionary Computing, pp. 94–100. IEEE Press, Piscataway (2001)
8. Clerc, M., Kennedy, J.: The particle swarm-explosion, stability, and convergence in a multidimensional complex space. *IEEE Transactions on Evolutionary Computation* 6, 58–73 (2002)
9. Mendes, R., Kennedy, J., Neves, J.: The fully informed particle swarm: simple, maybe better. *IEEE Transaction on Evolutionary Computing* 8(3), 204–210 (2004)
10. Ratnaweera, A., Halgamuge, S.K., Watson, H.C.: Self organization hierarchical particle swarm optimizer with time varying acceleration coefficients. *IEEE Transaction on Evolutionary Computing* 8(3), 240–255 (2004)
11. Janson, S., Middendorf, M.: A hierarchical particle swarm optimizer and its adaptive variants. *IEEE Transaction on System, Man and Cybernetics (Part B)* 35(6), 1272–1282 (2005)
12. Chatterjee, A., Siarry, P.: Non Linear inertia weight variation for dynamic adaptation in particle swarm optimization. *Computers and Operations Research* 33, 859–871 (2006)
13. Van Der Merwe, D.W., Engelbrecht, A.P.: Data clustering using particle swarm optimization. In: Proceedings of IEEE Congress on Evolutionary Computing 2003, Canberra, Australia, pp. 215–220 (2003)
14. Xiao, X., Dow, E.R., Eberhart, R.C., Miled, Z.B., Oppelt, R.J.: Gene Clustering Using Self-Organizing Maps and Particle Swarm Optimization. In: Proc of the 17th International Symposium on Parallel and Distributed Processing (PDPS 2003). IEEE Computer Society, Washington, DC (2003)
15. Chen, C.Y., Ye, F.: Particle swarm optimization algorithm and its applications to clustering analysis. In: Proceedings of IEEE International Conference on Networking, Sensing and Control, pp. 789–794 (2004)
16. Orman, M.G.H., Salman, A., Engelbrecht, A.P.: Dynamic clustering using Particle Swarm Optimization with application in image segmentation. *Pattern Analysis and Application* 8(4), 332–344 (2005)
17. Cohen, S.C.M., De Castro, L.N.: Data Clustering with Particle swarms. In: IEEE Congress on Evolutionary Computations, Vancouver, Canada (2006)
18. Jarboui, B., Cheikh, M., Siarry, P., Rebai, A.: Combinatorial particle swarm optimization(CPSO) for partitional clustering problem. *Applied Mathematics and Computation* 192, 337–345 (2007)
19. Karthi, R., Arumugam, S., RameshKumar, K.: Discrete Particle Swarm Optimization algorithm for Data Clustering. *Studies in Computational Intelligence, SCI*, vol. 236, pp. 75–88 (2009)
20. Karthi, R., Arumugam, S., Rameshkumar, K.: Comparative evaluation of Particle Swarm Optimization Algorithms for Data Clustering using real world data sets. *International Journal of Computer Science and Network Security* 8(1), 203–212 (2008)

Portability in Incremental Compilers

P.R. Mahalingam and C. Unnikrishnan

Rajagiri School of Engineering and Technology, Cochin, India
{prmahalingam,cnunnikrishnan}@gmail.com

Abstract. Compilers form the backbone of the field of computing nowadays. Computation has diversified from the initially acclaimed fields of research, to a tool for entertainment and application development. As computers gained popularity, users demanded more from it, and consequently, they began to develop applications that suited their needs on their own. For this, programming languages were developed. But as needs diversified, people needed more features, which led to more and more languages, stressing on the usability and friendliness. But in all this rapid development, languages began to branch out, with less common features and overlapping. So, it became difficult to integrate and incorporate features between languages. This resulted in development of cross-language interfaces. Also, with the diversification of operating systems, target languages began to vary, and interfaces found it difficult to cope with changes in platform. So, the search was on to find a technology that can solve the language and platform barrier, ultimately making applications robust, user-friendly and powerful. Here, we examine one possible solution for this. We stress upon the intermediate code which is the result of partial compilation. If we can make this code language and platform independent, we can easily solve the issue. This is particularly useful in incremental compilers, which face the challenge of patching up executable code modules.

Keywords: incremental compilers, intermediate code, portability.

1 Introduction

Compilers [1] is a concept familiar to all computer programmers. They are responsible for converting (translating) the input source code to the target machine language. In the broader sense, language translators [1] are vital in executing programs - may it be compilers or interpreters.

Now, for more complex execution [4], we may have to combine multiple source languages together. So, they may need compilation in parts [3]. As a result of this, it may not be feasible to combine all the executable modules together to get the final application. Incremental Compilers theoretically is supposed to do this - manipulate executable cde modules. ie, support modification of modules, adding new modules and removing existing ones.

But the major problem comes at this juncture. When we have to accommodate multiple languages, there arises the case that the compiled code may be

incompatible. So, we have to ensure that the source languages don't compile to conflicting formats, or incompatible forms. In many cases, the compatibility may be handled by interfaces that perform the necessary translation. But this is not always possible.

This is especially advantageous considering the case of software patches^[3]. If the complete Incremental Compilation can actually be implemented, we can actually embed the patch codes to the actual software package itself. But here too, another issue is prominent - portability. This is since the actual patch (or any increment) has to run on all platforms, with minimal rework. Here, we are analyzing the portability aspect of Incremental Compilation, without any loss of compatibility.

2 Related Work

Right now, there are methods that can implement the portability to a good degree. Some of them are:

1. **Recursive Neural Networks**^[5], which apply pattern recognition to train neural networks, and extract required information from the program. Here, the problem is the loss of precision.
2. **Intermediate Code Processing**, which uses language and architecture-specific code readers and writer modules to generate the required format. But here, the problem is the need for recompilation to implement portability. It has been implemented in one of the first Incremental Compilers, built for a functional language called "Hope", which has been named as "IncH" (Incremental Hope)^[6].
3. **Flexible Back-End**, which has a different back-end for each architecture, and can generate the corresponding code. This is popularly used in document processing, like UQ*^[7].
4. **Bytecode**, which was put forward by Java^[11] is now synonymous with portability. But the problem is that the bytecode format is cast for use with Java only, and cannot be extended for other languages, causing compatibility issues.

So, each of them have their share of advantages and disadvantages.

3 Proposed Solution

Here, we amalgamate all the existing solutions to get one that can give the advantages of the discussed ones to the maximum. This gives rise to a well-known concept called External Data Representation.

3.1 External Data Representation

The concept is derived from the methods discussed before. We take the concepts as follows:

The Neural Network can be used to identify patterns in the language that are vital to the syntax. Then, an implementation like IncH to generate a common intermediate representation, and the flexible back-end can generate the final architecture-dependent code. The formatting should contain all the necessary information, like in Bytecode. So, we get an implementation like in Fig 1.

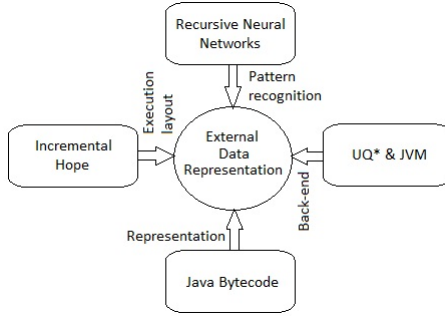


Fig. 1. Deriving proposed solution from exiting ones

But the next hurdle comes at the point where the actual representation has to be chosen. The representation should be such that it has a universal reach, and all the languages should be translated to this representation.

The search for such a representation ends at XML. This is since XML offers universal markup opportunities. Using markup, we can easily define semantic tags, which can help in representing the actual meaning behind every statement. For example, *printf("Hello World");* in C and *System.out.println("Hello World");* in Java mean the same to the underlying compiler - *Output "Hello World" to the screen*. So, by extracting such semantic information from the language, we can effectively perform language-independent translation.

3.2 Proposed Architecture

The above discussion leads to a possible architecture like in Fig 2.

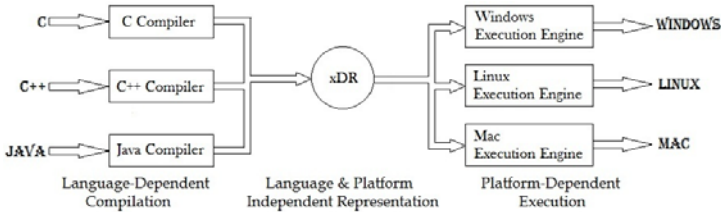


Fig. 2. Skeleton of operation

The advantage is evident from the diagram itself. The language compilers will generate the XML intermediate code instead of the machine code, which forms the input of the platform-dependent execution engines at each platform. This is something like bytecode, extended to all languages.

3.3 XML-Based Code

As seen before, the whole compilation process has been split into 2 phases. So, the intermediate code will be based on XML. But then, there arises a main issue - how to translate the source code into XML. This is important since XML is less related to the machine format, and the efficiency of this conversion decides how fast the code will run at the back-end. Also, use of good semantic tags will drastically improve the readability of the code. Also, the use of an efficient translation will help in patching up code easily.

Existing XML translations. There are some existing methods to translate the program into XML now itself.

1. **gccxml**[\[9\]](#) - directly translates C++ code into XML. The main problem here is that it also takes into consideration the headers included and a lot of memory usage details. Also, it is built for browsers, and doesn't facilitate further execution or patching.
2. **Annotated trees**[\[10\]](#) - the syntax tree is translated to the XML format, using the actual grammar using which it was generated.
For example, $expr \rightarrow expr \ op \ expr$; corresponds to the XML format

```
<expr>
<expr>..</expr>
<op>..</op>
<expr>...</expr>
</expr>
```

This is efficient, but too dependent on the source language grammar, making patching difficult.

3. **Direct translation**[\[9\]](#) [\[10\]](#) - the intermediate formats available like Triples, Quadruples, etc. can be directly formatted as XML, but that also cuts down the patching ability.

After seeing all this, we can find a new solution to efficiently translate programs to XML.

Proposed XML translation. Here, the XML code is generated from the AST (Abstract Syntax Tree) itself. So, it is independent of the actual grammar, and solely depends on the actual program tokens. For that, we have a set of stages[\[2\]](#) defined for developing a basic compiler, which is seen in Fig 3. They cover most of the basic features that have to be implemented in any compiler that is developed, alongwith the sequence that has to be followed in implementation.

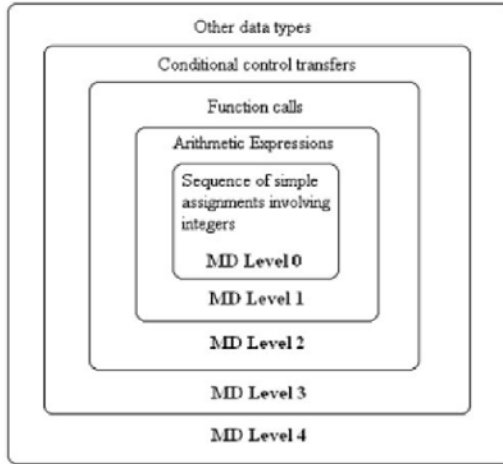


Fig. 3. Stages of implementation

So, for each stage, we have to define our own formats, that accommodate the required structures.

Some defined formats. Some of the formats can be defined like follows.

Assignment Statement

```
<EQU><LHS>...</LHS>
<RHS>...</RHS></EQU>
```

Variables -

```
<VAR>name</VAR>
```

Constants -

```
</CONST>value</CONST>
```

Arithmetic and Logical operations

```
<OPER_TYPE><OPD1>...</OPD1>
<OPD2>...</OPD2></OPER_TYPE>
```

Functions

```
<FUNCTION><NAME>function name</NAME>
<PARAM_LIST>
<PARAM>...</PARAM>
...
</PARAM_LIST>
<BODY>...</BODY></FUNCTION>
```

Function Calls

```

<CALL><FNAME>function name</FNAME>
<CALL_PARAMS>
<PARAM>...</PARAM>
...
</CALL_PARAMS></CALL>

```

IF Statement

```

<IF><CONDITION>...</CONDITION>
<THEN>...</THEN>
<ELSE>...</ELSE></IF>

```

FOR Statement

```

<FOR><INITIALIZATION>...</INITIALIZATION>
<CONDITION>...</CONDITION><INCREMENTS>...</INCREMENTS>
<BODY>...</BODY></FOR>

```

WHILE / DO-WHILE Statement

```

<{WHILE | DO-WHILE}><CONDITION>...</CONDITION>
<BODY>...</BODY></{WHILE | DO-WHILE}>

```

And finally, we can implement the Symbol Table to support datatypes, type checking and type matching. For that, we can define a format like

```

<SYMTAB>
<SYMBOL><NAME>...</NAME><TYPE>...</TYPE></SYMBOL>
...
</SYMTAB>

```

The SYMTAB is always preferred to be given as a separate file since that enables better patching. Also, the functions are totally independent. So, adding a new function to the code simply involves pasting the function on to the existing code.

Back-end. The back-end can be constructed from any language capable of reading XML, or even we can design custom XML readers. Then, the readers can be designed to perform necessary processing in case of corresponding tags. This is since, the innermost tags correspond to the leaves while outermost tags are for the root. So, executing inside-out performs correct execution.

Also, it is far easier executing the code if the tree is reconstructed from the XML, rather than executing as such.

This leads us to an extension of this concept.

So, a single Java execution engine [\[1\]](#) can also be made, eliminating the need of a separate execution system for each platform.

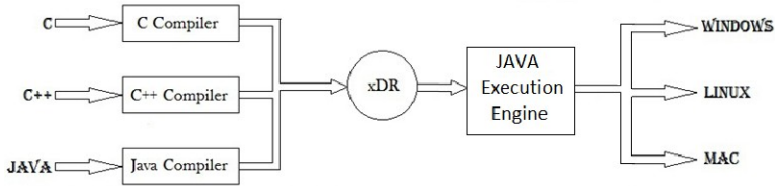


Fig. 4. Extension of the concept

4 Results and Observations

Simulations have been run on a new language¹ being defined specifically for this purpose. As of now, datatypes haven't been implemented. Consider a sample input code

```
if (a>b) {a=b+c;}
else {a=b-c;}
```

This generates the output

```
<IF><CONDITION><GT><OPD1><VAR>a</VAR></OPD1>
<OPD2><VAR>b</VAR></OPD2></GT></CONDITION>
<THEN><EQU><VAR>a</VAR>
<RHS><ADD><OPD1><VAR>b</VAR></OPD1>
<OPD2><VAR>c</VAR></OPD2></ADD></RHS>
</EQU></THEN>
<ELSE><EQU><VAR>a</VAR>
<RHS><SUB><OPD1><VAR>b</VAR></OPD1>
<OPD2><VAR>c</VAR></OPD2></SUB></RHS>
</EQU></ELSE></IF>
```

Observations - The code size is very high compared to the input code, and most conventional compilations. But this is one compromise that has to be coped with to avoid any loss of information. Also, as we increase the nesting of tags, the code becomes easier to optimize.

XML as such cannot be optimized easily since it is not related to the machine code side. But if we have enough tags, we can actually trace them to the exact operands, and perform some amount of optimization. In other cases, we can actually delegate the process to the back-end execution engine.

Advantages - As proposed, this enhances the portability and compatibility by a huge degree, especially since the functional modules can be made independent of each other. So, we can theoretically manipulate the executables at will. Also, the arrival of a Java execution engine, as an extension, will revolutionize the way we think of programming and execution in a completely different way.

¹ The language has restricted features. Support included for expressions, loops and conditional constructs only.

Also, the independent functions can be distributed across sites, where they execute in parallel, and pass calls and results using messages, like RPC, forming Parallel Incremental Compilation [8].

5 Conclusion and Future Work

So, a new representation that can aid in better portability and compatibility when it comes to Incremental Compilation has been proposed. This XML-based representation can be used as a semantic tree to understand the underlying execution of the programs, and help in porting them across platforms with ease. In the future, better optimization methods can be developed, and the compiled code size can be made comparable to conventional compilers.

References

1. Aho, A., Sethi, R., Lam, Ullmann, J.: *Compilers - Principles, Techniques and Tools*, pp. 15–67. Pearson Education, London (2008)
2. Deshpande, S., Khedker, U.P.: *Incremental Machine Description for GCC*, pp. 1–5. Department of Computer Science, Indian Institute of Technology, Bombay (2007)
3. Loudon, K.C.: *Compiler Construction Principles and Practice*, pp. 45–61. Prentice Hall, Englewood Cliffs (2006)
4. Taff, L.M., Sporrel, F.: An Incremental Compiler for Data Acquisition. *IEEE Transactions for Nuclear Science NS-26*(4), 19–26 (1979)
5. Costa, F., Frasconi, P., Lombardo, V., Sturt, P., Soda, G.: Ambiguity Resolution analysis in incremental parsing of Natural Language. *IEEE Transactions on Neural Networks* 16(4), 24–32 (2005); Department of Information Systems, Universit'a di Firenze, Italy,
6. de Lamadrid, J.G., Zimmerman, J.: IncH: an Incremental Compiler for a Functional Language, pp. 1–5. Bowie State University and Goucher College, Maryland (2007)
7. Cook, P., Welsh, J., Hayes, I.J.: Building a flexible incremental compiler backend. Technical Report SSE-2005-02, Division of Systems and Software Engineering Research, School of Information Technology and Electrical Engineering, The University of Queensland, pp. 1-8 (May 2005)
8. Gaffer, N.M.: Parallel Incremental Compilation. Technical Report 349, Department of Computer Science, University of Rochester, pp. 45–96 (June 1990)
9. GCC-XML: GCC-XML, An Introduction, <http://www.gccxml.org/HTML/Index.htmlgccxml> (2010)
10. Power, J.F., Malloy, B.A.: Program Annotation in XML: A parse-tree based approach. In: *Proceedings of the Ninth Working Conference on Reverse Engineering (WCRE 2002)*, pp. 1–9 (2002)
11. ByteCodes: “Java Bytecodes” Compilation, *JavaWorld* (2010), <http://www.javaworld.com/jw-09-1996/jw-09-bytecodes.html>

Test Case Optimization Using Artificial Bee Colony Algorithm

AdiSrikanth, Nandakishore J. Kulkarni, K. Venkat Naveen,
PuneetSingh, and PraveenRanjanSrivastava

Computer Science & Information System Department, Bits Pilani – 333031 (India)
{adisriikanth, kvenkatnaveen, nandkishorejk,
puneet.mir, praveenrsrivastava}@gmail.com

Abstract. Software Testing is one of the integral parts of software development lifecycle. Exhaustive testing on software is impossible to achieve as the testing is a continuous process. Using this as a constraint, software testing is performed in a way that requires reducing the testing effort but should provide high quality software that can yield comparable results. To accomplish this optimized testing, a software test case optimization technique based on artificial bee colony algorithm is proposed here. Based on intelligent behavior of honey bee, this method generates optimal number of test cases to be executed on software under test (SUT). This approach is qualified by self-organization, robustness and focuses on generation of paths derived from cyclomatic complexity. The resulting solution guarantees full path coverage.

Keywords: Software Testing, ABC (Artificial Bee Colony Optimization), Test-Case, Agents, Path-Coverage, optimal path, cyclomatic complexity, CFG (Control Flow Graph), test data.

1 Introduction

Software Testing [1] involves identifying the conditions where software deviates from its normal behaviour or exhibit different activity in contrast to its specification. Software testing has a major role in software development lifecycle (SDLC) [2] to develop the high quality software product [1]. At least, software testing consumes 50% of the development cost. Software testing [1] in the context of this paper is focused around path testing, a form of functional testing which concerns about path coverage [2] that is finding all possible execution sequence within code.

Software testing [1] is an optimization [2] process where multiple variables are taken into account to generate the efficient number of test cases and to provide the optimal path. Infinite number of test cases will be generated by use of exhaustive testing [1] [2]. However, only a part of them will be more effective if implemented in testing the software. This approach leads the tester to obtain the comparable results using reduced and optimal set of test cases.

Independent paths are useful in testing the path coverage and code coverage. Independent paths of SUT are identified using the CFG (control flow graph) [1]. CFG is a representation of different paths using the graphical notation. Paths generated by

CFG are traversed by the test cases [1] to be prioritized based on some test adequacy criteria or fitness value [3]. This perception drove the tester to develop an alternative approach based on search based intelligent selection and prioritization of test case. The meta-heuristic methods[5][7][9][10]tried in past to solve this problem yielded success, but with serious drawbacks such as operating on dynamic data sets may be difficult and can also converge early towards the solution which may not be useful later in further testing process. Using polynomial bounded computation many of the complex multivariable optimization problems [4] cannot be solved accurately. The optimal number of test-cases in covering the SUT provides better performance than other meta-heuristic approaches.

The ABC algorithm [3][11][12], a meta-heuristic approach [10]is used to generate the optimal number of test-cases which are sufficient to cover the paths generated by using the CFG (control flow graph)[1].The independent paths are considered to be an important criterion to generate the effective test cases for path coverage [1][2]. The effective test-cases are those which cover all the nodes of the path in less time.

The proposed tool focuses more on the independent paths generated by using CFG to find optimal path[3][11], the paths are traversed using the test data providing path coverage and these test data sets are optimized using ABC[3][11].

Finally the paper compares ABC with GA (genetic algorithm) [5][7][8] and ACO (ant colony optimization) [9][10][13] to differentiate the performance and efficiency issues.

2 Background Work

There are several research approaches that have been proposed which attempted to reduce the number of test-cases [2] using ACO (ant colony optimization) [9] [13] and GA (genetic algorithms) [5] [6] [7].

GA (genetic algorithms) [5][6] is a population based approach uses the genetic operators to derive the fitness function and calibrate the efficient test-cases. But, there are some demerits like the external optimization in the GA that provides a single solution where the local optimum can be modified for the worst cases rather than the global optimum, which makes result less stable. And GA not quite effectively solves the problems involving only one measure for judging purpose, as it is not sufficient to converge to a solution.

ACO [9] [13] shows the behaviour of the ants which is based on the randomization of the ant behaviour. The time to process and generate the effective number of test-cases wastes large amount of computing resources. It also produces inefficient higher length test sequences and repetition of nodes in same sequence and eliminates the use of fitness value.

The proposed approach is aimed to generate the optimal number of test cases and to achieve greater path coverage. To solve this problem ABC[3][11] approach is used where the bee agents [3] gather the food sources (test cases) and then calibrate the fitness function[11]which is in-turn used to identify the optimal test cases with highest path-coverage[2] using less amount of resources.

3 Artificial Bee Colony (ABC) Algorithm

The Artificial Bee Colony (ABC) [3] [11] algorithm is a population-based evolutionary, stochastic method that can be implemented on a wide range of problems, including global optimization. It's also a type of swarm intelligence algorithm for the same. ABC algorithm is derived and motivated by the behaviour of Bees [12]. Bees follow a certain protocol based searching and extracting honey. The search behaviour and its intelligence propelled to use ABC in rendering optimised Test Suite. The optimised Test suite will not only cover all the code, it also can be used for path testing.

The ABC algorithm is biologically enlivened technique of swarm intelligence for searching [12]. It's about honey bees' collective foraging and work distribution skill to amass surplus nectar for their survival [3]. In ABC, the colony of the artificial bees comprises three groups of bees: employed bees, onlookers and scouts. Food Positions are altered by Artificial Bees with time. Bees aim is to find out the places of food sources with high nectar amount. Employed and Onlookers bees select food sources depending on the experiences of themselves and their nest mates.

4 Proposed Strategy

The proposed approach attempts to solve the problem of Test Case Optimization by applying ABC algorithm. In the proposed approach each test case would represent a food source and the goal of the approach would be to find best food sources i.e. test cases with maximum coverage.

The food source position represents to a possible solution of the optimization problem and the nectar amount corresponds to the quality (fitness) of the associated solution. The basic steps of ABC algorithm [3] [11] [12] are:

Station the Search bees onto the food sources and ascertain their nectar amount.

- Compute the probability value of the sources with which they are opted by the Selector bee.
- The process of the traversing of the food sources by desolated bees is halted.
- The scouts are sent into the search area for exploring new food sources randomly.
- The best food source found so far is retained in the memory

As explained earlier the entire functionality is divided as the functions of three kinds of bees. The three kinds of bees act as [12]:

- Search bee which act as Search agent
- Selector bee which acts as Selector agent
- Scout bee which acts as Replace agent

The principle and working of the proposed approach:

The approach is a population based approach [9]. Here each test case comprises the possible solution in the optimization problem. The fitness value of each test case equates to the quality or the fitness of the colligated solution. An outline of the solution is presented the figure 1.

The work proposes that the Optimised test suite generated by using the algorithm will contain all possible independent paths along with its test data.

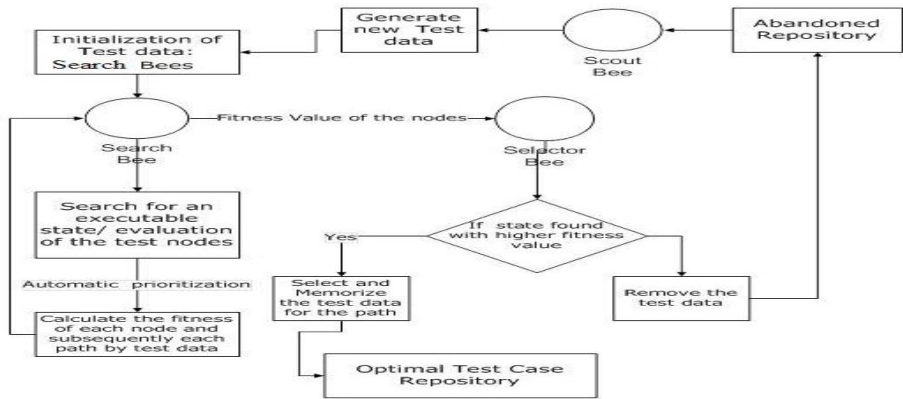


Fig. 1. Architecture/High level flow of the algorithm

The test data will be the required input to be given to the SUT, for travelling along the path and vice versa.

Initially the program is fed to the Test Case optimization tool, which converts the program into a corresponding Control Flow Graph (CFG). Now from control flow graph, the independent paths from the start node to the end node are generated. Each independent path would comprise number of normal nodes and predicate nodes. Every independent path would represent a Test Case [1]. Now the ABC algorithm is applied to generate an Optimal Test suite by generating optimal test data which would traverse through the independent paths and hence into to the test cases.

Consider the figure 1. Here the search bee acts as search agent which searches for the execution state of the SUT and also initialises the test cases with the initial test data with the help of equivalence partitioning and boundary value analysis [19]. Then the search agent calculates the fitness value of each test node by computing the coverage of each node. This process is repeated until an executable state of SUT is found.

Then the search bee passes the fitness value of the traversed nodes/neighbouring nodes to the selector agent. The selector bee compares the fitness value of nodes with the fitness value of the neighbouring nodes. If the node’s fitness value is found greater than the neighbouring node’s fitness value, the node’s information is stored in the optimal test case repository. The node whose fitness value is found less is pushed to the abandoned repository.

The scout bee generates the new set of test data from the abandoned repository and again the same procedure of search bee is repeated. The algorithm for test case optimization using Artificial Bee Colony optimization approach is presented in the next section.

5 ABC Optimization Algorithm

Initialize the Test Cases by initial test data generated by the tool. Initial test data is generated by the equivalence partitioning and boundary value analysis methods.

Step 1: The search bee employ test cases to the first executable node of the program. The fitness value of the test cases and the node information is returned to the employed bees.

Step 2: The Selector Bee takes the node information and fitness value of the test cases as the input and evaluates the fitness of each test case taken from the employed bees. Then the test case with the highest fitness value is stored in the memory or repository.

Step 3: Then the neighbouring nodes of the covered nodes are searched by the employed bees and the fitness value of the current test case is explored against the neighbouring nodes by the onlooker bee.

Step 4: The node with the highest fitness value is picked out and tagged with the current node which forms the test path. This information of the test path is stored in the Optimal Test repository.

Step 5: Rest of the nodes other than the covered node and the test cases other than the selected test case are abandoned and they are stored in a temporary abandoned repository.

Step 6:

Repeat the steps 3 to 6 till the test path is not complete

Else the nodes and the test cases from the abandoned repository are selected for the next test path generation by the employed bees

Step 7: The scout bee bring forth a new population of test cases and replaces the test cases of the abandoned repository with the new test cases if the onlooker bee finds the selected test cases are not efficient.

Until the user defined termination criterion is met.

In order to implement any algorithm, the algorithm must be converted into the pseudo code before programmatically developing an application. The detailed pseudo code of the Test Case Optimization algorithm using ABC approach is presented in the following section.

The following is the detailed algorithm

1. Initialise the test cases which is performed by the search bee(see figure 1)
2. Search for an executable state and evaluate the test nodes
3. Initialise the current traversal path as cycle=1
4. Repeat
5. Produce new test cases v_{ij} in the neighbourhood of x_{ij} for the the search agent using the formula

$$X_{ij} = \min_j + \text{rand}(0, 1) * (\max_j - \min_j) \quad [15]$$
 - X_{ij} is the initial test case
 - \min_j is the minimum no. of test cases
 - \max_j is the maximum no. of test cases
 - $\text{rand}()$ is a random number generation function which selects either 1 or 0 randomly.
6. Generate the test data using equivalence partitioning and boundary value analysis [19].
7. Apply greedy selection process [18] for the generated test data.
8. Traverse the SUT with the generated Test Data and calculate the fitness value.

9. The onlooker selects the test cases with the highest fitness value and abandons the rest
10. The traversal process is carried out till a particular test data with 100% fitness value and 0% fitness value is produced

$$P_i = (fit_i) / \sum (fit_i) \quad [16]$$

Where P_i is the probability function which signifies the probability with which the i^{th} test data traverses a independent test path successfully.

11. Add the test case to the optimal repository
12. In the next iteration scouts generate the new test data and go to step 5

To implement the above algorithm, our approach uses the ABC Test Suite Optimization tool to optimize the Test Cases by employing the ABC algorithm. The tool considers a program as an input to generate independent paths. Using the generated independent paths Test Cases are traversed along the paths with the help of ABC algorithm. By doing so the test cases with maximum coverage (High fitness Value) are recognized. Finally the optimal Test Suite is generated.

6 Case Study

The proposed algorithm is applied and tested for various real time applications, for illustration purpose this paper is used Quadratic Equation solving program

Two case studies are discussed in the paper.

Case Study

This Case Study’s SUT code is available in APPENDIX. The SUT uses 3 inputs of a quadratic equation and generates its roots. Nodes are represented in the code at APPENDIX. The Control Flow Graph (CFG) for the APPENDIX is given below.

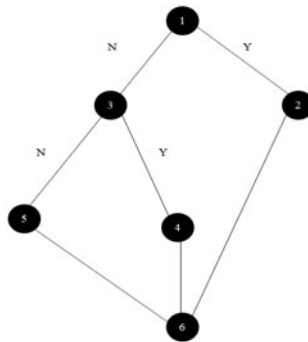


Fig. 2. CFG of the Quadratic Equation Code

The program in the appendix will find out the root of the quadratic equation and specify whether they are real or imaginary.

Source Code of the program is provided as input to the tool. Links to the different nodes are provided as a text document which also acts as an input to the tool. The onlooker bee generates different paths of the SUT. For the above program of 6 nodes following are the generated independent paths.

1. 1-2-6 2. 1-3-5-6 3. 1-3-4-6

According to proposed algorithm following steps happen.

1. All the 3 independent paths are initialized as test cases by the search bee
2. Initialise the current traversal path as cycle=1
3. Repeat

4. Produce new test cases, x_{ij} for the search agent using the formula

$$X_{ij} = \min_j + \text{rand}(0, 1) * (\max_j - \min_j) \quad (1)$$

X_{ij} is the initial test case

\min_j is 1 here (as minimum numbers of test cases are 1)

\max_j is 3 here (as maximum numbers of test cases are 1)

Hence, $X_{ij} = 1 + 0 * 3 - 1 = 1$

So, path number 1 i.e., 1-2-6 is selected by selector bee as it traverses the path 1-2-6 completely.

5. Generate the test data using equivalence partitioning and boundary value analysis [19] by analysing the code.

6. Traverse the SUT with the generated Test Data and calculate the fitness value.

7. The onlooker selects the test cases with the highest fitness value and abandons the rest

For the path 1-2-6, the test data (1,4,3), (1,4,4), (2,3,4) and (2,3,1) are tested by traversing along the path 1-2-6 which are generated using equivalence partitioning and boundary value analysis. In test data (1, 4, 3) covered node 1, node 2 and node 6 completely. Hence the fitness value of (1, 4, 3) with respect to path 1-2-6 is 100%.

Similarly the test data (1, 4, 4), (2, 3, 4) and (2, 3, 1) are traversed through the path 1-2-6. But they don't cover the test nodes 1, 2 and 6 completely. Hence the fitness value for the data (1, 4, 4), (2, 3, 1), (2,3,4) are 0%.

So test data for the case study will be 1, 4, 3 and 1, 4, 4 (One of the test data which didn't traversed the complete path 1-2-6).

8. In the next iteration scouts generate the new test data and go to step 5 of the above mentioned algorithm

9. Optimised test data is added to the repository

10. Go to Step 1 and continue the process for other test paths.

S.No.	Test Sequence	Test Cases	Coverage or Fitness Value
1	1-2-6	1,4,3	100
1	1-2-6	1,4,4	0
2	1-3-5-6	2,3,4	100
2	1-3-5-6	2,3,1	0
3	1-3-4-6	2,4,2	100
3	1-3-4-6	2,4,3	0

Where fitness values is 100 for complete traversal and 0 for incomplete

So, the optimal test suite is generated successfully for the above program.

7 Analysis and Comparison

With Bee System, the probability of falling into a local optimum is low because of the combination of local and global search since the aim of the algorithm is to improve the local search ability of genetic algorithm without degrading the global search ability.

From Figure 3, observe that, as the number of cycle’s increases, percentage of paths covered increases and after a particular threshold, it achieves complete coverage and maintains consistency. Path Coverage for the ABC testing based on cycles.

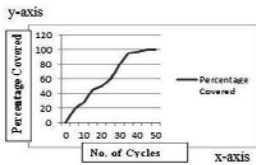


Fig. 3. Number of Cycles vs. Percentage Covered

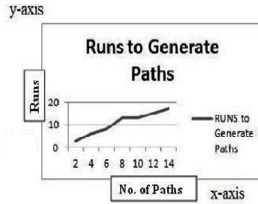


Fig. 4. Number of Paths vs. Runs

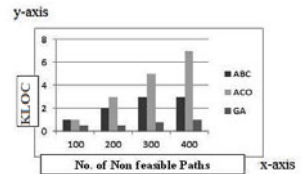


Fig. 5. Number of non-feasible paths vs. KLOC

The graph from Figure 4 plots the number of paths and runs taken to achieve the number of paths which increased incrementally. The line of graph is having lesser slope.

The Figure 5, graph plots comparison between ABC, ACO [9] [13], and GA [5] [6] for number of non-feasible paths against the LOC. This graph testifies that the ABC distinguishes and accounts less number of non-feasible paths. Thus more number of feasible paths with test data is covered. But GA doesn’t find any non-feasible paths. Non-feasible paths are the paths which cannot be traversed using any generated test data. Hence, the non-feasible path is technically accessible but logically cannot be traversed.

The previous approach[15] of ABC didn’t mention anything about feasible paths but the proposed approach in this paper will return non feasible paths if any path generated using ABC is not reachable using any test data.

8 Conclusions and Future Work

In this paper, test case optimisation is done using ABC. The proposed algorithm uses ABC approach to work with faster and efficient ways for generating optimised test suite. Also the algorithm is compared to previous algorithms and older ABC approaches, and the results confirm that the proposed algorithm is beating other algorithms and approaches. In the proposed ABC approach, optimised test suite is generated for each independent path of the program where each path will have two types of data. One is the test data which traverses the path with 100% coverage and the other which traverses the path with 0% coverage. The bee searches a path until we find the 100% fitness valued (coverage) test data. The bee searches a path until we find the 100 fitness valued test data. Prioritization in ABC is presently done using factors like code complexity, application feasibility and implementation complexity.

In future, more factors can be added to increase the efficiency of the ABC approach and give more prioritised and optimised test suite.

References

1. Pressman, R.S.: *Software Engineering: A practitioners Approach*, 6th edn., ch.1 (33-47),13(387-406),14(420-444). McGraw-Hill, New York (2007)
2. Sommerville: *Software Engineering*, 8th edn., ch.1 (27-42), 11(265-288), 23(561-589). Pearson, London (2007)
3. Basturk, B., Karaboga, D.: A powerful and efficient algorithm for numerical function optimization: artificial bee colony(ABC) algorithm. In: *Proceedings of the IEEE Swarm Intelligence Symposium*, pp. 459–471. IEEE, Indianapolis (2006)
4. KDeb: *Multi-Objective optimization using Evolutionary Algorithms*, 1st edn., ch.4 (140). John Wiley & Sons, UK (2001)
5. Christoph, C., Michael Gary, E., Michael, M., Schatz Curtis, A., Walton, C., et al.: *Genetic Algorithms for Dynamic Test Data Generation*. In: *Proceedings of the 12th International Conference on Automated Software Engineering (ASE)*, pp. 307–308. IEEE, Washington, DC, USA (1997)
6. Johnson, D.L., Pinglikar, J., Watkins, A., et al.: *Breeding Software Test Cases with Genetic Algorithms*. In: *Proceedings of the 36th Annual Hawaii International Conference on System Sciences*, p. 10. IEEE, Hawaii (2003)
7. Michael, C.C., McGraw, G., Schatz, M.A.: *Generating Software Test Data by Evolution*. *IEEE Transactions on Software Engineering*, 1085–1110 (2001)
8. Lin, J.C., Yeh, P.L.: *Using Genetic Algorithms for Test Case Generation in Path Testing*. In: *Proceedings of the 9th Asian Test Symposium (ATS)*, pp. 241–246. IEEE, Taiwan (2000)
9. Wagner, I.A., Lindenbaum, M., Bruckstein, A.M.: *ANTS: Agents, Networks, Trees, and Subgraphs*. In: Dorigo, M., Di Caro, G., Stützle, T. (eds.) *Special Issue on Ant Colony Optimization, Future Generation Computer Systems*, pp. 915–926. ACM, North Holland (2000)
10. Lawrence, S.: *Resource constrained project scheduling: an experimental investigation of heuristic scheduling techniques (Supplement)*. PhD thesis, Graduate School of Industrial Administration, Carnegie-Mellon University, pp. 1–15. ACM, Pittsburgh(1984)
11. Dahiya, S., Chhabra, J., Kumar, S.: *Application of Artificial Bee Colony Algorithm to Software Testing*. In: *Software Engineering Conference (ASWEC)*, pp. 149–154. IEEE, Auckland (2010)
12. Karaboga, D., Bahriye, A.: *A survey: algorithms simulating bee swarm intelligence*. *Artificial Intelligence Review*, 61–85 (2009)
13. Alaya, I., Solnon, C., Ghedira, K.: *Ant Colony Optimization for Multi-Objective Optimization Problems*. In: *Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence*, pp. 450–457. IEEE, Washington, DC, USA (2007)
14. Karaboga, D., Basturk, B.: *On The PerformanceOf Artificial Bee Colony (ABC) Algorithm*. In: *Applied Soft Computing*, pp. 687–697. Elsevier Science Publishers B. V., Amsterdam, ACM, Netherlands (2008)
15. Jeya Mala, D., Mohan, V.: *ABC Tester - Artificial Bee Colony Based Software Test Suite Optimization Approach*. *International Journal of Software Engineering, IJSE*, 1–33 (2009)
16. JeyaMala, D., Kamalapriya, M., Shobana, R., Mohan, V.: *A Non-Pheromone based Intelligent Swarm Optimization Technique in Software Test Suite Optimization*. In: *Intelligent Agent & Multi-Agent Systems (IAMA)*, pp. 1–5. IEEE, Chennai (2009)

17. Liu, Z., Gu, N., Yang, G.: An automate test case generation approach: using match technique. In: The Fifth International Conference on Computer and Information Technology (CIT), pp. 922–926. IEEE, Shanghai (2005)
18. Kodaganallur, V., Sen, A.K.: Greedy by Chance - Stochastic Greedy Algorithms. In: Sixth International Conference on Autonomic and Autonomous Systems, pp. 182–187. IEEE, Washington, DC, USA (2010)
19. Reid, S.C.: An empirical analysis of equivalence partitioning, boundary value analysis and random testing. In: Software Metrics Symposium, Proceedings, Fourth International, Albuquerque, NM, USA, pp. 64–73 (1997)
20. Korman, T.H.: Introduction to Algorithms, 2nd edn., ch.21. McGrawHill, New York (2001)

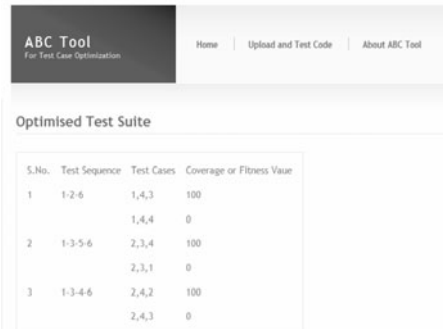
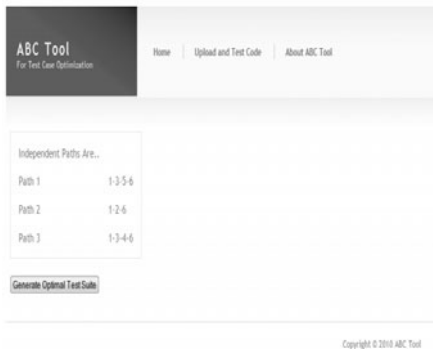
Appendix – 1

```

1. int main()
{
    int A,B,C;
    float disc,deno,x1,x2;
printf("\n\n PROGRAM TO FIND THE
ROOTS OF A QUADRATIC
EQUATION ");
    printf("\n\n\t ENTER THE
VALUES OF A,B,C...");
    scanf("%d,%d,%d",&A,&B,&C)
;
    disc = (B*B)-(4*A*C);
    deno = 2*A;
    if(disc > 0)
2.     {
        printf("\n\t THE
ROOTS ARE REAL ROOTS");
        x1 = (-
B/deno)+(sqrt(disc)/deno);
        x2 = (-B/deno)-
(sqrt(disc)/deno);
        printf("\n\n\t THE
ROOTS ARE...: %f and %f\n",x1,x2);
    }
3.     else if(disc == 0)
4.     {
        printf("\n\t THE
ROOTS ARE REPEATED ROOTS");
        x1 = -B/deno;
        printf("\n\n\t THE
ROOT IS...: %f\n",x1);
    }
5.     else
        printf("\n\t THE
ROOTS ARE IMAGINARY ROOTS\n");
        printf("\n-----");
        getch();
6. }

```

Screen Shots



Adaptive Power Allocation in CI/MC-CDMA System Using Genetic Algorithms

Santi P. Maity and Sumanta Hati

Department of Information Technology
Bengal Engineering and Science University, Shibpur, Howrah, India
santipmaity@it.becs.ac.in, Sumanta.hati@rediffmail.com

Abstract. Multicarrier code division multiple access (MC-CDMA) is one of the effective techniques for high data rate transmission with large user capacity. The use of carrier interferometry (CI) codes further improves this user capacity relative to the conventional spreading codes. However, CDMA is interference limited and needs proper power allocation for channel capacity (data transmission rate) improvement. According to Shannon's theorem, channel capacity i.e data transmission rate increases with the increase in transmit power due to an increased signal-to-noise-ratio (SNR). MC-CDMA system being multiuser communication system, both signal and interference power are increased with the increase in transmit power which in turn demands optimum power allocation for transmission capacity improvement. In this paper, we develop an adaptive transmit power allocation technique for carrier interferometry multicarrier code division multiple access (CI/MC-CDMA) system using Genetic Algorithms (GA). GA is used to find the optimum transmitted powers that maximize the channel capacity as well as reduce BER values. Objective function is defined as maximum channel capacity in a power constraint scenario. We have reported the performance of both the non-power adaptive and GA based proposed adaptive systems. Simulation results show that significant improvement in BER performance and transmission capacity are achieved in the present system.

Keywords: Multicarrier code division multiple access, carrier interferometry code, power allocation, genetic algorithms.

1 Introduction

On growing demand of data intensive applications such as interactive and multimedia services, the need of reliable and high rate data transmission over a wireless mobile channel becomes essential. However, high data rate transmission increases the inter symbol interference (ISI) caused by the time dispersive nature of the radio mobile channel. In code division multiple access (CDMA), multiple number of users share the same bandwidth at the same time through the use of (near) orthogonal spreading codes. So it has potential to provide higher user capacity compared to its other close competitors such as time division multiple access (TDMA) and frequency division multiple access (FDMA). Multicarrier code division multiple access (MC-CDMA), a

combination of orthogonal frequency division multiplexing (OFDM) and CDMA, provides a flexible multipath propagation and reduces ISI by introducing a guard interval which is a cyclic extension of any multicarrier signal. The usage of carrier interferometry (CI) code in MC-CDMA system enhances the user capacity relative to the other conventional spreading codes [1] [2].

According to Shannon channel capacity theorem [3], increase in signal power increases signal-to-noise ratio (SNR) for a given noise power. This in turn improves the system's channel capacity (c) i.e. data transmission rate for single user communication. However, in the presence of multiple users, if all users try to increase their data rates by increasing their transmit power, the users interfere and both the signal and interference power increase. Consequently, the signal-to-interference plus noise power-ratio (SINR) and therefore the users' rates saturate at a constant value [2]. In other words, high SNR values of other users act as strong interfering effect to the nearby users in multiuser environment transmitting in the same set of subcarriers. This suggests that optimum power allocation is essential in order to increase channel capacity and to reduce BER values. This is often done by exploiting the channel state information (CSI) that is fed back from transmitter to receiver [4].

In this paper, we propose an efficient power allocation scheme in CI/MC-CDMA system using genetic algorithms (GA) [5]. GA is used here to optimize the transmitted power for each user in respective subcarrier in order to reduce interfering effect with a hope to increase channel capacity. This is done assuming that CSI is fed back from the receiver to the transmitter end. Two antenna diversity is used at receiver and higher weight factor is assigned for the antenna system that offers higher value of signal-to-interference ratio (SIR). Experiment is carried out through simulation on frequency selective Rayleigh fading channel. Results show that channel capacity and bit error rate (BER) performance obtained for the adaptive system is significantly better than the non-power adaptive system.

The rest of the paper is organized as follows: Section 2 presents review of the prior works, limitations and scope of the present work. Section 3 describes system model for synchronous CI/MC-CDMA uplink with antenna diversity. Section 4 presents proposed GA based adaptive power allocation algorithm, while Section 5 reports performance evaluation. Conclusions are drawn in Section 6 along with the scope of future work.

2 Review of Prior Works, Limitations and Scope of the Present Work

Channel capacity improvement by adaptive power allocation in multicarrier system is an important research topic and several solutions are reported in literature. In [6], Shen et al propose an optimal power allocation algorithm in their OFDM system that maximizes the sum capacity and at the same time each user maintains a required data rate. Luo et al [7] studied the power allocation problem in decode and forward cooperative relaying system and propose an equal power allocation with a channel selection algorithm based on minimizing the outage probability. An algorithm for optimal transmitted power control is proposed by Zhang et al in [8]. They consider controlled transmit power for a two-band system as a linear function of the power attenuation difference (between the two bands) for a large range of these attenuation

differences. Jang et al [9] present an optimal transmit power adaptation method to maximize the channel capacity of a multicarrier system in a time-variant fading channel under the constraint of average transmit power.

It is reported in the literature that the joint capacity maximization and BER minimization in multiuser communication system is a non-convex integer problem and a closed form solution is difficult to find, suboptimal solutions are proposed in many cases. Moreover, majority of the above algorithms suffer from exponential computation cost with the increase of number of subcarriers and users. Attempt is also made for some of these works to reduce the computation complexity from exponential to linear with the number of subcarriers (users) in OFDM (CDMA) system. However, when the concepts are used for implementation in MC-CDMA system, power allocations are not efficient due to the presence of in-band multiple access interference (MAI). Hence, development of an efficient power allocation algorithm with low computation cost for multiuser multicarrier system is highly demanding and GA may be used to find optimal set of values.

3 System Model with Antenna Diversity

CI/MC-CDMA is a combination of CDMA and OFDM with the CI codes (spreading codes) applied in frequency domain. The CI code, for the k^{th} user, corresponds to $[1, e^{j\Delta_k}, \dots, e^{j(N-1)\Delta_k}]$ [1], $\Delta\theta_k = 2\pi k/N$. The CI codes [1] of length N have a unique feature that allows CI/MC-CDMA systems to support ‘ N ’ users orthogonally, and then as system demand increases, codes can be selected to accommodate up to an additional $(N-1)$ users pseudo-orthogonally. Additionally, there is no restriction on the length ‘ N ’ of the CI code (i.e., $N \in \mathbb{I}$) making it more suitable for diverse wireless environments.

3.1 Transmitter Model

In CI/MC-CDMA transmitter [10], the incoming data $a_k[n]$ which is the n -th bit for the k^{th} user, is transmitted over N narrowband subcarriers each multiplied with an element of the k -th user’s spreading code. Binary phase shift keying (BPSK) modulation is assumed here i.e., $a_k[n] = \pm 1$.

The mathematical form of total transmitted signal for K number of users is

$$S(t) = \sum_{k=1}^K \sum_{i=0}^{N-1} a_k[n] \beta_{k,i} \text{Cos}(2\pi f_i t + i\Delta\theta_k). P(t) \tag{1}$$

where $f_i = f_c + i\Delta f$ and $P(t)$ is a rectangular pulse of duration T_b , $\beta_{k,i}$ is the amplitude of k^{th} user and i^{th} subcarrier which controls the transmitted power. The value of Δf is selected in such a way that the carrier frequencies f_i , $i = 0, 1, \dots, N-1$ are orthogonal to each other, i.e. $\Delta f = 1/T_b$.

3.2 Channel Model

An uplink model has been considered where all the user’s transmissions are assumed to be synchronized for simplification of analysis, although, this condition is very

difficult to achieve practically. It is also assumed that every user experiences an independent propagation environment that is modeled as a slowly varying multipath channel. Multipath propagation in time translates into frequency selectivity in the frequency domain [3]. Frequency selectivity refers to the selectivity over the entire bandwidth of transmission and not over each subcarrier transmission. This is because $1/T_b \ll (\Delta f_C) < BW$, where Δf_C is the coherence bandwidth and BW is the total bandwidth.

3.3 Receiver Model

The received signal corresponds to

$$r(t) = \sum_{k=1}^K \sum_{i=0}^{N-1} a_k[n] \beta_{k,i} \alpha_{k,i} \cos(2\pi f_i t + i\Delta\theta_k + \Phi_{k,i}) \cdot P(t) + \eta(t) \quad (2)$$

where $\alpha_{k,i}$ is the Rayleigh fading gain, $\Phi_{k,i}$ is uniformly distributed phase offset, both the parameters indicate the value for k -th user in the i -th carrier and $\eta(t)$ represents additive white Gaussian noise (AWGN). The received signal is first projected onto N -orthogonal carriers and is despread using j^{th} users CI code resulting in $r^j = (r_0^j, r_1^j, r_2^j, \dots, r_{N-1}^j)$. The term r_i^j corresponds to

$$r_i^j = a_j[n] \beta_{j,i} \alpha_{j,i} + \sum_{k=1, k \neq j}^K a_k[n] \beta_{k,i} \alpha_{k,i} \cos(i(\Delta\theta_k - \Delta\theta_j) + \Phi_{k,i} - \Phi_{j,i}) + \eta_j \quad (3)$$

where η_j is a Gaussian random variable with mean 0 and variance $N_0/2$. Now, a suitable combining strategy is used to create a decision variable D_j , which then enters a decision device that outputs \hat{a}_j . Minimum mean square error combining (MMSEC) is employed as it is shown to provide the best performance in a frequency selective fading channel [11].

The decision variable D_j corresponds to [10]

$$D_n^j = \sum_{i=1}^N r_i^j W_{j,i} \quad (4)$$

$$\text{where } W_{j,i} = \frac{\alpha_{j,i} \beta_{j,i}}{\left(\text{Var}(a_j) A_{j,i} + \frac{N_0}{2} \right)} \quad (5)$$

and $A_{j,i} = \sum_{k=1, k \neq j}^K \beta_{k,i}^2 \alpha_{k,i}^2 \cos^2(i(\Delta\theta_k - \Delta\theta_j) + \Phi_{k,i} - \Phi_{j,i})$ with $\text{Var}(a_j) = 1$.

Thus the outputs of all the single user detectors of all users generate a decision vector $D = [D^1, D^2, \dots, D^K]$ which is used to obtain the initial estimates of the data $\hat{a} = [\hat{a}_1, \hat{a}_2, \dots, \hat{a}_K]$. These initial estimates are then used to evaluate the MAI experienced by each user in the interference cancellation technique [12].

3.4 Antenna Diversity

Diversity using two antennas, for simplicity, is used here although it can be extended for multiple antenna systems to achieve improved performance at the cost of greater computation. Received signal from two antennas are then weighted averaged according to their individual SIR. This diversity technique has the advantage of producing an acceptable weighted SIR even when none of the individual SIRs are themselves acceptable.

The received signal from antenna system 1, derived from Eq. (2) can be written as follows:

$$r(t) = \sum_{k=1}^K \sum_{i=0}^{N-1} a_k[n] \beta_{k,i} \alpha_{k,i} \text{Cos}(2\pi f_i t + i\Delta\theta_k + \Phi_{k,i}).P(t) + \eta(t)$$

So the received signal for the j^{th} user may be written as

$$\begin{aligned} r_j &= a_j[n] \beta_{j,i} \alpha_{j,i} + \sum_{k=1, k \neq j}^K \sum_{i=0}^{N-1} a_k[n] \beta_{k,i} \alpha_{k,i} \text{Cos}(i(\Delta\theta_k - \Delta\theta_j) + \Phi_{k,i} - \Phi_{j,i}) + \eta_{j,i} \\ &= a_j[n] \beta_{j,i} \alpha_{j,i} + I_{MAI} + N \end{aligned} \tag{6}$$

where multiple access interference I_{MAI}

$$\begin{aligned} I_{MAI} &= \sum_{k=1, k \neq j}^K \sum_{i=0}^{N-1} a_k[n] \beta_{k,i} \alpha_{k,i} \text{Cos}(i(\Delta\theta_k - \Delta\theta_j) + \Phi_{k,i} - \Phi_{j,i}) \\ &= [I_{MAI}^1, I_{MAI}^2, \dots, I_{MAI}^k] \end{aligned}$$

and the noise term $N = [n_{1j}, n_{2j}, \dots, n_{kj}]$

For the Rayleigh fading channel and the large value of 'K' i.e. the number of users, the distribution of MAI is approximately Gaussian (according to central limit theorem) and noise N is also a Gaussian random variable with mean 0 and variance $N_0/2$. Therefore the interference and noise power is equal to variance of the interference and noise terms of all the users. Now the interference power and noise power, respectively can be expressed as follows:

$$\sigma_{MAI}^2 = \text{Var}(I_{MAI}) \tag{7}$$

and

$$\sigma_N^2 = \text{Var}(N) \tag{8}$$

The SIR for the j^{th} user of antenna system 1

$$SIR_{j1} = a_j^2[n] \cdot \alpha_{j,i}^2 \cdot \beta_{j,i} / (\sigma_{MAI}^2 + \sigma_N^2) \tag{9}$$

The total SIR for the antenna system 1 and 2 can be written as follows:

$$SIR_1 = \sum_{j=1}^K SIR_{j1}$$

$$SIR_2 = \sum_{j=1}^K SIR_{j2}$$

Total weighted SIR of two-antenna receiver system is

$$SIR_{TW} = W_1 \cdot SIR_1 + W_2 \cdot SIR_2 \quad (10)$$

where

$$W_1 = 10 \cdot SIR_1 / (SIR_1 + SIR_2)$$

$$\text{and } W_2 = 10 \cdot SIR_2 / (SIR_1 + SIR_2)$$

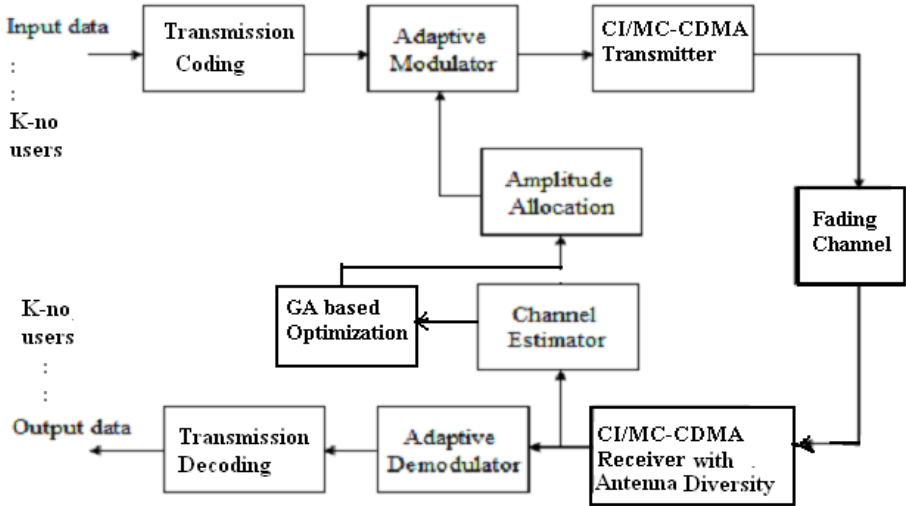


Fig. 1. Block diagram for GA based power adaptive system model

4 Proposed GA Based Power Allocation Algorithm

This section briefly describes proposed GA based adaptive power allocation for the antenna diversity assisted CI/MC-CDMA system described in previous section. The proposed system is also shown in Fig. 1 as block diagram representation. The instantaneous CSI is found from channel estimation and is used for power allocation. The main objective of this work is to maximize the overall channel capacity as well as maintaining BER performance within a limit under the constraint of transmit power. We first define objective function followed by GA implementation for power allocation.

4.1 Formulation of Objective Function

Wireless channel is highly nonlinear and random in nature. In the present system, it is assumed that CSI is available in transmitter (feedback from receiver) that is used to modify signal amplitude so that the channel capacity is maximized and the probability of bit error is maintained within the limit, subject to transmit power constraints.

Mathematically it is a nonlinear integer problem of conflicting nature and is difficult to find the close from solution. Hence GA may be used for sub-optimal solution to find adaptive power for each user on each subcarrier. The fitness (objective) function is considered as follows

$$\begin{aligned}
 F = & \lambda_1 \cdot \sum_{k=1}^K \sum_{i=0}^{N-1} \log_2 \left[1 + \frac{\alpha^2_{k,i} \cdot p_{k,i}}{\left(\sigma^2_{MAI} + \sigma^2_N \right)} \right] \\
 & + \lambda_2 \cdot \left(\sum_{k=1}^K \sum_{i=0}^{N-1} p_{k,i} - P_{total} \right)
 \end{aligned} \tag{11}$$

Subject to $b_e \leq B$

where b_e is the probability of bit error i.e BER, B is threshold or upper limit on acceptable BER value, $p_{k,i}$, P_{total} denote the transmitted power of k^{th} user on i^{th} subcarrier, and the maximum allowable transmit power, respectively. The symbols λ_1 and λ_2 are the weight factors such that $\lambda_1 + \lambda_2 = 1$. We assume equal value for both the weight factor i.e. $\lambda_1 = \lambda_2 = 0.5$. The expression of σ_{MAI} may be stated as follows:

$$\sigma_{MAI} = f(\rho_{jk}, \alpha_{k,i}, \beta_{k,i})$$

where ρ_{jk} is the cross-correlation between j -th and k -th users. Incorporating the correlation coefficient ρ , we can write the interference power $\sigma^2_{MAI} = \rho^2 \alpha^2_{k,i} \cdot \beta^2_{k,i}$ and noise power $\sigma^2_N = N_0$. Putting these values in Eq. (11), we get fitness function

$$\begin{aligned}
 F = & \lambda_1 \cdot \sum_{k=1}^K \sum_{i=0}^{N-1} \log_2 \left[1 + \frac{\alpha^2_{k,i} \cdot \beta^2_{k,i}}{\left(\rho^2 \alpha^2_{k,i} \cdot \beta^2_{k,i} + N_0 \right)} \right] \\
 & + \lambda_2 \cdot \left(\sum_{k=1}^K \sum_{i=0}^{N-1} \beta^2_{k,i} - P_{total} \right)
 \end{aligned} \tag{12}$$

Our goal is to maximize Eq. (12) subject to $b_e \leq B$.

4.2 GA Based Adaptive Power Allocation

Genetic Algorithms (GA) is one of the robust global optimization tool widely explored in solving complex optimization problems in numerous fields. The operations of GA depend on initial population, crossover and mutation. In the present study, adaptive power of each user on each subcarrier is calculated based on maximization of fitness function F defined in Eq. (12).

The experimental conditions of GA for the present problem are depicted as follows: Size of population is 20, number of generations 100, probability of crossover per generation is 0.8, and probability of mutation per bit is 0.09, upper limit set on MMSEC receiver’s BER value i.e. $B = 10^{-2}$.

Different steps for implementing GA based adaptive transmitter power allocation are described as follows:

Step 1: Initialization of twenty sets of random values of transmitted power within the maximum limit allowed by the power constraint channel.

Step 2: Calculate the fitness value F for the twenty sets of random transmit power obtained in step 1. A predefined threshold (F_u) value of F is assigned to identify the fit parameter sets.

Step 3: The particular set of transmit powers which produce the fitness value F above (F_u) are duplicated and the remaining sets are ignored from the population.

Step 4: A set of binary string is generated through decimal-to-binary conversion of all selected transmitted powers. Now crossover and mutation operations are done according to their respective probabilities stated above.

Step 5: A new set of transmitted power within the range are generated through binary-to-decimal conversion of the strings obtained in step 4.

Step 6: Repeat step 1 to step 5 for the desired number of iterations or till a predefined acceptable values for channel capacity and BER (b_e) are achieved.

5 Performance Evaluation

This section reports performance evaluation for the power adaptive synchronous CI/MC-CDMA system in terms of data transmission rate i.e. channel capacity vs number of users along with comparison with [1] and [2], MMSEC receiver BER performance for number of users in block parallel interference cancellation (PIC) system [10], non-adaptive system [1] and proposed power adaptive system integrated with 4 block PIC [12]. Simulations are done at SNR = 7dB, number of users varying from 10 to 80, the number of subcarrier is 24 and maximum allowable transmit power is 5 unit (mW). Fig. 2 (a) shows that the overall system channel capacity (data transmission rate) for the proposed system is much higher than the existing methods [1] and [2]. Fig. 2(b) represents BER performance for MMSEC receiver and it is observed that with the numbers of users increase, BER value increases at much higher rate for [1] than [10] and proposed one. Increase in BER value is the lowest for the present method.

It is seen that upper limit for BER value during simulation is set to 10^{-2} which is not always acceptable in many multimedia communication. To improve further BER performance several forms of multiuser detection in CDMA is reported like [2] and [10]. We finally integrate block PIC scheme in [12] with the proposed adaptive power allocation scheme to improve BER performance at high user capacity. In brief, users are classified into different groups, namely, very strong, strong, weak and very weak based on the magnitudes of decision statistics. The interferences within the individual groups of large decision magnitudes like very strong, strong, weak and very weak in order of sequence are removed to improve BER performance for the weakest group. Fig. 3(a) and 3(b) show that BER performance in block PIC system with proposed power adaptive scheme offers significantly much better results than the non-adaptive system. Simulation results also show that performance enhances with an increase in the number of blocks and the number of iterations of block PIC scheme at the cost of slight increase in computation complexity.

Numerical values reflect that four block PIC after three stage iterations for adaptive system can support number of users upto twice the number of subcarriers with BER

of the order of 0.00032. It can support users upto three times the number of subcarriers with BER of the order of 0.0054, while the similar values for non-power adaptive system are ~0.014 and 0.046, respectively.

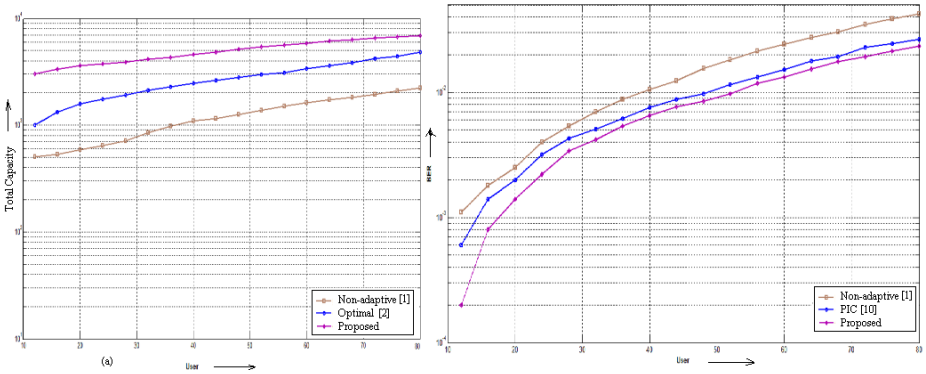


Fig. 2. (a) Performance comparison for capacity versus number of users ; (b) Performance comparison for BER versus number of users in MMSEC receiver

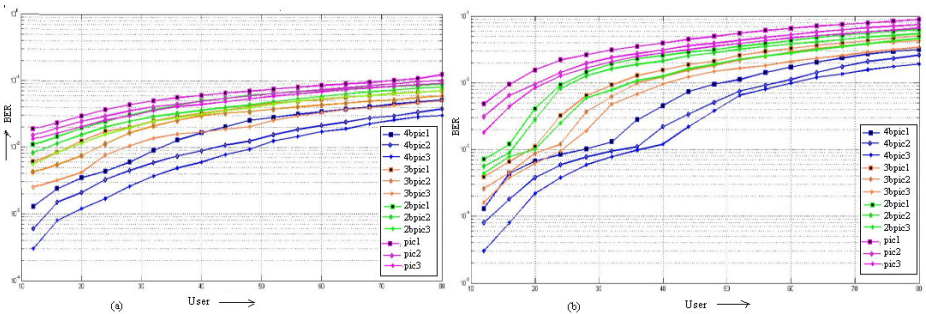


Fig. 3. (a) Block PIC BER performance for non-power adaptive system; (b) Block PIC BER performance for proposed power adaptive system

6 Conclusions and Scope of Future Works

The paper investigates the scope of usage of GA for adaptive power allocation in CI/MC-CDMA uplink system in order to increase channel capacity while maintaining BER value under transmit power constraint. Simulation results show that channel capacity and BER performance for the proposed power adaptive system have been significantly improved compared to non-power adaptive system. Simulation results show that four block PIC using proposed power allocation can support number of users upto three times the number of subcarriers at the end of third stage while BER value is of the order of 0.005. Future work would extend this concept for MC-CDMA based cognitive radio system design for capacity improvement in secondary user under the interference constraint to primary user.

References

1. Natarajan, B., Nassar, C.R., Shattil, S., Michelini, M., Wu, Z.: High performance mc-cdma via carrier Interferometry codes. *IEEE Trans. on Vehicular Tech.* 50, 1344–1353 (2001)
2. Maity, S.P., Mukherjee, M.: On Optimization of CI/MC-CDMA System. In: 20th IEEE Personal, Indoor and Mobile Radio Comm. Symp., Japan, pp. 3203–3207 (2009)
3. Proakis, J.G.: *Digital communications*, 3rd edn. McGraw Hill, New York (1995)
4. Caire, G., Taricco, G., Biglieri, E.: Optimum power control over fading Channels. *IEEE Trans. Inform. Theory* 45, 1468–1489 (1999)
5. Goldberg, D.E.: *Genetic Algorithms in Search, Optimization and Machine Learning*. Pearson Education, New Delhi (1999)
6. Shen, Z., Andrews, J.G., Evans, B.L.: Adaptive Resource Allocation in Multi-user OFDM Systems with Proportional Rate Constraints. *IEEE Trans. on Wireless Commun.* 42, 2726–2737 (2005)
7. Luo, J., Blum, R.S., Cimini, L., Cimini, L., Haimovich, A.: Power allocation in a transmit diversity system with mean channel gain information. *IEEE Commun. Lett.* 9, 2726–2737 (2005)
8. Zhang, J., Matolak, D.W.: Transmitted power allocation/control for multiband MC-CDMA. *Journal of Physical Commun.* 3, 139–146 (2009)
9. Jang, J., Lee, K.B., Lee, Y.H.: Frequency-Time Domain Transmit Power Adaptation for a Multicarrier System in Fading Channels. *IEEE Commun. Letter* 1, 100–103 (2001)
10. Thippavajjula, V., Natarajan, B.: Parallel Interference Cancellation Techniques for Synchronous Carrier Interferometry/MC-CDMA Uplink. In: *Proc. of IEEE Int. Conf. on Vehicular Technology*, pp. 69–78 (2004)
11. Cal, X., Akansu, A.N.: Multicarrier CDMA systems with transmit diversity. *IEEE Trans. on Vehicular Technology* 2, 658–661 (2000)
12. Maity, S.P., Hati, S., Maity, S.: Diversity-Assisted Block PIC for synchronous CI / MC-CDMA uplink system using Genetic Algorithm. In: *Proc. of Third IEEE Int. Conf. on Industrial and Information Systems (ICIIS 2008)*, pp. 1–6. Indian Institute of Technology, Kharagpur (2008)

PHY Abstraction for MIMO Based OFDM Systems

Tarun kumar Juluru¹ and Anitha Sheela Kankacharla²

¹ Department of ECE, Ramappa Engineering college, Warangal,
Andhra Pradesh, India

tarunjuluru@yahoo.com

² Department of ECE, JNTU College of Engineering, Hyderabad,
Andhra Pradesh, India

kanithasheela@gmail.com

Abstract. Now a days Computer simulations are often used to predict the performance of cellular networks. Link level simulations are used to model the link between the base and mobile stations, while system level simulators model the entire network. In order to predict the accurate performance of cellular network, a system level simulator, which includes the performance of the link between each base and mobile station, should be used, but this is computationally prohibitive. So to reduce the complexity caused by the system level simulator, “The Abstraction from the link level simulations” can be used. Thus, the objective of PHY abstraction is to accurately predict the link layer performance in a computationally easy way. The well known approach for link abstraction is the Effective Exponential SINR metric but the main disadvantage of this method is that it needs to compute a normalization parameter say β for various modulation and coding schemes and also it is difficult to extend this for Maximum likelihood detection (MLD). In this paper we propose a Link quality model, Received Bit Information Rate (RBIR) is used in system evaluations to simplify the simulation complexity, where the performance metric for PHY abstraction is to predict the Block error rate (BLER) under various channel realization across the OFDM sub carriers. In order to predict the BLER a post-processing SINR values are considered as an inputs to the PHY abstraction. From the simulation results it shows that this abstraction is computationally easy, accurate, simple, and independent to channel models, extensible to interference models and multi antenna processing. These simulations are very close to the actual simulations.

Keywords: BLER, Effective SINR, MIMO, OFDM, PHY abstraction, RBIR.

1 Introduction

Research for Broad band Wireless Access Standards is in progress worldwide from the past few years. The IEEE 802.16 working group on Broad band wireless standards, established by the IEEE standards board in 1999. The IEEE 802.16m is currently working on Advanced Air Interface [1]. This group is trying to achieve data rates of 100 Mbps for mobile applications and 1Gbps for fixed applications. In order to achieve these data rates so many features are evolved in the architecture of wireless

communications e.g. orthogonal frequency division multiple accesses (OFDM), multiple antennas, various packet switching protocols etc. The OFDM allows good performance in terms of maintaining orthogonality between the cells, protection against inter-symbol interference and effective utilization of band width. To obtain these high data rates in wireless networks such as World wide Inter-operability for microwave access (WiMAX), long term evolution (LTE) environment requires various modulation and coding schemes to utilize the available band width and multi user diversity.

The PHY Abstraction methods are used to predict the Block error rate (BLER) which is a performance indicator to realize the channel under different transmission schemes such as Single Input Single Output (SISO), Single Input Multiple Output (SIMO), and Multiple Input Multiple Output (MIMO) antenna schemes with antenna coding for different Modulation and coding schemes in the fading environment. These schemes uses link level simulations for abstraction which requires complex computations. The PHY abstraction is used to compute the performance of the link to avoid the complex simulations as used in link level simulations in such System- level simulations considers a large number of parameters taken into account and this should be very accurate, simple computations for various channel models.

1.1 Link Level Simulations

Link layer simulator model the PHY and the channel behavior Starting from the transmitter components and ending with the receiver components [(1)][(3)][(10)]. In Fig.1 shows the communication chain to perform link level simulations performing link level simulations has a high computational cost and also involves in complexity. Due to this, these simulations are normally performed in advance, and the results obtained are stored. Then, those results can be easily used to model the PHY behavior when other higher level issues want to be evaluated, avoiding lots of calculations.

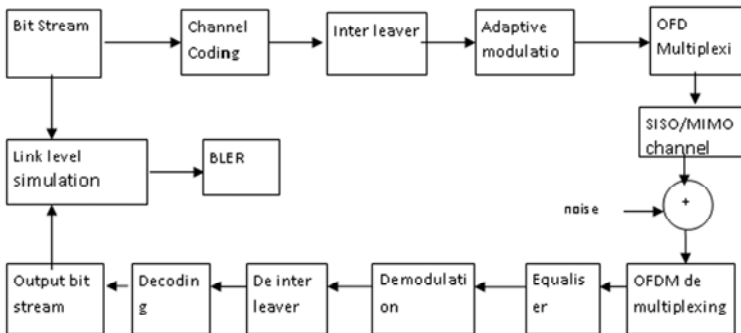


Fig. 1. Digital Communication Chain

1.2 System Level Simulations

System level simulations deal with higher level events or issues, like interference management, resource allocation management, power allocation, etc[(1)][(3)]. When

those issues want to be evaluated, the PHY behaviour is modelled by using the stored results from link level simulations performed in advance.

2 Abstraction Methods for Link to Systems

The Abstraction methods are used to predict the link layer performance. It includes simulating the physical layer links between multiple Base station and Mobile station in a network.

The Abstraction should be very accurate, with simple computations, various channel models and extensible to interference and Multi- antenna processing.

The system level simulations are used to predict the overall performance of a system, which makes easy to establish a system with proper frequency planning. In such simulations the performance of a system is estimated using the topology and channel characteristics to compute SINR (Signal to interference plus noise ratio) across the cell. Then each subscriber SINR are mapped to the highest modulation and coding. The abstraction methods are classified as

2.1 Static Methodology of Abstraction

In this method the average performance of system was quantified by using the topology and macro channel characteristics to compute a geometric or average SINR distribution across the cell [(1)][(3)]. Each subscriber geometric SINR was then mapped to the highest modulation and coding scheme (MCS), which could be supported based on the link level SINR tables that capture fast fading statistics. The link level SINR verses BLER (Block error rate) look up table served as the PHY abstraction for predicting average link layer performance.

2.2 Dynamic PHY Abstraction Methodology

The performance characterization solely based on SNR is sufficient if the SNR value remains constant during each coded block [2]. However in wireless mobile systems the assumption that the SNR value remains constant over a coded block usually doesn't hold. In system level simulations, an encoder packet may be transmitted over a time frequency selective channel [9]. For example OFDM systems may experience frequency selective fading and hence the channel gain of each sub carrier may not be equal.

The post processing values of SINR are considered as a input for PHY abstraction mapping as shown in Fig.2. An $SINR_{eff}$ (Effective SINR)–is required to accurately map the system level SINR on to the link level simulations[9]. $SINR_{eff}$ is approximated from a group of SINR's shown in step1 of Fig 2. This kind of mapping is termed as Effective SINR mapping (ESM). So many ESM approaches are used to predict the performance of link they are based on

- 1) Mean instantaneous capacity.
- 2) Exponential effective SINR mapping (EESM)
- 3) Mutual information effective SINR mapping (MIESM)

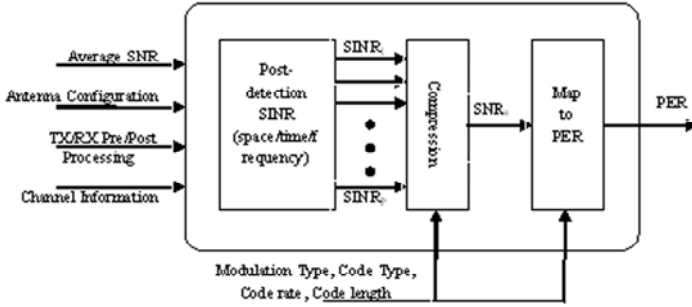


Fig. 2. Link Abstraction Model

In MIESM two major constraints one is based on mutual information per symbol received is used to compute the bit mutual information, another directly computes the bit mutual information per coded bit. The computation of the mutual information per coded bit is known to be Received symbol mutual information (RBIR). In RBIR the mutual information per symbol is used for mapping RBIR to BLER.

3 Received Symbol Mutual Information Rate

In a static channel with stationary additive white Gaussian noise (AWGN) [1], the Link performance of a coded digital communication system is generally characterized by the BLER versus SINR. This performance characterization solely based on SINR is sufficient if the SINR value remains constant during each coded block [2]. However in wireless systems, the assumption on that the SINR value remains constant over a coded block usually doesn't hold.

For a given Convolution code and frame length N , the relationship between the block error probability BLER and SINR can be determined for a conventional AWGN channel (with constant SINR) via analysis or simulation[8].

$$SINR_{eff} = \Phi^{-1} \left\{ \frac{1}{N} \sum_{n=1}^N \Phi(SINR_n) \right\} \quad (1)$$

Where $SINR_{eff}$ is the effective SINR, $SINR_n$ is the SINR in the n th sub-carrier; N is the number of symbols in a coded block, n is the number of sub-carriers used in an OFDM system and $\Phi(\cdot)$ is an invertible function. Received Bit Information Rate (RBIR) is an abstraction method; it will approximate the $SINR_{eff}$ by calculating the Mutual Information. The calculation of Mutual information between the transmitted and received symbols is calculated. For all the ESM methods, the following system model for describing the MIMO Relationship

$$Y = H X + U \quad (2)$$

Where Y is the received signal vector from the N_r antennas, H is the channel matrix, X is the transmitted Symbol stream which is a vector of dimensions $N_t \times 1$ (X is just

a scalar in the case of SISO/SIMO), and U is the noise vector of dimension $N \times 1$, modelled as zero-mean Complex Gaussian. The above equation demonstrates an OFDM system is equivalent to transmission of data over a set of parallel channels. In the case of the mutual information based ESM the function $\Phi(\cdot)$ is derived from the Constrained capacity; while in the case of EESM, the function $\Phi(\cdot)$ is derived from the Chernoff bound on the probability of error.

Assuming N sub carriers are used to transmit a coded block, and then the normalized Mutual information per received bit (RBIR) is given by

$$RBIR = \frac{\sum_{n=1}^N SI(SINR_n, m(n))}{\sum_{n=1}^N m(n)} \tag{3}$$

The symbol mutual information curves $SI(SINR_n, m(n))$ (where $m(n)$ is measured value of n^{th} OFDM symbol at n^{th} time) are generated once in the system Simulator

3.1 RBIR Mapping for MIMO Receiver

In a MIMO receiver the mutual information SI can be determined as

$$SI = \frac{1}{M} \sum_{n=1}^M \int_{-\infty}^{\infty} p(LLR_i) \log_2 \frac{M}{1 + e^{-LLR_i}} dLLR_i \tag{4}$$

Where the $p(LLR_i)$ is known as symbol level log-likelihood ratio (LLR) of i^{th} constellation point. The RBIR PHY abstraction is based on the fixed relationship between LLR distributions and BLER. Hence, a representative LLR distribution among M distributions is considered. By using the numerical integration method [5], the symbol information can be estimated as

$$SI \approx \log_2 M - \frac{1}{\log_e(2)} \cdot J \tag{5}$$

$$J = \left(\frac{J_A + J_B}{2} \right) + \left(\frac{J_A - J_B}{2} \right) \text{sign}(T - J_B) \tag{6}$$

$$J_A = \sqrt{VAR} \left\{ \frac{-\eta}{2} \cdot \text{Erfc} \left(\frac{\eta}{\sqrt{2}} \right) + \frac{1}{\sqrt{2\pi}} \cdot e^{\left(\frac{-\eta^2}{2} \right)} \right\} \tag{7}$$

The η is known as coefficient of variation and given as

$$n = \frac{AVE}{\sqrt{VAR}}, \tag{8}$$

$$J_B = \frac{2}{3} f(AVE) + \frac{1}{6} f(AVE) + \frac{1}{6} f(AVE - \sqrt{3VAR}), \quad (9)$$

Where $f(x)$ is given as

$$f(x) = \log_e(1 + e^{-x}) \quad (10)$$

and

$$sign(x) = \begin{cases} +1, & x \geq 0 \\ -1, & x < 0 \end{cases} \quad (11)$$

The mean (AVR) and variance (VAR) are computed as a function of intermediate random variable γ dB and it is given as

$$\gamma_{dB} = 10 \log_{10} \left(\frac{d^2 |H_k|^2}{\sigma^2} \right) \quad (12)$$

Where 'd' is the minimum distance between the decision points, normally it is assumed as

$$d = \begin{cases} \sqrt{2}, & \text{for QPSK} \\ 2/\sqrt{10}, & \text{for 16QAM} \\ 2/\sqrt{42}, & \text{for 64QAM} \end{cases} \quad (13)$$

H_k is a channel matrix of K^{th} column vector for a 2X2 MIMO the H is given as

$$H = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \quad (14)$$

σ^2 is known as variance of noise plus interference

3.2 Channel Estimation Error Modeling

In the practical scenario the exact information about the channel is not known, so this is one of the most important forms of actual receiver impairments. The Channel Estimation (CE) error at the receiver (MS) and its modeling is critical to the evaluation of system throughput [12]. In general, the CE error may impact the receiver performance differently for different types of receiver processing. Therefore the abstraction method has to work properly for CE errors. The key is to establish a common frame work that can be built upon to accommodate various receiver processing techniques. The CE error depends on type of channel estimator (e.g., MMSE, LSE), time-frequency pilot pattern (e.g., pilot number and positions, often power-boosted as well), Design parameters (e.g., assumed SNR for MMSE filter coefficients, 2D or two 1D MMSE) and actual channel behavior (e.g., Delay spread, Doppler).

In an OFDM frame assuming that the delay spread (τ) of channel is shorter than the cyclic prefix, the received symbol at pilot carrier is given by[11]

$$y[k, n] = x[k, n]H[k, n] + w[k, n] \quad (15)$$

where $x[k, n]$, $y[k, n]$, $H[k, n]$ and $w[k, n]$ are the transmitted symbol, the received symbol, the channel transfer function and complex zero mean white Gaussian noise with variance σ_w^2 at the k^{th} sub carrier in the n^{th} OFDM symbol.

The modified channel estimation can be adapted to NXM MIMO configuration. Since the loss of performance can primarily be attributed to the increase in effective noise variance. The post processing SINRs can be computed and then input to link abstraction methods. When system results are provided in a contribution with channel estimation schemes by considering the parameters such as ideal channel estimation, different filter designs could correspond to different permutation modes(PUSC, AMC), different pilot patterns (common pilots or dedicated pilots), SNRs, Doppler, channels etc. In the simulation when channel estimation is used, though they are specific to individual implementations, they have enough information to harmonize results.

4 Simulations

A system level simulator developed for cellular, user environment has been developed. Performance assessment of link performance models generally involves a comparison of the predicted/estimated block error rates BLER with the measured BLER which are derived from extensive link level simulation. The block diagram in Fig.1. is the validation approach including details about the considered link level chain is depicted. An OFDM-based data transmission employing random bit interleaved coded modulation in conjunction with linear spatial pre- and post-processing techniques is assumed. Any transceiver imperfections as e.g. synchronization and channel estimation errors are neglected. Every transmitted data (packet/block) is coded independently and the inter leaver is altered randomly from block to block. The link level chain generates channel specific error rates BLER (measured), keeping the (MIMO) channel characteristics as well as noise. The fraction of erroneous received blocks approximates the BLER for that specific channel.

Consider an OFDM based broadband wireless communication system with 2 transmits and 2 receive antennas. The channel on the K^{th} sub carrier at time n for the desired signal is assumed to be a flat fading Rayleigh channel and the path gain from transmit antenna m to the receive antenna j is denoted by $H_{jm}[K,n]$, where $j = 1, 2, \dots, N_r$. The path gains are modeled as independent samples of zero-mean, complex Gaussian variable with variance $1/2$ per real dimension. In order to mimic a practical channel, the variation of $H_{jm}[K,n]$ along frequency and time is modeled using the ITU and Jakes models respectively.

In the emerging OFDM-based IEEE 802.16m, data is allocated in groups of resource blocks (RBs) [1]. Each (RB) is composed of P sub carriers and Q OFDM symbols, and it is called a localized RB when P sub carriers are contiguous and distributed RB when the P sub carriers frequency band. The size of P and Q are 18 and 6 for IEEE 802.16m standard. And the bits are encoded in convolution and turbo codes (CTC)[7].

4.1 Simulation Parameter

The above simulation parameters are considered in the PHY abstraction the Fig 3. and 4 show that the information per bit for QPSK and 16 -QAM using RBIR mapping. It shows that the performance curves for a given code rate is independent of modulation schemes. The Fig.5 and 6 shows us the mean and variance for log likelihood ratio (LLR) to obtain the Mutual information per symbol. The simulations results shown in Fig.7 is the estimation results of PHY abstraction and which are very

Number of symbols in RB	6
No. of sub carriers in RB	18
Number of pilots in RB	6 or 12
Cyclic prefix length	128
FFT Size	1024
Bandwidth	10MHz
Sampling frequency	11.2 MHz
Sub carrier spacing	10.9375 kHz
Frame length	5 ms
Channel model	Rayleigh Channel
Channel estimator	MMSE
Delay spread	$3.7e^{-6}$
SINR	-20 to 27dB
Fading Model	Modified Ped B
Doppler	7 Hz or 70Hz
Modulation Scheme	QPSK or M -QAM
Pilot sequence	PRBS
Code rate	1/2
Coding Scheme	CTC
Antenna Configuration	1X1or1X2or2X2

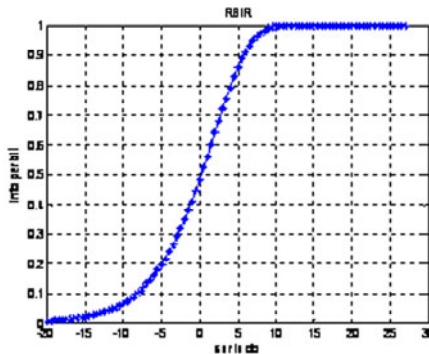


Fig. 3. Information per bit using RBIR for qpsk

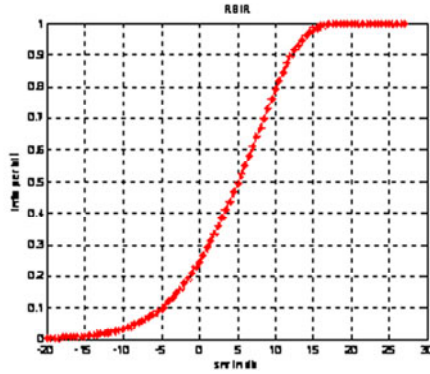


Fig. 4. Information per bit using RBIR for 16-qam

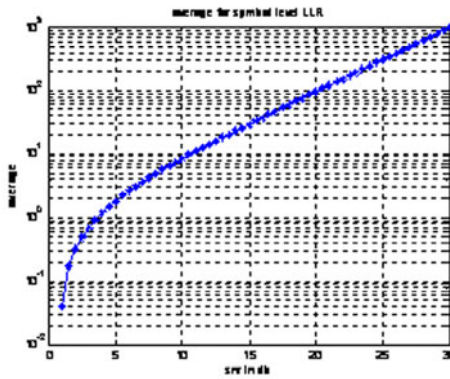


Fig. 5. Mean value for symbol level LLR

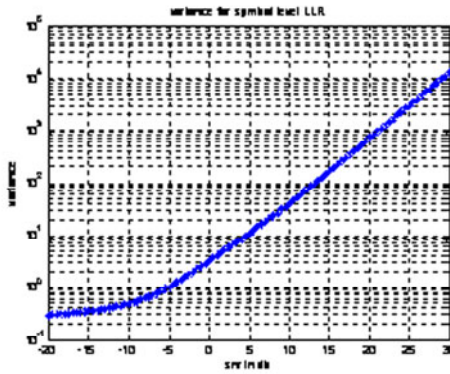


Fig. 6. Variance value for symbol level LLR

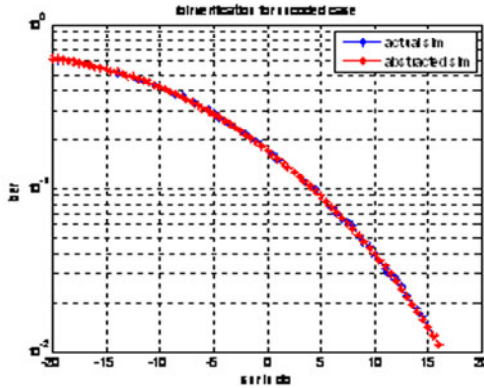


Fig. 7. SNR Vs BER for RBIR

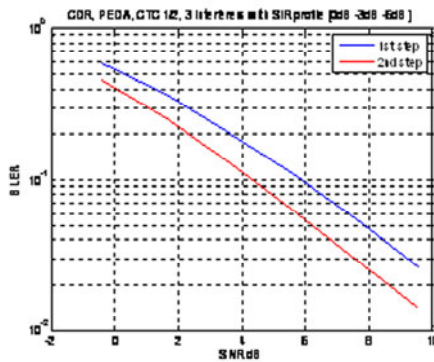


Fig. 8. SNR Vs BLER for 3 interferers with PED B channel with code rate 1/2

closure to the actual simulations which is performed for uncoded blocks, So the RBIR abstraction is very closure to the actual simulations. The BLER is estimated for various channel models and assuming 3 interferers effecting with a SIR (signal to interference ratio) of -0dB, -3dB, and -6dB for a code rate of $\frac{1}{2}$ in Fig.8. and which are very close to the actual simulations.

5 Conclusions

The link level simulations involve in complex calculations and ambiguity in networks and testing, whereas the simulations performed through the abstraction method RBIR are straight forward, simple and easy. The performance of RBIR is similar to that of actual stimulations (without interference) BLER Vs SNR is having more gain than to link level. When BER is performed in presence of AWGN vs. SNR is also accurate. RBIR graph is analogues to link level under stringent conditions and constraints

which is able to match for actual stimulations for wide range if MCS (Modulation and coding) and diversity schemes. Thus, RBIR abstraction method can predict link performance with great accuracy. This abstraction method can be used as the data base for various AMC (adaptive modulating and coding) schemes

References

1. IEEE 802.16m:Evaluation Methodology Document (EMD),
<http://ieee802.16.org>
2. Kliks, A., Zalonis, A., Dagres, I., Polydoros, A., Bogucka, H.: PHY Abstraction Methods for OFDM and NOFDM Systems. *Journal of Telecommunications and Information Technology*, 116–121 (March 2009)
3. Alamouti: A simple transmitter diversity scheme for wireless communications. *IEEE Journal on Selected Areas of Comm.*, 1451–1458 (October 1998)
4. Nortel: Effective SIR Computation for OFDM System-Level. Simulations, TSG-RANWG1 35 R03-1370 (November 2003)
5. Atheros, M.: ST Micro-Electronics and Marvell Semiconductors. Unified Black Box PHY Abstraction Methodology: IEEE Contribution 802.11-04/0218r1 (March 2004)
6. Jalloul, L.: On the Expected Value of the Received Bit Information Rate: IEEE C802.16m-07/195, Malaga, Spain (September 2007)
7. Lestable, T., Mourad, A., Jiang, M., Cho, Y., Yu, H., Kim, T., Cho, J., Jalloul, L.: Enhanced Approximation for RBIR PHY Abstraction in TGM.: IEEE C802.16m-08/067r4.pdf (March 2008)
8. Brueninghaus, Astely, K., Salzer, D., Visuri, T., Alexiou, S., Karger, A., Seraji, S.: Link Performance Abstraction for ML Receivers based on RBIR Metrics. United States Patent Application 20100064185.
9. He, X., Niu, K., He, Z., Lin, J.: Link layer abstraction in MIMO OFDM system. In: *International Workshop on Cross Layer Design, IWCLD*, pp. 41–44 (2007)
10. IEEE 802.16m.: Link To System Performance Mapping Based On Effective SINR,
<http://ieee802.16.org>
11. Li, Y.(G.): Simplified Channel Estimation For OFDM Systems With Multiple Transmit Antennas. *IEEE Transactions On Wireless Communications* 1(1) (January 2002)
12. Yushi Shen and Ed Martinez.: Channel Estimation In OFDM Systems, Free scale semiconductor, Application Note

Level Crossing Rate in Land Mobile Satellite Channel with a Modified Nakagami-Lognormal Distribution

Sayantana Hazra and Abhijit Mitra

Department of Electronics and Electrical Engineering,
Indian Institute of Technology, Guwahati 781039, India
(sayantan,a.mitra@iitg.ernet.in)

Abstract. A modified Nakagami-lognormal distribution has been proposed to represent statistical nature of land mobile satellite (LMS) channel. The normal or Gaussian variable associated with lognormal distribution is taken to be correlated with its time derivative. It does not change the probability density function but modifies the expression of level crossing rate (LCR), which is derived in this paper. The proposed model takes the conventional Nakagami-lognormal model as its special case which can be achieved by setting the correlation coefficient to zero. This correlation coefficient provides an extra degree of freedom in modeling the LMS channel. The effect of this correlation coefficient and other important parameters on LCR, computed from derived expression, has been depicted with LCR curves.

Keywords: land mobile satellite channel, small-scale fading, shadow fading, level crossing rate.

1 Introduction

With the advent of second generation mobile satellite receivers, land mobile satellite (LMS) communications have been popular through navigation services like Global Positioning System (GPS), aeronautical and personal communication services like MSAT (Mobile Satellite), VSAT (very small aperture terminal), Inmarsat, Iridium etc.

Among several propagation phenomena affecting the signal received at an LMS receiver, small-scale fading and shadow fading are considered to be analyzed by statistical channel modeling. Small-scale fading, also known as multipath fading, is small-scale variation or fading of amplitude and phase of the received signal over a short interval of time or travel distance. This is caused by the interference of scattered electromagnetic waves originated by reflections and scattering from obstacles around the receiver. Its power depends on the distribution of the scattering objects around the receiver.

Although the mean signal level or power seems to be constant over few meters of travel distance, a slow but random variation of the mean is observed over larger distance. This slow and random attenuation of received signal amplitude or power is termed as shadowing or shadow fading. The reason for this spatial variation is that, the distribution of the obstacles and scattering objects around the receiver may be vastly different at two different locations having the same Transmitter-Receiver separation [1]. This variation

Table 1. List of the models available in literature

Proposed by	complex channel process	Multipath fading	shadow fading	Comments
Loo [5]	$r = S e^{j\phi_0} + R e^{j\phi}$; S : log-normal, R : Rayleigh, ϕ_0, ϕ : uniform	Rayleigh	log-normal (LOS)	independent fading of LOS and diffuse components
Pätzold et al. [6]	$r = S e^{j(2\pi f_d t + \theta_d)} + x_1 + jy_1$; S : log-normal, x_1, y_1 : Gaussian, f_d and θ_d : constant	Rayleigh	log-normal (LOS)	same as Loo; but LOS component is Doppler shifted
Corazza-Vatalaro [7]	$r = R S e^{j\theta}$; S : log-normal, R : Rician, θ : uniform	Rician	log-normal (Multiplicative)	strongly correlated fading of LOS and scattered components
Vatalaro [8]	$r = R S e^{j\theta} + x_1 + jy_1$; S : log-normal, R : Rician, x_1, y_1 : Gaussian	Rician + Rayleigh	log-normal (Multiplicative)	a compromise between zero and strong correlation; includes an extra additive scatter component
Hwang et al. [9]	$r = A_c S_1 e^{j\phi} + R S_2 e^{j(\theta+\phi)}$; S_1, S_2 : log-normal, R : Rayleigh, A_c : constant	Rayleigh	log-normal (LOS)	LOS and scattered components undergo shadow fading independently

may be temporal also due to random attenuation through foliage or change of satellite elevation angle over time even when the receiver is not moving over large distance.

To represent small-scale or multipath fading in LMS channel, different researchers have exploited Rayleigh or Rician mostly; while Hoyt, Nakagami- q , Nakagami- m etc. have been in use in macrocellular or microcellular channels. The use of lognormal distribution for shadow fading is widely accepted and approved through empirical evidence based on measurement data [2-4]. Majority of the LMS models employ mixture distribution with Rayleigh, Rice, and lognormal statistics. Some models are shown in Table 1.

As Nakagami- m distribution can be used to epitomize the effect of different other distributions like Rayleigh, Rician or Hoyt by changing the parameter $m \in [0.5, \infty)$, and as it can represent the effect of multipath fading in open areas without LOS obstruction and urban areas with complete LOS obstruction as well; Nakagami- m has been our interest in exploitation for multipath fading. While $m = 1$ stands for only scattered signal components with absence of LOS; the range $0 \leq m \leq 1$, represents severe fading than Rayleigh channel [10]. A very large value of m signifies no obstruction of LOS i.e. it suits well with Rician channel. In the model being depicted here, the LOS and scattered multipath components are incorporated into a single Nakagami- m process and the lognormal shadow fading process is multiplicative to this Nakagami- m process. Independent fading of LOS and scattered components leading to an additive model, as shown in some models in Table 1, is not possible with Nakagami- m process in which fading of LOS and scattered components cannot be considered separately. The Gaussian or normal process associated with the lognormal process is considered here to be correlated with its time derivative. It will have an effect of modification in

the expression of joint probability density function (pdf) of shadow fading process and its time derivative and eventually on final expression of LCR. This modified model is a generalized version of conventional Nakagami-lognormal model (described in [11]) which can be achieved by setting the correlation coefficient to zero.

The rest of this paper is organized as follows. Section 2 provides the mathematical representation of envelope process and pdf of it. In section 3 expression for LCR is derived and computed LCR values are plotted. The effect of different parameters on LCR curves has been shown in section 3.3. Section 4 concludes the paper.

2 Envelope Process and Probability Density Function

The envelope (R) of received signal can be represented as a multiplication of a Nakagami- m (Z) and a lognormal (Y) process; i.e.

$$R = Z \cdot Y. \tag{1}$$

The probability density function (pdf) of R in terms of pdf of Z and Y is given as ([14], example 4.34)

$$p_R(r) = \int_{-\infty}^{\infty} \frac{1}{|z|} p_{YZ} \left(\frac{r}{z}, z \right) dz \tag{2}$$

where individual pdf of Z and Y are

$$p_Z(z) = \frac{2}{\Gamma(m)} \left(\frac{m}{\Omega} \right)^m z^{2m-1} \exp \left(-\frac{mz^2}{\Omega} \right), \quad z \geq 0 \tag{3}$$

and

$$p_Y(y) = \frac{1}{\sqrt{2\pi}\sigma_s y} \exp \left[-\frac{(\ln y - \mu_s)^2}{2\sigma_s^2} \right], \quad y \geq 0 \tag{4}$$

where $\Omega = E[z^2]$, $m = \Omega^2 / E[(z^2 - \Omega)^2]$, $m \geq 0.5$, and σ_s and μ_s are standard deviation and mean of $\ln y$ respectively. The Nakagami- m (Z) and lognormal (Y) process can be assumed independent of each other and as both are zero for negative values, $p_R(r)$ can be written as,

$$\begin{aligned} p_R(r) &= \int_0^{\infty} \frac{1}{z} p_Y \left(\frac{r}{z} \right) p_Z(z) dz \\ &= \sqrt{\frac{2}{\pi}} \frac{1}{\Gamma(m)\sigma_s r} \left(\frac{m}{\Omega} \right)^m \int_0^{\infty} z^{2m-1} \exp \left(-\frac{mz^2}{\Omega} \right) \exp \left[-\frac{\left(\ln \frac{r}{z} - \mu_s \right)^2}{2\sigma_s^2} \right] dz. \end{aligned} \tag{5}$$

3 Expression for Level Crossing Rate

3.1 Expression for Joint Pdf of Signal and Its Time Derivative

To derive the level crossing rate (LCR) in Nakagami-lognormal channel, the joint probability density function (pdf) of signal and its time derivative is required. Towards this,

we start with a Gaussian random process X with zero mean and standard deviation σ_s . Then the lognormal process Y can be written as $Y = e^{X+\mu_s}$, where μ_s is the mean of $\ln y$. The time derivative of X , \dot{X} is Gaussian with zero mean [12]. When X and \dot{X} are correlated, the joint pdf $p_{X\dot{X}}(x, \dot{x})$ is given by

$$p_{X\dot{X}}(x, \dot{x}) = \frac{1}{2\pi\sigma_s\sigma_{s'}\sqrt{1-\rho^2}} \exp\left\{-\frac{\left[\frac{x^2}{\sigma_s^2} - 2\rho\frac{x}{\sigma_s}\frac{\dot{x}}{\sigma_{s'}} + \frac{\dot{x}^2}{\sigma_{s'}^2}\right]}{2(1-\rho^2)}\right\} \tag{6}$$

where $\sigma_{s'}$ is standard deviation of \dot{X} and ρ is the correlation coefficient of X and \dot{X} . Correlation coefficient is related with the moments of the processes concerned or autocovariance and crosscovariance functions at zero time shift. This is justified, as to calculate LCR, observation must be taken at a single instant of time, not two different time instants.

Now the expression for joint pdf of the lognormal process Y and its time derivative \dot{Y} , can be derived through transformation of variables in $p_{X\dot{X}}(x, \dot{x})$ as

$$\begin{aligned} p_{Y\dot{Y}}(y, \dot{y}) &= p_{X\dot{X}}(\ln y - \mu_s, \dot{y}/y) / y^2 \\ &= \frac{1}{2\pi\sigma_s\sigma_{s'}\sqrt{1-\rho^2}y^2} \exp\left\{-\frac{\left[\frac{(\ln y - \mu_s)^2}{\sigma_s^2} - 2\rho\frac{\ln y - \mu_s}{\sigma_s}\frac{\dot{y}/y}{\sigma_{s'}} + \frac{(\dot{y}/y)^2}{\sigma_{s'}^2}\right]}{2(1-\rho^2)}\right\}. \end{aligned} \tag{7}$$

Now, the joint pdf of R and its time derivative \dot{R} , denoted as $p_{R\dot{R}}(r, \dot{r})$ can be derived from the joint pdf $p_{Z\dot{Z}Y\dot{Y}}(z, \dot{z}, y, \dot{y})$. As Nakagami multipath fading process and lognormal shadow fading process are assumed independent of each other,

$p_{Z\dot{Z}Y\dot{Y}}(z, \dot{z}, y, \dot{y}) = p_{Z\dot{Z}}(z, \dot{z})p_{Y\dot{Y}}(y, \dot{y})$, where $p_{Z\dot{Z}}(z, \dot{z})$ is given by [10, 13]

$$p_{Z\dot{Z}}(z, \dot{z}) = p_Z(z) \cdot \sqrt{\frac{4m}{\Omega}} \frac{1}{\sqrt{2\pi\psi''(0)}} \exp\left(-\frac{2m\dot{z}^2}{\Omega\psi''(0)}\right), \quad z \geq 0, -\infty < \dot{z} < \infty \tag{8}$$

where $\psi(\tau)$ is the normalized autocorrelation function of the process $X^2(t)$ and $\psi''(0)$ is its second derivative at $\tau = 0$. After transformation of variables we get

$$p_{R\dot{R}Y\dot{Y}}(r, \dot{r}, y, \dot{y}) = \frac{1}{y^2} p_{Z\dot{Z}}\left(\frac{r}{y}, \frac{\dot{r}}{y} - \frac{\dot{y}r}{y^2}\right) p_{Y\dot{Y}}(y, \dot{y}). \tag{9}$$

After integrating equation (9) for y and \dot{y} , we get $p_{R\dot{R}}(r, \dot{r})$.

$$\begin{aligned} p_{R\dot{R}}(r, \dot{r}) &= \frac{\sqrt{m}}{\pi\Gamma(m)\sigma_s} \left(\frac{m}{\Omega}\right)^m \int_0^\infty \frac{1}{y^2} \left(\frac{r}{y}\right)^{2m-1} \exp\left(-\frac{m}{\Omega}\left(\frac{r}{y}\right)^2\right) \frac{1}{\sqrt{\beta r^2 + ay^2}} \\ &\cdot \exp\left(-\frac{(\ln y - \mu_s)^2}{2\sigma_s^2 [1-\rho^2]}\right) \exp\left(-\frac{m\dot{r}^2\sigma_s^2 - 2\alpha r\dot{r}(\ln y - \mu_s) - \alpha\kappa y^2(\ln y - \mu_s)^2}{2\sigma_s^2 [\beta r^2 + ay^2]}\right) dy, \\ & \quad r \geq 0, -\infty < \dot{r} < \infty \end{aligned} \tag{10}$$

where $\beta = m\sigma_s^2(1-\rho^2)$, $\alpha = m\rho\sigma_s\sigma_{s'}$, $\kappa = \frac{\rho^2}{1-\rho^2}$ and $a = \frac{\Omega\psi''(0)}{4}$.

3.2 Level Crossing Rate

Level Crossing Rate (LCR) is the rate at which envelope of received signal crosses a specified threshold. Analytically, it is defined as

$$N_R(R) = \int_0^\infty \dot{r} p_{R\dot{R}}(R, \dot{r}) d\dot{r}. \tag{11}$$

Putting $p_{R\dot{R}}(r, \dot{r})$ from equation (10) in equation (11), we get

$$N_R(r) = \frac{1}{\pi \Gamma(m) \sigma_s} \left(\frac{m}{\Omega}\right)^m \int_0^\infty \frac{1}{y^2} \left(\frac{r}{y}\right)^{2m-1} \exp\left(-\frac{m}{\Omega} \left(\frac{r}{y}\right)^2\right) \exp\left(-\frac{(\ln y - \mu_s)^2}{2\sigma_s^2}\right) \cdot \left[\sqrt{\frac{mr^2\sigma_s'^2(1-\rho^2) + ay^2}{m}} \exp\left(-\frac{mr^2\rho^2\sigma_s'^2(\ln y - \mu_s)^2}{2\sigma_s^2 [mr^2\sigma_s'^2(1-\rho^2) + ay^2]}\right) + \frac{\sqrt{2\pi r}\rho\sigma_s'(\ln y - \mu_s)}{\sigma_s} \cdot Q\left(-\frac{\sqrt{mr}\rho\sigma_s'(\ln y - \mu_s)}{\sigma_s \sqrt{mr^2\sigma_s'^2(1-\rho^2) + ay^2}}\right) \right] dy. \tag{12}$$

By putting $\rho = 0$ at above equation we will get exactly the same formula of LCR in uncorrelated case [1]. Equation (12) can be normalized with respect to maximum doppler frequency (f_{\max}) and mean signal level ($S_0 = E[Y]; \mu_s = \ln S_0$) in a area large enough to experience shadow fading. $\psi''(0)$ is taken as $(2\pi f_{\max})^2$ [1, 12]. If the power spectral density (PSD) of shadow fading process is taken as a low-pass filter (as shadow fading has a very slow variation compared to the frequent variation of small-scale fading), then $\frac{\sigma_s'}{\sigma_s} = \frac{\sqrt{n}}{f_{\text{ratio}}} 2\pi f_{\max}$; where the constant n depends on the filter assumed and f_{ratio} is the ratio between f_{\max} and 3-dB cut-off frequency of the filter. These relations are used here for normalization with respect to maximum doppler frequency and the formula is also modified as a function of R/S_0 before plotting it. After normalization, we are left with four important parameters ($m, \Omega, \sigma_s, \rho$) influencing LCR curves at different values of R/S_0 .

3.3 Effects of Different Parameters on Computed LCR

Figure 1 compares Normalized LCR at $m = 1$ and for different values of ρ with LCR formulated in [1], where a Nakagami-lognormal distribution is used but with the assumption that the associated normal variable is uncorrelated with its time derivative (i.e. $\rho = 0$). It shows that as we increase ρ with other important parameters constant, the curve shifts to right side further. Not only that, increasing ρ has an effect of slight modification in the shape of LCR curve (prominent for $\rho = 0.9$). These two effects of ρ on LCR curve may be useful to fit LCR curves obtained from measurements to those obtained analytically.

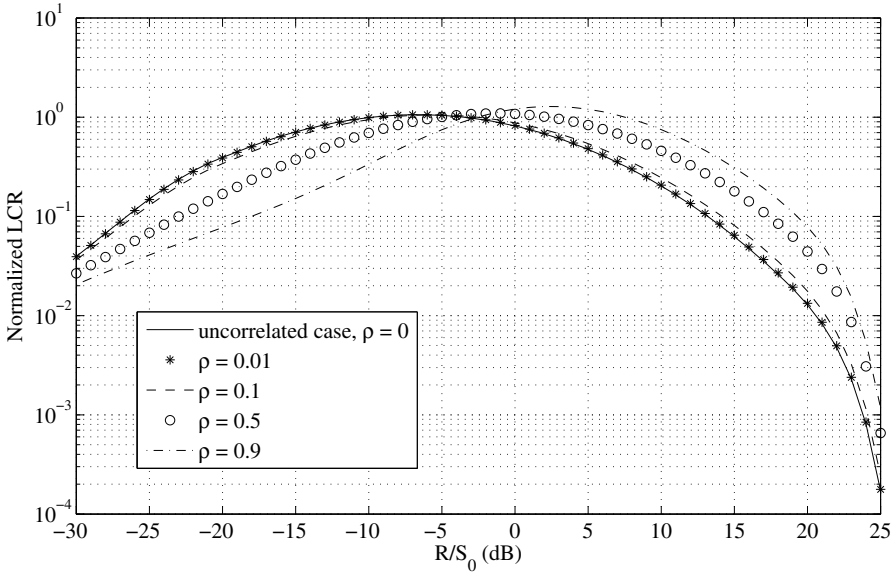


Fig. 1. LCR vs. R/S_0 (dB); $m = 1$, $\Omega = 0.1$, $\sigma_s = 3$ dB, $f_{ratio} = 1$

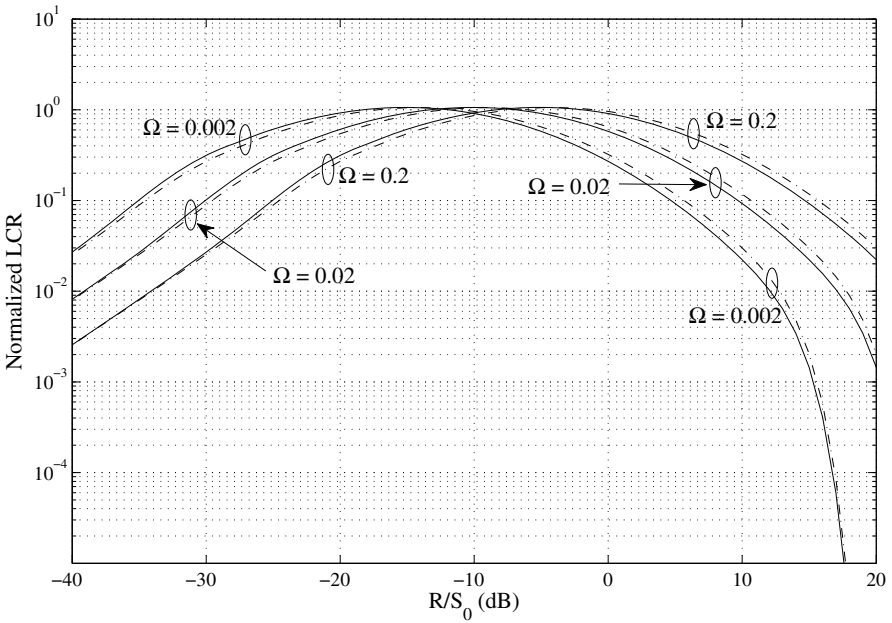


Fig. 2. LCR vs. R/S_0 (dB); $m = 1$, $\sigma_s = 3$ dB, $f_{ratio} = 1$; solid and dashed line for $\rho = 0$ and $\rho = 0.1$ respectively

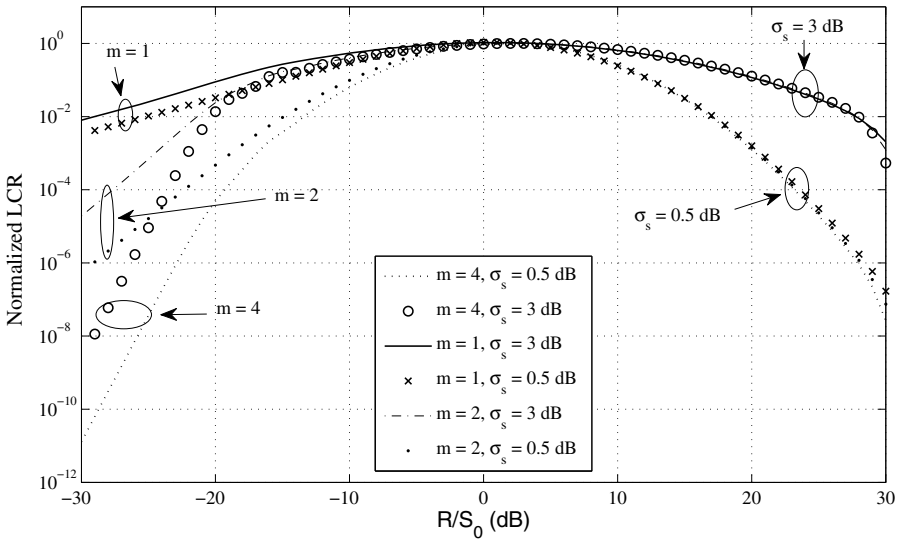


Fig. 3. LCR vs. R/S_0 (dB); $\Omega = 2$, $\rho = 0.1$, $f_{ratio} = 1$

In Figure 2 the effect of Ω , the mean power of Nakagami- m process, on LCR curve is shown at $m = 1$. It also has the effect of shifting but with no change in shape of the curve. Ω and σ (standard deviation of scattered components in Rayleigh fading) are related as [10],

$$\Omega = \frac{2\sigma^2}{\sqrt{1 - \sqrt{1 - m^{-1}}}}. \tag{13}$$

As the mean power Ω decreases, fading becomes severe and LCR at a given negative value of R/S_0 (dB) increases. That explains the shifting of LCR curve to left for decreasing Ω .

In Figure 3, LCR curves for different values of m and σ_s have been plotted. It is observed from the plots that for negative values of R/S_0 (dB), LCR decreases as m is increased, while it increases with σ_s . The effect of m is dominant here and increasing of m means fading effects are getting less severe. But for $R > S_0$, m seems to have a little effect on LCR curves and at a specific positive value of R/S_0 , LCR increases as σ_s is increased. This accords with the phenomenon that small-scale fading affects the received signal more when the envelope is below mean signal level in an area. On the other hand, shadow fading becomes dominant when envelope is well above the mean signal level (complying with equation (7) in [5]).

4 Conclusion

Analytical expression of level crossing rate has been derived for the proposed model. The effect of different important parameters like m , Ω , σ_s and ρ has been discussed

in detail. Vary small value of ρ leads to LCR curves same as that of conventional Nakagami-lognormal model given in [11] and increasing ρ provides right shift and little shape change of LCR curves. These attributes should act as extra degree of freedom in fitting curves obtained from measurements to curves based on analytical expression. The effect of Ω , the mean power of Nakagami- m process, is also shifting LCR curves. Another important finding is that, the small-scale Nakagami fading parameter m influences LCR values more compared to lognormal shadow fading parameter σ_s when received signal level is below the area mean signal level S_0 . For the reverse case, σ_s has dominant effect on LCR curves.

References

1. Rappaport, T.S.: *Wireless Communications*, 2nd edn. Prentice-Hall, NJ (2002)
2. Egli, J.J.: Radio propagation above 40 MC over irregular terrain. In: *Proc. IRE*, vol. 45(10), pp. 1383–1391 (1957)
3. Cox, D.C., Murray, R.R., Norris, A.W.: 800 MHz attenuation measured in and around suburban houses. *Bell Syst. Tech. J.* 63(6), 921–954 (1984)
4. Goldhirsh, J., Vogel, W.J.: Mobile satellite system fade statistics for shadowing and multipath from roadside trees at UHF and L-band. *IEEE Trans. Antennas Propag.* 37(4), 489–498 (1989)
5. Loo, C.: A statistical model for a land mobile satellite link. *IEEE Trans. Veh. Technol.* 34(3), 122–127 (1985)
6. Pätzold, M., Li, Y., Laue, F.: A study of a land mobile satellite channel model with asymmetrical doppler power spectrum and lognormally distributed line-of-sight component. *IEEE Trans. Veh. Technol.* 47(1), 297–310 (1998)
7. Corazza, G.E., Vatalaro, F.: A statistical model for land mobile satellite channels and its application to nongeostationary orbit systems. *IEEE Trans. Veh. Technol.* 43(3), 738–742 (1994)
8. Vatalaro, F.: Generalised Rice-lognormal channel model for wireless communications. *Electron. Lett.* 31(22), 1899–1900 (1995)
9. Hwang, S.-H., Kim, K.-J., Ahn, J.-Y., Whang, K.-C.: A channel model for nongeostationary orbiting satellite system. In: *Proc. IEEE Veh. Technol. Conf.*, vol. 1, pp. 41–45 (1997)
10. Nakagami, M.: The m -distribution - a general formula of intensity distribution of rapid fading. In: Hoffman, W.G. (ed.) *Statistical Methods in Radio Wave Propagation*, pp. 3–36. Pergamon Press, Oxford (1960)
11. Tjhung, T.T., Chai, C.C.: Fade statistics in Nakagami-lognormal channels. *IEEE Trans. Wireless Commun.* 47(12), 1769–1772 (1999)
12. Yacoub, M.D., Bautista, J.E.V., Guerra de Rezende Guedes, L.: On higher order statistics of the Nakagami- m distribution. *IEEE Trans. Veh. Technol.* 48(3), 790–794 (1999)
13. Youssef, N., Munakata, T., Takeda, M.: Fade statistics in Nakagami fading environments. In: *Proc. IEEE 4th Int. Symp. Spread Spectrum Techniques and Applications*, Mainz, Germany, vol. 3, pp. 1244–1247 (1996)
14. Leon-Garcia, A.: *Probability and Random Processes for Electrical Engineering*, 2nd edn. Pearson Education, NJ (2007)

Cache Invalidation for Location Dependent and Independent Data in IVANETS

Anurag Singh, Narottam Chand, and Lalit Kr Awasthi

Computer Science & Engg Dept, NIT Hamirpur
anuraglkw@gmail.com, {nar,lalit}@nitham.ac.in

Abstract. Internet-based vehicular ad hoc network (IVANET) is an emerging technique that combines a wired Internet and a vehicular ad hoc network (VANET) through roadside infrastructures for providing universal information and service accessibility to the vehicles. A key optimization technique in IVANETS is to cache frequently accessed data items in a local storage of vehicles. Since vehicles are not critically limited by the storage space, it is a less of a problem which data items to cache. Rather, a critical design issue is how to invalidate them when data items are updated. This is particularly a concern due to vehicles' high-speed mobility. In this paper we proposed a scheme in which we classify data in two categories and use two different Invalidation Reports (IRs) to invalidate them.

Keywords: cache invalidation, IR, IVANET.

1 Introduction

A vehicular ad hoc network (VANET) consists of a set of high-speed mobile vehicles equipped with communication facilities [1]. It supports inter-vehicle communications through a multi-hop message relay without the assistance of any fixed infrastructure as in adhoc networks. In order to provide a flexible connectivity, accessibility, and a rich set of services through internet, it is imperative to consider the integration of a VANET with a wireless infrastructure, such as a wireless local area network (e.g. IEEE 802.11) and a wireless wide area network (e.g. 3G). It is envisaged that such an IVANET, will prevail to become a ubiquitous communication infrastructure in the coming times.

A key optimization technique in improving the communication performance of IVANETS is to cache the frequently accessed data items in a local storage. In an IVANET, it is less of a problem to determine which data items to cache because the storage space in a vehicle is not critically limited but large. However, a critical design issue is the cache invalidation scheme, since applications require data consistency with the server to avoid using any stale data. When a data item in a server is updated, it is necessary to invalidate the cached copies of this data item stored at the vehicle cache by broadcasting an invalidation report (IR). However, due to fast roaming vehicles, cache invalidation schemes developed for cellular networks and mobile ad hoc networks (MANETS) may not work well. Unlike these conventional networks, energy conservation is not an issue in IVANETS because a vehicle is supported by its

own built-in battery. Rather, our concerns are query delay and cache hit ratio in the presence of mobility and frequent data updates.

The following observations characterize the IVANETs in the context of cache invalidation: (i) Due to high-speed mobility, vehicles reside in a coverage area for a short period of time. For example, the connection time within a coverage area ranges from 5 to 24 seconds at city driving [2]. Therefore, when a data server broadcasts an IR, it is very difficult to recognize which coverage area(s) to target. And, since multiple coverage areas are involved in a broadcast operation, the cost of broadcasting IRs becomes non-negligible; (ii) Since it is unlikely that adjacent vehicles have common data items in a real IVANET environment, it is wasteful to broadcast the same content of IRs to different vehicles since most of contents may not be relevant to the some vehicles; and (iii) As pointed in [3], a Web proxy caching may reduce network traffics in a data server but it does not reduce the traffics in wireless links. In order to support a scalable invalidation operation with the minimized IR traffics in both data server and wireless links, it is essential to coordinate with network agents of location management. To address these problems we propose a scheme in which we try to effectively deal with cache invalidation of fast moving vehicles without incurring significant overhead. Our work contributes to following:

- (1) We suggest a new infrastructure consisting of data servers, Home Agents and the Access points only. This in comparison to existing infrastructure don't have extra layer of gateways as all the mobility and location dependency is taken care by home agents.
- (2) We classify data into two categories location dependent and location independent. Location dependent data consist of information that depend upon the current location of vehicles, for example, nearest hospital, nearest petro pump etc. If the current location changes, the value of such data also changes. While location independent data are general services and data that doesn't address the location dependency.
- (3) We then give an IR based cache invalidation scheme which invalidates both location dependent and independent data.

2 Related Work

In this section we analyze the previous related work. There are two strategies given in [4] and [5]. In first scheme, Sunho Lim et al give cooperative cache method (CCI) and its enhanced version (ECCI). They proposed a state-aware cooperative approach, where both a server and location management agents coordinate for cache invalidation. By maintaining a list of data items and the access history by vehicles, a server asynchronously sends an IR to an HA rather than blindly broadcasts to vehicles. Then the HA judiciously refines and distributes the IR to appropriate GFAs. In this scheme, GFAs do not pro-actively send the IR to the individual vehicles, but reply a vehicle's validity request by a on-demand-basis. Then GFAs pass the IR to vehicles.

Second scheme by Sunho Lim et al works on triangular routing system based on Mobile IP. They proposed cache invalidation scheme integrated with a mobile IP based location management. The server asynchronously sends an IR to a home agent (HA) rather than blindly broadcasts it to the vehicles. Then the HA judiciously refines and distributes the IR to appropriate gateway foreign agents (GPAs) based on triangular routing method similar to Mobile computing. When a vehicle moves into a coverage area within the same regional network, it sends the location update to the GFA. When a vehicle moves into a different regional network, however, it sends the location update to the HA for correct forwarding the packets through the GFA and AP. Both these schemes work take all the data as the same where as we classify the data into two categories and apply invalidation to both the data differently.

3 Proposed Scheme

3.1 System Model

The System model is as represented in figure below:

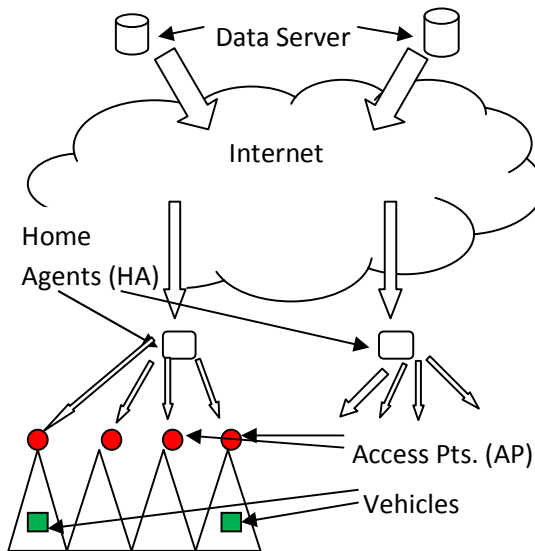


Fig. 1. System model of proposed IVANET

In this scheme we propose a system model that is similar to the existing model but doesn't have the extra GFA (Gateway Foreign Agent) layer. As in our all the location management work is done by the Home Agent. Therefore we removed the extra infrastructure in the existing model. Thus our model now consists of data servers at top of the hierarchy with the internet connecting them to the Home Agent. The home agent is connected to number of Access points each having its own coverage area.

3.2 Proposed Approach

In this paper our approach is to classify data into location dependent and independent data. Let's assume that location dependent data is 40% while location independent data is 60%. Now as the vehicle enters new coverage area under different region, it registers to new HA. On receiving the advertisement from new HA the location dependent data automatically invalidates. This is due to change in location. Thus 40% of data which depends on location of the vehicle does not need IR from the server. Therefore server only needs to send IR for only 60% data. This reduces the traffic at the server and increases the utilization of the bandwidth of the channel. Now IR from server is broadcasted to HA which is connected via internet (wired). The HA at which the vehicle is currently registered, passes the IR to access points that are under it. Rest of the HA discard the IR. Access points broadcasts the IR in their coverage area on the wireless network. Each vehicle is uniquely identified by a mobile IP. Thus server doesn't have to keep record of location of vehicle.

3.3 Proposed Cache Invalidation Technique

Let the total number of data be D and the total number of vehicle be N . Then 40% data will be location dependent $(d_x^{ld}, x \in D)$ and 60% data will be location independent $(d_x^{li}, x \in D)$. Server will maintain a list in register (R_s) , $[d_x^{li}, vid(v_y), t_{xy}]$, where t_{xy} is time of access of data d_x^{li} by vehicle having id (v_y) . If the data changes the server will generate an IR $(IR_s), \langle d_x^{li}, vid(v_{y1}, v_{y2}, \dots), t_{curr} \rangle$, and send it or broadcast it to all HA. Here, t_{curr} is the current timestamp and (v_{y1}, v_{y2}, \dots) are vehicle which accessed (d_x^{li}) . Every vehicle must be registered at the HA under whose coverage area it is present moving. This is maintained in register (R_{ha}) , $[vid(v_y)]$. On receiving the IR_s HA compares vid in IR_s and R_{ha} . If the match is not found IR_s is discarded, otherwise HA will broadcast another IR $(IR_{ha}), \langle d_x^{li}, t_{curr} \rangle$, to its coverage area. Therefore, invalidating the location independent data in the vehicle cache.

Now each location dependent data is cached in vehicles register (R_v) as $[id(HA_z), d_x^{ld}]$ where HA_z is the HA under which d_x^{ld} was cached. HA will broadcast advertisement, $\langle id(HA_z), t_{curr} \rangle$, at regular time period. As the vehicle enters another coverage area it will hear the advertisement and send register request to new HA. It will also match for HA id. In case of mismatch the location dependent data will be automatically invalidated.

While the vehicle request for new data (query request), it sends a query packet, $\langle d_x^{li}, id(HA_z), vid(v_y), t_{curr} \rangle$ to the server. Server update it register with d_x^{li} and $vid(v_y)$. It sends back reply to HA having id same as in the query packet.

In other case the request packet is $\langle d_x^{ld}, id(HA_z), vid(v_y), t_{xy} \rangle$ but server do not update entry in its register for location dependent data. This is mechanism of our proposed scheme.

4 Theoretical Analysis

The theoretical analysis has been done by us to support our work. We assumed a random map for the road links. It was covered by 30 APs (Access Points). The network would be connected by three Protocols:

1. Car to car adhoc network which was ranged for about 500 mts.
2. Car to AP and AP to Car via wireless network again ranged for about 500 mts.
3. AP to server via wired network working on TCP/IP connection. This is more secured and reliable than the other two connections. The parameters were taken into account is given in Table 1.

The result achieved was better than the previously two given algorithms. It was better in Query processing delay and giving better Cache hit ratio. A better query delay was achieved with growing data size. Heterogeneous IR provided more bandwidth and less server traffic.

Table 1. Parameter for the network analysis of the network

Parameter	Value
Network size (km)	24.5
Diameter of AP coverage (m)	500
Diameter of Car coverage (m)	500
Velocity (km/h)	60
Number of vehicle	250
Database size (items)	1,000
Data item size (KByte)	1 - 500
Hot data items	100
Cold data items	Remainder of DB
Hot data item access prob.	0.8
Number of Access Pts	30
Number of Home Agents	7

5 Conclusion

In this paper, we propose the cache invalidation issue in an IVANET, where minimizing the impact of high-speed mobility. We proposed a hierarchical network model and a cache invalidation scheme to maintain the cache consistency. In comparison to two existing schemes, in the proposed schemes server has send invalidation for only 60% of data, while other 40% invalidates itself. This increases the efficiency of query processing and bandwidth.

References

1. Dedicated Short Range Communications (DSRC) Home,
<http://www.leeearmstrong.com/DSRC/DSRCHomeset.htm>
2. Bychkovsky, V., Miu, B.H.A., Balakrishnan, H., Madden, S.: A Measurement Study of Vehicular Internet Access Using in Wi-Fi Networks. In: Proc. ACM MOBICOM, pp. 50–61 (2006)
3. Chen, H., Xiao, Y.: Cache Access and Replacement for Future Wireless Internet. IEEE Communications Magazine, 113–123 (2006)
4. Lim, S., Yu, C., Das, C.R.: Cooperative Cache Invalidation Strategies for Internet-based Vehicular Ad Hoc Networks. IEEE Transactions (2009) 978-1-4244-4581-3/09/\$25.00 © 2009
5. Lim, S., Chae, S.R., Yu, C., Das, C.R.: On Cache Invalidation for Internet-based Vehicular Ad Hoc Networks. IEEE Transactions (2008) 1-4244-2575-4/08/\$20.00 @2008
6. Wang, Z., Kumar, M., Das, S.K., Shen, H.: Investigation of Cache Management Strategies for Multi-cell Environments. In: Proc. 4th International Conference on Mobile Data Management (MDM), pp. 29–44 (2003)
7. Khurana, S., Kahol, A., Gupta, S., Srimani, P.: A Strategy to Manage Cache Consistency in a Distributed Mobile Wireless Environment. In: Proc. IEEE International Conference on Distributed Computing Systems (ICDCS), pp. 530–537 (2000)
8. Perkins, C.E.: IP Mobility Support, Request for Comments (RFC) (2002-2006)
9. Xie, J., Akyildiz, I.F.: A Distributed Dynamic Regional Location Management Scheme for Mobile IP. In: Proc. IEEE INFOCOM, pp. 1069–1078 (2002)
10. The CSIM User Guides,
<http://www.mesquite.com/documentation/index.htm>
11. Yoon, S., Ngo, H.Q., Qiao, C.: On “Shooting” a Moving Vehicle with Data Flows. In: Proc. Mobile Networking for Vehicular Environments (MOVE-2007), pp. 49–54 (2007)
12. Cao, G.: A Scalable Low-Latency Cache Invalidation Strategy for Mobile Environments. IEEE Transactions On Knowledge And Data Engineering 15(5) (September/October 2003)

VCAR: Scalable and Adaptive Data Dissemination for VANET

K. Naveen and Komathy Karuppanan

Department of Computer Science & Engineering,
Easwari Engineering College, Anna University, Chennai, India
knaveencs@gmail.com, gomes1960@yahoo.com

Abstract. Reliable data dissemination becomes a critical issue in Vehicular Ad Hoc Network (VANET) due to its high dynamic nature. Moreover, data delivery ratio degrades after a road junction as the vehicles depart in different directions. Existing schemes such as static node assisted adaptive data dissemination (SADV) has attempted to increase the delivery ratio for medium and low vehicular density. However in a real time scenario, the vehicular density at the road junction is higher than that of the roads. Our proposed system manages to improve the data dissemination at the junction even if the vehicular density is high. This paper introduces a routing protocol called Vehicular Cluster Assisted Routing (VCAR) for scalable networks. The objective of the proposal is to upgrade the performance of VANET under high traffic scenario. To support data dissemination at the junction, the associated network is effectively partitioned into groups called clusters and they tend to move along with the vehicles towards the destination. It is observed that the moving cluster significantly improves the connection time and thereby delivery ratio too. Scalability is also considered while designing the routing.

Keywords: VANET, V2V, V2I, cluster, scalability, data delivery.

1 Introduction

Vehicular Ad Hoc Network (VANET) is an integral part of the intelligent transportation system (ITS), which aims to improve road safety, optimize traffic flow and reduce congestion, and so on. Topographical structure of a VANET is more dynamic when compared to MANET where an end-to-end connection is usually assumed. MANET routing protocols fails at high mobility and radio obstacles as well. High mobility leads to frequent broken routes in VANET. Also, connectivity time solely depends on the vehicular density. Moreover, vehicular network brings additional problems related to the junction regions. For lower vehicular density, a sender at the intersection of the roads may not be able to find a connecting path to the destination that may be out of coverage region. This problem is attempted to solve by using static-node assisted adaptive data dissemination protocol for vehicular networks (SADV) [9], which enables packets to be buffered at the intersection until a connecting path is available.

Vehicular density remains high always in urban roads. In urban VANET, traffic signal further influences the vehicle movement and thereby network connectivity. A red signal abruptly stops the vehicles from moving, which may result in network partitioning or loss of connectivity. After the signal turns on, vehicles at the intersection of roads will depart in all possible directions. Vehicles will further look for new networks to join after the junction, which initiate an expensive route discovery process.

This paper makes an effort to improve the performance of static node placed at the junction. A novel routing mechanism called Vehicular Cluster Assisted Routing (VCAR) is proposed here to counteract mainly the problems arising out of high vehicular density. It is designed with the existing traffic infrastructures to partition the network effectively, for a better delivery ratio and to reduce broadcast storms. A message is to be relayed by a minimum of intermediate nodes to the destination. To do so, nodes are organized as a set of clusters in which one node called *cluster head* gathers data from the static node and send them to the actual destination. Cluster based routing will aim to provide less propagation delay and high delivery ratio with bandwidth fairness too. Our proposed model combines cluster based dissemination mechanism with the static node approach to tackle the high vehicular density.

2 Related Work

The following section surveys the research works on routing and data dissemination under various vehicular densities. Works reported in this vehicular ad hoc routing are brought out to study the impact of the models and schemes incorporated within.

Miguel Luis et al. [5] presented a new method to improve OLSR routing protocol by decreasing the probability of path breaks. The performance result explicitly confirms that the proposed protocol outperforms the original protocol, recommending its use for high mobility and high density scenarios. Manvi et al. [7] gave a comparison of the routing protocols such as AODV, DSR, and SWARM, and analyzed the differences in the performance. An important observation was that the examined routing protocols showed highly heterogeneous performance results and concluded that AODV and DSR may not be suitable for vehicular environments.

Jing Zuo et al. [2] in their work analyzed the properties of the two mobility models in high density urban areas and the results compared the performance of reactive and proactive routing protocols. It indeed, informs drivers and other passengers of potentially dangerous traffic situations while there is still time to avoid them. Vehicle Assisted Data Delivery (VADD) [10] implemented a predictable mobility model by taking the traffic pattern and road layout into account. It is based on carry-and-forward method and it performs well even in sparse road conditions.

Benslimane et al. [11] presented a new method to improve the high data rates of IEEE 802.11p based VANETs and the wide coverage area of 3GPP networks (e.g., UMTS) and the work envisions VANET-UMTS integrated network architecture. In that architecture, vehicles are dynamically clustered according to different related metrics. From these clusters, a minimum number of vehicles equipped with IEEE 802.11p and UTRAN interfaces are selected as vehicular gateways to link VANET to UMTS. In GSR (Geographic Source Routing) [4], the shortest path between source

and destination is calculated based on the street map. While calculating the shortest path, GSR does not consider whether there are enough vehicles or not on a street to provide connectivity between two involved junctions. OLSR [1] only perform well in high vehicular density.

From the above survey of related works, it is observed that the data dissemination at the road junction when the direction of the vehicle changes becomes a major issue in the routing of the vehicular communication. Cluster based dissemination mechanism with the static node approach is proposed in this paper to tackle the high vehicular density.

3 Proposed Protocol Model

The proposed Vehicular Cluster Assisted Routing (VCAR) mechanism considers the roadmap of an urban area that mainly deals with road junctions and high vehicular density. Each node facilitates the positioning services, such as GPS, which enables the position of the node, directional elements and its velocity. The VCAR protocol involves the operation of two modes namely junction mode and highway mode.

3.1 Junction Mode

As the vehicle approaches a road junction it switches to junction mode if the destination node is disconnected from the communication link. Then the sender hands over the packets to the static node placed at the road junction. The static node periodically scans the immediate environment and spots the network partitions called clusters as displayed in Fig 1. Cluster formation is covered in detail in Section 3.3. The cluster in the direction of the destination is chosen to forward the packets. One of the vehicles within the chosen cluster is selected as the cluster head (CH) and the data is disseminated to it.

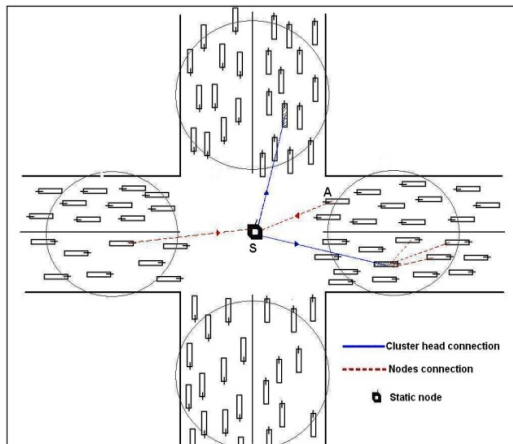


Fig. 1. Junction mode operation

3.2 Highway Mode

At regular interval of time the static node broadcasts a beacon signal. If the vehicle falls within the communication range of the static node, it stays in junction mode otherwise it switches to highway mode. In highway mode, the vehicles invoke V2V communication with multi-hop forwarding to deliver data to the actual destination. A mobile cluster carrying a swarm of vehicles is illustrated in Fig 2. This situation prevails till the mobile cluster reaches the next junction.

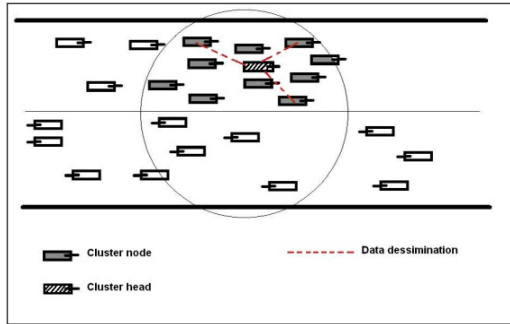


Fig. 2. Highway mode operation

3.3 Cluster Formation

Let us consider the junction mode in which N number of vehicles and a static node are present. The static node offers a high bandwidth data communication to the vehicles. Consider a vehicle node A having the velocity V_a , a directional vector $(\overrightarrow{dx}, \overrightarrow{dy})$ and the cartesian coordinates (X_a, Y_a) . Now consider the static node as an arbitrary point having cartesian co-ordinates (X_s, Y_s) . The steps involved in forming the cluster are listed below.

Step 1: Directional vector of vehicle A towards static node S is given by the equ (1).

$$\overrightarrow{SA} = (x_A - x_s, y_A - y_s) \tag{1}$$

Step 2: The velocity vector of vehicle A is calculated as in equ (2).

$$\overrightarrow{V}_a = v_a \overrightarrow{dx}_a + v_a \overrightarrow{dy}_a \tag{2}$$

Step 3: dsv is the directional similarity value, which is the dot product of direction vector and velocity vector is estimated as given in equ (3).

$$dsv = \overrightarrow{V}_a \cdot \overrightarrow{SA} \tag{3}$$

Similarly, the directional similarity value dsv for the rest of the vehicles in the junction is estimated. The vehicles having related dsv are grouped together as a single static cluster. The dsv of the vehicle which travels in the opposite direction will be omitted by the cluster. Each static node maintains a list of cluster IDs and a corresponding list of vehicles as given in Table 1. In this sample, cluster C1 has to

handle packets for two separate destinations. The source node indicates the origin of the packets and the destination points to the node where the packets are to be destined. This table gets renewed after every beacon by the static node and the beacon duration coincides with the traffic signaling.

Table 1. Static node maintaining a cluster table

Cluster ID	Source Vehicle ID	Destination Vehicle ID	Cluster Nodes	Time of Installation	No of packets
c1	n10	n2	n12,n8,n15,n17,n20	10.02s	4
c1	n40	n2	n12,n8,n15,n17,n20	10.02s	2

3.4 Cluster Head Announcement

Each cluster has its own cluster ID and it covers several vehicles. Once the cluster has been formed, cluster head (CH) is chosen using Algorithm-1.

```

Algorithm-1:
Input :  $N_a, \{ \eta(n_y), \xi(n_y), t_i(n_y) \} \forall n_y \in N_a$ 
Output :  $R(n_a)$ 
   $R(n_a) \leftarrow \phi$ 
  coverage  $\leftarrow 0$ 
  for each_neighbor  $n_y \in N_a$  do
    if  $\eta(n_y) \geq k_{est}$  and  $n_y$  is CH then
       $R(n_a) \leftarrow R(n_a) \cup n_y$ 
      update (coverage)
      remove_from_list( $n_y, N_a$ )
    end if
   $N_{ord} = \text{sort by descendent } (N_a)$ 
  while coverage < 100% do
    remove_from_list( $n_y, N_{ord}$ )
     $R(n_a) \leftarrow R(n_a) \cup n_y$ 
    update(coverage)
  end
    
```

The distance between the static node and the each of the vehicles in the cluster is estimated from equ (4).

$$D_{as} = \sqrt{(x_a - x_s)^2 + (y_a - y_s)^2} \tag{4}$$

Let R_s denotes the coverage range of the static node. If D_{as} is the closest distance within the coverage area, then the static node relieves the packet handling thereon to CH stating the source and the destination of the packets.

3.5 Mobile Cluster in Highway Mode

Once the traffic signal turns on, each cluster starts moving in its own direction. When the cluster moves away from coverage range of the junction region, cluster head starts announcing its status to all vehicles present within its cluster. VCAR protocol now

employs a multi-hop relaying technique within a mobile cluster to forward the packets to the actual destination. When compared to flooding, it reduces the number of redundant packets in the network. Cluster head (CH) starts selecting a set of relay nodes R within one hop region from CH using Algorithm-2. If enough relay nodes are not available in R , then two-hop neighbors are also added in the mobile cluster. Thus, using the relay nodes R in the form of a cluster, CH disseminates the packets to the respective destinations within the cluster.

Algorithm-2:

Input : $N_a, \{ \eta(n_y), \xi(n_y), t_i(n_y) \} \forall n_y \in N_a, OLD_R(n_a)$

Output : $R(n_a), OLD_R(n_a)$

$R(n_a) \leftarrow \phi$

enough_additional_R $\leftarrow 0$

coverage $\leftarrow 0$

for each_neighbor $n_y \in N_a$ do

if $\eta(n_y) \geq k_{est}$ and n_y is CH then

$R(n_a) \leftarrow R(n_a) \cup n_y$

update(coverage)

remove_from_list(n_y, N_a)

for each_neighbor $n_y \in OLD_R(n_a)$ do

if $n_y \in neighbor(n_a)$ then

$R(n_a) \leftarrow R(n_a) \cup n_y$

update(coverage)

enough_additional_R++

else

remove_from_list($n_y, OLD_R(n_a)$)

$N_{ord} = sort_by_ascendent \eta(N_a)$

while coverage $\neq 100\%$ and enough_additional_R $\neq 2$ do

remove from list(n_y, N_{ord})

$R(n_a) \leftarrow R(n_a) \cup n_y$

$OLD_R(n_a) \leftarrow OLD_R(n_a) \cup n_y$

update(coverage)

enough_additional_R++

end

4 Simulation and Performance Evaluation

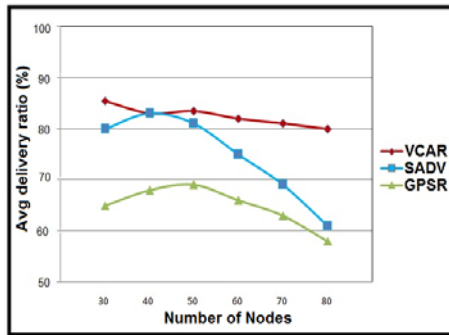
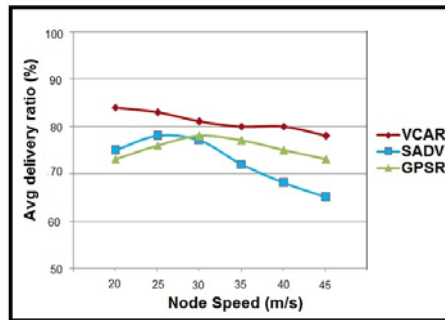
In order to observe the vehicle movements in the metropolitan scenario the Manhattan mobility model is preferred, which contains a grid road topology. The traffic simulator MobiSim [6] and the network simulator NS2 version 2.34 [8] are deployed to simulate the proposed protocol VCAR. The simulation parameters that are configured for the experiment are summarized in Table 2.

4.1 Simulation Results and Analysis

To analyze the protocol comprehensively, VCAR is compared with other two typical routing protocols namely SADV and GPSR. In this paper, in order to evaluate the protocol's flexibility to node density, the number of nodes is varied from 30 to 80. Similarly, to evaluate the protocol's adaptability to node speed, the node speed is varied from 15mps to 40mps. The simulation results are shown in Fig 3 and Fig 4.

Table 2. Simulation Parameters

Parameter	Value
Simulation area	2000m×2000m
Number of intersections	3
Distance between intersections	600m
Number of vehicles	30 – 80
Speed of vehicles	15 – 45 m/s
Simulation time	200s
MAC protocol	IEEE802.11p
Data rate	2 Mbps
Data packet size	512bytes

**Fig. 3.** Average delivery ratio for node speed 30 mps**Fig. 4.** Average delivery ratio for 50 nodes

It is observed from Fig 3 that VCAR and SADV have performed well to display a high delivery ratio at lower node density. GPSR fails to deliver the packets as this does not consider the junction characteristics. When nodes in the network are sparse, say between 30 and 40, VCAR performs better with as much as 22% increase compared to GPSR and as much as 16% increase compared to SADV. When node density is comparatively high say, from 50 to 80, the delivery ratio of VCAR remains high. Moreover, VCAR performs well in terms of scalability than other ones and its

delivery ratio reaches up to 80%. Fig 4 shows that the average delivery ratio of VCAR, which has higher adaptability while changing the node speed and it performs better than other protocols. Higher node density does not necessarily mean an improved performance for protocols such as GPSR, which do not consider the road conditions. The delivery ratio of GPSR improves evidently when there is an increasing node speed whereas SADV drops to about 65% as the speed increases. This implies that SADV is not able to cope up with high mobility.

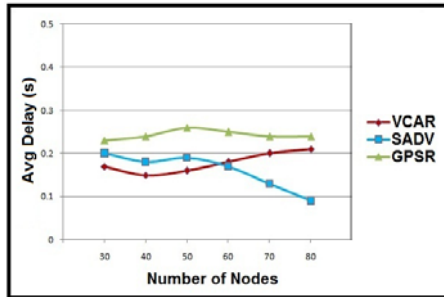


Fig. 5. Average end-to-end delay for node speed is 30 mps

Fig 5 shows the average end-to-end delay of the protocols versus node density. It is observed that GPSR and VCAR exhibit a moderate variation in the delay while increasing the number of nodes. The degrading performance of SADV compared to VCAR in the case of average end-to-end delay is due its dependence on node density and speed. The drop in delay for SADV on the higher side of the density is because of its behavior shown in Fig 3. SADV drops packets heavily at higher node density that results packets-in-hand to deliver are low and it is obvious that the delay also gets lower. GPSR tries to initiate the route discovery more often, the delay increases considerably. VCAR handles a high packet delivery as the node density and speed increases and the delay shoots up. Mobile cluster, most of the time helps to stay connected for a longer period at least till the next road intersection.

5 Conclusion

This paper proposed a new technique VCAR, a routing protocol for vehicular networks, which improves the performance of VANET even if the vehicular density is high. Static node mechanism combined with the features of moving clusters has been taken advantage of enhancing the packet delivery ratio and delay. It is evaluated the performance of VCAR in an urban roadmap with more cross roads. Under MobiSim and NS2 simulators, VCAR performance in terms of node density and speed is compared with two other routing protocols. It is observed from the analysis of the simulation results that VCAR routing mechanism is able to adapt to scalability and frequent topological change by extending the connection lifetime of mobile nodes using moving clusters so as to improve the performance of VANET. It can adapt to VANET in urban scenarios and mainly in junction regions commendably.

References

1. Clausen, T., Jacquet, P.: Optimized Link State Routing Protocol. IETF Internet Draft (2003), <http://www.ietf.org/internet-drafts/draft-ietf-manetolsr-10.txt>
2. Zuo, J., Wang, Y., Liu, Y., Zhang, Y.: Performance Evaluation of Routing Protocol in VANET with Vehicle-node Density. *IEEE Trans. Veh. Technol.* 38(5), 3709–4244 (2010)
3. Karp, Kung, H.: GPSR: Greedy perimeter stateless routing for wireless networks. In: Proceedings of the 6th International Conference on Mobile Computing and Networking, MobiCom, New York, pp. 243–254 (2000)
4. Lee, K.C., Haerri, J., Lee, U., Gerla, M.: Enhanced Perimeter Routing for Geographic Forwarding Protocols in Urban Vehicular Scenarios. In: Proc. of Globecom Workshops, November 26–30, pp. 1–10. IEEE, Los Alamitos (2008)
5. Luis, M., Oliveira, R., Bernardo, L., Pinto, P.: Improving Routing Performance in High Mobility and High Density ad hoc Vehicular Networks. In: IEEE Symposium, Portugal, pp. 1–6 (2010)
6. MOBISIM Tool, <http://sourceforge.net/projects/mobisim/files>
7. Manvi, S., Kakkasageri, M.S., Mahapurush, C.V.: Performance Analysis of AODV, DSR, and Swarm Intelligence Routing Protocols In Vehicular Ad hoc Network Environment. In: IEEE Conference on Future Computer and Communication, Kuala Lumpur, pp. 21–25 (2009)
8. NS2 Website, <http://www.isi.edu/nsnam/ns/ns-build.html>
9. Ding, Y., Xiao, L.: SADV: Static-Node-Assisted Adaptive Data Dissemination in Vehicular Networks. *IEEE Trans. Veh. Technol.* 59(5) (June 2010)
10. Zhao, J., Cao, G.: VADD: Vehicle-assisted data delivery in vehicular ad hoc networks. *IEEE Trans. Veh. Technol.* 57(3), 1910–1922 (2008)
11. Benslimane, A., Taleb, T., Sivaraj, R.: Dynamic Clustering-Based Adaptive Mobile Gateway Management in Integrated VANET- 3G Heterogeneous Wireless Network. *IEEE Trans. Veh. Technol.* 29(3), 0733–8716 (2011)

Efficient Distributed Group Authentication Protocol for Vehicular Ad Hoc Network

Priya Karunanithi and Komathy Karuppanan

Department of Computer Science and Engineering, Easwari Engineering College,
Anna University, Chennai, India
kv.priya06@gmail.com, gomes1960@yahoo.com

Abstract. Intelligent Transportation System (ITS) has evolved to support Vehicular Ad Hoc Network (VANET) for communicating road safety messages while driving on the road. However, VANET faces the vital security challenges such as key and certificate management and privacy preservation when scalability and mobility become the essential part of this network. Privacy service prevents the original identification of the vehicle to be used by other vehicles from tracing. In this paper, a distributed authentication protocol called as Efficient distributed Group Authentication (E-GAP) is proposed to resolve the most conflicting security requirements in authentication and conditional privacy. The group signature and batch verification schemes proposed in the E-GAP protocol aim to control the overall message delay during authentication process. Session based pseudonym strengthens further the privacy preservation process that is decided on the anonymous communication between vehicles. Trusted Authority is capable of tracing both benign and malign vehicles using the conditional privacy feature of the protocol.

Keywords: IVC, RVC, group signature, privacy, bogus messages.

1 Introduction

VANET is a promising technology that employs wireless communication networks to facilitate vehicles to communicate with one another, and with a fixed infrastructure such as Road Side Units (RSU). IEEE 802.11p task group is working on Dedicated Short Range Communication (DSRC) Standard, which aims to enhance the 802.11 protocol to support wireless data communications for RSU to vehicle communications (RVC) and Inter vehicle communication (IVC) [1]. Application of IVC and RVC falls into two categories: safety applications and infotainment applications. In this paper, only security issues of safety-related messages are focused as they are crucial for VANET applications. Vehicles periodically broadcast safety messages to provide important information to other vehicles. So even in the presence of attackers, it is necessary to continue operation of safety-related applications.

In this paper, an efficient distributed group authentication protocol (E-GAP) is proposed to secure vehicular communications. E-GAP protocol can efficiently deal with growing revocation list while achieving conditional traceability by trusted authority. Rather than depending on a huge storage space at each vehicle, the

proposed protocol can keep key storage minimal without losing privacy. Meanwhile, proposed protocol achieves a fast verification on safety messages and an efficient conditional privacy mechanism.

2 Related Work

IEEE 802.11p task group is working on the DSRC Standard, which aims to enhance the 802.11 protocol to support wireless data communications for IVC and RVC [1]. WAVE technology is the next generation DSRC technology, which provides high-speed IVC and RVC and has major applications in ITS, vehicle safety services and Internet access. Security and Privacy issues discussed in various works [2-9] have been brought out here. Habaux et al [2] have introduced anonymous authentication. Each anonymous key pair is employed for certain number of messages, which introduces complexity in key storage. Raya et al [3] have proposed a security scheme based on PKI and for tracing by authority using GPS (Global Positioning System). Gollan et al [4] proposed use of digital signatures along with GPS technology to securely identify cars, improve fleet management, and provide new applications for private and public sectors. Habaux et al [5] in another proposed the security architecture for VANET and he applied digital signature for message integrity. In scheme [6], vehicle uses short-lived keys to sign messages used for VANET communication. Changing temporary keys impacts applications in two major ways: interrupting routing and interrupting ongoing end-to-end communication.

Zhang et al [7] proposed a RSU-aided message authentication scheme called RAISE in which RSUs are responsible for verifying authenticity of messages sent from vehicles and for notifying results back to vehicles. This scheme has high message loss ratio because all validation is performed only by RSU. In GSIS [8], group signatures for OBUs and identity-based signatures for RSUs have been proposed in order to maintain security and privacy. Lu et al. [9] proposed an alternative way to overcome limitation of pre-storing a large number of anonymous certificates while preserving conditional privacy. Since a vehicle should change anonymous certificate quite often to avert tracing of messages, it should frequently interact with RSUs. In [13], protocol maintains two key sets: anonymity key set, is used by vehicles for signing the safety messages and emergency key set, is used for signing purposes in emergency situations such as certificate revocation caused by a misbehaving vehicle. This protocol has high message loss ratio due to slow validation of messages and has high key storage complexity due to short-time validity of key. Sun et al [14] employs an identity-based cryptosystem where certificates are not needed for authentication and uses PLT (pseudonym lookup table) for each registered vehicle in its domain. Zhang et al [15] sets up secure channel between the RSU and vehicle using signcryption and uses batch verification. Jiun-Long et al [16] used batch verification method to simultaneously authenticate multiple requests sent from different vehicles by RSU and communication is V2I.

The proposed protocol mainly focuses on reduction of message delay, loss ratio and certificate revocation list (CRL) size. E-GAP also considers issue in key management and conditional anonymity preservation of vehicle.

3 Protocol Design

This section lays out the salient features of the proposed protocol in terms of system architecture, functional model and related algorithms. The components of distributed model of the proposed protocol are TA, RSU and vehicle. TA generates digital certificate for vehicles and RSUs and distributes through RSUs. TA also invalidates the certificate of malicious vehicle who is broadcasting the false safety information and stores IDs of malevolent vehicles in a black list. TA broadcasts the black list to RSUs in order to prevent network from further injecting fake safety messages. In this paper, RSUs issue pseudonyms to vehicles for communication and also maintains CRL and black list. RSUs connect with TA by wired links and act as relay nodes between TA and vehicles. RSU is assumed to have higher communication bandwidth. IEEE802.11p protocol is used for IVC and RVC. The vehicle requests for pseudonym from RSU for communicating with other vehicles and thereon vehicle anonymously broadcasts the safety message with other vehicles.

3.1 Security Requirements

The requirements of E-GAP protocol while embedding security services are studied and emphasized as:

- Message integrity and authentication: A vehicle should be able to verify that a message is indeed sent and signed by another vehicle without being modified.
- Identity privacy preserving: The real identity of a vehicle should be kept anonymous from other vehicles and a third-party should not be able to reveal a vehicle's real identity by analyzing multiple messages sent by it.
- Traceability and revocability: Although a vehicle's real identity should be hidden from other vehicles, if necessary, TA should have ability to obtain a vehicle's real identity and to revoke it for future usage.

3.2 E-GAP Protocol

The functional components of E-GAP protocol are displayed in Fig 1. TA has two main components namely certificate management and tracing. RSU, as part of authentication, involves in pseudonym generation and CRL maintenance. Vehicles are responsible for signature generation and batch verification of safety messages. RSU also maintains a database containing CRL and pseudonym details. TA preserves the certificate list, CRL and black list. This section further details the primary functions of E-GAP protocol.

Certificate Management. TA performs certificate generation, distribution and revocation. A vehicle requests for a certificate to TA and sends it through RSU. RSU waits for such requests from vehicles for a period of time, consolidates and sends to TA as a single request.

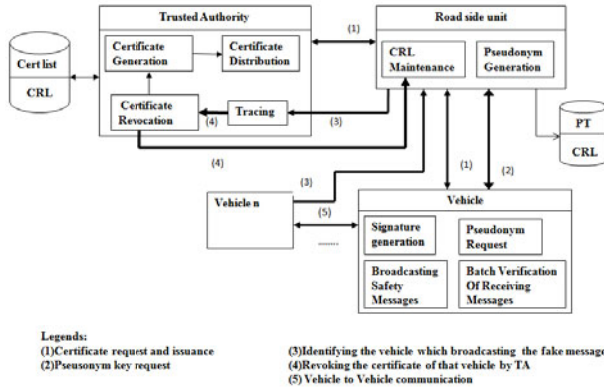


Fig. 1. Functional Architecture

The format of consolidated request message of the vehicles is given by $M_1 = \{VIDs, RID, N\}$ where VIDs is the set of vehicle IDs, RID is the ID of requesting RSU and N refers to the number of requests that are packed within message. TA generates certificates for all requesting vehicles as per equation (1).

$$Certificate_v = E_{ku-TA}(id \parallel Lifetime \parallel Public Key) . \tag{1}$$

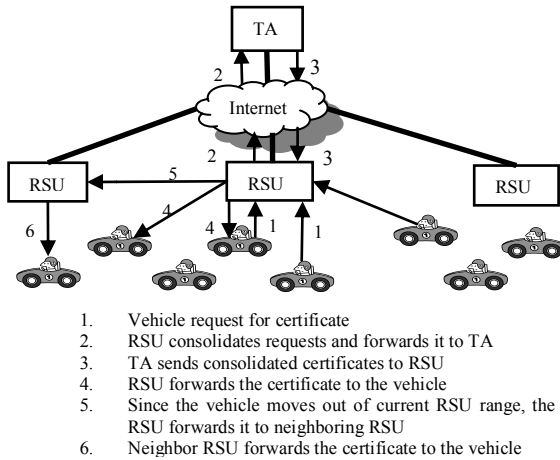


Fig. 2. Certificate Request and Reply

In order to protect certificate from adversary, certificate of the vehicle is encrypted using TA’s public key. The certificate contains original identity of the vehicle, lifetime and a public key. Bilinear pairing algorithm [10] is used for generating the public key. At time instant t1, a vehicle forwards its request to RSU for a certificate and is displayed in Fig 2. On generating the certificates for all the vehicles, TA transfers a consolidated message to RSU. The message from TA comprises the

certificates in the format $M_2 = \{\text{VIDs, certificates, RID, N}\}$, where certificates refer to certificates of VIDs in the list. RSU distributes certificates individually to vehicles. If the vehicle has moved out of its range, then RSU forwards the certificate to its neighboring RSU and the certificate successfully reaches the corresponding vehicle. Fig 2 also represents the forwarding of certificate to a vehicle at time instant t_2 . TA and RSU maintain two types of lists called revocation and adversary. The revocation list (RL) contains IDs of all expired certificates. The adversary list (AL) contains IDs of misbehaving vehicles.

Pseudonym Generation and Distribution. RSU forms a group containing on-the-fly vehicles within its communication range. RSU periodically broadcasts its public key to the passing vehicles. When the vehicle enters the communication range of RSU and if the vehicle is a new group member, RSU runs the pseudonym generation procedure; otherwise it does nothing. Steps involved in pseudonym generation and distribution are listed in Fig 3.

Algorithm: Pseudonym Generation and Distribution

- V_i is a new group member of RSU, R_i .
1. V_i selects timestamp TS, certificate Cert_v and public key of R_i , K_{u-R_i}
 2. Compute signature $\sigma = E_{K_{u-R_i}}(H(\text{Cert}_v \parallel \text{TS}))$
 3. Send $\rho = (\text{Cert}_v \parallel \text{TS} \parallel \sigma)$ to R_i
 - On receiving ρ from V_i , R_i generates the pseudonym by checking the following steps:
 4. $\text{Cert}_v \parallel \text{TS} = E_{K_{u-R_i}}(H(\text{Cert}_v \parallel \text{TS}))$
 5. If encrypted value of step 4 is equal to the signature then
 6. accept the message
 7. If certificate is valid then
 8. Generator $g_2 \in G_2$
 9. pseudonym_rk, $p_{rk} = x, x \in Z_q$
 10. pseudonym_uk, $p_{uk} = (g_2)^x$
 11. send $(\text{TP} \parallel p_{rk} \parallel p_{uk})$
 12. else
 13. add Cert_v in RL
 14. update RL of TA
 15. else
 16. revokes the certificate
 17. add Cert_v in AL and RL
 18. update RL and AL of TA
 19. Once the pseudonym is received, check whether the value of TS is equal to the original TS. If it is equal accept the pseudonym; otherwise reject the message

Fig. 3. Procedure for pseudonym generation and distribution

The pseudonym request is described in the steps 1 and 2. Every RSU periodically broadcasts its public key to the passing vehicles and neighboring RSUs. The vehicle uses this public key for signing the request message. The steps from 4 to 18 represent pseudonym generation procedure. RSU uses its private key for verifying signature in step 5. For each vehicle, RSU maps original identity with generated pseudonym using pseudonym table (PT) and later this PT is used to track a vehicle by TA. In step 19, the vehicle accepts the pseudonym. After receiving pseudonym, a vehicle checks the timestamp (TS) for redundancy.

Group Signing. A vehicle uses its pseudonym to broadcast the safety message with group signature. The group is formed with identification. Only the members of the group can send and receive message among them. The message format used by the vehicle for broadcasting safety message is $M_3=\{M,\sigma,T,p_uk\}$. Lifetime represents expiry duration of safety message. The signature [14] is generated as in equation (2).

$$\sigma \rightarrow H(M)^{p_rk} . \quad (2)$$

Signature Verification. Message verification employs either by a single signature verification method or a batch verification method. A vehicle checks the validity of the received message before verification and the signature is verified by pairing operations [11]. The verification of messages should be performed extremely faster because the vehicle broadcasts a safety message almost for every 300ms. To verify a large number of safety messages within a short time, batch verification is preferred. If a vehicle receives only a single message, it verifies the message as per the steps given in Fig 4.

Algorithm: Single Message Verification

Input: Message= $\{M, T, \sigma, p_uk\}$ and $g1, g2, Zq$

1. Check the validity of T value
2. If T expires reject the message
3. $C=e(H(M), p_uk)$
4. if $C==e(\sigma, g2)$
then accept the message
else reject the message

Fig. 4. Signature Verification

Algorithm: Batch Verification

Input: Message= $\{M, T, \sigma, p_uk\}$ and $g2$

1. Check the validity of T value
2. for $i=1; i<s; i++$
3. If T expires reject the message
4. End for
5. $C=[\prod_i e(\prod_{j=1}^i (H(M_j), p_pk_j))$, where $1<s<n-1, 1<l<n$
6. if $C==e(\prod_{j=1}^i \sigma_j, g2)$
then accept all the message
else reject the message

Fig. 5. Batch Verification

Batch Verification: The batch verification requires $(s-1)$ pairing operations, where s represents number of signers [12]. Fig 5 gives the steps executed when a vehicle receives more than one message. The steps from 1 to 4 check the validity of each message.

Identity Tracing. Malicious vehicles existing in VANET may send fake safety messages to other vehicles to divert the traffic intentionally. If this occurs, TA can disclose the identity of vehicle by invoking following steps.

- i. Recover the vehicle’s p_uk from safety message
- ii. Get (ID, p_uk) from RSU’s database
- iii. Revoke the certificate of that vehicle and update RL and AL list

After revealing the identity of vehicle, TA revokes the certificate and puts that vehicle’s ID into adversary list and forwards the list to all RSU’s.

4 Simulation

The proposed protocol is experimented using network simulator tool, NS2 version 2.34. The simulation parameters are listed in Table1. RSU is able to manage more number of vehicles since it has higher communication range of about 1000m. Finally, to facilitate higher range of communication at RSU, IEEE802.11p protocol is used.

Table 1. Simulation Parameters

Simulation Time	200s
Immobile nodes	4
Mobile nodes	40 to 60
Min speed	60kmph
Max Speed	120kmph
Mac Protocol	IEEE 802.11p
Mobility Model	Free Wave
Routing Protocol	AODV

Table 2. CRL Table

Original Vehicle ID	Certificate Expiry Time	Pseudonym
17	170.648	61389026929261741771378717295432846613636
3	170.852	43541829658445194304395258395667437969462627
15	181.241	16486079511470902412731328954526044290254792
12	183.029	10148751248186918714752029999330926838606496038

RSU maintains the certificate expiry time and pseudonym in a table. The expiry time serves as an input for renewal of the certificate. The table also maps the original vehicle id with pseudo id. The current pseudonym available in the table is used for tracing the vehicle by TA. CRL table with sample values is depicted in the Table 2. Privacy of the vehicles is thus preserved through pseudonyms and our proposed model introduces a session based pseudonym. Use of session-based pseudonym sampled from the simulated results is depicted in Table 3. Node 17 uses three different pseudonym values during three sessions and as a sender it transfers the safety message to destination node 6. Similarly node14 has been shown with different pseudonym values.

Table 3. Session based Pseudonym

Vehicle ID		Session	Pseudonym of the Source
Source	Dest		
17	6	1	613890269203569261741737871295432846613636087
17	6	2	6731191100184764731766577920596720673740882
17	6	3	103600464468072398516253802887237601485855109
14	8	1	204040142927713982204941879162987612977249
14	8	2	1816855413968434172273889122360304277491882

4.1 Performance Evaluation

The observations of E-GAP are compared with existing protocols such as ECPP [9] and GSIS [8] used in VANET. ECPP requires two RTT for transferring a short-time anonymous certificate. With GSIS, the time requirement for verifying n safety messages increases with the number of revoked certificates in the revocation list.

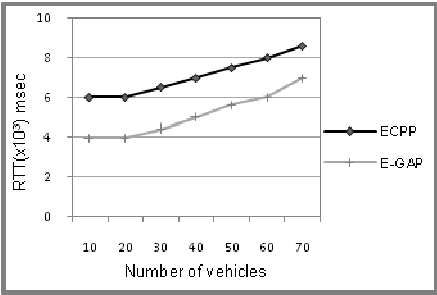


Fig. 6. RTT between vehicle and TA

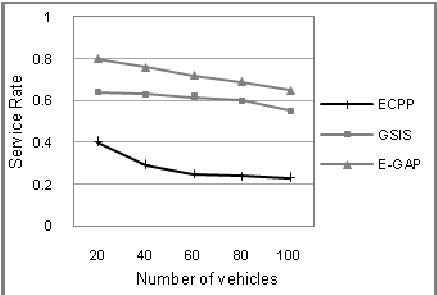


Fig. 7. Service Rate of RSU

Fig. 6 compares ECPP and E-GAP with respect to RTT between a vehicle and TA. RTT between TA and vehicle is based on the processing delay while generating a certificate by TA and delay in transmitting the certificate via RSU to the vehicle from

TA. E-GAP considerably keeps the RTT under control. The service rate of RSU is calculated as per equation (3). Fig 7 shows the service rate of ECPP, GSIS and E-GAP. E-GAP protocol shows high service rate compared to ECPP and GSIS. ECPP needs to check the certificate from heavily loaded revocation list because of short-lifetime certificate.

$$service\ rate = \frac{\text{number of delivered messages}}{\text{number of received messages}} \tag{3}$$

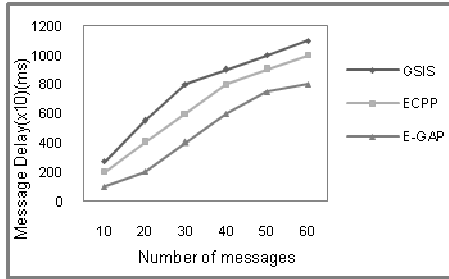


Fig. 8. Overall Message Delay

Fig 8 describes overall message delay of different protocols. The overall message delay depends on signing, transmission and verification of message as per the equation (4). Duration for signing and verification is more for ECPP [9] and GSIS [8]. Since E-GAP uses group signature and batch verification, the delay gets reduced. E-GAP protocol produces a signature size of 160bytes, which reduces transmission time.

$$Message\ Delay = T_{sign} + T_{trans} + T_{verify} \tag{4}$$

5 Conclusions and Future Work

E-GAP protocol for VANET satisfies the major security requirements such as authentication, traceability and anonymity preservation. The proposed protocol also scales well when the number of messages has increased. For future work, the protocol can be tested with different batch verification period and can also be simulated in traffic simulator.

References

1. National highway traffic safety administration, <http://www-nrd.nhtsa.dot.gov/pdf/nrd-12/060419-0843/PDF/TOC.html>
2. Raya, M., Hubaux, J.P.: Securing VANET. J. of Comp. Security 15(1), 39–68 (2007)
3. Raya, M., Papadimitratos, P., Hubaux, J.P.: Securing Vehicular Networks: Laboratory of Computer communications and Applications (2004)

4. Gollan, L., Meinel, C.: <http://www.hpi.unipostsdam.de/fileadmin/hpi/FG ITS/papers/DigitalSignaturesAuto02.pdf>
5. Habaux, J.P., Capkun, S., Luo, J.: Security and Privacy of Smart Vehicles. *IEEE Security and Privacy Magazine* 2(3), 49–55 (2004)
6. Studer, A., Shi, E., Bai, F., Perrig, A.: TACKing Together Efficient Authentication, Revocation, and Privacy in VANETs. *IEEE Comm. Society, Secon 1*, 1–9 (2009)
7. Zhang, C., Lin, X., Lu, R., Ho, P.: RAISE: An efficient RSU-aided message authentication scheme in vehicular communication networks. In: *IEEE ICC*, pp. 1451–1457 (2008)
8. Lin, X., Sun, X., Ho, P.H., Shen, X.: GSIS: a secure and privacy-preserving protocol for vehicular communications. *IEEE Trans. on Veh. Tech.* 56, 3442–3456 (2007)
9. Lu, R., Lin, X., Zhu, H., Ho, P.-H., Shen, X.: ECPP: Efficient Conditional Privacy Preservation Protocol for Secure Vehicular Communications. In: *IEEE INFOCOM*, pp. 1903–1911 (2008)
10. Boneh, D., Lynn, B., Shacham, H.: Short Signatures from the Weil Pairing. In: Boyd, C. (ed.) *ASIACRYPT 2001*. LNCS, vol. 2248, pp. 514–532. Springer, Heidelberg (2001)
11. Camenisch, J., Hohenberger, S., Pedersen, M.: Batch verification of short signatures. In: *Advance in Crypt-Eurocrypt*, pp. 246–263. Springer, Heidelberg (2007)
12. Qin, X., Zhang, S., Jia, L.: Research on Pairing-Based Batch Verification. In: *IEEE CMC*, vol. 1, pp. 46–50 (2010)
13. Bayrak, A.O., Acarman, T.: A Secure and Privacy Protecting Protocol for VANET. In: *IEEE Intelligent Vehicles Symposium*, pp. 579–584 (2010)
14. Sun, J., Zhang, C., Zhang, Y., Fang, Y.: An Identity-Based Security System for User Privacy in VANET. In: *IEEE Trans. on Parallel and Distributed Systems*, pp. 1227–1239 (2010)
15. Zhang, L., Wu, Q., Solanas, A., Domingo-Ferrer, J.: A Scalable Robust Authentication Protocol for Secure Vehicular Communications. *IEEE Trans. on Vehicular Technology* 59, 1606–1617 (2010)
16. Huang, J.L., Yeh, L.Y., Chien, H.Y.: ABAKA: An Anonymous Batch Authenticated and Key Agreement Scheme for Value-Added Services in Vehicular Ad Hoc Networks. *IEEE Trans. on Vehicular Tech.* 60, 248–262 (2011)

Opportunistic Dissemination for Accident Management Using Vehicular Networks

R. Namritha and Komathy Karuppanan

Department of Computer Science and Engineering
Easwari Engineering College, Anna University, Chennai, India
namrithar88@gmail.com, gomes1960@yahoo.com

Abstract. While handling catastrophes like fire accidents, road crashes and other life threatening emergencies, delivery time is very important. This paper presents a novel routing approach for speedy accident recovery using Vehicular AdHoc Networks (VANETs). The accident is reported to hospitals, police stations, fire stations, paramedics etc. with minimum notification delay. In the proposed system, end-to-end delay is reduced by minimizing the propagation delay in two stages. In the first stage, the directionality of the vehicles is estimated. In the second stage, an innovative method of choosing the shortest path among the feasible routing channels is introduced. Other preferences of using vehicular networks for accident management are unlimited power source, low cost communication and the availability of location information. The proposed system was tested under simulated traffic of an urban road map. It is observed that the delay is adequate for accident relief applications.

Keywords: Accident recovery, VANET, shortest path, propagation delay, location information.

1 Introduction

Road traffic crashes are the leading cause of death among Indian citizens, especially in urban areas. While studying how these issues are tackled in advanced countries, speedy rescue services is observed to have reduced the effects of injuries or even save lives. For example, eCall[1] brings rapid assistance to accident victims by propagating accident notifications to the nearest emergency resources via cellular networks. Emergency resource (ER) may refer to hospitals, fire stations, police station, paramedics etc. An alternative to overcoming the critical issues such as power source, location discovery, bandwidth with eCall is recommending Vehicular AdHoc Networks (VANETs) for rescue services. Vehicular networks are endowed with unlimited battery power and storage thus guarding the system against power failures. Recent advances in vehicular technology such as the provision of DSRC specifically designed for automotive use has provided efficient channels for inter vehicular communications. Also, vehicles fitted with GPS devices provide accurate location information. Thus the inherent characteristics of vehicular networks can be exploited to provide fail safe rescue services for accidents and other mishaps.

The major focus of this paper is to provide an emergency routing with stringent delay constraints in city environments under high density of vehicles. Formally, the design goal of the proposed system can be described as: *the accident management system must be adaptable to a highly dynamic topology of vehicular network and enforce severe delay restrictions*. To achieve this goal, the following issues related to VANETs are to be considered:

- High mobility of the vehicles leads to frequent link breakages, making end to end routing unsuitable. In the worst case, if two cars with the speed of 60 kmph are driving in opposite directions, the link will last only for at most 10 seconds [2].
- Due to the random mobility of the vehicles, it is impossible to maintain the hop length uniformly. So the shortest path, traditionally defined using the no. of hops is inefficient. This unique characteristic of vehicles demands an innovative routing scheme that uses real time information to perform route selection.

The objectives of the proposed system are to overcome the above problems and to provide successful delivery of the accident notification with minimum delay. The following main functions are involved while designing the proposed system.

Opportunistic Routing. It allows an end to end path to be built dynamically using a chain of forwarders. The forwarder is chosen using local positional information. Hence, vehicles need not keep their routing table up to date. Also, opportunistic routing performs better for high density environments in which each vehicle has many neighbours, making it ideal for implementing in an urban environment.

Compass Routing. Forwarders are chosen based on their directionality. The vehicles, which are moving towards the AN are chosen to progress the accident notification. The direction of motion is calculated using real time information of the vehicles obtained from their GPS.

Shortest Path Routing. It is built on the concept that the shortest distance between the AV and AN is a straight line. The vehicles which have minimum deviation from this straight line are chosen as forwarders. This approach has used real time location information instead of hop count for routing.

Mobility. The proposed system has taken advantage of the mobility of the vehicles and uses them as mobile carriers to tow the accident notification closer to the anchor node by virtue of their high speed. Thus the proposed system banks on the mobility of the participating vehicles, usually considered to be an inconvenience to provide efficient routing that satisfies the hard delay constraints of accident management.

The rest of the paper is organised as follows: In Section 2, the review of the related works is presented. In Section 3, the system design is outlined and the design aspects of the proposed routing algorithms are covered. This is followed by the experimental results and a brief analysis in Section 4. Section 5 concludes the paper by presenting the future works.

2 Related Works

One of the prevalent accident management systems using vehicular network in the current literature is Network on Wheels (NoW) [3][4]. In [3], Debopam Acharya et al. introduced GPS and stationery RDS(Radio Data System) to inform EMS (Emergency Medical Services) services on accidents. RDS are placed at accident prone zones. When an accident is detected by the RDS, it automatically sends a message to the nearest EMS. In [4], the same authors present an improved version of the accident management system presented in [3]. This system comprises of a sensing unit on the vehicle, a NoW server and a DB. When an accident is detected by the sensing unit, an automatically generated message is sent to the NOW server, where qualified medical personnel are present to notify the emergency services. The DB stores details of the accident for future reference. Prasanth et al [5] presented Edge Based Greedy Routing (EBGR) in which the potential score of a neighbour moving towards the destination is calculated based on shortest distance to the destination, closeness to the next hop, velocity vector and source-destination vector. The neighbour with largest potential score is chosen as the forwarder. Wenjing Wang et al [6] and Jason LeBrun et al [7] introduced directional forwarding in which control and data packets are routed via vehicles moving towards the destination. The direction of motion of vehicles is calculated using directional antennas and motion vectors respectively. Jiayu Gong et al [8] presents Predictive Greedy Directional Routing that combines Position First Forwarding (PFF) and Direction First Forwarding (DFF). A score is calculated by each vehicle for both PFF and DFF. The routing method with the largest score is chosen. Josiane Nzouonta et al [9] introduced Road Based Vehicular Traffic (RBVT) in which the end to end path is defined in terms of intersections. Link breaks are repaired at the intermediate intersection, not the source thus reducing delay.

The major issues in the existing accident management systems are the use of infrastructure to handle accident management, presence of human-machine interface which increases notification latency. The issues in routing schemes are the use of hop count as a decision criteria for routing, establishment of end-to-end path using RREQ and RREP, dependability on direction and mobility of the vehicles alone to provide delay constrained routing. Also, the existing systems use either directional routing or position based routing. The above issues have been overcome in the proposed system.

3 Proposed Model

The important components of the system architecture for the proposed system are specified in Fig 1. It consists of vehicles whose direction of motion is indicated using an arrow, a fixed node called Anchor Node (AN) and ER. Anchor node (AN) is used in case the ER lies outside the VANET communication range and all positional information of ER is mapped to AN. The anchor node routes the accident notification to the nearest ER via a wired network. All vehicles are assumed to know their positions using GPS and they use Dedicated Short Range Communication/Wireless Access in Vehicular Environment (DSRC/WAVE) for efficient V2V communication with a coverage of 1000 metres for inter vehicular communications. The vehicles are assumed to be moving at a speed of 60 to 100 kmph in a dense urban area. Thus vehicles are always available to detect and forward the accident notification. The

vehicles are of three types namely, the vehicle at accident site (AV), which generates the accident notification, Forwarding Vehicles (FV), which are intermediate nodes to forward the accident notification and the non-participant nodes whose forwarding functions are unused.

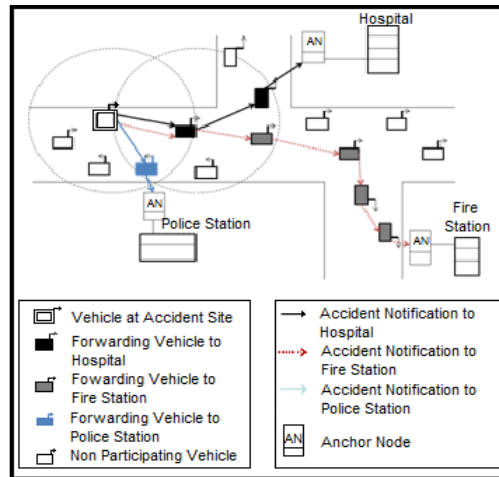


Fig. 1. System Architecture

When an accident is detected, the location of the nearest ER is dynamically obtained using GPS map service available in the vehicle. AV chooses an optimal forwarder among the one hop neighbours based on two selection criteria namely, (i) the direction of motion of FV must be towards AN [5] and (ii) FV should have the minimum deviation from the shortest path between the AV and AN. The above criteria have brought a novel routing algorithm called Least Length Path Routing (LLPR) and are described in detail in the section below.

3.1 Least Length Path Routing (LLPR)

LLPR is an on demand and an opportunistic routing protocol, which briefly incorporates the concept of direction resolution from compass routing. LLPR attempts to provide a successful transmission of accident notification to AN with minimum delay. As the end-to-end delay is highly dependent on propagation delay, the notification latency can be effectively controlled by minimizing the propagation delay. The proposed algorithm limits the propagation delay by means of two mechanisms namely, (i) directional forwarding based on the motion vector of the forwarding vehicles and (ii) minimizing the transmission distance of the packets.

Directional Forwarding based on the motion vector of the forwarding vehicles.

This approach utilises node mobility, an inherent characteristic of vehicular networks. Node mobility has usually been regarded as an inconvenience in routing. However using compass routing, vehicular mobility can be used to enhance routing decisions by taking advantage of their directionality. Using the motion vectors, vehicles whose

direction of motion is towards the AN are chosen as Potential Forwarders (PF). This serves to avoid routing loops and fractionally reduces the residual distance to the AN.

Minimizing the transmission distance of the packets. Here, the routing path length is minimized, thereby minimizing the propagation delay. Propagation delay is defined as the amount of time taken by a packet to travel from the source to the destination over a medium[10]. This step majorly reduces the propagation delay by minimizing the distance travelled by the packets. Equation (1) uses the transmission distance as a routing criteria instead of hop count and uses realtime information from the GPS to limit delay.

$$\text{Propagation Delay} = \frac{\text{Transmission distance}}{\text{Speed}} \tag{1}$$

LLPR incorporates the above approaches using two algorithms namely Approach/Depart Algorithm and Near/Far Algorithm, which are executed sequentially.

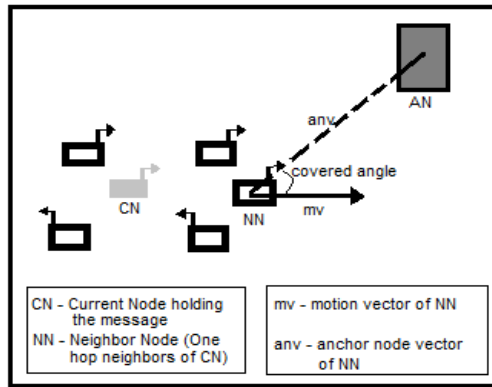


Fig. 2. Approach/Depart Algorithm

Approach/Depart Algorithm. The algorithm controls the propagation delay by directional forwarding. The direction of motion of the vehicles is calculated using the motion vector of the vehicles. The vehicles moving towards AN are chosen as PF. An illustration of the algorithm is given in Fig 2 and the steps are given in Fig 3.

In Fig 2, CN queries the location details of its one hop neighbors. When it receives a response, the mv and anv are calculated. The covered angle is the angle subtended between the motion vector and the anchored node vector. It is used to calculate the direction of motion of the vehicle, i.e. whether the vehicle is moving towards the ER or away from the ER. This step is crucial in the proposed algorithm because the forwarder is chosen from the subset calculated in this algorithm. The formula for covered angle is given in step 3 of the Algorithm-1. Based on the covered angle, the direction of motion is identified. The direction resolution of the neighbor nodes using the covered angle is depicted in step 4. If the covered angle < 90, the direction is set as *approach*, otherwise it is set as *depart*. After direction resolution, the forwarding subset composed of PF is computed.

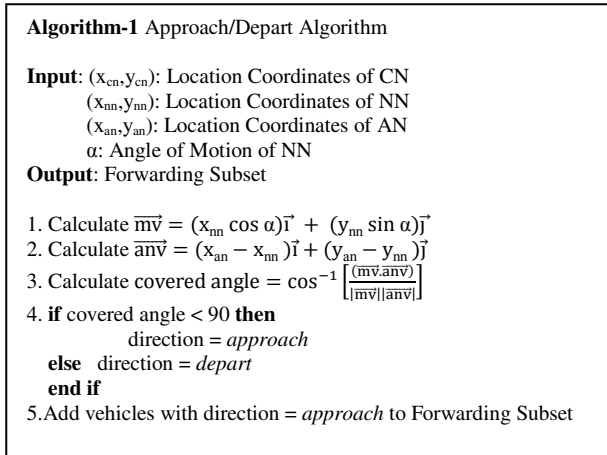


Fig. 3. Steps of Approach/Depart Algorithm

Near/Far Algorithm. This algorithm aims to reduce the delay by minimizing the transmission distance of the packets. It is based on the premise, *the shortest path between two points is a straight line* [11]. Theoretically, the shortest path may go through buildings and other obstacles and may not have usable vehicles to account. In the absence of vehicles along the hypothetical shortest path, the key operation in this stage is choosing an optimal vehicle among the available vehicles using angular deviation from the shortest path. Formally, the operation in Near/Far Algorithm is described as,

Direction of V_i ($Dir(V_i)$) is not sufficient to identify an optimal neighbour, we also need to consider the length of the routing path through that node. Deviation of V_i ($Dev(V_i)$) outputs V_i 's deviation from the shortest path w.r.t. other approaching neighbours V_{i+1} . $Dev(V_i) < Dev(V_{i+1})$ if V_{i+1} has greater angular divergence from the shortest path.

$Dir(V_i)$ and $Dev(V_i)$ are calculated by applying geometric properties to the vehicular environments using Approach/Depart and Near/Far Algorithm respectively. Near/Far algorithm operates on the forwarding subset calculated in Approach/Depart Algorithm. After applying the algorithm to the PF, the forwarder is chosen based on the deviation angle. An illustration of the selection of forwarder is given in Fig 4. The deviation angle is calculated using the Near/Far algorithm given in Fig 5.

The steps of Near/Far algorithm are as follows, Initially the equation of the shortest path from the accident site to the anchored node is calculated using the step 1. Next, the perpendicular distance (PD_{v_i}) of the potential forwarder v_i to the shortest path and the lapsed distance from the source (LS_{PF_i}) of PF_i are given by step 2 and 3 respectively. The deviation angle θ_{PF_i} of the PF_i is given by step 4. The PF which has the minimum θ is chosen as the forwarder.

For example, in Fig 4, $\theta_{pf1} < \theta_{pf2}$ where θ_{pf1} and θ_{pf2} are the deviation angles of PF1 and PF2 respectively. Hence PF1 is chosen as the forwarder. By repeating the same steps at each of the vehicles holding the accident notification, the same is transmitted from the accident site to the destination via a chain of FVs. The above

protocol faces many challenges due to the high mobility and dynamic topology of vehicular networks such as (1) network partitioning and (2) maintenance of routing table. The above problems lead to following two test scenarios.

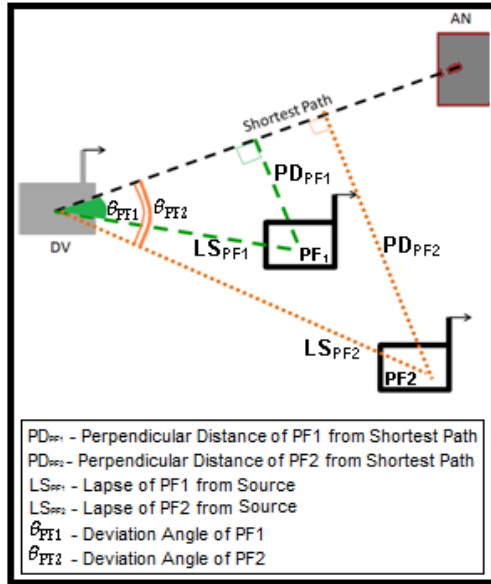


Fig. 4. Near/Far Algorithm

Algorithm: Near/Far Algorithm

Input: (x_{dv}, y_{dv}) : Coordinates of AV,
 (x_{an}, y_{an}) : Coordinates of AN,
 (x_{fvi}, y_{fvi}) : Coordinates of $FV_i, 1 \leq i \leq n$
 (x_{pfi}, y_{pfi}) : Coordinates of $PF_i, 1 \leq i \leq n$

Output: FV_{i+1}

1. $(y_{an} - y_{dv})x + (x_{dv} - x_{an})y + (x_{an} y_{dv} - y_{an} x_{dv}) = 0$.
2. $PD_{PFi} = \frac{|(y_{an} - y_{dv})x_{pfi} + (x_{dv} - x_{an})y_{pfi} + (x_{an} y_{dv} - y_{an} x_{dv})|}{\sqrt{(y_{an} - y_{dv})^2 + (x_{dv} - x_{an})^2}}$.
3. $LS_{PFi} = \sqrt{(x_{dv} - x_{pfi})^2 + (y_{dv} - y_{pfi})^2}$.
4. $\theta_{PFi} = \sin^{-1}(PD_{PFi} / LS_{PFi})$.
5. Select PF_i with minimum θ as FV_{i+1}

Fig. 5. Steps of Near/Far Algorithm

ER Approachable. In this scenario, vehicles are available to forward the disaster notification to the nearest ER. This scenario occurs when the density of vehicles is high during the daylight hours in an urban road.

ER Non-approachable. This scenario arises under two situations such as: there is no vehicle available to forward the disaster message; and the current forwarder crosses the ER. The vehicle holding the notification first calculates the covered angle between its motion vector and the ER using Approach/Depart algorithm in Fig 3. If the covered angle is greater than 90, the vehicle has crossed the ER. In this case, the current forwarder should select an alternative ER that incorporates lesser end-to-end delay. To perform this operation, the vehicle locates an alternative ER using the navigating services, which is the closest to the respective vehicle and thereon the routing process is initiated. Since LLPR routing attempts to achieve a successful delivery of accident notification, it adapts effectively the above two scenarios. Design of the routing table has been simplified to overcome the two challenges by incorporating per-hop routing where the vehicle holding the message alone needs to store the details of its one hop neighbours. The LLPR routing table contains current time, node ID, approaching neighbor ID and angle of deviation.

4 Performance Evaluation

The proposed system has been tested using VANET MobiSim and NS 2 tools. The results are compared with an existing vehicular routing algorithm called MoVe. We consider a simulation area of 2000 x 2000 with around 150 nodes. Nodes move at a speed of 60 to 100 kmph. Traffic simulation of the vehicles is done using VANETMobiSim. Freeway model is used, where each vehicle is restricted to its lane. Vehicles change course at the intersections. The accident notification is transmitted in term of packets of the size 512 bytes. The simulation is run for 300 seconds and the results are recorded in Fig 8, 9 and 10.

Fig 6 represents the packet delivery per accident notification with respect to the number of nodes. At low density, the packet delivery is low due to the restricted availability of forwarders. In such a situation, the selection of forwarder for the delivery of the notification is dependent on the spatial arrangement of the vehicles, which is highly dynamic. At high density, a sufficient no. of vehicles are available for forwarder selection. Hence the packet delivery is higher by around 20%. Also, an increase in the no. of ERs triggers a corresponding increase in the packet delivery, because of the availability of alternative ERs. If the no. of ERs is 1, and it is not accessible, the packets are dropped leading to lower packet delivery. In the case of multiple ERs, packets can be alternatively routed to other accessible ERs thereby increasing the packet delivery.

Fig 7 represents the end-to-end delay with respect to the number of nodes. Three patterns of end-to-end delay have been observed such as (1) the delay decreases with increase in the no. of nodes for both MoVe and LLPR, (2) LLPR incurs lesser delay compared to MoVe, and (3) the delay increases as the no. of ERs increases.

It is inferred from pattern (1) that the increase in the no. of nodes leads to an increase in the no. of potential forwarders along the shortest path. Hence the optimal forwarder can be chosen based on the angle of deviation. This leads to a decrease in the no. of hops and hence reduced delay. Pattern (2) is due to LLPR selecting the forwarder along the feasible shortest path whereas MoVe opting its forwarder closest to the destination. Hence the transmission distance is reduced and lesser end-to-end delay is accounted in LLPR. Similarly requirement of multiple ERs, if the first ER is

not accessible, implies that the increase in the no. of ERs leads to successful delivery of accident notification. This step invokes the decision processes namely, detecting for subsequent FV available towards ER and deciding an alternative ER. In such cases, the accident notification is buffered until an alternative ER is chosen leading to the increase in end-to-end delay. The increase in delay is due to double buffering time accounted with each unit increase in the ER.

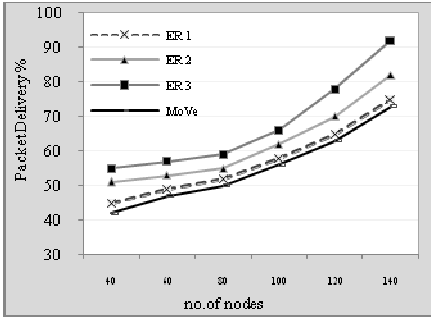


Fig. 6. Packet delivery per disaster notification vs no. of nodes

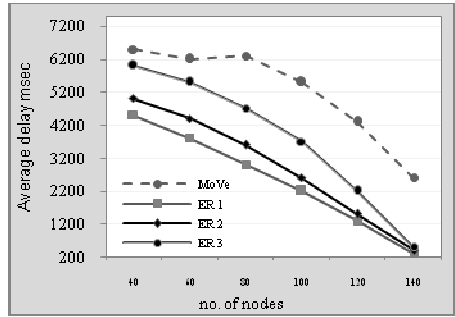


Fig. 7. End-to-end delay vs no. of nodes

The above results proves that the transmission distance is reduced when angle based forwarding is implemented. This in itself is an important conclusion. Since mobility is simulated using VANETMobiSim, this idea can be utilised for other accident management systems using vehicular networks.

5 Conclusion

This paper introduced a novel routing algorithm which uses the mobility and directionality of vehicles to enhance routing. Our proposed mechanism LLPR incorporates opportunistic, compass and shortest path routing strategies. It considers the dependence of delay on transmission distance and the geometrical properties of vehicular networks to reduce delay. However, link dependant features like load balancing and future position prediction have not been considered. Incorporating these factors as a decision criteria will enhance the performance. In future works, the use of transmission distance as a criterion to choose an optimal forwarder can be adopted as a key idea for emergency routing.

References

1. Pinart, C., Carlos, C.J., Nicholson, L., José, V.A.: ECall-compliant early crash notification service for portable and nomadic devices. In: IEEE 69th Vehicular Technology Conference, pp. 1-5. IEEE Press, Barcelona (2010)
2. Jerbi, M., Senouci, S.-M., Ghamri-Doudane, Y., Cherif, M.: Handbook of research on Next Generation Networks and Ubiquitous Computing. Networks and Multimedia Systems Research Group, France (2009)

3. Acharya, D., Kumar, V., Garvin, N., Greca, A., Gary, G.M.: A Sun SPOT based Automatic Vehicular Accident Notification System. In: International Conference on Information Technology and Application in Biomedicine, China, pp. 296–299 (2008)
4. Agarwal, A., Goel, N., Alcatel-Lucent: GPD-RDS enabled location based smart transit. In: 13th International Conference on Intelligence in Next Generation Networks, pp. 1–5. IEEE Press, India (2009)
5. Prasanth, K., Duraisamy, K., Jayasudha, K., Chandrasekhar, C.: Minimizing End to End delay using Edge Node Based Greedy Routing. In: 1st International Conference on Advanced Computing, pp. 135–140. IEEE Press, India (2009)
6. Wang, W., Xie, F., Chatterjee, M.: Small Scale and Large Scale Routing in Vehicular AdHoc Networks. *IEEE Transactions on Vehicular Technology* 58, 229–241 (2009)
7. LeBrun, J., Chuah, C.-N., Ghosal, D., Zhang, M.: Knowledge-Based Opportunistic Forwarding in Vehicular Wireless Ad Hoc Networks. In: 61st Vehicular Technology Conference VTC, China, pp. 2289–2293 (2010)
8. Gong, J., Xu, C.-Z., Holle, J.: Predictive Directional Greedy Routing in Vehicular AdHoc Networks. In: 27th International Conference on Distributed Computing Systems Workshops, Detroit, p. 2 (2007)
9. Nzouonta, J., Rajgure, N., Wang, G.G., Borcea, C.: VANET Routing on City Roads Using Real-Time Vehicular Traffic Information. *IEEE Transactions on Vehicular Technology* 58, 3609–3626 (2009)
10. Asokan, R., Natarajan, A.M.: An Approach for Reducing the End to End Delay and Increasing Network Lifetime in Mobile Ad Hoc Networks. *International Journal of Information and Communication Engineering* 4, 431–437 (2008)
11. James Kurose, F., Keith Ross, W.: *Computer Networking*. Pearson Education, India (2007)

Machine Learning Approach for Multiple Misbehavior Detection in VANET

Jyoti Grover, Nitesh Kumar Prajapati, Vijay Laxmi, and Manoj Singh Gaur

Department of Computer Engineering
Malaviya National Institute of Technology, Jaipur, India
{jyoti.grover, prajapati5nitesh, vlgaur, gaurms}@gmail.com

Abstract. The motivation behind Vehicular Ad Hoc Networks (VANETs) is to improve traffic safety and driving efficiency. VANET applications operate on the principle of periodic exchange of messages between nodes. However, a malicious node may transmit inaccurate messages to trigger inevitable situations. Each transmitted packet contains the status of sender like its identity, position and time of sending the packet in addition to safety message. A misbehaving node may tamper with any information present in the propagated packet. Fake messages may be created by attacker node itself or it may force another node to create fake messages. In this paper, we present a machine learning approach to classify multiple misbehaviors in VANET using concrete and behavioral features of each node that sends safety packets. A security framework is designed to differentiate a malicious node from legitimate node. We implement various types of misbehaviors in VANET by tampering information present in the propagated packet. These misbehaviors are classified based upon multifarious features like speed-deviation of node, received signal strength (RSS), number of packets delivered, dropped packets etc. Two types of classification accuracies are measured : Binary and Multi-Class. In Binary classification, all types of misbehaviors are considered to be in a single “misbehavior” class whereas, Multi-class classification is able to categorize misbehaviors into particular misbehaving classes. Features of packet sending nodes are extracted by performing experiments in NCTUns-5.0 simulator with different simulation scenario (varying the number of legitimate and misbehaving nodes). Proposed framework for classification of misbehavior is evaluated using WEKA. Our approach is efficient in classifying multiple misbehaviors present in VANET scenario. Experiment result shows that Random Forest and J-48 classifiers perform better compared to other classifiers.

1 Introduction

Vehicular Ad Hoc Networks (VANETs) applications are based upon the frequent exchange of safety messages by vehicles. This is also required to facilitate route planning, road safety and e-commerce applications. VANET security is an essential component for each of these applications. VANET is a typical kind of Mobile Ad Hoc Network (MANET). High mobility, large scale and frequent topology changes are the main characteristics of vehicular networks. Vehicle passenger safety can be improved by means of inter-vehicle communication. For example, in case of an accident, VANET communication can be used to warn other vehicles approaching the site.

VANETs are facing a number of security threats [1] which may degrade the performance of VANET and even life safety. For example, an attacker node may create an illusion of traffic congestion by pretending multiple vehicles simultaneously and launch Denial of Service (DoS) attack by impairing the normal data dissemination operation.

In this paper, we introduce different kind of attacks based upon the physical properties and safety packet distribution behavior of nodes in VANET. Usually, a node observing a safety event triggers safety warning packets. All the nodes receiving this message forward it to other nodes in their transmission range. In this manner, it is the responsibility of honest nodes to forward each received safety packet to other nodes in its neighborhood. Terms *attackers*, *malicious node*, *misbehaving node*, *illegitimate node* are used interchangeably in this paper.

We perform a series of experiments with different number and type of misbehaviors and derive features from physical and packet transmission characteristics of honest and misbehaving nodes. We construct various instances of honest and misbehaving nodes. Machine learning method is used to classify the behavior as *honest* or *malicious*. Our classification algorithms are able to differentiate different kinds of misbehaviors during classification phase. Features used for classification are speed deviation, distance, received signal strength (RSS), number of packets generated, delivered, dropped, collided. These features are calculated by nearby observer nodes. All observer nodes exchange their observation with other nodes in its vicinity. The experiments are evaluated using Naive Bayes, IBK, J-48, Random Forest and Ada Boost1 classifiers [15] supported by Waikato Environment for Knowledge Analysis (WEKA) [2], which is a data mining tool. As per our knowledge, machine learning method has not been previously applied to distinguish different misbehaviors in VANET. Specifically, the major contributions of our paper are:

1. Implementation of variants of misbehaviors in VANET.
2. Extract the features that may be used in differentiating misbehavior instances.
3. Classification of misbehaviors using machine learning method.

The rest of this paper is organized as follows. Section 2 discusses related work on misbehaviors in VANET and their detection methodologies. Section 3 provides overview of VANET and attacker model. In Section 4, a brief outline of proposed methodology is introduced. It discusses different features, feature extraction module used in classification of misbehaviors in VANET. Experimental setup and results are discussed in Section 5 and 6 respectively. Concluding remarks with future work are covered in Section 7.

2 Related Work

Various types of attacks on an inter-vehicle communication system are presented by Aijaz *et al* [1]. They analyze how an attacker may manipulate the input of an OBU and sensor readings. The authors proposed plausibility checks using constant system examinations. However, this paper does not discuss how to apply plausibility checks in detail.

M. Raya and J.P. Hubaux [3] discuss a number of unique challenges in VANETs. They describe how adversaries use safety applications to create various attacks and

security problems. However, they have not analyzed how to achieve these attacks and do not provide solutions to address these security threats.

Golle *et al* [4] propose an approach for detecting and correcting malicious data in VANET. In this approach, every vehicle builds a model of VANET in which specific rules and statistical properties of VANET environment are implemented and stored. When a node receives a message, it compares the received message with the VANET model. If the received message does not comply with the VANET model, it is considered to be an invalid message. The VANET model used in this paper is predefined and does not provide the flexibility to switch to a new one. Ghosh *et al* [5] [6] present a misbehavior detection scheme (MDS) for post crash notification (PCN) applications. They consider various parameters such as tuning and impact of mobility on performance of MDS. In their work, OBUs determine the presence of bogus safety message by analyzing how the driver behaves in response to an event.

Raya *et al* [7] have formulated a misbehavior detection system to exclude malicious vehicles from the communication system. It is based on the deviation of attacker node from normal behavior. Cryptographic keys belonging to a malicious node are revoked upon its detection. Normal and abnormal behavior is differentiated by using a clustering algorithm. Yan *et al* have proposed a position verification approach for detection of position related misbehaviors in [8]. Xiao *et al* [9] illustrate a localized and distributed scheme to detect ID spoofing attack in VANETs. This approach takes advantage of VANET traffic patterns and road side base stations. Their detection approach uses statistical analysis of signal strength distribution.

Raya *et al* [10] present their work on “data centric trust” in VANETs. They confirm the occurrence of an event based upon the messages received from multiple vehicles. They propose that vehicles use a decision logic system like Dempster-Shafer, Bayesian inference in their framework. However, their work assume that vehicles leave their radio range rapidly. Hence, it removes the use of reputation and ignores information from local sensors. Schmidt *et al* [11] construct reputation models for other vehicles based on the claims from sending vehicles. In this way, they create a model for normal behavior of nodes in VANET. If the behavior of a node differs from the normal behavior, it is marked as suspicious.

Kim *et al* [12] have introduced a message filtering model that leverages multiple complementary sources of information. They constructed a multi-source detection model so that drivers are alerted after multiple sources prove the existence of a certain event. Our framework is somewhat similar to the above with an exception that we are also able to classify different types of misbehaviors apart from just being able to identify them.

3 System and Attacker Model

VANET consists of two basic components: (1) Road Side Unit (RSU) and (2) On Board Unit (OBU). RSU is a fixed unit while OBUs are installed in vehicles and hence these are mobile. Each node in VANET consists of an Event Data Recorder (EDR), Global Positioning System (GPS) receiver, computing platform and a radar. A conceptual architecture of VANET is shown in Figure 1. There is a hierarchy of central authorities (CA) that is responsible for managing vehicles identities registered in its respective

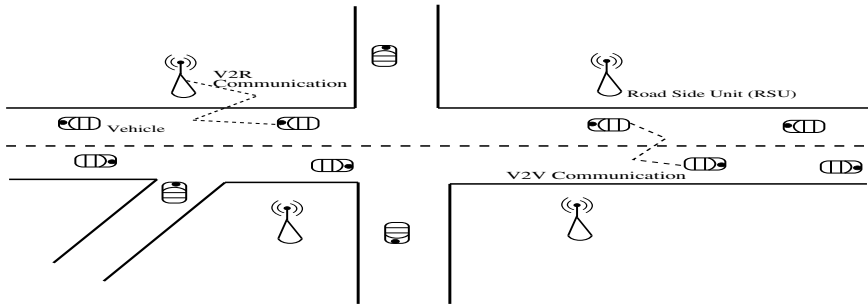


Fig. 1. Architecture of Vehicular Ad Hoc Network

geographic region. As communication framework, dedicated short range communication (DSRC) protocol [13], currently being standardized as IEEE 802.11p is used at data link layer. This protocol provides transmission range of 250 to 1000m and data rates in the 6-27Mbps range.

In VANET, all information is publicly available as safety messages are meant for each vehicle. Safety communication system of VANET differs from the usual approach in information systems where restrictions can be applied to access the system. VANET vulnerabilities originate from openness of wireless communication channel (can be accessed by anyone) and unencrypted exchange of information.

We consider active attackers only and they possess capability to compromise the integrity of messages. Each node transmitting a safety/alert packet dispatch its identity, position and time-stamp with the packet. However, a misbehaving node may tamper with any of the information sent in the packet. For example, attackers may create bogus alerts or suppress legitimate messages. Malicious nodes may generate wrong traffic warning message and propagate it to other vehicles in the network. These active attackers may force legitimate drivers to change their driving behavior. For example, honest vehicles may slow down or take alternate routes if fake messages regarding harmful events are distributed in the network. As a result, malicious nodes succeed in disruption of normal driving behavior of vehicles. Attackers may also suppress valid messages of critical safety information by dropping or capturing them. Malicious node may generate fake packets in place of forwarding valid packets. We briefly discuss different types of VANET misbehaviors implemented in this paper:

1. **Packet Suppression Attack:** In this attack, whenever a vehicle receives safety packets from a neighboring node, it does not forward these packets unlike the normal functionality of VANET. All the legitimate nodes in VANET forward every received safety packet, but malicious nodes do not follow this approach. Misbehaving nodes can also insert fake safety packets in the network.
2. **Packet Replay Attack:** Replay attack is a form of attack in which a normal data transmission is fraudulently repeated or delayed. This operation is carried out by a malicious node that intercepts the safety packet and retransmits it. Replay attack is usually performed by malicious or unauthorized node to impersonate a legitimate vehicle or RSU. It creates an illusion of apparently valid though non-existing events.

3. **Packet Detention Attack:** This attack is a subset of the packet replay attack. In this attack, a vehicle delays the packet forwarding process by certain time duration in the network. It is more dangerous than replay attack, as vehicles do not get enough time to respond to a particular emergency situation. For example, a honest vehicle $V1$ may broadcast *TRAFFIC JAM* safety packet at time-stamp t_0 to its neighboring nodes. This message is sent after some delay by a malicious vehicle. After receiving another *TRAFFIC JAM* packet, vehicles may change their path to nearby road-segment thereby leading to real congestion on this route even though the jam may have cleared by this time.
4. **Identity Spoofing Attack:** This attack is a impersonation or spoofing attack where an attacker spoofs the identity of another node in the network and hence, all the messages directed to the victimized node are received by the attacker. The attacker can feign safety message by using multiple identities simultaneously to create illusion of non-existing events.
5. **Position Forging Attack:** In this attack, an attacker broadcasts timely coordinated wrong traffic warning messages with forged positions, leading to illusion of a car accident, a traffic jam or an emergency braking.
6. **Combination of Identity and Position Forging Attack:** In this attack, attacker may use multiple identities while launching position forging attacks.

We have created variants of these attacks. First three misbehaviors are temporal, i.e. related to *time* of sending the packets, which plays a key role for VANET applications. Invalid value of *Position* and *Identity* fields in packet is also responsible for creating illusion of non-existing events.

In all attacks, the main objective of malicious node is to distrust VANET traffic and part thereof. By assuming false identity or spoofing identities, malicious node's aim is to evade detection. The main contribution of this paper is detection of misbehavior instances, detecting node responsible for this misbehavior is beyond the current scope of the paper.

We assume that vehicles communicate using the Dedicated Short Range Communication (DSRC) technology. Majority of vehicles are honest. In our case, only up to a 25% of vehicles are assumed to be illegitimate. In our model, it is assumed that RSUs are always honest. All the vehicles trust the message generated by RSUs.

4 Proposed Methodology

Our proposed methodology uses classification algorithms supported by WEKA [2]. It is used to train and test malicious and legitimate instances. Figure 2 describes our detection approach. It shows different steps involved in classification of samples (either as legitimate or particular type of misbehavior).

In *Feature extraction module*, we extracted features from different attack cases which are able to differentiate various types of misbehaviors. Different inputs of feature extraction module are: (1) VANET model, (2) Attack model and (3) VANET application which is affected by attack. A series of experiments are performed in NCTUns-5.0 [14] by mutation of varied number of legitimate and misbehaving nodes to extract the indiscriminated features. These extracted features are applied to the classification algorithms

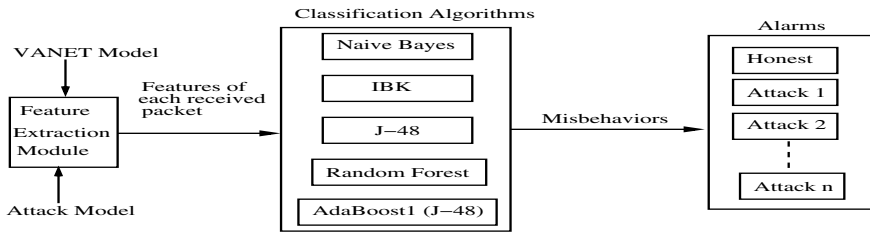


Fig. 2. Proposed Design for classification of legitimate and misbehaving nodes

like Naive Bayes, IBK, AdaBoost1 (with J-48 as base classifier), J-48 and Random Forest (RF) algorithms [15] in WEKA [2]. Finally, various misbehaviors are differentiated based upon these features. We briefly describe features used for classification.

Features are the attributes used to classify different types of misbehavior in VANET. Classifications are performed on relevant attributes or patterns existing with higher frequency to classify a sample as legitimate node or misbehaving node. We derive following features after performing sequence of experiments with different combinations of above defined attacks in VANET scenario. These features are normalized after their node-wise extraction at each interval of time.

- **Geographical Position validation:** Nodes receiving the safety messages verify whether the message is generated by the node located within the area of critical situation or not.
- **Acceptance Range Verification w.r.t observing RSUs:** Uniformly deployed RSUs in VANET scenario verify the acceptance range of each received packet. This operation is performed by measuring the difference between the sending vehicle and RSU position at each time interval. If this difference is greater than the acceptance range, the received message is discarded.
- **Speed Deviation verification:** Consistency in speed at consecutive time interval is verified for each vehicle. This is required to verify fake position broadcasts in the network.
- **Received Signal Strength (RSS):** This parameter is required to verify ID spoofing attacks. RSS value of attacker vehicle and spoofed vehicles is same for the attack duration.

Features related to verification of geographical position, acceptance range, speed and RSS are required to classify position and identity spoofing (Sybil) attacks. Features related to safety packet distribution behavior (Number of packets transmitted, received, dropped, captured) are responsible to classify temporal attacks. These features may be difficult to estimate accurately in the real scenario of VANET. It is assumed that nodes transmit information about packet transmitted/received and packet delivery ratio (PDR) periodically to RSUs within range. All RSUs exchange this information to estimate these features. Though malicious nodes may falsify this information, it is assumed that low number of such nodes in a large network like VANET may not have much impact on the performance of network.

Though we collected legitimate and malicious samples through simulations, it must be noted that features related to delivery time of packets can't be accurately determined in realistic VANET scenario. The features used to detect ID and position spoofing attacks are equally applicable in realistic scenarios. Following are the features used to detect temporal attacks in VANET.

- **Packets Transmitted:** It is the total number of packets generated by each node.
- **Packets Received:** It is the total number of packets received by each node.
- **Packet Delivery Ratio:** It is defined as the ratio of number of packets received and number of packets transmitted. This parameter is required to determine the probability of presence of temporal attacks (packet detention, replay and suppression attacks).
- **Packet Drop Ratio:** It is defined as the ratio of number of packets dropped and number of packets transmitted.
- **Packet Capture Ratio:** It is the ratio of number of captured packets and transmitted packets.
- **Packet Collision Ratio:** It is the ratio of number of packet collisions and transmitted packets.
- **Packet Retransmission Error Ratio:** It is the ratio of number of packet retransmission errors and transmitted packets.

5 Experimental Setup

We conducted our experiments using NCTUns-5.0 simulator [14], an integrated network and traffic simulation platform. This platform provides multiple features including road network simulation, communication and network protocol simulation, vehicular traffic simulation and feedback loop among vehicles. In our experiments, we simulated a two-direction 6 km highway with multi lanes in each direction. Simulation time is varied between (20–2000 seconds). Varying number of Vehicles are randomly deployed and have different behavior including speed (10–50 m/s), acceleration and deceleration. Inter vehicle distance is also varied for each experiment i.e. density is also a variable factor. Traffic arrival rate is 500 vehicles/hour and transmission range is 250 meters. Each experiment is run 5-10 times with different seed values. Average of these results is calculated. We vary different parameters for deriving multiple samples used in classification. We applied a widely used radio propagation model – ‘shadowing model’ – to consider the multi-path propagation effects of the real world communication system.

Packet replay, suppression and detention attacks are basically related to delivery time of packets. If the packets are not delivered in time, it can lead to serious mishappenings. Packet may contain some important information like traffic jam ahead or any information related to road blockage due to accidents or any other natural calamities like landslides. In a packet suppression attack, the number of packets which are to be delivered get reduced. There may be some other reason for the delay in the delivery of packets such as collision and congestion in the network. For the above reason, we calculated a threshold value of the packet delivery ratio by running a number of experiments and training these results using Weka.

ID and position spoofing attacks are implemented as follows. Whenever a misbehaving node is about to send a beacon packet to announce its present ID and position, it selects an ID randomly on the field and applies it to the beacon packet (rather than applying its real ID and position). Whenever a malicious node receives any safety message, it drops this packet. Number of fake identities bounded with illegitimate node is varied in each experiment.

6 Results and Analysis

In this section, we analyze the results produced after classification. All classifiers use 10-fold cross-validation for training and testing samples. Our experimental results are evaluated using the following evaluation metrics [16]. True positive (TP) is the number of malicious vehicles correctly identified as malicious. False positive (FP) is the number of legitimate nodes incorrectly identified as malicious vehicle. True negative (TN) is the number of legitimate nodes correctly identified as a legitimate vehicle. False negative (FN) is the number of malicious nodes incorrectly identified as legitimate. The performance of classifiers can be measured by analysing the following parameters.

- **True Positive Rate (TPR):** It is the ratio of malicious vehicles correctly classified as malicious. It is defined as $TPR = TP / (TP + FN)$
- **False Positive Rate (FPR):** It is the proportion of legitimate vehicles incorrectly classified as malicious. It is also called false alarm rate and defined as $FPR = FP / (FP + TN)$.
- **True Negative Rate (TNR):** It is the proportion of legitimate vehicles identified as legitimate. It is defined by the expression $TNR = TN / (TN + FP)$.
- **False Negative Rate (FNR):** It is defined as the proportion of malicious nodes incorrectly identified as legitimate. It is evaluated as $FNR = FN / (FN + TP)$.

The classifiers are trained with various features consisting of 3101 legitimate and 1427 malicious samples (184 packet detention + 200 replay + 370 suppression + 300 Identity forging + 373 Position forging samples). Table 1 shows the binary-class and multi-class classification accuracy using attributes of legitimate and misbehaving vehicles. In binary-class classification, all types of misbehaviors are considered to in a single "misbehavior" class. Multiclass classification is a special case of statistical classification. It is used to assign one of several misbehaving class labels to an instance. Multiclass classification is more complex as compared to binary classification. We observe from this table that Random Forest (RF) and J-48 classifier outperform rest of the classifiers in terms of high values of TPR , TNR and small values of FPR , FNR . The reason for better classification is the *bagging* and *boosting* properties of these classifiers. Whereas, Naive Bayes classifier shows poor results as compared to other classifiers. Improved multi-class classification accuracies can also be seen in Table 1.

We have shown through simulations that our proposed approach shows promising results on legitimate and malicious instances gathered from the simulation process. However, in a realistic VANET scenario, the proposed approach may be unsuitable to detect temporal attacks because of unavailability of packet transmit/receive information by individual node.

Table 1. Classification accuracy using various attributes of legitimate and misbehaving vehicles

Classifiers	Binary-Class				Multi-Class			
	TPR	FPR	TNR	FNR	TPR	FPR	TNR	FNR
Naive Bayes	0.42	0.03	0.96	0.57	0.32	0.01	0.98	0.67
IBK	0.56	0.11	0.88	0.43	0.58	0.04	0.95	0.41
J-48	0.92	0.01	0.98	0.07	0.93	0.004	0.99	0.06
RF	0.92	0.01	0.98	0.07	0.93	0.005	0.99	0.06
AdaBoost1	0.92	0.02	0.97	0.07	0.93	0.008	0.99	0.06

A distributed approach in realistic scenario can be used to evaluate the parameters required for temporal attacks detection. In this approach, all VANET nodes (vehicles and RSUs) observe the functionality of their neighboring nodes (nodes present in their vicinity). Each node that generates a safety packet overhears the further forwarding of that packet by all receiving nodes. Features of VANET like mobility, presence of RSUs in the form of traffic and road lights and traffic pattern can be used in this approach. RSUs can record the trajectories of vehicles moving across them and exchange it with others RSUs or vehicles. In this manner, all nodes exchange observations regarding the send and receive time of each packet, thereby contributing to generation of parameters required to detect temporal attacks in VANET. We use the benefit of availability of this information in the simulator for temporal attack detection in our implemented approach.

7 Conclusion

Misbehaving vehicles are disastrous for any VANET application. Misbehavior detection mechanisms are becoming prominent area of research in VANET. In this paper, we implemented various forms of misbehaviors in VANET. Features of each form of misbehavior is extracted. These experiments are performed using various combinations of misbehaviors. We have designed a framework for binary and multi-class classification to differentiate between (legitimate and malicious behavior) and (legitimate and particular class of misbehavior) respectively. The results are validated using evaluation metrics computed by various classifiers. We observe that Random forest and J-48 classifiers signify the perfect classification of malicious vehicles with high values of *TPR*, *TNR* and small values of *FPR* and *FNR*. The basic framework proposed in this paper can be made compatible with any type of misbehaviors induced in a particular application. As a part of future work, we would like to implement detection approach for temporal attacks in realistic scenario.

References

1. Aijaz, A., Bochow, B., Dtzer, F., Festag, A., Gerlach, M., Kroh, R., Leinmller, T.: Attacks on Inter Vehicle Communication Systems - an Analysis. In: Proc. WIT, pp. 189–194 (2006)
2. University of Waikato: Open Source Machine Learning Software Weka, <http://www.cs.waikato.ac.nz/ml/weka>

3. Raya, M., Hubaux, J.P.: Securing Vehicular Ad Hoc Networks. *Journal of Computer Security* 15(1), 39–68 (2007)
4. Golle, P., Greene, D., Staddon, J.: Detecting and Correcting Malicious Data in VANETs. In: *Proceedings of the 1st ACM International Workshop on Vehicular Ad Hoc Networks (VANET 2004)*, pp. 29–37. ACM, New York (2004)
5. Ghosh, M., Varghese, A., Kherani, A.A., Gupta, A.: Distributed Misbehavior Detection in VANETs. In: *Proceedings of the 2009 IEEE Conference on Wireless Communications and Networking Conference*, pp. 2909–2914. IEEE, Los Alamitos (2009)
6. Ghosh, M., Varghese, A., Kherani, A.A., Gupta, A., Muthaiah, S.N.: Detecting Misbehaviors in VANET with Integrated Root-cause Analysis. *Ad Hoc Netw.* 8, 778–790 (2010)
7. Raya, M., Papadimitratos, P., Aad, I., Jungels, D., Hubaux, J.P.: Eviction of Misbehaving and Faulty nodes in Vehicular Networks. *IEEE Journal on Selected Areas in Communications, Special Issue on Vehicular Networks* 25(8), 1557–1568 (2007)
8. Yan, G., Olariu, S., Weigle, M.C.: Providing VANET Security Through Active Position Detection. *Comput. Commun.* 31(12), 2883–2897 (2008)
9. Xiao, B., Yu, B., Gao, C.: Detection and localization of Sybil nodes in VANETs. In: *Proceedings of the 2006 Workshop on Dependability Issues in Wireless Ad Hoc Networks and Sensor Networks (DIWANS 2006)*, pp. 1–8. ACM, New York (2006)
10. Raya, M., Papadimitratos, P., Gligor, V.D., Hubaux, J.P.: On data centric trust establishment in ephemeral ad hoc networks. In: *IEEE INFOCOM* (2008)
11. Schmidt, R.K., Leinmuller, T., Schoch, E., Held, A., Schafer, G.: Vehicle Behavior Analysis to Enhance Security in VANETs. In: *Vehicle to Vehicle Communication, V2VCOM* (2008)
12. Kim, T.H., Studer, H., Dubey, R., Zhang, X., Perrig, A., Bai, F., Bellur, B., Iyer, A.: VANET Alert Endorsement Using Multi-Source Filters. In: *Proceedings of the Seventh ACM International Workshop on Vehicular InterNetworking*, pp. 51–60. ACM, New York (2010)
13. Jiang, D., Delgrossi, L.: IEEE 802.11p: Towards an International Standard for Wireless Access in Vehicular Environments. In: *Vehicular Technology Conference (VTC-2008)*, pp. 2036–2040. IEEE, Los Alamitos (2008)
14. NCTUns 5.0, Network Simulator and Emulator, <http://NSL.csie.nctu.edu.tw/nctuns.html>
15. Witten, I.H., Frank, E.: *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Francisco (1999)
16. Fawcett, T.: An introduction to ROC analysis. *Pattern Recogn. Lett.* 27(8), 861–874 (2006)

Improved Position-Based Routing in Vehicular Ad Hoc Networks Using P-DIR Method

Ram Shringar Raw and D.K. Lobiyal

School of Computer and Systems Sciences,
Jawaharlal Nehru University, New Delhi, India
rsrao08@yahoo.in, lobiyal@gmail.com

Abstract. Due to high mobility of nodes in multi-hop Vehicular Ad hoc Networks (VANETs), choice of next hop forwarding node plays a vital role to improve the performance of routing protocol. In this paper, we have proposed a position-based routing protocol, that we call Projection based Directional Routing (P-DIR) protocol. In this protocol, a neighbor node with maximum potential gain (PG) is selected as a next hop node. The potential gain PG is estimated in terms of Link Duration (LD) between the source and next hop node, and Direction (DR) and Position metric (PN) of the next hop node towards the closest direction of the destination. We have simulated the proposed protocol in the MATLAB and results have been computed for link duration and position metric with the varying distance between one hop neighbor nodes.

Keywords: VANET, Position based routing, MFR, DIR, P-DIR, Potential gain.

1 Introduction

Vehicular Ad hoc Networks (VANETs) is composed of a large number of vehicles providing connectivity to each other. We assume that vehicles move in any direction with high mobility. Neighboring vehicles that are within a transmission range directly communicate over a wireless links. End-to-end connectivity between source and destination vehicle in the VANET requires multi-hop packet forwarding by intermediate vehicles. We focus on spontaneous vehicular ad hoc networks that begin to appear in highly dense environments to provide full network connectivity.

Unlike traditional ad hoc and other routing protocols, position based routing protocols presents challenging and interesting properties of VANETs [1]. Position based routing protocols does not require any information about the global topology, but uses the local information of neighboring nodes and position information of destination node to forward packets. Generally position based routing is based on greedy forwarding scheme that guarantees loop-free operation. Position based routing only requires local routing tables based on local information of neighboring nodes that are restricted to the transmission range of any forwarding node. Due to this restrictions, it gives low overhead of their creation and maintenance. The local information about the physical location of nodes can be obtained by using Global Positioning System (GPS), provided the nodes are equipped with GPS receiver [2].

The task of finding and maintaining routes in VANETs is challenging and difficult due to fast mobility of vehicular nodes and frequent network topology change. We assume a circular transmission range to place vehicle nodes and study routing performance for different network sizes and node densities. In this paper, we propose a position based routing protocol, that we call Projection based Directional Routing (P-DIR) protocol. We also propose a mathematical model for P-DIR and use it to maximize the routing performance in terms of link duration (connectivity) between nodes within the transmission range.

The rest of this paper is organized as follows. We discuss the related work in section 2. In section 3, the design of P-DIR routing protocol is introduced. Section 4 presents the simulation result and performance analysis of the proposed protocol. Finally, we conclude this paper in section 5.

2 Related Work

In VANET, position based routing protocols are more promising routing solutions than other routing protocols. In a position based routing scheme, forwarding node currently holding the packet, is aware only about the position of its neighbors within the transmission range and destination (see Fig. 1). Examples of such routing protocols includes Greedy Perimeter Stateless Routing (GPSR), Most Forward within Radius (MFR) routing, Location-Aided Routing (LAR), Directional Routing (DIR) etc.

Takagi and Kleinrock [3] proposed the first position based routing protocol that is Most Forward within Radius (MFR) routing protocol which is based on the notion of progress. MFR is a well-known loop-free method for finding a route in a network by utilizing position information of nodes [4]. The neighbor with the greatest progress on the straight line joining the source and destination is chosen as next-hop for sending packets further. Therefore MFR forwards the packet to the node that is closest to the destination node in an attempt to minimize the number of hops [5]. In Fig.1, *S* and *D* denote the source and destination nodes and the circle with radius *R* indicates maximum communication range of source node *S*. The source node *S* has five neighbors within its communication range. It selects node *A* as next hop for forwarding packet to the destination since the projection *A'* of *A* on the line *SD* is closest to destination *D*.

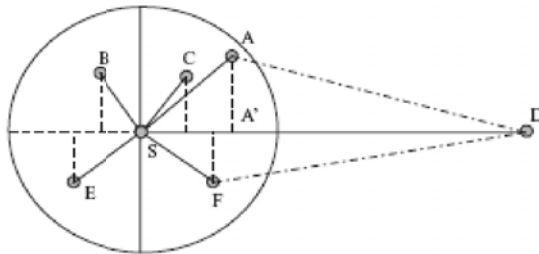


Fig. 1. Selection of neighbor nodes F and A as the next hop node in DIR and MFR method

Kranakis [6] proposed the Directional Routing (DIR) (referred as the Compass Routing) is based on the greedy forwarding method in which the source node uses the position information of the destination node to calculate its direction. Then the message is forwarded to the nearest neighbor having direction closest to the line drawn between source and destination. Therefore, a message is forwarded to the neighboring node minimizing the angle between itself, the previous node, and the destination. In Fig. 1, node F has the closest direction (node F has minimum angle among all the neighbors within the communication range) to line SD and is the selected next hop node for further packet forwarding.

In DIR [4], next-hop neighbor node is decided through unicast forwarding by using the position information of the sender node, its next neighbor nodes, and the destination node. To obtain the position information of the next neighbor node, each node (vehicle) within the communication range sends a beacon or a Hello packet containing its identity (ID), current position and other important information in the network. On the reception of a Hello packet from the neighbor node, the receiving node obtains the position information of its neighbor node. The performance of all these routing protocols depends on the network density. Further, all these routing schemes have high packet delivery rates for dense vehicular networks and low delivery rates for sparse vehicular networks.

3 Proposed Work

In this work, we have proposed a position based routing protocol that we call Projection based Directional Routing Protocol (P-DIR) designed for VANET. It uses the characteristics of both MFR and DIR. P-DIR is greedy forwarding protocol that uses movement of vehicular nodes. By knowing the movement information of the nodes (vehicles) involved in the route, we can estimate their future positions used to predict the link duration between each pair of nodes in the route. In this routing protocol, the neighbor node is selected as the next-hop node with maximum progress and minimum angle with full connectivity between nodes within the transmission range. Hence, a packet is sent to the neighbor node with maximum progress as the distance between source and destination projected onto the line drawn from source to destination and direction source to selected neighbor node is closest to direction SD . Therefore, when a source node have more than one neighbors having the same maximum progress (projection on the line SD , source to destination) towards destination within the transmission range, our proposed protocol P-DIR is useful.

In Fig. 2, S and D are source and destination nodes. R is the transmission radius of the source node S . Source node S has six neighbors in its transmission range. Among these neighbors, nodes A , B , and C have same projection A' on the line SD . Node B has smallest angle (direction SB is closest to the direction SD) that is, the angle BSD is minimum among angle node ASD and node CSD . Therefore, source node S selects node B as the next-hop node for further transmission.

In P-DIR, we improve MFR protocol using directional method and link duration between the nodes to increase its reliability. Let B is a potential forwarder (PF) of source node S , randomly located at a distance D and angle α , then the potential gain (P_G) can be calculated to find the closest neighbor as a next-hop node to forward packet from source to destination.

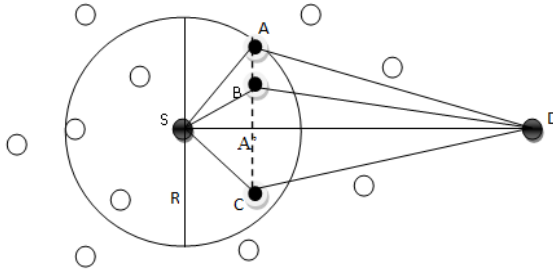


Fig. 2. Selection of node B as the next hop node in P-DIR method

$$Potential\ Gain\ (P_G) = a L_D + b D_R + c P_N \tag{1}$$

Where a , b , and c be the three potential factors of the metrics link duration (L_D), direction (D_R), and position metric (P_N) respectively.

3.1 Link Duration (L_D)

Let N nodes be placed inside a square area having side length k and m is the average number of neighbors of each node (average node density) in the given area. Let R is the radius of the circle which is equal to the transmission range of the any node inside the square area [7] (see Fig. 3). We assume that half of the area of the circle (transmission range), where nodes are randomly distributed in the direction of destination. The transmission range of any node needed to achieve the expected number of nodes inside a circle centered at one of nodes can be estimated as follows:

$$m = \left(\frac{\pi R^2}{2k^2} \right) \cdot (N - 1) \tag{2}$$

Therefore, the transmission range R is as:

$$R = k \sqrt{\frac{m}{\pi (N - 1)}} \tag{3}$$

This is the transmission range having the desired average node density.

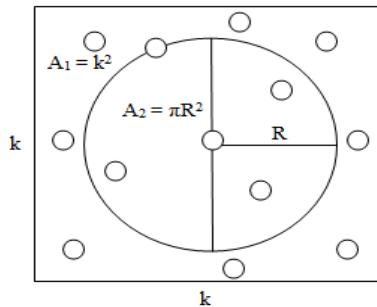


Fig. 3. Node distribution in square area

We assume that the nodes are uniformly distributed within the transmission range R and most of time the network is connected. A path that consists of links between pair of nodes in the path with maximum link duration form a stable path. Link duration is defined as the maximum time of connectivity between any two one-hop neighboring nodes within the transmission range. Let A and B be two neighbors within the transmission range R of any forwarding node (source node). (X, Y) and (X_1, Y_1) be the coordinates of nodes A and B moving with node velocity V_1 and V_2 . Further, it is assumed that nodes A and B are moving in α and β direction towards destination respectively. After time T the new coordinates for node A and B will be (X_2, Y_2) and (X_3, Y_3) , respectively [8] [9] (see Fig. 4). Now for time T , we can calculate the distance D_1 and D_2 traveled by node A and B as follows:

$$D_1 = V_1 T \text{ and } D_2 = V_2 T$$

The new coordinates (X_2, Y_2) and (X_3, Y_3) of node A and B , respectively can be calculated as:

$$X_2 = X + X_2 = X + D_1 \cos \alpha = X + T (V_1 \cos \alpha)$$

$$Y_2 = Y + Y_2 = Y + D_1 \sin \alpha = Y + T (V_1 \sin \alpha)$$

$$X_3 = X_1 + X_3 = X_1 + D_2 \cos \beta = X_1 + T (V_2 \cos \beta)$$

$$Y_3 = Y_1 + Y_3 = Y_1 + D_2 \sin \beta = Y_1 + T (V_2 \sin \beta)$$

Now we can calculate the distance D between node A and node B at time T as follows:

$$d^2 = (x + d_1)^2 + (y + d_2)^2$$

$$d = \sqrt{(x + d_1)^2 + (y + d_2)^2} \tag{4}$$

Where $x = (X - X_1)$, $y = (Y - Y_1)$, $d_1 = T (V_1 \cos \alpha - V_2 \cos \beta)$, and $d_2 = T (V_1 \sin \alpha - V_2 \sin \beta)$.

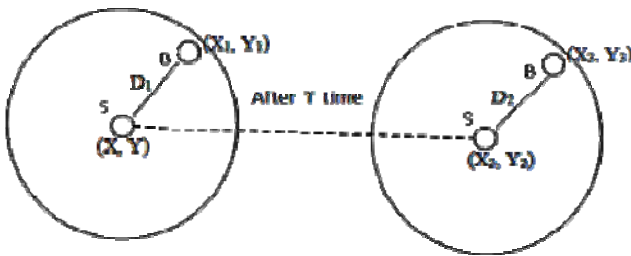


Fig. 4. New coordinates during node movements

For transmission range R of any forwarding node, we can calculate the Link Duration (L_D) between forwarding node and selected next-hop node over time T as follows:

$$L_D = \frac{\text{Transmission range of forwarding node (R)}}{\text{Distance between forwarding node and selected Next - hop node(d)}}$$

$$L_D = \frac{\sqrt[k]{\frac{m}{\pi^{(N-1)}}}}{\sqrt{(x + d_1)^2 + (y + d_2)^2}} \tag{5}$$

When the distance between forwarding node and selected next-hop node becomes larger than the maximum transmission range R of the forwarding node, both the nodes will be disconnected.

3.2 Direction towards Destination (D_R)

We assume that the next-hop forwarding node selected in the optimal direction towards destination may result into minimum end-to-end delay due to small number of hops. Therefore, the nodes placed in the half of the communication range having direction towards destination are selected for packet forwarding. Let θ be the angle between the direction of the movement of next-hop node and the straight line drawn between source and destination (SD). Therefore, direction of next-hop node towards destination can be calculated as follows:

$$\cos \theta = \frac{d^2 + D_{SD}^2 - d'^2}{2 d D_{SD}}, \text{ where } \theta \leq \frac{\pi}{2}$$

As the direction of node may be both positive and negative, to make its magnitude a single quantity whichever side of the base axis it is, the direction of next-hop node can be represented as:

$$D_R = \cos^2 \theta = \left[\frac{d^2 + D_{SD}^2 - d'^2}{2 d D_{SD}} \right]^2 \tag{6}$$

Where $d = \sqrt{(x + d_1)^2 + (y + d_2)^2}$.

3.3 Position Metric for Next-Hop Node

MFR protocol selects a node as next hop node that makes maximum progress towards the destination. In our proposed protocol, we have used a position metric defined in term of the progress towards the destination. The progress made by a next-hop node is calculated as:

$$P_{Prog} = \frac{d^2 + D_{SD}^2 - d'^2}{2 D_{SD}}$$

Where P_{Prog} is the maximum progress towards destination and it is defined as the distance between the forwarding node (source node) and its projection B' on the line SD , between source and destination (see Fig. 5). Further the position metric P_N of the selected next-hop node which is closer to destination is defined as follows:

$$P_N = \frac{P_{Prog}}{D_{SD}} = \frac{d^2 + D_{SD}^2 - d'^2}{2 D_{SD}^2} \tag{7}$$

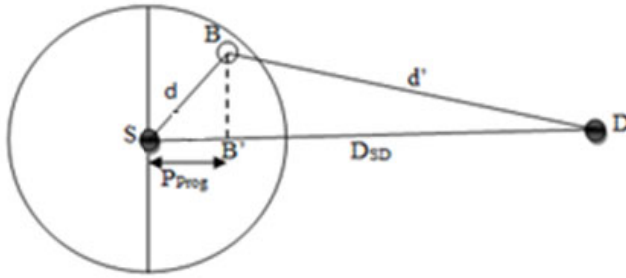


Fig. 5. Position of Next-hop node calculation

Now the potential gain of a neighboring node with stable link between forwarding and next-hop node in the closest direction of the destination with maximum progress can be expressed as:

$$P_G = a \left(\frac{R}{d} \right) + b \left(\frac{d^2 + D_{SD}^2 - d'^2}{2dD_{SD}} \right)^2 + c \left(\frac{d^2 + D_{SD}^2 - d'^2}{2D_{SD}^2} \right) \tag{8}$$

Therefore, a node with highest potential gain among neighbors of the forwarding node can be chosen as next-hop node.

4 Results and Performance Analysis

To evaluate the performance of P-DIR protocol, it is implemented in MATLAB. The performance of P-DIR is computed analytically and numerically both. Based on the simulation parameters given below in table 1, we have simulated the protocol for varying distance between the source and next-hop node from 0 to 200. In the simulations, results have been computed in terms of link duration (connectivity between nodes) and position metric of next-hop node for the effect of varying distance between nodes within transmission range of each other.

Table 1. Simulation parameter

Parameter	Value
Simulated Area	2000m × 2000m
Transmission Range	200m
Number of Nodes	1000
Distance	0-200m
Simulation time	100s

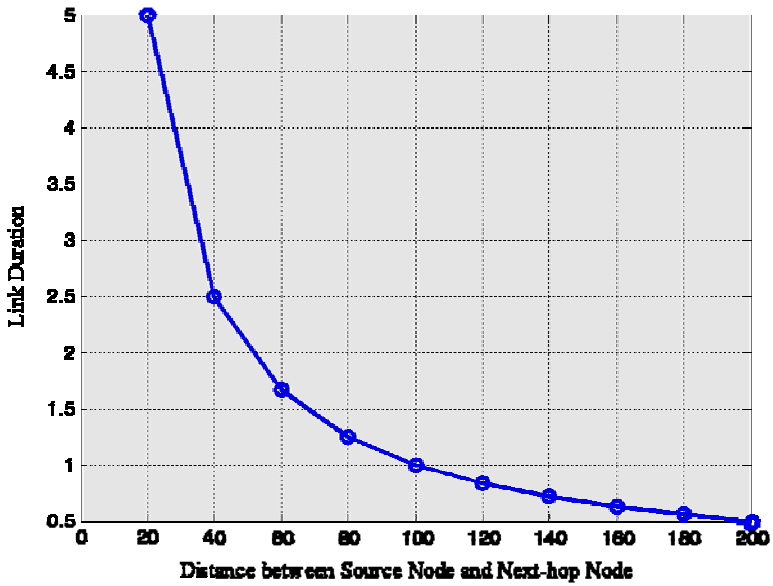


Fig. 6. Link duration between nodes

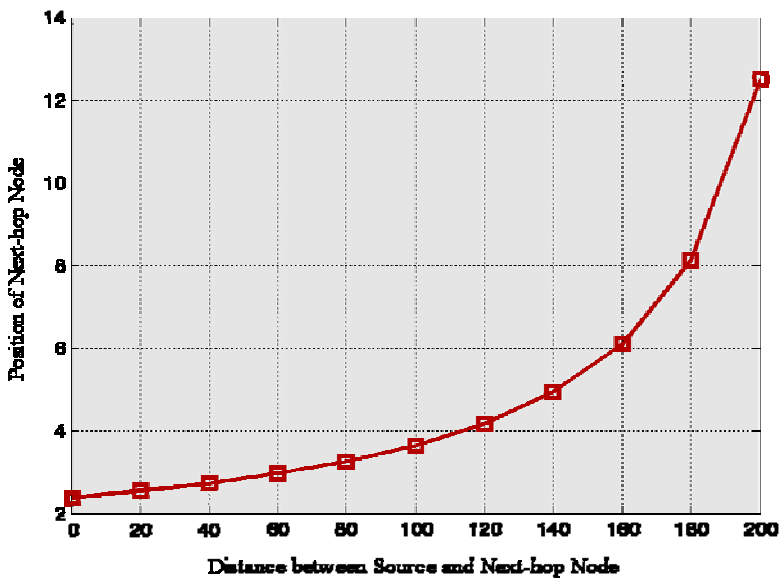


Fig. 7. Position of next-hop node

4.1 Link Duration

High mobility of vehicular nodes in the VANET changes the network connectivity rapidly and it affects the performance of routing protocol. The link duration is one of the most useful metrics to predict this behavior. Link duration is the amount of time a node has an active link to its one-hop neighbor. Fig. 6 shows the link duration with the different distance between one hop neighbor nodes for the link duration potential factor $\alpha = 0.5$. Link duration between nodes (vehicles) decreases as distance between nodes within the transmission range increase.

4.2 Position of the Next-Hop Node

In the P-DIR scheme the packet is forwarded to a neighbor node whose progress is maximal towards destination. The position metric is defined in terms of the ratio of the progress the distance between the source and destination.

Therefore, position metric ensures that preference is given to neighbor node that is closer to destination. In Fig. 7, we can see as the distance between source node and next-hop node is increasing, the position metric for the next-hop node is also increasing. This increase is rapid as the distance increases beyond 40.

5 Conclusion

In this work, we have proposed a routing protocol that we call Projection based Directional Routing (P-DIR) protocol based on greedy forwarding scheme. Our approach is based on mathematical analysis. At every timeslot, each node in the network has a potential gain. Based on this estimate, P-DIR selects a neighbor node as a next-hop node to forward the packet from source to destination. In the simulation part, link duration metric is computed, so that at each timeslot there is end-to-end path (link) that connects every pair of source and destination. As shown in the figure, link duration is increases as node density increases.

References

1. Yousefi, S., Mousavi, M.S., Fathy, M.: Vehicular Ad Hoc Networks (VANETs): Challenges and Prospectives. IEEE, Los Alamitos (2006)
2. Capkun, S., Hamdi, M., Hubaux, J.P.: GPS-Free Positioning in Mobile Ad-hoc Networks. Cluster Computing 5(2), 157–167 (2002)
3. Takagi, H., Kleinrock, L.: Optimal transmission ranges for randomly distributed packet radio terminals. IEEE Transactions on Communications 32(3), 246–257 (1984)
4. Stojmenovic, I., Ruhil, A.P., Lobiyal, D.K.: Voronoi diagram and convex hull based Geocasting and routing in wireless networks. In: Wireless Communications and Mobile Computing. John Wiley & Sons Ltd., Chichester (2006)
5. Raw, R.S., Lobiyal, D.K.: E-DIR: A Directional Routing Protocol for VANETs in a City Traffic Environment. International Journal of Information and Communication Technology (IJICT) 3(2) (2011) ISSN: 1466-6642

6. Kranakis, E., Singh, H., Urrutia, J.: Compass routing on geometric networks. In: Proc. 11th Canadian Conference on Computational Geometry, Vancouver (August 1999)
7. Contla, P.A., Stojmenovic, M.: Estimating Hop Counts in Position Based Routing Schemes for Ad Hoc Networks. *Telecommunication Systems* 22, 1–4 (2003)
8. Brahmi, N., Boussejra, M., Mouzna, J., Bayart, M.: Adaptive Movement Aware Routing in Vehicular Ad Hoc Networks. In: *IWCMC 2009*. ACM, New York (2009)
9. Prasanth, K., Duraiswamy, K., Jayasudha, K., Chandrasekar, C.: Improved Packet Forwarding Approach in Vehicular Ad hoc Networks Using RDGR Algorithm. *International Journal of Next Generation Network (IJNGN)* 2(1) (March 2010)

IMS and Presence Service Integration on Intelligent Transportation Systems for Future Services

Andrés García, José Santa, Antonio Moragón, and Antonio F. Gómez-Skarmeta

Departamento de Ingeniería de la Información y las Comunicaciones,
Universidad de Murcia, 30100 Murcia, España
{asgarcia, josesanta, amoragon, skarmeta}@um.es

Abstract. Once some stable solutions appear to be found at physical and network layers for cooperative Intelligent Transportation Systems (ITS), deploying vehicular services and managing their operation over a common framework is becoming a key issue. Apart from proprietary and ad-hoc solutions, this paper presents a platform based on the IP Multimedia Subsystem (IMS) for deploying ITS services. The proposed architecture decouples IMS from a specific communication technology, to provide a common framework where to deploy and access both peer-to-peer and infrastructure-oriented ITS services. IMS capabilities have been used as an overlay network to provide vehicular services. Standard features of IMS, such as authentication, authorization, accounting, instant messaging, and session or subscription management, are exploited as a common basis for ITS services. In particular, the work presented shows how the standardized IMS Presence Service can be extremely useful for services which follow a publish-subscribe scheme, very common in vehicle-to-infrastructure applications. The whole IMS-based platform has been both designed and developed, even presenting reference implementations of real services.

Keywords: Vehicular Services, IMS, NGN, SIP, ITS.

1 Introduction

According to [1], vehicle telematics can be defined as “the use of computers and telecommunications to enhance the functionality, productivity and security of both vehicles and drivers”. Vehicular networks are becoming essential for telematics services within the Intelligent Transportation Systems (ITS) field. Nowadays, there are more and more vehicular services which needs communications with both the surrounding vehicles and the infrastructure. The literature about ITS has many works about communication technologies used to connect vehicles with other vehicles and with local or remote infrastructure points, and multitude of works dealing with routing problems at network level. However, apart from physical, MAC and network routing issues, an ITS telematic platform needs a

suitable framework to access, delivery and manage services oriented to drivers, occupants, road operators, administrations, etc.

In the above described scenario is where Next Generation Networks (NGN), and the solution given by the 3GPP, by means of the IP Multimedia Subsystem (IMS), are found useful to deploy ITS telematic services. IMS is a proposal standardized by 3GPP for the provision and deployment of multimedia services to final users, enabling the convergence of voice, video and real time data over IPv4 and IPv6. IMS works over the IP layer, and the technologies in which it relies, such as SIP (Session Initiation Protocol) [2] for session control, allow a fast development of services and reduces the operation and infrastructure costs. Furthermore, there are several standardized services within IMS specification, such as the Presence Service, location functions or Instant Messaging (IM). Among them, the Presence Service provides a mechanism able to dynamically manage information about the status of users, equipments and services, thus posing a great opportunity to deploy feature-rich services in the ITS context.

In this paper, we propose an architecture that integrates IMS in the ITS domain, to provide telematic services over a common framework, and decoupling the IMS architecture defined by 3GPP from cellular networks. The idea has been to develop an overlay network, communication technology agnostic, for the management, deployment and delivery of services in ITS field. Among other functionalities, the proposal has been extended with the design and implementation of a service deployment scheme based on the Presence Service. The whole work comprises both design and implementation issues, including various service prototypes in order to demonstrate the potential of the solution.

The structure of the paper is as follows. Section 2 introduces IMS and its application in the vehicular field. After that, several related works are presented in Section 3. Section 4 details the architecture of the solution, dealing with the integration of IMS entities and basic services. Section 5 explains the development of the presence-based service platform, and details how ITS telematic services can benefit from the solution. Section 6 describes the service prototypes implemented as a proof of concept. Finally, the work is concluded in Section 7 with future works and emphasizing the importance of the IMS-based platform for the provision of new services in vehicular environments.

2 IMS Background

2.1 3GPP IMS Core

The 3GPP IMS is an architectural framework for delivering IP services on a common platform. It consists of a standardized set of functions that cooperate and are usually matched to different nodes. These are common recognized entities in most of the the IMS implementations that perform a well-defined task. The IMS specification has been designed to enable operators to provide a wide range of real-time and packet-based services by which users can be charged. It also provides a framework for deploying both common calling capabilities and advanced telematic services [3]. The heart of the IMS network are the Call Session

Control Functions (CSCF), typically a SIP server with a customizable behavior. CSCFs process SIP signaling providing registration and session establishment capabilities. It could be said that CSCFs conform the routing machinery of the IMS [4]. IMS core entities are:

- Proxy CSCF (P-CSCF). It is the entry point to the IMS domain and serves as the outbound proxy server for clients.
- Home Subscriber Server (HSS). It is the central repository for user information, including identity management and user status. It holds any required information to handle user registration and session establishment.
- Serving CSCF (S-CSCF). It is the brain of the IMS core, by processing registrations of terminals with the IMS domain and managing calls to service endpoints. The operation of the S-CSCF can be configured by modifying the policies stored in the HSS.
- Interrogating CSCF (I-CSCF). It acts as an inbound SIP proxy server. On IMS registrations, I-CSCF queries the HSS to select the appropriate S-CSCF which must serve each user. The I-CSCF routes the incoming session requests to the S-CSCF of the called party.
- Subscriber Location Functions (SLF). It is a database that maps users' URI to the associated HSS (in case more than one exists).
- SIP Application Server (SIP-AS). It is the native IMS application server, and serves as host for the execution of IP multimedia or non multimedia services based on SIP signaling.

2.2 3GPP IMS Emergency Sessions

Recently, 3GPP has extended the IMS architecture to provide Emergency Call services (e-Call). To do so, new IMS functions have been included:

- Emergency CSCF (E-CSCF). E-CSCF is usually implemented as a SIP Server, as the rest of CSCFs, and it is responsible for handling emergency session establishments and termination [5].
- Location Retrieving Function (LRF). This functional entity handles the retrieval of location information for the User Equipment (UE).

When UE registers with the IMS network with a special emergency flag, it tries to identify its own location by asking the LRF. With that information, UE sends a call initiating message to P-CSCF, which ignores the roaming restrictions of the subscriber (since it could be connecting with a visited IMS domain) and whether the user is registered or not, and then forwards the request to the nearest E-CSCF. E-CSCF determines then the address of an appropriate Public Service Access Point (PSAP) and routes the call to it, thus completing the call establishment. The PSAP is the demarcation point between the IMS network and the Public Switched Telephone Network (PSTN), and will be in charge of managing the emergency situation according to the data provided.

2.3 3GPP IMS Presence Architecture

Over the IMS core functions, 3GPP also defines the Presence Service (PS) based on the IETF presence model specifications known as “SIP for Instant Messaging and Presence Leveraging Extensions” (SIMPLE) [6]. PS accepts, stores and distributes presence information based on SIP messages. In that model, presence is understood as information published by users (presence entities, abbreviated as presentities) to other users (watchers) to indicate their ability and willingness to communicate [7]. Therefore, a presentity provides information about its own presence to the PS. This presence information is then redistributed by the PS to all the IMS entities subscribed to this presentity. A presentity can have several devices capable of sending information about the user, these are known as Presence User Agents (PUA). In fact, PUAs send presence information to a Presence Agent (PA), which is the part of the PS responsible for storing and forwarding Presence Information to all the involved watchers.

In the proposal given in this paper, developed services can also use IM for sending/receiving messages during a session, also following the SIMPLE framework. In fact, one of the most powerful advantages of our platform is the integration with advanced NGN enablers. These enablers (also known as SIP-AS) provide common functionalities to services that they can reuse, delegating critical operations to a trusted reliable entity. Some examples of enablers are: identity provisioning, authentication, authorization, location server, broadcast service, user profile, multi-conference, instant messaging and presence. These enhance consistency and reliability of new services, promoting re-use, reducing development costs and improving usability.

3GPP have defined the Presence Information Data Format (PIDF) [8] and the Rich Presence Information Data Format (RPID) [9] for exchanging presence information, although they have been extended in this proposal. PIDF defines a compact model to describe the presence information of a presentity, whereas RPID allows a presentity to send detailed and rich presence information to her watchers.

Going a step further, 3GPP has also extended the base PS with new optimizations. Usually, each time a watcher wants to subscribe to a presentity, it have to initiate a subscription transaction. This signalization follows a path from the UE of the potential watcher to the UE of the presentity, what can implies too many hops if the terminals are in different domains. To solve this problem, the IETF created the concept of resource list. A resource list (also called presence list) is a list of SIP URIs that are stored in a new functional entity called Resource List Server (RLS). RLS receives watcher subscriptions to presence list URIs (since a presence list has its own URI) and automatically performs the subscription process to all the presentities in the list. When using an RLS, the presence list is stored in an independent reliable network entity rather than being stored locally in the presence user application. Moreover, the RLS not only carries out all individual subscriptions, but it also combines presence information from presentities in a single message delivered to watchers of resource lists.

This work integrates PS, the RLS function, and the XML Configuration Access Protocol (XCAP) to manage resource lists. XCAP is a XML-based protocol which provides capabilities to add, modify, and delete XML configuration data stored in a server, such as users in a presence list, authorization policies or a list of participants in a conference. XCAP is transported over Hypertext Transfer Protocol (HTTP). According to the Open Mobile Alliance (OMA), XCAP is the protocol used for the XML Document Management (XDM). The current work integrates an XDM Server (XDMS) to manage contacts lists and presence rules for each presentivity, as it is described later.

3 Related Work

The use of IMS in ITS as a platform to deliver services has not been sufficiently exploited yet, although some solutions in this line are available in the literature. The iRide [10] proposal comprises an IMS-based service, with an associated client application, for warning drivers about hazardous situations on the road. For this purpose, a multi-hop wireless sensor network is deployed along with a wireless infrastructure for connecting IMS clients to the core system. Once in the main system, data provided by the sensor network is transmitted to an application server that hosts the IMS service. This processes the received information and alert drivers about dangerous situations, although this communication flow is not specified in the proposal. A similar work is presented in [7], where a wireless sensor network is also used to monitor health status data in order to establish an emergency call when needed. The example given here tracks heart conditions, and automatically establishes a 911 call with the PSAP upon the detection of an incoming stroke. The main problem found in this work is that a wireless sensor network is used to collect health data, and further service-level mechanisms would be needed to assure the establishment of an emergency call when needed.

The work presented in [5] focuses on IMS e-Call capabilities. It proposes to enhance the emergency functionality of IMS to provide context-aware emergency services. It proposes the management of contextual information (also captured in this work by a wireless sensor network) by the system, and the integration of this information in emergency service operations by extending the PIDF format. The retrieved information is processed and stored in a database entity, which is also responsible of managing and disseminating more contextual information provided by sensor gateways. Besides, the LRF is extended to be able to retrieve contextual information related to users. The main drawback is that the emergency service is isolated from the rest of the IMS core, and it cannot interact with other services, like the presence one.

4 Reference Architecture

The proposed architecture is based on the 3GPP IMS Release 7 [11] and the main entities are depicted in Fig. 1. Starting from the bottom part of the diagram, the clients of the services are vehicle terminals mounted on the on-board equipment.

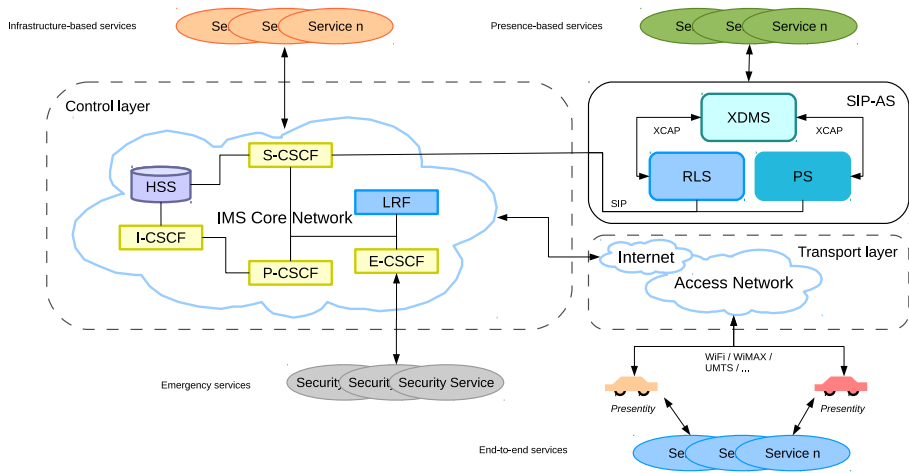


Fig. 1. Reference architecture

In the proposal, each vehicle has its own IMS identity provided in advance by the administrator of the IMS network. The IMS infrastructure is thus independent of any telecom operator. The separation of the communication technology from the potential services to be offered, leads to an access technology agnostic system. Examples of possible access technologies could be 802.11p, WiFi, UMTS, LTE or WiMAX. Additionally, the architecture provides reliable authentication and authorization to users. This feature will be essential for accessing future vehicular services.

In the proposed architecture, signaling is based on well-known and widely used protocols, such as SIP; SDP (Session Description Protocol), for capacity negotiation; RTP (Real-Time Transport Protocol) for multimedia data transmissions; or Diameter, for authentication and authorization purposes. All parties connected this overlay IMS network are identified by an unique URI.

Independently of the communication technology used, every vehicle communicates with the IMS core, composed of the three classical CSCFs (as explained before), and the E-CSCF/LRF functions, for providing emergency services. Once registered in the network, each vehicle can benefit from a series of services. The E-CSCF has an interface with both S-CSCF and P-CSCF to receive emergency registrations and emergency call session establishments, and it relies in LRF for user localization purposes. In case of a new emergency session establishment, both LRF and E-CSCF expect the GEOPRIV PIDF location object format [12] and the civic one [13]. If SIP messages do not contain this information, LRF can be configured to retrieve it form a LOCSIP server, defined by OMA Location in SIP/IP Core.

On the top of the IMS Core Network, there can be several SIP-AS hosting several IMS basic services. Those services will be described later, but one of them is the presence one. Regarding PS, the architecture presented in Fig. 1

includes some extensions. To improve the PS performance an RLS entity has been added to reduce the number of SIP messages exchanged, by acting as a proxy when a new watcher subscription is done and by merging presence notifications. Note that watchers can be far away of the IMS Core Network and connected by a limited wireless or expensive connection. To cope with this objective, RLS needs previously configured contact lists with presentities, to perform the subscription process. As can be noted, by adding this presence capabilities it is possible to easily deploy services that follow a publish-subscribe scheme, such as fleet management, vehicle monitoring and, in general, services which require the collection of data from the road side.

The architecture presented enables operators, authorities and final users to take advantage of four different types of services:

- Emergency services. These derive from the *eCall* idea, i.e. a service for accepting safety calls.
- End-to-end services. These are peer-to-peer services that extend common calls, such as videoconference, sessions for exchanging context-related information for a collaborative goal, or instant messaging between terminals. In the session establishment phase the user terminal initiates the SIP dialog and it is supervised by the IMS Core. For these kind of services, ASs do not take part of the negotiation process.
- Infrastructure-based services. These services are also supported by the 3GPP specifications, and they need both the core and a SIP-AS where to install service ends. Examples of this kind of services can follow a common client-server approach, such a contextual on-demand tourism service or a ticketing service for spectacles/museums.
- Presence-based services. These are more sophisticated services which use the presence-based architecture proposed here, such as the prototypes presented in following sections.

Presence-based services act as watchers of a resource list, this one composed of presentities (vehicle terminals) that send status information. Since vehicle terminals only maintain a connection with one PS, all information regarding all presence-based services comes in the same notification. This behavior presents both performance and privacy issues, since all watchers would receive all status data (its own ones and the ones for the rest of services). This problem is solved by using presence rules to filter in RLS the information for each presence-based service (watcher). This reduces the amount of data sent to each watcher and avoid privacy concerns, by choosing which attributes or nodes will be notified to each watcher/service. As with presence lists, presence rules can be managed only by registered users or entities, who must be authenticated and authorized to modify a certain presence rules document; hence, to make this process easier, both RLS tasks and presence rules are managed by an XDMS entity, as can be seen in Fig. 11. Presence lists are matched with concrete presence-based services with an unique SIP URI defined in RLS, which can be managed through XDMS. The service itself communicates with the XDMS to add, modify or delete presentities in its own presence list. In general, the XDMS server accomplishes the

management of the XML documents and it is fully XCAP compliant, including some enhanced features. It provides authentication methods and constraint the introduction of information for each existing user within the IMS network.

5 Development of the Presence-Based Service Platform

This section explains in detail the implementation of the proposed architecture. As it has been mentioned, one of the main novelties of the proposal is the integration of new services in a presence-based platform, specially designed for the ITS field. According to our design, PS manages status publications from multiple vehicles. This includes refreshing or replacing existing presence information with newly-published information. PS is fully compatible with the IMS Core. The integration of the IMS core and this enabler has been made by inserting specific trigger points in HSS thus, enabling the S-CSCF to redirect presence SIP messages (REGISTER, SUBSCRIBE, PUBLISH and its responses) to PS.

Regarding the IMS Core Network, the Fraunhofer FOKUS *Open IMS Core* has been used. This system is largely based on the SIP Express Router (SER) project, which includes an implementation of P-CSCF, I-CSCF, S-CSCF and E-CSCF. A lightweight implementation of an HSS has been added, which works as a repository for application server settings and user profiles. A basic LRF for retrieving user location information is also given. These elements comprise the IMS Core Network and they are responsible of the session control, which is essential for supporting services.

In the first stages of the development, the Mobicents SIP Presence Service was used as SIP-AS. Mobicents is an Open Source VoIP Platform programmed in Java, which enables the convergence of voice, video and data for NGN applications. Mobicents SIP Presence Service is composed of three separate but interrelated servers, which can be deployed separately or in the same host: a SIP presence server, an XML server implementation suitable to develop our XDMS, and an RLS. These cover the necessities marked by the architecture previously presented.

Due to some instabilities in the Mobicents PS, it was later substituted by Kamailio. Kamailio, called before OpenSER, is a fork of the SER (SIP Express Router) project. Kamailio aims to deliver a similar level of flexibility and high performance as SER. It is written in C programming language and is based on several modules. The presence related modules needed are XCAP, RLS and some modules for managing XML documents. Since Kamailio can provide PS, XDMS and RLS, we can swap Mobicents for Kamailio with low effort.

Each IMS/SIP entity of our solution has been deployed on a separated Xen virtual machine with 256 MB of RAM memory, 4GB of hard disk, and two 2GHz CPUs shared among all of them. Only the SIP-AS has an increment of 1GB of RAM memory to maintain an acceptable performance when several services are installed.

For the on-board equipment, we have chosen JAIN SIP 1.2 for SIP signaling processing. This provides a standardized interface through a Java API to develop SIP-based applications and, in this case, to access the IMS infrastructure.

For the provision of status information from vehicles, the RPID model used includes extensions to support the defined presence-based services. An example of the document is shown next. The extension includes the new fields needed for a service later explained in detail. For an accurate determination of the position of the vehicle, apart from the geographical coordinates, an integrity factor about the current position is included [14]. Together with these fields, the current speed and time are also added.

```
<?xml version='1.0' encoding='UTF-8'?>
<presence xmlns='urn:ietf:params:xml:ns:pidf'
  xmlns:c='urn:ietf:params:xml:ns:pidf:cipid'
  xmlns:dm='urn:ietf:params:xml:ns:pidf:data-model'
  xmlns:rpidf='urn:ietf:params:xml:ns:pidf:rpidf'
  entity='sip:vehicle@open-ims.test'>
  <tuple id='t288c55b7'>
    <status>
      <basic>open</basic>
    </status>
  </tuple>
  <dm:person id='pb4cedcfd'>
    <rpidf:activities>
      <rpidf:emergency>
        <dm:position>
          <dm:latitude>38.0240442</dm:latitude>
          <dm:longitude>-1.1743681</dm:longitude>
          <dm:integrity>1.0</dm:integrity>
        </dm:position>
        <dm:speed>22.372652952823422</dm:speed>
        <dm:timestamp>1145534561736</dm:timestamp>
      </rpidf:emergency>
    </rpidf:activities>
  </dm:person>
</presence>
```

6 Service Prototypes

In the frame of the developed platform, some implementations have been made to validate the architecture for each of the four kind of services covered: as end-to-end service, a videoconference application has been created and installed in the on-board computer; as infrastructure-based service a traffic information channel has been implemented taking advantage of the instant messaging capabilities of IMS; as security service a reference implementation of an eCall-equivalent system has been developed to assist vehicles on emergency situations; and, finally, as a presence-based service we have developed the Access to Restricted Areas tool.

Access to Restricted Areas service allows authorities or road operators to control the access of vehicles, above all ambulances or official vehicles, to special



Fig. 2. Access to Restricted Areas service reference architecture

places or road segments. When an authorized vehicle is close to a previously configured restricted area, the Traffic Control Centre (TCC) will open the access barrier. For this purpose TCC and the vehicle terminal must be registered in the service, so that any variation on the location of the vehicle will be notified to TCC when the vehicle is in "Emergency State".

In this service, the approaching vehicle will act as a presentity, thus using the service to dynamically publish its state ("Emergency" or "Regular") as well as its location and other contextual information relevant for TCC. The presence-base service in the platform, i.e. TCC, is configured as a watcher of every presentity in a contact list of potential emergency vehicles. The presence and location information of the presentity is encapsulated according to the model presented in the previous section, and it is periodically sent to PS, which is in charge of filtering all presence data for each service and forward notifications to them. This status information enables TCC to monitor the location of emergency vehicles, check their state, and detect whether they are approaching to a certain restricted area by matching location information with a digital map. When the logic of the service, attending these presence events, realizes a vehicle is close to a restricted area, it opens barrier automatically. The trigger distance of the vehicle towards the barrier can be configured at the TCC application, which can be seen in Fig. 2. In this screenshot an authorized operator is monitoring a vehicle in emergency state, which has just access a restricted area in the Espinardo Campus at University of Murcia. As can be seen, TCC has been implemented as a Web application that can be opened on a Web browser after an authentication stage.

7 Conclusions and Future Work

In this paper an IMS-based platform for deploying services in ITS is presented. This platform is composed of an IMS Core, basic services deployed over an AS (above all, the enriched presence service) and high-level services that take advantage of the platform. The work also analyses the many advantages found in IMS for service provisioning in vehicular scenarios. Readers must also notice that this proposal is in line with current standardization works regarding communication architectures in ITS, such as CALM (Communications Access for Land Mobiles) or the ETSI European ITS Communication Architecture, since the necessary IMS middleware on ITS stations located on vehicles, personal devices, roadside or centrals, can be placed on the facilities layer, i.e. working over transport protocols.

The deployment of advanced functionalities for vehicular services, such as service-level network authentication/authorization, accounting or the usage of base services such as presence or instant messaging, are offered with no extra efforts and efficiently by means of the proposed IMS architecture. Thus, the proposal not only enriches significantly the logic of services, but also becomes a very attractive framework for a successful growth of vehicular services by using current communication technologies. Based on the developed IMS presence service, together with the RLS functionality, resource lists, the XDMS proposal and the extension of presence notification messages, we have successfully tested real prototypes of services, such as the Access to Restricted Areas, demonstrating the feasibility of the proposal.

Regarding the performance of the platform, we must notice that ASs may collapse due to the number of notifications when presence services are used by many vehicles. However, this load can be reduced by offering more than one IMS Core domain and even including load balancing strategies with several ASs on each domain. Our work will continue by developing more services, such a novel radio-equivalent system implemented with the base Push-to-Talk functionality of IMS, and testing their performance in a wider context, involving more vehicles. Some works have also already carried out in this line, by emulating the operation of services with traffic traces generated with SUMO (Simulation of Urban Mobility). Our plans also consider integrating this IMS proposal with an IP-based mobility environment for in-vehicle networks, to test services over several network access technologies (including vertical and horizontal handovers among them) and providing network-level security by means of IPSec (Internet Protocol Security) supported by IKEv2 (Internet Key Exchange Version 2).

Acknowledgements

This work has been sponsored by the Spanish Ministry of Education and Science, through the SEISCIENTOS Project (TIN2008-06441-C02), the European Commission, through the ITSSv6 project (270519), and the Seneca Foundation, by means of the Excellence Researching Group Program (04552/GERM/06).

References

1. Wrigh, M.J.: What are vehicle telematics and vehicle tracking? In: UK Telematics Online website (August 2009)
2. Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M.: SIP: Session Initiation Protocol. RFC 3261 (June 2002)
3. Lequerica, I., Ruiz, A.J., Garcia, A.S., Gomez-Skarmeta, A.F.: An IMS based Vehicular Service Platform. In: IEEE Vehicular Technology Conference Fall, Ottawa, pp. 1–5 (October 2010)
4. Poikselkä, G.M.M.: The IMS: IP multimedia concepts and services, 3rd edn. Wiley, John Wiley and Sons Ltd., The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, United Kingdom (2009)
5. El Barachi, M., Kadiwal, A., Ghitho, R., Khendek, F., Dssouli, R.: An Architecture for the Provision of Context-Aware Emergency Services in the IP Multimedia Subsystem. In: IEEE Vehicular Technology Conference Spring, Singapore, pp. 2784–2788 (May 2008)
6. Campbell, H.K.B.: SIMPLE: SIP for Instant Messaging and Presence Leveraging Extensions. RFC 3856 (March 2011)
7. El Barachi, M., Kadiwal, A., Ghitho, R., Khendek, F., Dssouli, R.: A Presence-based Architecture for the Integration of the Sensing Capabilities of Wireless Sensor Networks in the IP Multimedia Subsystem. IEEE Communications Society 3116–3121 (April 2008)
8. Klyne, G., Sugano, H., Fujimoto, S., et al.: Presence Information Data Format (PIDF). RFC 3863 (August. 2004)
9. Schulzrinne, H., Gurbani, V., Kyzivat, P., Rosenberg, J.: RPID: Rich Presence Extensions to the Presence Information Data Format (PIDF). RFC 4480 (July 2006)
10. Elkotob, M., Osipov, E.: iRide: A Cooperative Sensor and IP Multimedia Subsystem Based Architecture and Application for ITS Road Safety. Communications Infrastructure. Systems and Applications in Europe 16, 153–162 (2009)
11. Technical Specification Group Services and System Aspects. IP Multimedia Subsystem (IMS); Stage 2. Technical report, 3rd Generation Partnership Project, Valbonne, France (December 2009)
12. Tschofenig, H., Winterbottom, J., Thomson, M.: GEOPRIV Presence Information Data Format Location Object (PIDF-LO). RFC 5491 (March 2009)
13. Winterbottom, J., Thomson, M.: Revised civic location format for presence information data format location object (pidf-lo). Rfc 5139 (February 2008)
14. Toledo-Moreo, R., Santa, J., Zamora-Izquierdo, M.A., Ubeda, B., Gomez-Skarmeta, A.F.: A study of integrity indicators in outdoor navigation systems for modern road vehicle applications. In: IEEE International Conference on Intelligent Robots and Systems, Workshop on Planning, Perception and Navigation for Intelligent Vehicles, Nice, France (September 2008)

Author Index

- Abbadi, Imad M. IV-406, IV-557
Abbas, Ash Mohammad II-307
Abraham, Anuj III-503
Abraham, John T. III-168
Abraham, Siby I-328
Achuthan, Krishnashree I-488, II-337
AdiSrikanth, III-570
Aditya, T. I-446
Adusumalli, Sri Krishna IV-572
Agarwal, Vikas II-595
Aghila, G. II-327, IV-98
Agrawal, P.K. IV-244
Agrawal, Rohit II-162
Agrawal, Shaishav III-452
Agushinta R., Dewi II-130, II-138,
II-146
Ahmed, Imran II-317
Ahn, Do-Seob II-595
Aishwarya, Nandakumar II-490, II-498,
III-269
Akhtar, Zahid II-604
Al-Sadi, Azzat A. II-535
Alam, Md. Mahabubul III-349
Alam Kotwal, Mohammed Rokibul
II-154
Ananthi, S. I-480
Andres, Frederic IV-79
Anisha, K.K. III-315
Anita, E.A. Mary I-111
Anju, S.S. II-490, II-498, III-269
Annappa, B. IV-396
Anto, P. Babu III-406
Anusiya, M. IV-155
Aradhya, V.N. Manjunath III-289,
III-297
Arifuzzaman, Md. III-349
Asif Naeem, M. II-30
Asokan, Shimmi IV-63
Athira, B. II-80
Awais, Muhammad II-374
Awasthi, Lalit Kr III-609
Azeem, Mukhtar II-525
Azeez, A.A. Arifa IV-145
Babu, Korra Sathya II-1
Babu, K. Suresh II-636
Babu, L.D. Dhinesh I-223
Babu, M. Rajasekhara I-182
Baburaj, E. I-172
Badache, N. IV-593
Bagan, K. Bhoopathy IV-524
Bajwa, Imran Sarwar II-30
Bakshi, Sambit III-178
Balasubramanian, Aswath I-411
Banati, Hema II-273
Banerjee, Indrajit III-68
Banerjee, Joydeep III-82
Banerjee, Pradipta K. II-480
Banerjee, Usha II-648
Bansal, Roli III-259
Banu, R.S.D. Wahida II-545
Baruah, P.K. I-446
Basak, Dibyajnan I-519
Basil Morris, Peter Joseph II-577
Baskaran, R. II-234, IV-269
Bastos, Carlos Alberto Malcher
IV-195
Batra, Neera I-572
Bedi, Punam II-273, III-259
Bedi, R.K. II-397
Behl, Abhishek II-273
Bhadoria, P.B.S. IV-211
Bhardwaj, Ved Prakash II-568
Bharti, Brijendra K. IV-358
Bhat, Veena H. III-522
Bhattacharyya, Abhijan I-242
Bhosale, Arvind IV-512
Bhuvanagiri, Kiran Kumar IV-293
Bhuvanewary, A. II-327
Biji, C.L. IV-300
Binu, A. I-399
Biswas, G.P. II-628
Biswas, Subir III-54
Biswas, Suparna II-417
Biswas, Sushanta II-612, II-620
Biswash, Sanjay Kumar I-11
Boddu, Bhaskara Rao II-296
Borah, Samarjeet III-35

- Borkar, Meenal A. IV-25
 Boutekkouk, Fateh II-40

 Chaganty, Aparna IV-19
 Chaitanya, N. Sandeep IV-70
 Chakraborty, Suchetana II-585
 Chakravorty, Debaditya III-35
 Challa, Rama Krishna IV-608
 Chanak, Prasenjit III-68
 Chand, Narottam III-122, III-609
 Chandra, Deka Ganesh II-210
 Chandra, Jayanta K. II-480
 Chandran, K.R. I-631
 Chandrika, I-704
 Chaniyani, S.S. Mozaffari III-289
 Chaoub, Abdelaali I-529
 Chaudhary, Ankit III-488
 Chauhan, Durg Singh I-21
 Chawhan, Chandan III-35
 Chawla, Suneeta II-430
 Chia, Tsorng-Lin III-334
 Chintapalli, Venkatarami Reddy IV-455
 Chitrakala, S. III-415
 Chittineni, Suresh III-543
 Choudhary, Surendra Singh I-54
 Chouhan, Madhu I-119
 Chowdhury, Chandreyee I-129
 Chowdhury, Roy Saikat II-577

 Dadhich, Reena I-54
 Dahiya, Ratna III-157
 Dandapat, S. IV-165
 Das, Madhabananda IV-113
 Das, Satya Ranjan II-172
 Das, Subhalaxmi IV-549
 Datta, Asit K. II-480
 Dawoud, Wesam I-431
 Deb, Debasish II-577
 Dedavath, Saritha I-34
 Deepa, S.N. III-503
 Dehalwar, Vasudev I-153
 Dehuri, Satchidananda IV-113
 Desai, Sharmishta II-397
 Devakumari, D. II-358
 Devani, Mahesh I-213
 Dhanya, P.M. IV-126
 Dhar, Pawan K. I-284
 Dharanyadevi, P. II-234
 Dhavachelvan, P. II-234
 Dhivya, M. II-99

 Dilna, K.T. III-185
 Dimililer, Kamil III-357
 Diwakar, Shyam II-337
 Doke, Pankaj I-607, II-430
 Dongardive, Jyotshna I-328
 Donoso, Yezid II-386
 Doraipandian, Manivannan III-111
 Dorizzi, Bernadette III-20
 Durga Bhavani, S. III-1
 Dutta, Paramartha I-83
 Dutta, Ratna IV-223

 El Abbadi, Jamal I-529
 El-Alfy, El Sayed M. II-535
 Elhaj, Elhassane Ibn I-529
 Elizabeth, Indu I-302
 Elumalai, Ezhilarasi I-1

 Ferreira, Ana Elisa IV-195
 Ferri, Fernando IV-79

 Gadia, Shashi II-191
 Gaiti, Dominique II-471
 Ganeshan, Kathiravelu IV-501
 Garcia, Andrés III-664
 Garcia, Anilton Salles IV-195
 Gaur, Manoj Singh I-44, I-162, I-562,
 II-183, II-452, III-478, III-644
 Gaur, Vibha II-284
 Gautam, Gopal Chand I-421
 Geetha, V. II-48
 Geevar, C.Z. III-460
 Ghosh, Pradipta III-82
 Ghosh, Saswati II-620
 Giluka, Mukesh Kumar I-153
 Gindi, Sanjyot IV-349
 Gireesh Kumar, T. II-506
 Giuliani, Alessandro I-284
 Godavarthi, Dinesh III-543
 Gómez-Skarmeta, Antonio Fernando
 III-664
 Gondane, Sneha G. II-99
 Gopakumar, G. I-320
 Gopalan, Kaliappan IV-463
 Gore, Kushal I-607
 Gosain, Anjana I-691
 Gosalia, Jenish IV-378
 Govardhan, A. I-581
 Govindan, Geetha I-294
 Govindarajan, Karthik I-192

- Grifoni, Patrizia IV-79
 Grover, Jyoti III-644
 Gualotuña, Tatiana IV-481
 Guerroumi, M. IV-593
 Gunaraj, G. I-192
 Gunjan, Reena III-478
 Gupta, Ankur I-501
 Gupta, B.B. IV-244
 Gupta, Deepika II-183
 Gupta, J.P. I-260
 Gupta, Juhi IV-205
 Gupta, Priya IV-512
- Habib, Sami J. II-349
 Hafizul Islam, SK II-628
 Harivinod, N. III-396
 Harmya, P. II-490, II-498, III-269
 Harshith, C. II-506
 Hassan, Foyzul II-154, III-349
 Hati, Sumanta III-580
 Hazarika, Shyamanta M. II-109, II-119
 Hazra, Sayantan III-601
 Hemamalini, M. IV-175
 Hivarkar, Umesh N. IV-358
 Hsieh, Chaur-Heh III-334
 Huang, Chin-Pan III-334
 Huang, Ping S. III-334
- Ibrahim, S.P. Syed I-631
 Indira, K. I-639
 Isaac, Elizabeth IV-145
- Jaganathan, P. I-683
 Jagdale, B.N. II-397
 Jain, Jitendra III-326
 Jain, Kavindra R. III-239
 Jain, Kavita I-328
 Jalal, Anand Singh II-516, IV-329
 Jameson, Justy II-693
 Janani, S. IV-175
 Jaya, IV-233
 Jayakumar, S.K.V. II-234
 Jayaprakash, R. II-656
 Jena, Sanjay Kumar II-1
 Jia, Lulu IV-421
 Jiménez, Gustavo II-386
 Jisha, G. IV-1, IV-137
 Joseph, Shijo M. III-406
 Juluru, Tarun Kumar I-34, III-590
- Kacholiya, Anil IV-205
 Kahlon, K.S. II-58
 Kakoty, Nayan M. II-119
 Kakulapati, Vijayalaxmi IV-284
 Kalaivaani, P.T. III-143
 Kale, Sandeep II-604
 Kanade, Sanjay Ganesh III-20
 Kanavalli, Anita I-141
 Kancharla, Tarun IV-349, IV-368
 Kanitkar, Aditya R. IV-358
 Kanivadhana, P. IV-155
 Kankacharla, Anitha Sheela I-34, III-590
 Kanmani, S. I-639, II-69
 Kannan, A. II-19
 Kannan, Rajkumar IV-79
 Kapoor, Lohit I-501
 Karamoy, Jennifer Sabrina Karla II-138
 Karande, Vishal M. IV-386
 Karmakar, Sushanta II-585
 Karthi, R. III-552
 Karthik, S. I-480
 Karunanithi, D. IV-284
 Karunanithi, Priya III-624
 Karuppanan, Komathy III-425, III-615, III-624, III-634
 Katiyar, Vivek III-122
 Kaur, Rajbir I-44, I-162
 Kaushal, Sakshi IV-445
 Kavalcioglu, Cemal III-357
 Kayarvizhy, N. II-69
 Keromytis, Angelos D. III-44
 Khajaria, Krishna II-9
 Khalid, M. I-182
 Khan, Majid Iqbal II-471, II-525
 Khan, Srabani II-620
 Khan Jehad, Abdur Rahman III-349
 Khanna, Rajesh III-205
 Khattak, Zubair Ahmad IV-250
 Khilar, P.M. I-119
 Kim, Pansoo II-595
 Kimbahune, Sanjay I-607, II-430
 Kiran, N. Chandra I-141
 Kishore, J.K. II-460
 Ko, Ryan K.L. IV-432
 Kolikipogu, Ramakrishna IV-284
 Kopparapu, Sunil Kumar II-317, IV-293
 Koschnicke, Sven I-371
 Kothari, Nikhil I-213

- Krishna, Gutha Jaya I-382
 Krishna, P. Venkata I-182
 Krishna, S. III-522
 Krishnan, Saranya D. IV-63
 Krishnan, Suraj III-374
 Kopparapu, Sunil Kumar III-230
 Kulkarni, Nandakishore J. III-570
 Kumar, Chiranjeev I-11
 Kumar, C. Sasi II-162
 Kumar, G.H. III-289
 Kumar, G. Ravi IV-70
 Kumar, G. Santhosh I-399
 Kumar, Ishan IV-205
 Kumar, K.R. Ananda I-704
 Kumar, K. Vinod IV-19
 Kumar, Manish I-44
 Kumar, Manoj II-9
 Kumar, Naveen I-461
 Kumar, Padam I-461
 Kumar, Ravindra II-307
 Kumar, Santosh I-619
 Kumar, Santhosh G. III-93
 Kumar, Saumesh I-461
 Kumar, Sumit I-619
 Kumaraswamy, Rajeev IV-339
 Kumari, M. Sharmila III-396
 Kumari, V. Valli IV-572
 Kumar Pandey, Vinod III-230
 Kumar Sarma, Kandarpa III-512
 Kurakula, Sudheer IV-165
 Kussmaul, Clifton III-533

 Lachiri, Zied IV-318
 Lal, Chhagan II-452
 Latif, Md. Abdul II-154
 Laxmi, V. II-183, II-452
 Laxmi, Vijay I-44, I-162, I-562, III-478,
 III-644
 Lee, Bu Sung IV-432
 Li, Tiantian IV-421
 Limachia, Mitesh I-213
 Lincoln Z.S., Ricky II-130
 Linganagouda, K. III-444
 Lingeswarara, C. II-19
 Liu, Chenglian IV-534
 Lobiyal, D.K. III-132, III-654
 Londhe, Priyadarshini IV-512
 López, Elsa Macías IV-481

 Madheswari, A. Neela II-545
 Madhusudhan, Mishra III-365

 Mahalakshmi, T. I-310
 Mahalingam, P.R. III-562, IV-137
 Maheshwari, Saurabh III-478
 Maiti, Santa II-172
 Maity, G.K. III-249
 Maity, Santi P. I-519, III-249, III-580
 Maity, Seba I-519
 Majhi, Banshidhar III-178
 Majhi, Bansidhar IV-549
 Maji, Sumit Kumar I-649
 Malay, Nath III-365
 Malaya, Dutta Borah II-210
 Malik, Jyoti III-157
 Mallya, Anita I-302
 Manan, Jamalul-lail Ab IV-250
 Mandava, Ajay K. I-351
 Mannava, Vishnuvardhan I-250
 Manomathi, M. III-415
 Maralappanavar, Meena S. III-444
 Marcillo, Diego IV-481
 Marimuthu, Paulvanna N. II-349
 Mary, S. Roselin IV-9
 Masera, Guido II-374
 Mastan, J. Mohamedmoideen Kader
 IV-524
 Mehrotra, Hunny III-178
 Meinel, Christoph I-431
 Mendiratta, Varun II-273
 Menta, Sudhanshu III-205
 Mishra, A. IV-244
 Mishra, Ashok II-223
 Mishra, Dheerendra IV-223
 Mishra, Shivendu II-407
 Misra, Rajiv I-101
 Missaoui, Ibrahim IV-318
 Mitra, Abhijit III-512, III-601
 Mitra, Swarup Kumar III-82
 Mittal, Puneet II-58
 Modi, Chintan K. III-239
 Mohammadi, M. III-289
 Mohandas, Neethu IV-187
 Mohandas, Radhesh II-685, III-10
 Mohanty, Sujata IV-549
 Mol, P.M. Ameera III-193
 Moodgal, Darshan II-162
 Moragón, Antonio III-664
 More, Seema I-361
 Moussaoui, S. IV-593
 Mubarak, T. Mohamed III-102
 Mukhopadhyay, Sourav IV-223

- Mukkamala, R. I-446
 Muniraj, N.J.R. I-270, III-168
 Murthy, G. Rama IV-19

 Nadarajan, R. II-366
 Nadkarni, Tanusha S. II-685
 Nag, Amitava II-612, II-620
 Nagalakshmi, R. I-683
 Nagaradjane, Prabagarane III-374
 Nair, Achuthsankar S. I-284, I-294,
 I-302, I-320
 Nair, Bipin II-337
 Nair, Madhu S. III-193, III-276
 Nair, Smita IV-368
 Nair, Vrinda V. I-302
 Namboodiri, Saritha I-284
 Namritha, R. III-634
 Nandi, Sukumar I-619
 Narayanan, Hari I-488
 Nasiruddin, Mohammad II-154
 Naskar, Mrinal Kanti III-82
 Nataraj, R.V. I-631
 Naveen, K. Venkat III-570, III-615
 Naveena, C. III-297
 Nazir, Arfan II-525
 Neelamegam, P. III-111
 Neogy, Sarmistha I-129, II-417
 Nigam, Apurv II-430
 Nimi, P.U. IV-46
 Niranjana, S.K. III-297
 Nirmala, M. I-223
 Nirmala, S.R. III-365
 Nitin, I-21, II-568, IV-25
 Noopa, Jagadeesh II-490, II-498, III-269
 Nurul Huda, Mohammad II-154, III-349

 Oh, Deock-Gil II-595
 Okab, Mustapha II-40
 Oliya, Mohammad I-232
 Olsen, Rasmus L. IV-37

 Padmanabhan, Jayashree I-1, IV-541
 Padmavathi, B. IV-70
 Pai, P.S. Sreejith IV-339
 Pai, Radhika M. II-460
 Paily, Roy IV-165
 Pais, Alwyn R. II-685, IV-386
 Pais, Alwyn Roshan III-10
 Pal, Arindarjit I-83
 Palaniappan, Ramaswamy IV-378

 Palaty, Abel IV-56
 Pandey, Kumar Sambhav IV-56
 Panicker, Asha IV-300
 Panneerselvam, S. I-223
 Pappas, Vasilis III-44
 Parasuram, Harilal II-337
 Parmar, Rohit R. III-239
 Parthasarathy, Magesh Kannan I-192
 Parvathy, B. I-204
 PatilKulkarni, Sudarshan III-342
 Patnaik, L.M. I-141, II-636, III-522
 Patra, Prashanta Kumar I-649
 Pattanshetti, M.K. IV-244
 Paul, Anu II-201
 Paul, Richu III-213
 Paul, Varghese II-201
 Paulsen, Niklas I-371
 Pavithran, Vipin I-488
 Pearson, Siani IV-432
 Perumal, V. I-471
 Petrovska-Delacrétaz, Dijana III-20
 Phani, G. Lakshmi IV-19
 Ponpandiyan, Vigneswaran IV-541
 Poornalatha, G. II-243
 Povar, Digambar I-544
 Prabha, S. Lakshmi I-192
 Prabhu, Lekhesh V. IV-339
 Pradeep, A.N.S. III-543
 Pradeepa, J. I-471
 Prajapati, Nitesh Kumar III-644
 Prakasam, Kumaresh IV-541
 Pramod, K. III-444
 Prasad, Ramjee IV-37
 Prasath, Rajendra II-555
 Prasanna, S.R. Mahadeva III-326
 Prasanth Kumar, M. Lakshmi I-11
 Pratheepraj, E. III-503
 Priya, K.H. I-471
 Priyadharshini, M. IV-269
 Priyadharshini, V. IV-175
 Pung, Hung Keng I-232

 Qadeer, Mohammed Abdul II-442

 Radhamani, A.S. I-172
 Rafsanjani, Marjan Kuchaki IV-534
 Raghavendra, Prakash S. II-243
 Raghuvanshi, Rahul I-153
 Rahaman, Hafizur III-68

- Raheja, J.L. III-488
 Raheja, Shekhar III-488
 Rahiman, M. Abdul III-304
 Rahman, Md. Mostafizur II-154
 Rai, Anjani Kumar II-407
 Rai, Anuj Kumar III-111
 Rai, Mahendra K. III-469
 Raja, K.B. II-636
 Rajapackiyam, Ezhilarasie III-111
 Rajasekhar, Ch. I-78
 Rajasree, M.S. III-304
 Rajendran, C. III-552
 Rajesh, R. III-497
 Rajeswari, A. III-143
 Rajimol, A. II-253
 Rajkumar, K.K. III-435
 Rajkumar, N. I-683
 Raju, C.K. II-223, IV-211
 Raju, G. I-671, II-253, III-435
 Ramachandram, S. IV-70
 Ramamohanreddy, A. I-581
 Ramaraju, Chithra I-661
 Ramasubbareddy, B. I-581
 Ramaswamy, Aravindh I-411
 Ramesh, Sunanda I-1
 Ramesh, T. I-250
 Rameshkumar, K. III-552
 Rana, Sanjeev I-91
 Rani, Prathuri Jhansi III-1
 Rao, Appa III-102
 Rao, Avani I-213
 rao, D. Srinivasa I-78
 Rao, Prasanth G. III-522
 Rastogi, Ravi I-21
 Rathi, Manisha I-260
 Rathore, Wilson Naik II-676
 Razi, Muhammad II-146
 Reddy, B. Vivekavardhana IV-309
 Reddy, G. Ram Mohana IV-473
 Reddy, P.V.G.D. Prasad III-543
 Reddy, Sateesh II-460
 Regentova, Emma E. I-351
 Reji, J. III-276
 Revathy, P. IV-284
 Revett, Kenneth IV-378
 Roberta, Kezia Velda II-146
 Rodrigues, Paul IV-9, IV-269
 Rokibul Alam Kotwal, Mohammed
 III-349
 Roopalakshmi, R. IV-473
 Roy, J.N. III-249
 Roy, Rahul IV-113
 Sabu, M.K. I-671
 Saha, Aritra III-35
 Sahaya, Nuniek Nur II-138
 Sahoo, Manmath Narayan I-119
 Sahoo, Soyuj Kumar III-326
 Saikia, Adity II-109, II-119
 Sainarayanan, G. III-157
 Sajeev, J. I-310
 Sajitha, M. III-102
 Saljooghinejad, Hamed II-676
 Samad, Sumi A. III-93
 Samanta, Debasis II-172
 Sambyal, Rakesh IV-608
 Samerendra, Dandapat III-365
 Samraj, Andrews IV-378
 Samuel, Philip II-80, IV-1
 Sandhya, S. II-88
 Santa, José III-664
 Santhi, K. III-221
 SanthoshKumar, G. II-263
 Santhoshkumar, S. I-223
 Saralaya, Vikram II-460
 Sarangdevot, S.S. I-592
 Saraswathi, S. IV-155, IV-175
 Sardana, Anjali IV-233
 Saritha, S. II-263
 Sarkar, D. II-612, II-620
 Sarkar, Partha Pratim II-612, II-620
 Sarma, Monalisa II-172
 Saruladha, K. II-327
 Sasho, Ai I-340
 Sasidharan, Satheesh Kumar I-552
 Satapathy, Chandra Suresh III-543
 Sathisha, N. II-636
 Sathishkumar, G.A. IV-524
 Sathiya, S. IV-155
 Sathu, Hira IV-491, IV-501
 Satria, Denny II-138
 Sattar, Syed Abdul III-102
 Savarimuthu, Nickolas I-661
 Sayeesh, K. Venkat IV-19
 Schatz, Florian I-371
 Schimmler, Manfred I-371
 Sebastian, Bhavya I-302
 Sehgal, Priti III-259
 Selvan, A. Muthamizh III-497

- Selvathi, D. IV-300
 Sen, Jaydip IV-580
 Sendil, M. Sadish I-480
 Senthilkumar, Radha II-19
 Senthilkumar, T.D. III-185
 Shah, Mohib A. IV-491, IV-501
 Shahram, Latifi I-351
 Shajan, P.X. III-168
 Sharma, Amita I-592
 Sharma, Dharendra Kumar I-11
 Sharma, Divya I-511
 Sharma, H. Meena I-162
 Sharma, Neeraj Kumar II-284
 Sharma, Ritu I-511
 Sharma, Sattvik II-506
 Sharma, Sugam II-191
 Sharma, Surbhi III-205
 Sharma, T.P. I-421
 Shekar, B.H. III-396
 Shenoy, P. Deepa I-141, III-522
 Shenoy, S.K. III-93
 Sherly, K.K. II-693
 Shringar Raw, Ram III-654
 Shukla, Shailendra I-101
 Shyam, D. II-99
 Sikdar, Biplab Kumar III-68
 Singal, Kunal III-488
 Singh, Anurag III-609
 Singh, Ashwani II-374
 Singh, Jai Prakash IV-89
 Singh, Jyoti Prakash I-83, II-612, II-620
 Singh, Manpreet I-91, I-572
 Singh, Puneet III-570
 Singh, Rahul I-340
 Singh, Sanjay II-460
 Singh, Satwinder II-58
 Singh, Vijander I-54
 Singh, Vrijendra II-516, IV-329
 Singh, Preety II-183
 Sinha, Adwitiya III-132
 Sivakumar, N. II-88
 Skandha, S. Shiva IV-70
 Smith, Patrick II-191
 Sojan Lal, P. III-460
 Song, Jie IV-421
 Soni, Surender III-122
 Sood, Manu I-511
 Soumya, H.D. I-361
 Sreenath, N. II-48
 Sreenu, G. IV-126
 Sreevathsan, R. II-506
 Srikanth, M.V.V.N.S. II-506
 Srinivasan, Avinash IV-260
 Srinivasan, Madhan Kumar IV-269
 Srivastava, Praveen Ranjan III-570
 Srivastava, Shweta I-260
 Starke, Christoph I-371
 Suaib, Mohammad IV-56
 Suárez-Sarmiento, Alvaro IV-481
 Subramaniam, Tamil Selvan Raman IV-541
 Suchithra, K. IV-339
 Sudarsan, Dhanya IV-137
 Sudhansh, A.S.D.P. IV-165
 Sujana, N. I-361
 Sukumar, Abhinaya I-1
 Sulaiman, Suziah IV-250
 Sundararajan, Sudharsan I-488
 Swaminathan, A. II-648
 Swaminathan, Shriram III-374
 Swamy, Y.S. Kumara IV-309
 Tahir, Muhammad II-471
 Takouna, Ibrahim I-431
 Thakur, Garima I-691
 Thampi, Sabu M. I-64, IV-126, IV-145, IV-187
 Thangavel, K. II-358
 Thilagu, M. II-366
 Thiyagarajan, P. IV-98
 Thomas, Diya I-64
 Thomas, K.L. I-544, I-552
 Thomas, Likewin IV-396
 Thomas, Lincy III-425
 Thomas, Lisha III-221
 Thukral, Anjali II-273
 Tim, U.S. II-191
 Tiwary, U.S. III-452, III-469
 Tobgay, Sonam IV-37
 Tolba, Zakaria II-40
 Tripathi, Pramod Narayan II-407
 Tripathi, Rajeev I-11
 Tripathy, Animesh I-649
 Tripti, C. IV-46
 Tyagi, Neeraj I-11
 Tyagi, Vipin II-568
 Uma, V. II-656
 Umber, Ashfa II-30
 Unnikrishnan, C. III-562

- Usha, N. IV-309
 Utomo, Bima Shakti Ramadhan II-138
- Vanaja, M. I-78
 Varalakshmi, P. I-411, I-471
 Varghese, Elizabeth B. III-383
 Varshney, Abhishek II-442
 Vasanthi, S. III-213
 Vatsavayi, Valli Kumari II-296
 Venkatachalapathy, V.S.K. II-234
 Venkatesan, V. Prasanna IV-98
 Venugopal, K.R. I-141, II-636, III-522
 Verma, Amandeep IV-445
 Verma, Chandra I-284
 Verma, Gyanendra K. III-452, III-469
 Verma, Rohit I-21
 Vidya, M. I-361
 Vidyadharan, Divya S. I-544
 Vijay, K. I-78
 Vijaykumar, Palaniappan I-411
 VijayLakshmi, H.C. III-342
 Vinod, P. I-562
 Vipeesh, P. I-270
- Vishnani, Kalpa III-10
 Vivekanandan, K. II-88
 Vorungati, Kaladhar I-488
 Vykopal, Jan II-666
- Wadhai, V.M. II-397
 Wankar, Rajeev I-382
 Wattal, Manisha I-501
 William, II-130
 Wilscy, M. III-315, III-383
 Wirjono, Adityo Ashari II-130
 Wisudawati, Lulu Mawaddah II-146
 Wu, Jie IV-260
- Xavier, Agnes I-328
- Yadav, Gaurav Kumar IV-368
 Yu, Fan III-54
 Yuvaraj, V. III-503
- Zaeri, Naser II-349
 Zheng, Liyun IV-534
 Zhu, Shenhaochen I-340
 Zhu, Zhiliang IV-421