Sabu M. Thampi · Oge Marques
Sri Krishnan · Kuan-Ching Li
Domenico Ciuonzo
Maheshkumar H. Kolekar (Eds.)

Communications in Computer and Information Science 968

Advances in Signal Processing and Intelligent Recognition Systems

4th International Symposium SIRS 2018
Bangalore, India, September 19–22, 2018
Revised Selected Papers

Springer

Communications in Computer and Information Science 968

Commenced Publication in 2007
Founding and Former Series Editors:
Phoebe Chen, Alfredo Cuzzocrea, Xiaoyong Du, Orhun Kara, Ting Liu, Dominik Ślęzak, and Xiaokang Yang

Editorial Board

More information about this series at http://www.springer.com/series/7899

Sabu M. Thampi · Oge Marques
Sri Krishnan · Kuan-Ching Li
Domenico Ciuonzo · Maheshkumar H. Kolekar (Eds.)

Advances in Signal Processing and Intelligent Recognition Systems

4th International Symposium SIRS 2018
Bangalore, India, September 19–22, 2018
Revised Selected Papers

Springer

Editors
Sabu M. Thampi
Indian Institute of Information Technology and Management
Kerala, India

Oge Marques
Department of Computer and Electrical Engineering and Computer Science
Florida Atlantic University
Boca Raton, FL, USA

Sri Krishnan
Department of Electrical and Computer Engineering
Ryerson University
Toronto, ON, Canada

Kuan-Ching Li
Department of Computer Science and Information Engineering
Providence University
Taichung, Taiwan

Domenico Ciuonzo
University of Naples Federico II
Naples, Italy

Maheshkumar H. Kolekar
Electrical Engineering Department
Indian Institute of Technology Patna
Patna, India

ISSN 1865-0929 ISSN 1865-0937 (electronic)
Communications in Computer and Information Science
ISBN 978-981-13-5757-2 ISBN 978-981-13-5758-9 (eBook)
https://doi.org/10.1007/978-981-13-5758-9

Library of Congress Control Number: 2018966824

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This edited volume contains a selection of refereed and revised papers originally presented at the 4th International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS 2018). The symposium was held at the PES Institute of Technology (South Campus), Bangalore, India, during September 19–22, 2018. SIRS 2018 was co-located with the International Conference on Applied soft computing and Communication Networks (ACN 2018). The symposium provided a forum for the sharing, exchange, presentation, and discussion of original research results in both methodological issues and different application areas of signal processing, computer vision, and pattern recognition.

In response to the call for papers, 92 papers were submitted for presentation and inclusion in the proceedings. The papers were evaluated and ranked on the basis of their significance, novelty, and technical quality. A double-blind review process was conducted to ensure that the author names and affiliations were unknown to the TPC. Each paper was reviewed by the members of the Program Committee. The authors were asked to address each and every comment made by the referees for improving the quality of their papers. The technical program of SIRS 2018 consisted of 28 regular papers and 11 short papers (acceptance rate of 42.4%).

This symposium would not have been possible without the efforts of many people. First of all, we are very grateful to all authors who contributed their research work and shared their knowledge with the scientific community. We would like to express our gratitude to the Program Committee and external reviewers, who worked very hard in reviewing papers and providing suggestions for their improvements. Many thanks to all the chairs for their involvement and support that added greatly to the quality of the symposium. We also wish to thank all the members of the Advisory Committee, whose work and commitment were invaluable.

We thank the PES Institute of Technology, Bangalore, for hosting the conference. Sincere thanks to Dr. M. R. Doreswamy, Chancellor, PES University and Dr. D. Jawahar, Pro-Chancellor, PES University for their valuable suggestions and encouragement. We would like to thank the Organizing Committee, and the many other volunteers who worked behind the scenes to ensure the success of this symposium. Dr. J Surya Prasad, Dr. Sudarshan T. S., and Dr. Shikha Tripathi deserve special mention for their unwavering support. The EDAS conference system proved very helpful during the submission, review, and editing phases.

We wish to express our thanks to Suvira Srivastav, Associate Editorial Director, Springer, New Delhi, for her help and cooperation. Finally, we would like to

acknowledge Springer for thier active cooperation and timely production of the proceedings.

September 2018

Sabu M. Thampi
Oge Marques
Sri Krishnan
Kuan-Ching Li
Domenico Ciuonzo
Maheshkumar H. Kolekar

Organization

Chief Patron

M. R. Doreswamy (Chancellor)	PES University, India

Patrons

D. Jawahar (Pro-chancellor)	PES University, India
Ajoy Kumar (COO)	PES Institutions, India

Honorary General Chair

Sankar Kumar Pal	Indian Statistical Institute, Kolkata, India

General Chairs

Oge Marques	Florida Atlantic University (FAU), USA
Sri Krishnan	Ryerson University, Toronto, Canada
Kuan-Ching Li	Providence University, Taiwan

General Executive Chair

Sabu M. Thampi	IIITM-Kerala, India

Steering Committee Chairs

J. Surya Prasad	PESIT, Bangalore South Campus, India
Sudarshan T. S. B.	PESIT, Bangalore South Campus, India

Technical Program Chairs

Michal Wozniak	Wroclaw University, Poland
Domenico Ciuonzo	University of Naples Federico II, Italy
Maheshkumar H. Kolekar	Indian Institute of Technology Patna, India

Workshop Chairs

Sergey Mosin	Kazan Federal University, Russia
Bartosz Krawczyk	Virginia Commonwealth University, USA

Organizing Chair

Shikha Tripathi	PESIT, Bangalore South Campus, India

Organizing Co-chairs

Annapurna D.	PESIT BSC, India
Subhash Kulkarni	PESIT BSC, India
Sandesh B. J.	PESIT BSC, India

Advisory Committee

Swagatam Das	Indian Statistical Institute, Kolkata, India
Rajesh M. Hegde	Indian Institute of Technology Kanpur, India
David Zhang	The Hong Kong Polytechnic University, Hong Kong, SAR China
Selwyn Piramuthu	University of Florida, USA
Millie Pant	Indian Institute of Technology Roorkee, India
El-Sayed El-Alfy	King Fahd University of Petroleum and Minerals, Saudi Arabia
Ranjan Gangopadhyay	LNMIIT Jaipur, India
Vipin Chaudhary	University at Buffalo, SUNY, USA

Organizing Secretaries

Ajay S. N. R.	PESIT – BSC, India
Pooja Agarwal	PESIT – BSC, India
Krishna Srikanth K. S. V.	PESIT – BSC, India

Technical Program Committee

Emmanuel Agu	Worcester Polytechnic Institute, USA
Ali Qusay Al-Faris	University of the People, USA
Mehmet Celenk	Ohio University, USA
Harishchandra Dubey	University of Texas at Dallas, USA
Rajgopal Kannan	University of Southern California, USA
Hossein Malekinezhad	Michigan Technological University, USA
V. B. Surya Prasath	Cincinnati Children's Hospital Medical Center, USA
Peng Zhang	Stony Brook University, USA
Zhe Zhang	George Mason University, USA
Natacha Gueorguieva	City University of NY, USA
Kaveh Ahmadi	University of Toledo, USA
Abdelbaset Hamza	Appalachian State University, USA
Doyel Pal	LaGuardia Community College, CUNY, USA
Pin-Yu Chen	AI Foudations, IBM T. J. Watson Research Center, USA

Anustup Choudhury	Dolby Laboratories, USA
Kyriakos Manousakis	Applied Communication Sciences, USA
Vesh Raj Sharma Banjade	Intel Corporation, USA
Upasna Vishnoi	Marvell Semiconductor Inc., USA
Yingmeng Xiang	GEIRI North America, USA
Chi Zhang	Midea America Corporation, USA
Eduard Babulak	National Science Foundation, USA
Sergey Biryuchinskiy	Vigitek, Inc., USA
Ali Al-Sherbaz	The University of Northampton, UK
Ognjen Arandjelovic	University of St. Andrews, UK
Md Moinul Hossain	University of Kent, UK
Quoc-Tuan Vien	Middlesex University, UK
Hector Zenil	Oxford University and Karolinska Institute, UK and Sweden
Ahmed Zobaa	Brunel University London, UK
George Magoulas	Birkbeck College, UK
Reyer Zwiggelaar	Aberystwyth University, UK
Kamran Arshad	Ajman University, United Arab Emirates
Burhan Gulbahar	Ozyegin University, Turkey
Yasin Kabalci	Nigde Omer Halisdemir University, Turkey
Mourad Fakhfakh	University of Sfax, Tunisia
Elyas Rakhshani	Delft University of Technology, The Netherlands
Thaweesak Yingthawornsuk	King Mongkut's University of Technology Thonburi, Thailand
Grienggrai Rajchakit	Maejo University, Thailand
Uei-Ren Chen	Hsiuping University of Science and Technology, Taiwan
Wen-Liang Hwang	Institute of Information Science, Academia Sinica, Taiwan
Kuan-Chieh Huang	National Cheng Kung University, Taiwan
Bao Rong Chang	National University of Kaohsiung, Taiwan
I-Cheng Chang	National DongHwa University, Taiwan
Ying-Ren Chien	National I-Lan University, Taiwan
Tzung-Pei Hong	National University of Kaohsiung, Taiwan
Mu-Chun Su	National Central University, Taiwan
Jiachyi Wu	National Taiwan Ocean University, Taiwan
Jia-Chin Lin	National Central University, Taiwan
Ching-Min Lee	I-Shou University, Taiwan
Ali Abbas	Higher Institute for Applied Science and Technology, Syria
Juan Pedro Lopez Velasco	Universidad Politecnica de Madrid, Spain
Gema Piñero	Universitat Politècnica de València, Spain
Manuel Gil Pérez	University of Murcia, Spain
Emilio Jiménez Macías	University of La Rioja, Spain
Addisson Salazar	Universidad Politécnica de Valencia, Spain
Anna Antonyová	University of Prešov in Prešov, Slovakia

Er Joo	University, Singapore
Dragana Bajić	University of Novi Sad, Serbia
Zoran Bojkovic	University of Belgrade, Serbia
Sameera Abar	King Khalid University, Saudi Arabia
El-Sayed El-Alfy	King Fahd University of Petroleum and Minerals, Saudi Arabia
Yassine Khlifi	Umm Al-Qura University, Saudi Arabia
Sergey Mosin	Kazan Federal University, Russia
Salvatore Distefano	University of Kazan, Russia
Andrey Krylov	Lomonosov Moscow State University, Russia
Anh Huy Phan	Skolkovo Institute of Science and Technology, Russia
Mihaela Albu	Politehnica University of Bucharest, Romania
Simina Emerich	Technical University of Cluj-Napoca, Romania
Dan Milici	University of Suceava, Romania
Radu Vasiu	Politehnica University of Timisoara, Romania
Radu Arsinte	Technical University of Cluj-Napoca, Romania
Doru Florin Chiper	University Gh. Asachi of Iasi, Romania
Bogdan Dumitrescu	University Politehnica of Bucharest, Romania
Radu-Emil Precup	Politehnica University of Timisoara, Romania
Faycal Bensaali	Qatar University, Qatar
Rodolfo Oliveira	Nova University of Lisbon, Instituto de Telecomunicações, Portugal
Ramiro Barbosa	Institute of Engineering of Porto, Portugal
Joao Paulo Carvalho	Technical University of Lisbon, Portugal
Isabel Jesus	Institute of Engineering of Porto - ISEP, Portugal
Adão Silva	Instituto de Telecomunicações (IT)/University of Aveiro, Portugal
Grzegorz Debita	Wroclaw University of Technology, Poland
Marek Miskowicz	AGH University of Science and Technology, Poland
Ryszard Tadeusiewicz	AGH University of Science and Technology, Poland
Anna Bartkowiak	University of Wroclaw, Poland
Wael Salah	Palestine Technical University - Kadoorie, Palestine
Belal Amro	Hebron University, Palestine
Muhammad Nasir Khan	The University of Lahore, Pakistan
Rujun Chen	Central South University, P.R. China
Lin Gao	Harbin Institute of Technology (Shenzhen), P.R. China
Chao Zhai	Shandong University, P.R. China
Jiayu Chen	Wuhan University, P.R. China
Boyang Zhou	Zhejiang University, P.R. China
Xiangguo Li	Henan University of Technology, P.R. China
Bin Cao	Harbin Institute of Technology Shenzhen Graduate School, P.R. China
Tsung-Hui Chang	The Chinese University of Hong Kong, Shenzhen, P.R. China
Chi-Hua Chen	Fuzhou University, P.R. China
Pingyi Fan	Tsinghua University, P.R. China

Wei-Chiang Hong	Jiangsu Normal University, P.R. China
Jiayin Wang	Xi'an Jiaotong University School of Management, P.R. China
Ruqiang Yan	Southeast University, P.R. China
Kui Xu	Army Engineering University of PLA, P.R. China
Waleed Abdulla	The University of Auckland, New Zealand
Raouyane Brahim	Faculty of Sciences Ain Chock, Morocco
Ahmed El Oualkadi	Abdelmalek Essaadi University, Morocco
Rodrigo Montufar-Chaveznava	Universidad Nacional Autonoma de Mexico, Mexico
Rosaura Palma-Orozco	Instituto Politécnico Nacional, Mexico
Badrul Hisham Ahmad	Universiti Teknikal Malaysia Melaka, Malaysia
Mohd Ashraf Ahmad	Universiti Malaysia Pahang, Malaysia
Yun Huoy Choo	Universiti Teknikal Malaysia Melaka, Malaysia
Hong Seng Gan	Universiti Kuala Lumpur, Malaysia
Mohd Shahrieel Mohd Aras	Universiti Teknikal Malaysia Melaka, Malaysia
Kok-Lim Alvin Yau	Sunway University, Malaysia
Abdul Halim Ali	Universiti Kuala Lumpur, Malaysia
Kamarulafizam Ismail	Universiti Teknologi Malaysia, Malaysia
Marwan Nafea	Universiti Teknologi Malaysia (UTM), Malaysia
Fazirulhisyam Hashim	Universiti Putra Malaysia, Malaysia
Yen Mei Chee	Sirim Measurements Technology, Malaysia
Gunasekar Thangarasu	Linton University College, Malaysia
Mohammad Faiz Liew Abdullah	Universiti Tun Hussein Onn Malaysia (UTHM), Malaysia
Simon Fong	University of Macau, Macau
Ibrahim Nasir	Sebha University, Libya
Artis Mednis	Institute of Electronics and Computer Science, Latvia
Mike Asiyo	Technical University of Mombasa, Kenya
Galymzhan Nauryzbayev	LN Gumilyov Eurasian National University, Kazakhstan
Alex Pappachen James	Nazarbayev University, Kazakhstan
Nasser Abdellatif	Applied Science University, Jordan
Toshio Tsuji	Hiroshima University, Japan
Noriko Etani	Peach Aviation Limited, Japan
Paolo Crippa	Università Politecnica delle Marche, Italy
Eleonora D'Andrea	University of Pisa, Italy
Gianluigi Ferrari	University of Parma, Italy
Ugo Fiore	University of Naples Federico II, Italy
Giuseppe Ruggeri	University of Reggio Calabria, Italy
Domenico Rotondi	FINCONS SpA, Italy
Igor Bisio	University of Genoa, Italy
Stavros Ntalampiras	Università degli studi Milano, Italy
Giovanni Pau	Kore University of Enna, Italy
Sergio Saponara	University of Pisa, Italy
Gaetano Vacca	Politecnico di Bari, Italy

Adel Obed	Middle Technical University, Iraq
Majida Alasady	University of Tikrit, Iraq
Dheyaa Sabr Al Azzawi	Wasit University, Iraq
Hamed Mojallali	University of Guilan, Iran
Seyyed Amir Hadi Minoofam	Islamic Azad University, NazarAbad Centre, Iran
Mas Irfan Hidayat	Institut Teknologi Sepuluh Nopember, Indonesia
Shikha Agrawal	Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal, India
Shamim Akhter	Jaypee Institute of Information Technology, Noida, India
Shajith Ali	SSN College of Engineering, Chennai, India
Ramesh Babu	DSCE, Bangalore, India
Vinayak Bairagi	University of Pune, India
Shilpi Birla	Manipal University, India
Nabendu Chaki	University of Calcutta, India
Chinmay Chakraborty	Birla Institute of Technology, Mesra, India
Nishant Doshi	PDPU, India
Omar Farooq	Aligarh Muslim University, Aligarh, India
Ashish Goel	JIIT, India
B. Harish	S. J. College of Engineering, India
Madhu Jain	Jaypee Institute of Information Technology, Noida, India
Samudravijaya K.	Indian Institute of Technology Guwahati, India
Raveendranathan Kalathil Chellappan	College of Engineering Thiruvananthapuram, India
Rahul Kher	G. H. Patel College of Engineering and Technology, India
Mofazzal Khondekar	B. C. Roy Engineering College, India
Hareesha K. S.	Manipal Institute of Technology, India
Ajey Kumar	Symbiosis International Deemed University (SIU), India
Dhananjay Kumar	Anna University, India
Anirban Kundu	Netaji Subhash Engineering College, India
Dnyanesh Mantri	Pune University, India
Joycee Mekie	IIT Gandhinagar, India
Deepak Mishra	IIST, India
Dipti Patil	Pune University, India
Parul Puri	Jaypee Institute of Information Technology, India
Jaynendra Kumar Rai	Amity University Uttar Pradesh, India
Harikumar Rajaguru	Bannari Amman Institute of Technology, India
Shanmuganathan Raman	Indian Institute of Technology, Gandhinagar, India
G. Ramachandra Reddy	Vellore Institute of Technology, India
Jibendu Roy	School of Electronics Engineering, KIIT University, India

Shubhajit Roy Chowdhury	School of Computing and Electrical Engineering, IIT Mandi, India
Muthukumar S.	Indian Institute of Information Technology, Tamil Nadu, India
Kandasamy Selvaradjou	Pondicherry Engineering College, India
Raghvendra Sharma	Amity School of Engineering and Technology, Gwalior, India
Pushpendra Singh	Bennett University, Greater Noida, India
Sajai Singh	Jaypee Institute of Information Technology, India
China Sonagiri	IARE JNTUH Hyderabad, India
Madhur Upadhayay	Shiv Nadar University, India
Ajay Verma	Hindustan Institute of Technology and Management, India
Dinesh Vishwakarma	Delhi Technological University, India
Rajesh Vishwakarma	Jaypee University of Engineering and Technology, India
Krishna Battula	Jawaharlal Nehru Technological University Kakinada, India
Nikisha Jariwala	VNSGU, India
Deepa Jose	KCG College of Technology, Karapakkam, Chennai, India
Lajish V. L.	University of Calicut, India
Vanidevi M.	IIST, India
K. Mahantesh	SJBIT, India
Varun Menon	SCMS School of Engineering and Technology, India
Durgesh Nandan	National Institute of Technology, Patna, India
Nitish Ojha	DIT University, Dehradun, India
Ravi Subban	Pondicherry University, Pondicherry, India
Rohit Thanki	C. U. Shah University, India
Arjav Bavarva	RK University, India
Rupam Bhaduri	Dayananda Sagar Institutions- 2nd Campus (DSATM), Bangalore, India
Ravi G.	Sona College of Technology, India
Sudhanshu Gonge	Savitribai Phule Pune University, India
Suneet Gupta	Bennett University Greater Noida, India
Muzzammil Hussain	Central University of Rajasthan, India
Ramachandran K. I.	Amrita Vishwa Vidyapeetham, India
Manikandan J.	PES University (PESU), India
D. John Aravindhar	Hindustan Institute of Technology and Science, India
Kouser K.	GFGC Gundlupet (University of Mysore), India
Navin Kumar	Amrita University, India
Rakesh Kumar	MMM University of Technology, India
Abraham Mathew	NIT Calicut, India
Rajeswari Mukesh	Hindustan University, India
Rajarshi Pal	Institute for Development and Research in Banking Technology (IDRBT), Hyderabad, India

Sanjaya Kumar Panda	Veer Surendra Sai University of Technology, Burla, India
Anitha Pillai	Hindustan University, Chennai, India
Munaga Prasad	IDRBT, India
Kamatchi R.	Amity University, Mumbai, India
Purushothama R.	National Institute of Technology Goa, India
R. Rajalakshmi	Vellore Institute of Technology, Chennai Campus, India
Natarajan S.	VTU, India
Hemraj Saini	Jaypee University of Information Technology, India
Abhijit Sharma	National Institute of Technology, Durgapur, India
Shina Sheen	PSG College of Technology, India
Ananthi Sheshasaayee	Quaid-E-Millath Government College for Women, India
Preety Singh	LNM Institute of Information Technology, India
Vir Singh	Delhi College of Arts and Commerce University of Delhi, India
Sudha T.	University of Calicut, India
Shikha Tripathi	PES University, India
Ashwani Yadav	Amity University Rajasthan, India
Satya Dubey	CSIR-National Physical Laboratory, India
Arun Jana	Centre for Development Advanced Computing, India
Rajeev Shrivastava	MPSIDC, India
Anand Kumar M.	National Institute of Technology - Karnataka, India
Muzammil Hussain	Indian Institute of Science, India
Manjunath Ramachandra	Wipro Technologies, India
Hemant Kumar Rath	Tata Consultancy Services, India
Hrishikesh Sharma	Innovation Labs, Tata Consultancy Services Ltd., India
Vahida Attar	College of Engineering Pune, India
Ravichandra Bagalatti	Visveshwaraya Tehnological University, Belgaum, India
Surajeet Ghosh	Indian Institute of Engineering Science and Technology, India
Bharat Gupta	National Institute of Technology, Patna, India
Ravibabu Mulaveesala	Indian Institute of Technology Ropar, India
Shashikant Patil	SVKMs NMiMS Mumbai, India
S. s. Prabhu	International Institute of Information Technology, India
Ankita Pramanik	Indian Institute of Engineering Science and Technology, Shibpur, India
Priya Ranjan	Amity University, India
Andrews Samraj	Mahendra Engineering College, India
Poonam Singh	National Institute of Technology, Rourkela, India
Ciza Thomas	College of Engineering Trivandrum, India
Snehashish Chakraverty	National Institute of Technology Rourkela, India
Ratnadeep Deshmukh	Babasaheb Ambedkar Marathwada University, India
Surya Durbha	I. I. T. Bombay, India

Shriram K. Vasudevan	Amrita University, India
Prasanta Panigrahi	IISER Kolkata, India
Alok Mittal	AICTE, New Delhi, India
Kalman Palagyi	University of Szeged, Hungary
Jozsef Vasarhelyi	University of Miskolc, Hungary
Dimitris Ampeliotis	University of Patras, Greece
Katerina Kabassi	TEI of the Ionian Islands, Greece
Georgios Drakopoulos	MDAKM Lab, Greece
Sotiris Karachontzitis	University of Patras, Greece
George Tambouratzis	Institute for Language and Speech Processing, Greece
Michael Vrahatis	University of Patras, Greece
Ilka Miloucheva	Media Applications Research, Germany
Rolf Würtz	Ruhr University of Bochum, Germany
Sebastian Feld	LMU Munich, Germany
Munir Georges	Intel, Germany
Larbi Boubchir	University of Paris 8, France
Mohammed Nabil El Korso	Paris 10 University, France
Salah Bourennane	Ecole Centrale Marseille, France
Lotfi Chaari	University of Toulouse, France
Paul Honeine	University of Rouen, France
Kester Quist-Aphetsi	University of Brest, France
Patrick Siarry	University of Paris XII, France
Maki Habib	The American University in Cairo, Egypt
Ivo Bukovsky	Czech Technical University in Prague, Czech Republic
George Dekoulis	Aerospace Engineering Institute, Cyprus
Marina Ivasic-Kos	University of Rijeka, Croatia
Camilo Franco	Andes University, Colombia
Herbert Rojas	Universidad Distrital Francisco Jose de Caldas, Colombia
Santhana Krishnan Boopalan	University of Waterloo, Canada
Nizar Bouguila	Concordia University, Canada
Stefka Fidanova	Institute of Information and Communication Technologies, Bulgaria
Minh-Son Dao	Universiti Teknologi Brunei, Brunei Darussalam
Silvana Costa	Instituto Federal de Educação, Ciência e Tecnologia da Paraíba, Brazil
Marcio Eisencraft	Escola Politécnica da Universidade de São Paulo, Brazil
Felipe Henriques	Celso Suckow da Foseca Federal Center of Technological Education - CEFET/RJ, Brazil
Lisandro Lovisolo	Universidade do Estado do Rio de Janeiro, Brazil
Otavio Teixeira	Universidade Federal do Para (UFPA), Brazil
Joao Paulo da Costa	University of Brasília, Brazil
Aratã Saraiva	UESPI, Brazil
Suryakanthi Tangirala	University of Botswana, Botswana

Geert Verdoolaege	Ghent University, Belgium
Dimitri Papadimitriou	University of Antwerp, Belgium
Federico Menna	EIT Digital, Belgium
José Luis Hernandez Ramos	European Commission, Belgium
Mohammad Nasir Uddin	American International University-Bangladesh, Bangladesh
Peter Balazs	Austrian Academy of Sciences, Austria
Edwin Lughofer	University of Linz, Austria
Ali Rafiei	University of Technology Sydney, Australia
Keyurkumar Patel	Australian Defence Force, Australia
Steven Guan	Xian Jiatong-Liverpool University, Australia
Ouarda Assas	University of Batna 2, Algeria
Moussa Diaf	Université Mouloud Mammri, Algeria
Samir Ladaci	National Polytechnic School of Constantine, Algeria
Belal Alshaqaqi	University of Science and Technology of Oran, Algeria
Zahéra Mekkioui	University of Tlemcen, Algeria
Fatiha Merazka	USTHB University, Algeria
Lahcène Mitiche	University of Djelfa, Algeria

Additional Reviewers

Alireza Ghasempour	ICT Faculty, USA
Amanpreet Kaur	The NORTHCAP University, India
Ambrish Kumar	Netaji Subhas Institute of Technology, Delhi, India
Amel Serrat	USTO MB, Algeria
Anjali Sharma	National Physical Laboratory, India
Anupama S.	Visvesvaraya Technological University, India
Anurag Yadav	CDOT, India
Archana Nawandhar	CMRIT, VTU, India
Arihant Jain	Jaipur Engineering College and Research Centre, India
Ashutosh Dubey	Trinity Institute of Technology and Research Bhopal, India
B. Borah	Tezpur University, India
B. Tripathy	VIT University, India
Baljit Singh	Universiti Teknologi MARA, Malaysia
Bharathi Pilar	Mangalore University, India
Bhaskar Belavadi	SJB Institute of Technology, India
Bhupendra Fataniya	Sarkhej Gandhinagar Highway, India
Boon Chong Ang	Intel, Malaysia
Byju C.	CDAC, India
Carlo Galuzzi	TU Delft, The Netherlands
Ch. Raghava Prasad	KL University, India
Charu Gandhi	Jaypee Institute of Information Technology, India
Debanjan Konar	Sikkim Manipal Institute of Technology, India
Deepak Balram	National Taipei University of Technology, Taiwan
Duc-Lam Nguyen	Seoul National University, Korea

Eric Simon	University of Lille, France
Galandaru Swalaganata	Institut Agama Islam Negeri Tulungagung, Indonesia
Ganga Holi	VTU, India
Gustavo Fernández Domínguez	AIT Austrian Institute of Technology, Austria
Hari C. V.	Rajagiri School of Engineering and Technology, Kochi, India
Ibrahim Missaoui	National Engineering School of Tunis, Tunisia
Imranul Chowdhury	Pedrollo nk Limited, Bangladesh
Jagadeesh Gunda	Faraday Grid Ltd., UK
Jens Ahrens	Chalmers University of Technology, Sweden
Jitender Grover	International Institute of Information Technology, India
Joseph Mathew	C-DAC, Thiruvananthapuram, India
Josip Music	University of Split, Croatia
Kesavaraja D.	Dr. Sivanthi Aditanar College of Engineering, India
Kharuna Chenglerayen	Ministry of Technology, Communication and Innovation, Mauritius
Kothandapani Sudhakar	Ganadipathy Tulsi's Jain Engineering College, India
Kriti Saroha	CDAC, India
Lacrimioara Grama	Technical University of Cluj-Napoca, Romania
Mahesh Babu Jayaraman	Ericsson, India
Manuj Aggarwal	NSIT, Delhi, India
Md. Rabiul Islam	Nanyang Technological University, Singapore
Megha Khanna	Delhi Technological University, India
Meghdad Aynehband	Mahshahr Branch, Islamic Azad University, Mahshahr, Iran
Merry K. P.	Gujarat Technological University, India
Michel Owayjan	American University of Science & Technology, Lebanon
Muharrem Tümçakır	Aselsan Inc. Defense Systems Technologies Division, Turkey
P. Pablo Garrido Abenza	Miguel Hernandez University, Spain
P. Ram Sukumar	ISRO, India
Padma Prasada	VTU Belgaum, India
Panneerselvam Ramesh Kumar	V. R. Siddhartha Engineering College, India
Partha Mangipudi	Amity University, Noida, India
Peter Noel	TSMC, Canada
Poming Lee	NCTU, Taiwan
Pournami N.	NIT Calicut, India
Pradeep Tiwari	Bharat Electronics Ltd., India
Prasanth Vaidya Sanivarapu	Gayatri Vidya Parishad College of Engineering, Visakhapatnam, India
Prathap C.	SIT, Tumakuru, India
Preecha Somwang	Rajamangala University of Technology Isan, Thailand
Prem Sankar C.	University of Kerala, India

Prema V.	RV College of Engineering, India
R. Prasad Kodaypak	AT&T Labs, USA
Rakesh Shirsath	University of Pune, India
Ram Ratan	SAG, DRDO, India
Ram Shukla	OPJIT, Raigarh, India
Ramkumar Jaganathan	VLB Janakiammal College of Arts and Science, India
Rana Khudhair Ahmed	Al-Rafidain University College, Iraq
Ravindra Bhat	NMIMS, India
Roberto Carlos Herrera Lara	National Polytechnic School, Ecuador
S. Agrawal	Delhi Technological University (DTU), India
Sameerchand Pudaruth	University of Mauritius, Mauritius
Sandeep Kakde	Y C College of Engineering, India
Sangeetha Rajesh	K. J. Somaiya Institute of Management Studies and Research, India
Sanjay Sahay	BITS, Pilani, India
Sapan Mankad	Nirma University, Ahmedabad, India
Sasmita Parida	C. V. Raman College Of Engineering BPUT, India
Seera Dileep Raju	Indian Institute of Technology Hyderabad, India
Seng Hansun	Universitas Multimedia Nusantara, Indonesia
Shankar Setty	K. L. E. Technological University, Hubballi, India
Sharada Chougule	Finolex Academy of Management abd Technology, Ratnagiri, India
Sheeba Kumari	VTU, Bangalore, India
Shrinivas Desai	B. V. B. College of Engineering and Technology, India
Shyamalendu Kandar	Indian Institute of Engineering Science and Technology, Shibpur, India
Sibghatullah Khan	Babasaheb Naik College of Engineering PUSAD, India
Siddhartha Bhattacharyya	RCC Institute of Information Technology, India
Smain Femmam	University UHA, France
Sreekumar K.	College of Engineering Poonjar, India
Sunil Semwal	Graphic Era Deemed University, India
Surya Uduga	Acharya Nagarjuna University, India
Swapna R.	Amrita Vishwa Vidyapeetham, India
Tameem Ahmad	Aligarh Muslim University, Aligarh, India
Tarun Yadav	Defence Research and Development Organisation, Ministry of Defence, GOI, India
V. Sowmya	Amrita Vishwavidyapeetham, India
Vaishali Maheshkar	CDAC, India
Vikram Seshasai (Principal Engineer)	India
Vinaya Gohokar	Maharashtra Institute of Technology, India
Wan Hussain Wan Ishak	Universiti Utara Malaysia, Malaysia

Organized by

Contents

Intelligent Recognition Techniques and Applications

Intelligent Recognition Techniques and Applications

Deep Learning Locally Trained Wildlife Sensing in Real Acoustic Wetland Environment

Clement Duhart(✉), Gershon Dublon, Brian Mayton, and Joseph Paradiso

Responsive Environment Group, MIT Media Lab, Cambridge, USA
{duhart,gershon,bmayton,joep}@mit.de

Abstract. We describe 'Tidzam', an application of deep learning that leverages a dense, multimodal sensor network installed at a large wetland restoration performed at Tidmarsh, a 600-acre former industrial-scale cranberry farm in Southern Massachusetts. Wildlife acoustic monitoring is a crucial metric during post-restoration evaluation of the processes, as well as a challenge in such a noisy outdoor environment. This article presents the entire Tidzam system, which has been designed in order to identify in real-time the ambient sounds of weather conditions as well as sonic events such as insects, small animals and local bird species from microphones deployed on the site. This experiment provides insight on the usage of deep learning technology in a real deployment. The originality of this work concerns the system's ability to construct its own database from local audio sampling under the supervision of human visitors and bird experts.

Keywords: Wildlife acoustic identification · Signal processing
Deep learning · Wetland environment

1 Introduction

In an era of increasingly ubiquitous sensing, Paradiso et al. [PP16] discuss how it is now possible to document natural ecosystems and record ecological change over longer periods and at significantly higher resolutions than ever before. These new capabilities are of particular interest to restoration scientists and practitioners, who seek to create conditions for complex ecosystems to flourish where human activity previously eradicated them. The restoration context offers researchers the unique opportunity to embed sensor networks directly into new ecosystems as they form, and the resulting data gives the public a chance to learn about ecological functions and environmental impacts. With these dual goals in mind, we developed comprehensive network and sensing infrastructure on a wetland restoration site in southern Massachusetts, called Tidmarsh. Once a 600+ acre industrial cranberry farm, Mass Audubon's recently opened Tidmarsh Wildlife Sanctuary and the Town of Plymouth's Foothill Preserve now host the largest freshwater wetland restoration in New England. We deployed a large number of

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 3–14, 2019.
https://doi.org/10.1007/978-981-13-5758-9_1

custom-designed wireless sensor devices, microphones, and cameras on the site to monitor its transition from industrial farm to protected wetland. The data from our sensors are recorded and streamed in real-time for use in scientific studies, as well as for new immersive experiences for the remote and visiting public. Those experiences include both traditional web applications and augmented reality tools for landscape exploration. A significant challenge we face in this work is in the automated analysis and classification of our data, particularly of the streaming audio and video. To process the audio, we developed a system called *Tidzam* that analyzes large numbers of live streams, recognizes ambient acoustic scenes, and identifies the sources of transient sonic events from an array of wildlife (including dozens of bird species, frogs, and insects), vehicles, and visitors. Recognizing the enormous potential for visitors to submit audio from their mobile devices, our system can also flexibly process temporary streams. In both cases, the resultant classifications are made available to end-user applications in real time. This paper presents the end-to-end Tidzam acoustic wildlife sensing system, its technical underpinnings, and its novel applications to both environmental science and public outreach. A review of the broader vision, as well as details of the sensing and environmental interaction projects, can be found in [MM18]. The locations of the different sensors, microphones and the network infrastructure can be found in Fig. 1.

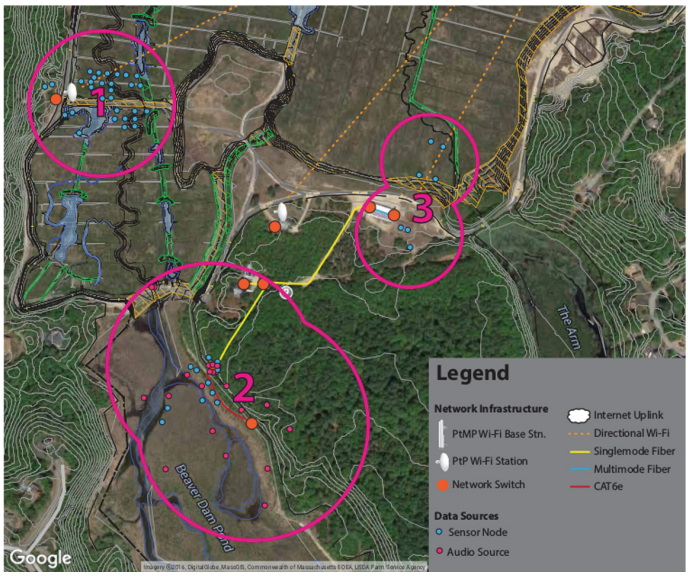


Fig. 1. Locations of the deployed sensors and microphones on Tidmarsh site.

A unique challenge of our acoustic classification task in the restoration context is the appearance and disappearance of numerous different sound sources

both seasonally and over the long environmental recovery period. As a result, Tidzam is scalable to new classification tasks to accommodate newly resident and migrating wildlife; the system is able to detect when a new kind of sound appears and builds new classifiers as needed. A second distinguishing feature of the Tidzam system is our focus on real-time processing for online use with minimal latency. Finally, in addition to identifying bird calls and other wildlife, it has also been designed to record/playback requested samples for expert scientists to review/use.

Automated Recorded Systems (ARS) are crucial tools for wildlife monitoring, estimating bird species abundance and diversity as discussed in Celis et al. [CMDA09]. However, while ARS have evolved from manually triggered to time scheduling systems as presented in Acevedo et al. [AVR06], automated stream segmentation and real-time signal identification remain difficult in noisy and unpredictable outdoor environments. Even in a wildlife sanctuary, human activity appears across the spectrum, with noise caused by cars, aircraft, visitors, and abutters. In real deployment, a wildlife identification system must be coupled with an acoustic scene analysis system.

Acoustic Scene Classification (ASC) is an active research area that has seen significant recent advancement with Deep Learning algorithms. Li et al. [LDM+17] compare such approaches with classical ones, and conclude that temporal specialized models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNN) produce better results than resolution-specialized models like Gaussian Mixture Model (GMayton) and i-vector. A combination of both model types also improves the results by some percent points. A number of researchers are focused on efficient deep learning architectures that are robust to noise or improve diversity of learned features, such as Han et al. [HP17] and Xu et al. [XHW+17]. Those improvements allow expert classifiers for bird calls to be cascaded in realistic outdoor deployments.

Acoustic Bird Identification (ABI) systems in outdoor deployments face two main challenges: first, calls from same and different species frequently overlap in time, and second, many species sound similar to one another in addition to have multiple calls. Kojima et al. [KSH+17] propose a source separation approach which allows the classifier to be processed on independent sources for each possible target bird. This promising approach requires multiple microphones, and we intend to experiment with it in future work. The issue of large output classes is investigated by Hershey et al. [HCE+17], who propose a bottleneck model similar to a de-noising auto-encoder. Their approach forces the system to learn efficient kernels of features in CNNs, and the feature compression may increase diversity and robustness. Cakir et al. [CAP+17, ADCV17] present a Convolutional Neural Networks (RCNN) approach, which allows the classifier to learn on both the acoustic texture and its evolution over time, with possible long term dependencies.

Our system has a deep learning expert architecture combining an ASC with an ABI. Section 2 presents the overall architecture as applied to our real-time classification task on 24 live microphone streams and visitor-submitted audio at

Tidmarsh, as well as the sample extraction strategy we developed to build our database. Section 3 presents an evaluation of classifier accuracy, which progressively improves with sample extraction during database construction. Section 4 discusses current limitations, diverse applications, and planned future work, leading into our concluding remarks in Sect. 5.

2 Tidzam

The Tidzam architecture is composed of several processing pipelines presented in Fig. 2. A stream manager receives the different input audio streams produced by the fixed microphones onsite (illustrated in Fig. 3) or from visitor-submitted mobile audio streams. To reduce bandwidth requirements, the microphone streams are compressed in software running on an embedded server in the marsh; by bundling the channels together into a single Opus-encoded stream, that application also maintains sample-accurate synchronization. The Tidzam server decodes the multichannel stream and sends individual channels to the classifier, which in turn sends its analysis to remote clients and a database server for logging. To grow the sample database, a rules-based engine extracts samples automatically according to classifier confidence. Samples are automatically forwarded to a web interface used by bird identification experts to label unknown samples and further refine the classifier for the next cycle.

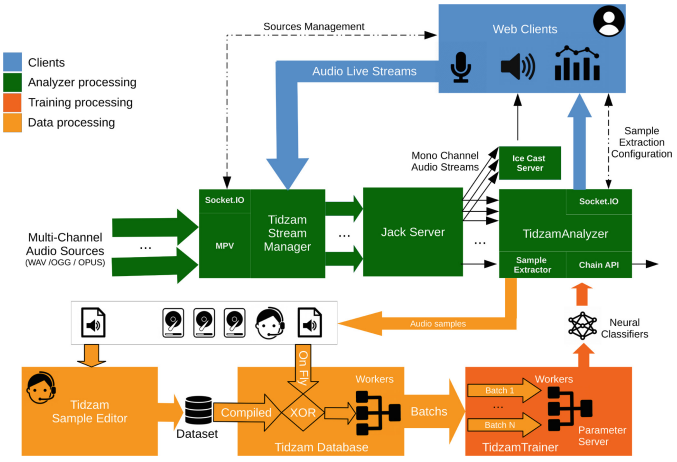


Fig. 2. Overview of Tidzam architecture with the three main pipelines. The green modules represent the classification flows of the different microphone streams loaded by the StreamManager, dispatched by the JACK server to the classifiers in the Analyzer. The results are transmitted to subscribed data consumers in the blue module, which also configure the input streams. The orange modules represent the processing chain for the classifier training, from extraction of a poorly identified sample through the upgrade of the classifier. (Color figure online)

2.1 Audio Live Capture Infrastructure

On Site. A set of 24 microphones, specially designed for harsh environmental conditions, were deployed at Tidmarsh (Fig. 3). The audio channels are synchronized and compressed before their transmission to the Tidzam server. Additional audio channels are collected from the streaming wildlife cameras.

Fig. 3. Fixed audio capture at Tidmarsh: 24 microphones are connected to a 32-channel mixer and audio interface, whose output is compressed into a Ogg/Opus stream.

Data Sampling. By default, input audio streams are split into segments of 500 ms with a half-overlapping window. A Mel-Filter Banks (MFB) spectrogram is computed on each sample, followed by a median filter for background noise reduction in the presence of brief sonic events. A 50 Hz to 12 KHz bandpass filter is then applied to select the frequency range of the major local bird vocalizations. Finally, samples are normalized for the classifier. Some result examples can be observed in Fig. 4. Our choice of a simple preprocessing step over the state of the art reflects a trade-off between the real-time constraints and available compute resources.

2.2 Convolutional Expert Based Classifier

The classifier is a Convolutional Neural Network (CNN) with an expert architecture, as illustrated in Fig. 5. The classifier layer is separated into an Acoustic Scene Classification (ASC) and an Acoustic Bird Identification (ABI). The ASC computes a general classification that weights the inhibition of the ABI according to its probabilistic estimation of bird presence. The final probabilities are fused by multiplying ABI outputs with the ASC bird class probability. The classifiers

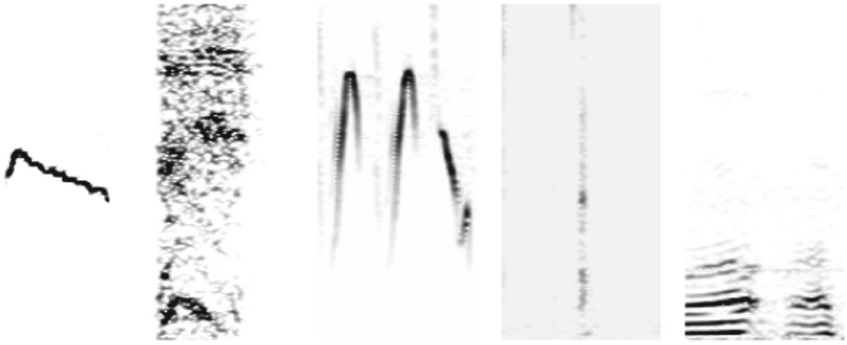


Fig. 4. Illustration of preprocessed sample spectra before their transmission to the classifier: (a) Blue Jay (b) American Crow (c) Fox Sparrow (d) Frog (e) Human Voice.

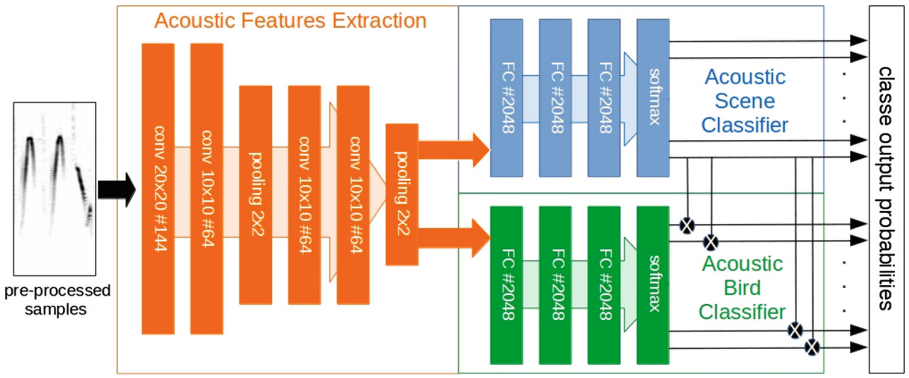


Fig. 5. Neural expert architecture

share the same stack of convolutional layers responsible for the acoustic feature learning. The cost function is a regular multi-class cross-entropy with parameter regularization as defined in Eq. 1. The ASC was trained on 15 different acoustic scene samples (e.g. rain, crickets, aircraft) as well as brief generic sonic events for inhibition control (e.g. human voice, frog, birds, etc). The ABI has been trained on 50 local bird species.

$$C = -\sum_{i} y'_i \log y_i + (1 - y'_i) \log (1 - y_i) + \gamma \sum_{k} ||w_k||_2^2 \text{ with } \gamma \in]0, 1] \qquad (1)$$

Eq. 1: Cross-entropy cost function used during classifier training with L2 regularization on the weights w. y_i and y'_i are the neural network response probability and the expected value for the class i, respectively.

Note: The design of our expert architecture is a result of early experiments, which demonstrated the difficulty of learning both ambient sounds and punctual sonic

events in the same classifier. We observed ambient sounds stimulating a large number of neurons, which can overwhelm the subset of activations produced by smaller patterns of punctual sonic events. In the expert architecture, the separation of the classification task into two independent layers allows the ABI classifier to retain details in the information flow for bird species identification without being overwhelmed by ASC sensitivity.

2.3 Semi-supervised Database Generation

The design of a relevant ASC and ABI database is strongly dependent on the site-specific acoustic scene and the local bird species. Even if a list of bird species is provided by experts, the specific acoustic environment of Tidmarsh must be learned to isolate wildlife sounds from rustling tree leaves, rain, wind, etc. In addition, new species appear over time and throughout the year, requiring constant classifier updates. Semi-supervised database generation from a combination of local ambient sound and off-site samples provides a flexible framework for iterative classifier updates. Based on a confidence rule, the classifier extracts current local samples in order to refine its database. If a sound is labeled as unknown, it is brought to human attention, or can become a candidate for a database augmentation if found to be misclassified as such.

Classifier Confidence Function. The confidence function F defined in Eq. 2 evaluates the reliability of the classifier response for the current sample. Based on the comparison between the confidence value and two threshold hyperparameters $d_u, d_a \in [0; 1]$, the system can decide to extract the sample for a human consideration if lower than d_u (unidentified sample), or for database augmentation (with a preset label) if lower than d_a.

Note: This confidence function tends to extract samples which do not produce enough sparse responses between the output classes. In practice, the confidence thresholds start with low values, which are manually and progressively increased according to the classifier versions. As the diversity of samples in the database increases, the output classifier probabilities tend to become more saturated.

$$F(y_1, ..., y_n) = b_1 - \sum_{j=2}^{n} b_j \tag{2}$$

Eq. 2: Confidence function evaluating how distant the predicted class output b_1 is from the others, where b_j is the j^{th} largest value of the classifier outputs y_i.

Extraction Rules. The extraction rules produce samples for both *human consideration* and *database augmentation*, according to the confidence function and the distribution of samples among the classes as defined in Eq. 3. The well-identified samples are used for the next classifier training (after validation by

expert operator), whereas unidentified samples are extracted for manual labeling. Unidentified samples are considered as a class, so their extraction probability depends on the number of samples awaiting human consideration.

Note: At the same time, the system favors samples for classes that are not well represented in the database, regulating the flow of samples for manual labeling.

$$P(s_i) = \lambda_1(1 - \lambda_2^i \frac{|D^i|}{\max_{j \neq i} |D^j|}) \text{ with } s \in D, \lambda_1, \lambda_2 \in [0; 1[\quad (3)$$

Eq. 3: Extraction probability for a sample s of class i according to the database composition D. λ_1 is a flow control parameter used to smooth the number of extracted samples over time, whereas λ_2 controls a probabilistic margin of extraction dependent on class maturity in the database.

2.4 Novel Class Bootstrapping and Training

The classifier is periodically retrained from scratch according to the database augmentation. The training and testing sets, respectively 80% and 20% of the database, are composed of downloaded samples from online databases, mixed with samples extracted on the Tidmarsh site. The bootstrapping of a new class is based on downloaded or manually extracted samples for a rough pretraining and setup of the confidence function. At the first database update, the sample diversity is low, which in turn favors very similar samples by the confidence function. At the same time that the classifier is trained with similar samples (which tends to saturate its output probability), the confidence function threshold is reduced in order to increase the diversity of possible extracted samples. Hence the system extracts samples increasingly diverse from its initial bootstrapped sample references.

Note: After running for one year, Tidzam has generated a database of approximately 300,000 samples, composed of a set of acoustic scenes such as rain noise, water flowing, aircraft, etc, and a set of bird calls from local species. Bootstrapping a new class requires at least 500 samples, which are selected to cover the largest sound diversity for each species' vocalizations or each acoustic scene. The class is then progressively augmented with extracted samples cross-validated with human oversight.

3 Evaluation

The following experimentation has been conducted in order to evaluate whether the self-extraction of samples directly from the acoustic environment can produce effective sample databases. Over one year, a new classifier is trained from scratch each month on a new dataset composed of the previous dataset version augmented with the collected samples of the current month.

3.1 Datasets

During the database construction, new classes were added over one year. For an objective comparison in terms of classifier improvements due to database augmentation, this study has been reduced from 75 to 25 classes (10 acoustic scenes and 15 bird species) that were present in the first version of the database. The samples were added in order of their timestamps of extraction and in order to conserve the balance between the different classes of each training and testing dataset versions.

3.2 Experimental Results

The curves in Fig. 6 illustrate the classifier improvements on the testing datasets through the sequence of database updates each two months. The curve V0 is the initial classifier trained on the bootstrap dataset composed of samples from the online Cornell database and some audio recordings from the acoustic scene of Tidmarsh e.g. quiet sound, rain, wind, etc. Each classifier version has been tested on each database version in order to evaluate the improvement resulting from the addition of the newly extracted samples. Parallel trainings are operated in order to get the best classifier depending of the regularization hyper-parameter γ. For a given database version, usually with a success rate greater than 93%. It can be observed that each generation of the classifier outperforms the previous one. Indeed each classifier version gets lowest scores on following augmented

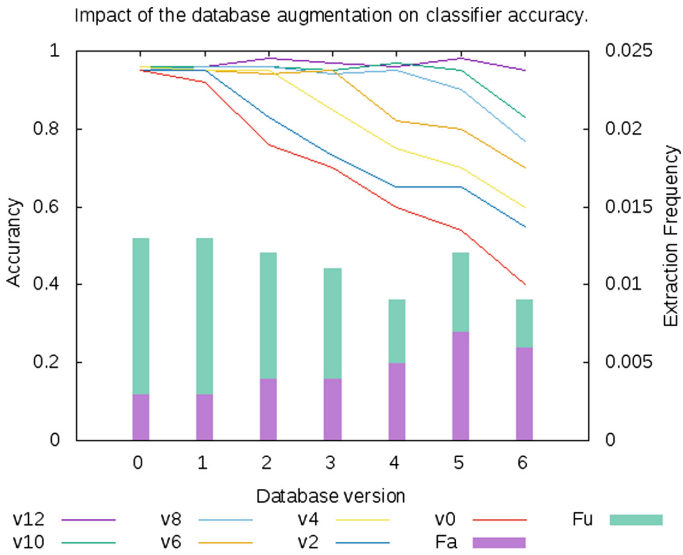


Fig. 6. Evaluation of the impact of extraction strategy on classifier accuracy. Curves represent the accuracy, on the test dataset, of the different upgraded versions of the classifier according to the iterations of database augmentation each two months. Bars represent the extraction frequencies of samples.

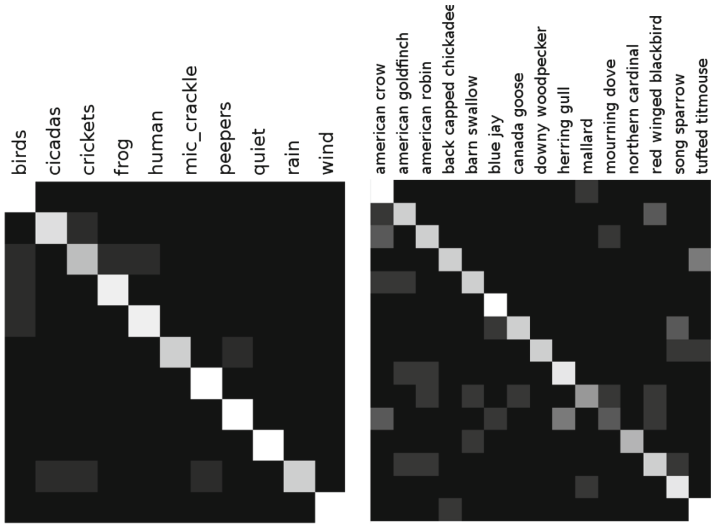


Fig. 7. Confusion matrices of the ASC and ABI classifiers on version V12 of the testing dataset with $\gamma = 0.01$. Main errors occur in the bird species classification.

databases. Hence the database augmentation by sample extraction improves the inference generalization. The histograms represent the sample extraction frequencies F_u for unidentified samples and F_a for database augmentation. At the beginning, the system tends to extract mainly the unidentified samples, but after several generations of training, this proportion starts to invert. The average frequency of extraction is maintained using the flow control parameter of λ_1. The Fig. 7 presents the confusion matrices of the ASC and ABI of the last classifier version on the testing dataset.

Note: The inversion of extraction frequency from unidentified samples to data augmentation results in an improvement of the confidence function, potentially due to an improvement in sound diversity of the database. The system has learned more different kinds of bird calls, but still needs to improve its sensitivity according to the extraction rules in the confidence function.

4 Limitations, Discussions and Applications

As Tidzam is an online learning system which builds its database over time, it is not yet possible to draw firm conclusions about the choices made regarding the neural network architecture for this application. However, we can observe the generic sample extraction strategy constantly improving the database by examining the classifier accuracy over time. This improvement has been achieved thanks to the confidence function, which allows the system to automatically extract samples cross-validated by a human using the Tidzam web interface.

In the interest of expediency, the web application was designed to be as simple as possible, consisting of an automatic sound player, spectrogram visualization, and validation buttons. Given our 24/7 monitoring, validation can be extremely time consuming, with as many as 700 cross-validation samples per day. In the future, this web interface will be replaced by a more engaging, gamified interface, in which both remote experts and Tidmarsh visitors would be able to learn more about the wildlife sanctuary. At the same time, they would be able to correct misclassifications and aid in the labeling of new samples. We see this closed learning loop as beneficial both to amateur enthusiasts interested in learning about bird calls and to experts investigating animal behavior. Crowd-sourcing the validation and labeling process would help Tidzam refine its database.

This approach is in line with the goals of the overall Tidmarsh project, in which immersive technologies are used to augment the visitor experience (see Mayton et al. [MM18]). In one example, called HearThere, a custom-designed smart headphone allows users to hear a dynamic spatial rendering of real-time or recorded sound from all the microphones on the landscape as they walk through it. Output from Tidzam is used to adjust the mix of microphones to promote channels where interesting wildlife has been observed, and suppress channels carrying undesirable sound, such as wind or the voices of other visitors. In another example, called DoppleMarsh, a 3d model of the Tidmarsh terrain is used as the basis of a virtual reality (VR) world driven by the sensing on the physical site. Users can navigate the site remotely in VR, and Tidzam classifications are used to render virtual wildlife and other scenic dimensions. However Tidzam is not able to localize precisely the acoustic sources which would be a very interesting assets in terms of rendering for DoppleMarsh as well as ecological studies. In future works, geo-localization based on source separation on the microphone array will be investigated after an acoustic propagation study on site. This will lead a location update/new deployment of microphones on Tidmarsh.

5 Conclusion

This contribution presents the technique used by the Tidzam project for acoustic scene analysis, wildlife detection, and bird species identification in the outdoor acoustic environment of a wetland. Based on 24 microphones deployed on site, a classifier expert architecture based on deep learning techniques is used to analyze in real-time multiple audio streams. The system is able to automatically extract the samples in which its confidence is too low so they can be identified by human experts on a web interface. Then the classifier is refined, incorporating the new samples. The preliminary results of this contribution are promising - along 12 training iterations, the system performance has significantly improved. The system has now been building its own database of 300,000 samples over the past year. It is currently used as a wildlife tracker in a large wetland restoration project.

Additional Information: Tidzam is running on two NVIDIA Titan X GPUs and available online at http://tidzam.media.mit.edu/. All source code can be downloaded from https://github.com/mitmedialab/tidzam.

Acknowledgement. The authors would like to acknowledge Living Observatory and the Mass Audubon Tidmarsh Wildlife Sanctuary for the opportunity to realize the audio deployment at this location. The NVIDIA GPU Grant Program has provided the two TITAN X which are used by Tidzam. Clement DUHART has been supported by the PRESTIGE Fellowship of Campus France and the Pôle Léonard de Vinci. We also thank the Elements Collaborative and the sponsors of the MIT Media Lab for their support of this work.

References

[ADCV17] Adavanne, S., Drossos, K., Cakir, E., Virtanen, T.: Stacked convolutional and recurrent neural networks for bird audio detection. In: 25th European Signal Processing Conference (EUSIPCO), pp. 1729–1733, August 2017

[AVR06] Acevedo, M.A., Villanueva-Rivera, L.J.: From the field: using automated digital recording systems as effective tools for the monitoring of birds and amphibians. Wildlife Soc. Bull. **34**(1), 211–214 (2006)

[CAP+17] Cakir, E., Adavanne, S., Parascandolo, G., Drossos, K., Virtanen, T.: Convolutional recurrent neural networks for bird audio detection. In: 25th European Signal Processing Conference, EUSIPCO 2017, Kos, Greece, 28 August–2 September 2017, pp. 1744–1748 (2017)

[CMDA09] Celis-Murillo, A., Deppe, J.L., Allen, M.F.: Using soundscape recordings to estimate bird species abundance, richness, and composition. J. Field Ornithol. **80**(1), 64–78 (2009)

[HCE+17] Hershey, S., et al.: CNN architectures for large-scale audio classification. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 131–135 (2017)

[HP17] Han, Y., Park, J.: Convolutional neural networks with binaural representations and background subtraction for acoustic scene classification. In: Proceedings of the Detection and Classification of Acoustic Scenes and Events 2017 Workshop (DCASE 2017), November 2017

[KSH+17] Kojima, R., Sugiyama, O., Hoshiba, K., Nakadai, K., Suzuki, R., Taylor, C.E.: Bird song scene analysis using a spatial-cue-based probabilistic model. J. Robot. Mechatron. (JRM) **29**, 236–246 (2017)

[LDM+17] Li, J., Dai, W., Metze, F., Qu, S., Das, S.: A comparison of deep learning methods for environmental sound detection. In: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 126–130, March 2017

[MM18] Mayton, B., et al.: Networked sensory landscape: capturing and experiencing ecological change across scales. To appear in Presence (2018)

[PP16] Paradiso, J.: Our extended sensoria - how humans will connect with the internet of things. Next Step Exponential Life Open Mind Collect. **1**(1), 47–75 (2016)

[XHW+17] Xu, Y., et al.: Unsupervised feature learning based on deep models for environmental audio tagging. IEEE/ACM Trans. Audio Speech Lang. Process. **25**(6), 1230–1241 (2017)

Classification of Plants Using GIST and LBP Score Level Fusion

Pradip Salve[1(✉)], Milind Sardesai[2], and P. Yannawar[1]

[1] Vision and Intelligence Lab, Department of Computer Science and IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India
pradipslv@gmail.com, pravinyannawar@gmail.com

[2] Floristic Research Lab, Department of Botany, Savitribai Phule Pune University, Pune, India
sardesaimm@gmail.com

Abstract. Plant Leaf retains many characteristics that can be used to achieve automatic plant classification, Leaf also contains other traits like shape, size, color, texture but these are not enough to distinguish one plant species from another. Extraction of prominent leaf features involves enhancement and feature normalization. The leaf venation may explore as a promising feature due to its correlation similarity between intra-classes. Recently numerous plant classification systems have been proposed. Most of the plant recognition systems rely on single feature but automatic plant leaf classification system that uses single feature often faces several limitations in terms of accuracy. The limitations of system can be overcome by building multimodal plant classification systems that fabricates multiple features together using feature level fusion as well as score level fusion. This paper presents score level fusion of LBP (Local Binary Patterns) and GIST (Global descriptors) features towards building more robust automatic plant classification system. The results shows that, the score level fusion has contributed towards efficient plant classification with the 87.22% genuine accept rate (GAR) for GIST features, LBP features with 78.39% GAR and GIST + LBP scores 89.23% of GAR were observed.

Keywords: Leaf vein · Plant recognition · Multimodal plant classification

1 Introduction

Plants are significant part of our ecosystem. Plants present everywhere and they used in our day to day life in various form viz. in medicines and pharmaceuticals, furniture industries similarly plant also provides oils, rubber, timber, dyes etc. for human use. Life would not be possible without plants. The biodiversity studies particularly plant taxonomy has contributed significantly towards identification and marking plant on digital medium, however many plants yet to be recognized. Many plants are considered a Critically Endangered. For the sake of environment protection plant must be preserved digitally for recognition purpose. Automatic plant classification may help to monitor plants. Numerous attempts have been made by initiating interdisciplinary study with association of computer and plant taxonomist to propose automatic plant

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 15–29, 2019.
https://doi.org/10.1007/978-981-13-5758-9_2

classification system. Traditionally, plants can be also distinguished by other traits like *flowers, buds, color, roots* etc. Although many of them are depend on CBIR (Content Based Image Retrieval) using textual and shape features of leaf. Dhondt et al. [1] developed an online framework, designated Leaf Image Analysis Interface (LIMANI), in which venation patterns are automatically segmented and measured on dark-field images. Leaf venation segmented by semi-automatically by providing user interface, allowing supervision and rectification steps in the automated image analysis pipeline and ensuring high-fidelity analysis. The framework was used to study vascular differentiation during leaf development and to analyse the venation pattern in transgenic lines with contrasting cellular and leaf size traits. The results presented in [1] shows the evolution of vascular traits during leaf development and suggested a self-organizing map mechanism for leaf venation patterning, and reveal a tight balance between the number of end-points and branching points within the leaf vascular network that does not depend on the leaf developmental stage and cellular content, but on the leaf position on the rosette. Chai et al. [2] proposed structurally-coherent solution contrary to the pixel-based and object-based methods as a resultant to the graph in which each node represents either a connection or an ending in the line-network. Based on stochastic geometry, [2] have developed a new family of point processes consisting in sampling junction-points in the input image by using a Monte Carlo mechanism. The quality of a configuration is produced by a probability density which takes into account both image consistency and shape priors.

APIS (Advanced Plant Identification System) is a plant identification system developed by Rankothge et al. [3] was primarily based on image processing and neural network techniques. In [3], image processing methods were used to de-noise the leaves, its normalization, cleaning white background and to scale the leaf image. 2D-FFT technique is used in APIS system to provide rotational invariant to the input image. If APIS neural network returns more than one result image, APIS system will go further and Color (RGB values), width, height matching will be done. Finally concluded APIS achieved an accuracy of 95% and it gives results within few minutes. An automatic initialization of vein search parameter using Standard Hough Transform was proposed by Rahmadhani et al. [4], shape feature extraction with Hough transform and Fourier descriptor were implemented. Effectiveness on leaf shape recognition were compared Recall-precision evaluation showed that leaf shape represented by Fourier descriptors is more effective than represented by Hough transform. Plant leaf image vein extraction using b–spline representation was implemented. Sun et al. [5] have presented a method for extracting leaf veins and contour from point cloud data, curvature information of the point cloud were used to extract the leaf venation structure followed by the mesh model of plant leaf was constructed and the leaf contour was also extracted using mesh algorithm. Clarke et al. [6] scale space analysis and an automated technique that includes a combination of edge detectors are compared with a manual technique. A small data set of five images is considered in this initial exploratory work and the results are qualitatively evaluated [6] authors have adopted two methods to analyse venation patterns on fresh leaves of *Ivy, Monstera, Nettle* and *Ribes* and dried leaves of *Ochna*. These were (i) an application of an algorithm that includes a scales pace approach (ii) an application of a simple smoothing and edge detection combination algorithm. Both methods were compared with a manually edited image of the

venation pattern and the results of the three techniques were presented independently for evaluation to few botanists scale space analysis and region growing were used to extract blood vessels by Martnez-Prez et al. [6], similar approach have been used by [7]. Leaf venation was extracted by Valliammal et al. [8] by combination of a thresholding and H-maxima transformation method. An intensity histogram was used for initial segmentation and thresholding was used to determine leaf venation. Followed by a fine segmentation using H maxima transformation based method for object pixel as its inputs. Salve et al. [9] used a HOG (*Histogram of Gradients*) and ZM (*Zernike moments*) as a features. The experiments done on *VISLeaf* dataset. The features were passed to Euclidean distance classifier for recognition purpose. Deng et al. [10] have done the experiments on weeds and crop for detection of the weeds in the crop. Authors have used wavelength as a character to distinguish crop and weed using Principal Component analysis as a classifier. Anjomshoae and Rahim [11] focused on a template-based method to classify rubber tree. Key points have been extracted and used as features using Scale invariant feature transform (SIFT). De Souzaa et al. [12] adopted silhouette measure based methods like the simulated annealing, differential evolution and particle swarm optimization methods, to achieve the set of optimal parameters for leaf shape description.

2 Proposed Methodology

The *VISleaf* dataset was used in this research work and all the experiments were tested on the same dataset during the study. The dataset was constituted with 60 plant species collected from botanical garden of Dr. B. A. M. University, Aurangabad Campus. Ten leaf (10) samples from each plant have been collected at the time for database formation. The scientific nomenclature was verified from experts experienced taxonomist during this study, the system performance is verified using False Acceptance Rate (FAR), False Rejection Rate (FRR), Equal Error Rate (ERR) and Genuine Accept Rate (GAR) evaluation criteria's. The process flow of the proposed system is shown in the Fig. 1.

List of Plants with Scientific name used in the experimental work:

Acalypha indica L, Aegle marmelos (L.) Corrêa, Anacardium occidentale L, Antigonon leptopus Hook. Arn., Argemone mexicana L, Azadirachta indica A.Juss, Barleria prionitis L, Bauhinia racemosa Lam, Boerhavia diffusa L, Bougainvillea spectabilis Willd, Butea monosperma (Lam.) Taub, Caesalpinia bonduc (L.) Roxb, Calotropis procera (Aiton) Dryand, Cassia siamea Lam, Cinnamomum verum J.Presl, Cissus repanda Vahl, Cissus repens Lam, Citrus aurantiifolia (Christm.) Swingle, Coccinia grandis (L.) Voigt, Cocculus hirsutus (L) W Theob, Corchorus olitorius L, Euphorbia geniculata Ortega, Euphorbia hirta L, Ficus benghalensis L, Ficus virens Aiton, Gliricidia sepium (Jacq.) Walp, Grewia hirsuta Vahl, Hamelia patens Jacq, Hibiscus rosa-sinensis L, Hiptage benghalensis (L.) Kurz, Holoptelea integrifolia Planch, Ipomoea nil (L.) Roth, Ipomoea pes-tigridis L, Jasminum nudiflorum Lindl, Jatropha integerrima Jacq, Kigelia africana (Lam.) Benth, Lantana camara L, Leonotis nepetifolia (L.) R.Br, Mangifera indica L, Mimusops elengi L, Morinda pubescens Sm, Nerium oleander L, Parthenium hysterophorus L, Peristrophe bicalyculata (Retz.) Nees, Pithecellobium

dulce (Roxb.) Benth, Plectranthus mollis (Aiton) Spreng, Polyalthia longifolia (Sonn.) Thwaites, Punica granatum L, Senna tora (L.) Roxb, Spathodea campanulata P.Beauv, Syzygium cumini (L.) Skeels, Tamarindus indica L, Tecoma stans (L.) Juss. ex Kunth, Tephrosia villosa (L.) Pers, Terminalia bellirica (Gaertn.) Roxb, Tinospora sinensis (Lour.) Merr, Trigonella foenum-graecum L, Ventilago maderaspatana Gaertn, Withania somnifera (L.) Dunal, Ziziphus jujuba Mill.

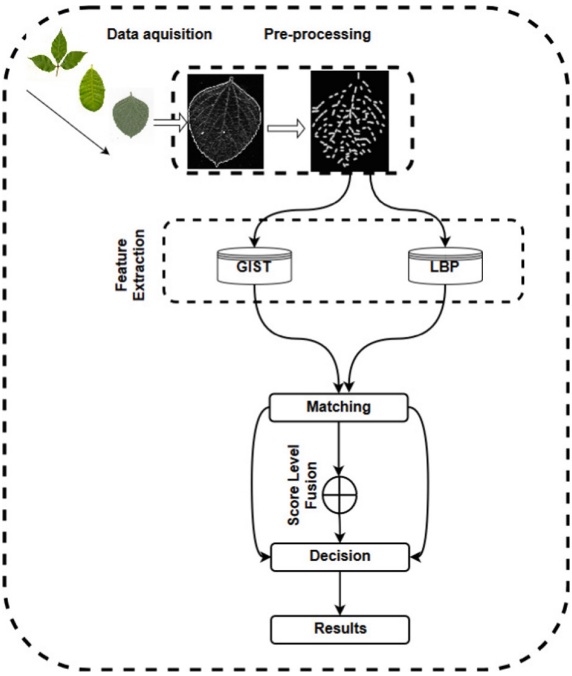


Fig. 1. Block diagram of proposed methodology.

Fig. 2. Sample leaf from *VISLeaf* dataset

The classification system begins with input color image. The color image is composed of three channels i.e. R + G + B. The green channel (G) were extracted due to its color dominancy leaf images. Furthermore, gradient of an image by using Eq. (1) were calculated (Fig. 2).

$$\nabla f = \begin{bmatrix} gx \\ gy \end{bmatrix} = \begin{bmatrix} \frac{\partial f}{\partial x} \\ \frac{\delta f}{\delta y} \end{bmatrix} \tag{1}$$

Where $\frac{\partial f}{\partial x}$ is the derivative with respect to x (gradient in the x direction) and $\frac{\delta f}{\delta y}$ is the derivative with respect to y (gradient in the y direction). $\frac{\partial f}{\partial x}$ is calculated by convolving the $\begin{bmatrix} -1 \\ 0 \\ +1 \end{bmatrix}$ filter onto the image.

$$\frac{\partial f}{\partial x} = \begin{bmatrix} -1 \\ 0 \\ +1 \end{bmatrix} * G \tag{2}$$

Where G is the green channel. The result of the gradient image is as shown in Fig. 3

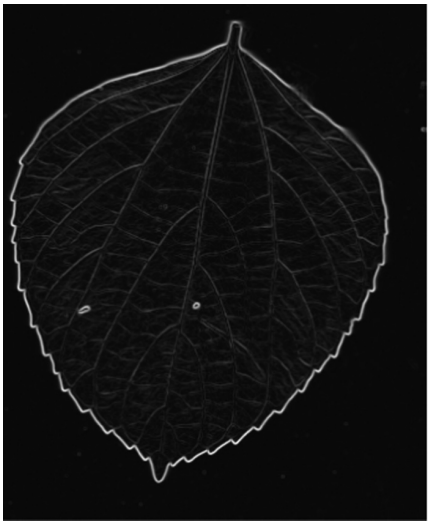


Fig. 3. Sample of gradient image

In order to extract venation of leaf, Otsu thresholding method were adopted for representation of venation structure more precisely and clearly. The Otsu threshold was

computed by Eq. (3). Figure 4, shows the sample image of venation extracted from gradient image.

$$\sigma_{\omega}^{2}(t) = \omega_0(t)\sigma_0^2(t) + \omega_1(t)\sigma_1^2(t) \tag{3}$$

Where ω_1 and ω_2 are the probabilities of classes separated by threshold t.

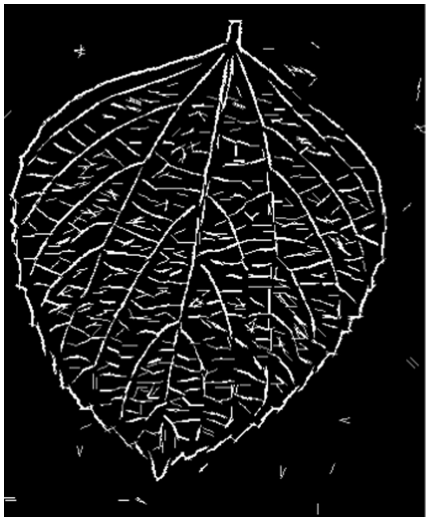


Fig. 4. Sample image of leaf venation pattern.

3 Feature Extraction

3.1 Gist Features

GIST summarizes the gradient information (scales and orientations) for different parts of an image. GIST uses gradient information of different level of scale and orientation from different part of leaf images. After the pre-processing step, the input leaf image convolved by 32 Gabor filters at 4 different scales and 8 type of orientation. Resultant 32 features map then each feature map was divided into 16 (4 * 4) grids further averaging the feature value. Afterwards, concatenating the 16 averaged value of all 32 feature maps produces 512 (16 * 32) GIST descriptor features which can be used in further process of classification. In order to extract GIST features input leaf images were filtered into gray scale and orientation channels. Centre-surround operation within each channel ***i*** were performed between filter output maps, **O*i*(s)**, at different scales 's'. This produces feature maps M*i*(c, s), given centre scale 'c' and surround scale 's'. Feature, maps were used to detect conspicuous regions in gray scale channel and were concatenated to yield a saliency map. Information from the orientation channel is incorporated by employing Gabor filters to the grayscale image using Eq. (3) at four spatial centre scales (c = 0, 1, 2, 3) and at four different angles ($\boldsymbol{\theta_i}$ = 0°, 45°, 90°, 135°) [13] (Fig. 5).

Orientation channels, feature maps are computed using Eq. 4:

$$M_i(c) = Gabor(\theta_i, c) \tag{4}$$

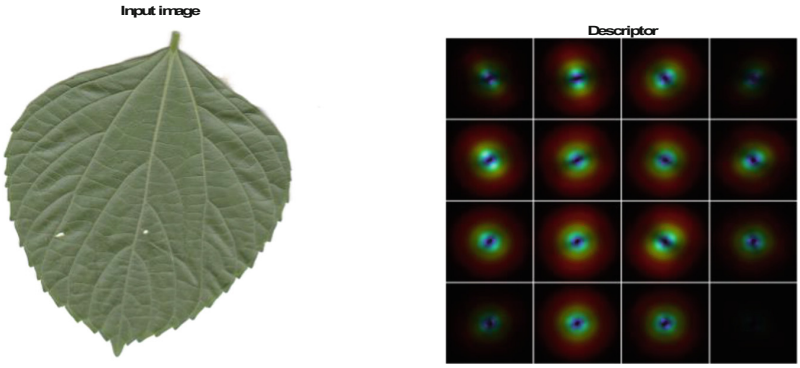


Fig. 5. Original image and visualization of its orientation channel descriptors used to compute feature maps fabricated using Eq. (4).

After the computation of low-level centre-surround features, each sub-channel extracts a gist vector from its corresponding feature map. Averaging operations are applied in a 4 × 4 grid of sub regions over the map. Sixteen raw gist features $G_i^{k,1}(c,s)$ are computed per feature map using Eq. 5:

$$G_i^{k,1}(c,s) = \frac{1}{16WH} \sum_{u=\frac{kW}{4}}^{\frac{(k+1)w}{4}-1} \sum_{v=\frac{lH}{4}}^{\frac{(i+1)H}{4}-1} [M_i(c,s)](u,v) \tag{5}$$

Where, W and H are width and height of the image respectively. k and l are the indices in horizontal and vertical directions respectively. Finally, extracted gist features were stored on disk (Fig. 6).

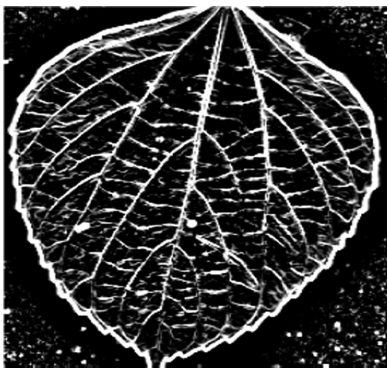


Fig. 6. LBP image.

3.2 LBP Feature Extraction

The LBP are the local image descriptors which captures local texture information, initially proposed by Ojala et al. [14], LBP transform image into integer labels or an array of integers which describes small scale appearance of an image. The statistics of integer labels can be represent as a histograms, these histograms further can be used as a texture descriptor. Local binary patterns uses 8 pixels from the block of 3 * 3 pixels remaining 1 is the centre pixel of the block. Initially the LBP operators convolves on image and finds the centre pixel value of 3 * 3 pixel block of an image. The value of centre pixel then used as a threshold. The $s(x)$ is the thresholding function shown in Eq. 6.

$$s(x) = 1, \; if \; x \geq 0, \; \text{otherwise} \; s(x) = 0 \tag{6}$$

Where $\boldsymbol{x}$ is the neighbourhood pixel of centre pixel from the 3 * 3 block. The generic LBP operator can be obtained from the concatenation of result from $\boldsymbol{s(x)}$. The LBP operator can be expand to utilize different size of neighbourhoods by creating the circle of radius R from the centre pixel. To acquire neighbourhood pixels the sampling points P on the perimeter of the circle is added. The centre pixel then compared with the points on the perimeter of the circle to get the threshold value. It can be derived from the differences of threshold adding the weight to the power of two. The LBP operator shown in Eq. 7 below.

$$LBP_{P,R} = \sum_{p=0}^{P-1} s(g_p - g_c)2^p \tag{7}$$

$$g_c = I(x, y) \tag{8}$$

Where $I(x, y)$ is gray image and g_c indicate the gray level of an arbitrary pixel (x, y).

$$g_p = I\left(x_p, y_p\right), p = 0 \ldots p - 1 \tag{9}$$

$$x_p = x + Rcos(\frac{2\pi p}{p}) \tag{10}$$

$$y_p = y + Rsin(\frac{2\pi p}{p}) \tag{11}$$

Where, g_p in Eq. (9) indicate the gray value of a sampling point in an evenly spaced circular neighbourhood of P sampling points and radius R around point (x, y). In this research work LBP operator was applied on each leaf sample where each image was divided into 8 * 8 and 16 * 16 of blocks. The final LBP feature vector was calculated by combining the histogram of each block (Figs. 7 and 8).

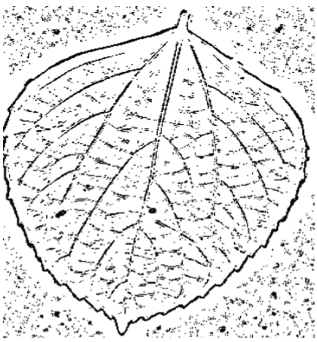


Fig. 7. LBP image obtained by circle of Radius 8

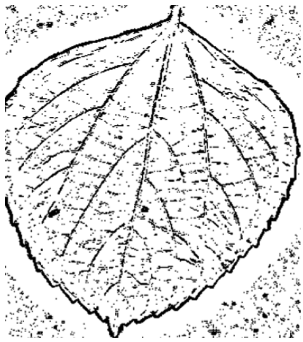


Fig. 8. LBP image obtained by circle of Radius 16

4 Matching Criteria

In this section the feature vectors from each feature extraction technique was received to perform the classification. The feature vectors was divided in to training set and testing set in the ratio of 70%:30% respectively. The size of training set was 420 samples and 180 samples for testing from different plant classes. To obtain the score matrix the Euclidean distance measure was used and calculated by using the Eq. 12.

$$S(i,j) = \sqrt{\sum_{i=1}^{n} (q_i - p_j)^2} \tag{12}$$

The score matrix was generated and used for taking decision, this process is useful in determining class of test sample. To generate this the threshold value for each leaf sample we have driven the steps to generate threshold value from [15] which can summarize as the following:

$$\Delta = \frac{\max(FV_i) - \min(FV_j)}{\beta} \tag{13}$$

Where $\boldsymbol{\beta}$ is a pre-determined constant which is used to divide threshold value into N parts.

$$\theta i = \min(FV) + \Delta i \tag{14}$$

$\theta i (i = 1, 2, \ldots, N)$ is selected when the value approaches the False Rejection Rate (FRR) or False Acceptance Rate (FAR) value is very small depending on the specifications required.

4.1 Decision Criteria

Finally the decision has been made from Threshold value ***(T)***. System accepts the identity of the leaf sample ***(if DS $\leq$ T)*** or rejects ***(if DS > T)***. Where DS is the distance score between samples.

4.2 Evaluation Criteria

The evaluation has been carried out using False acceptance rate (FAR), False rejection Rate (FRR), Equal error rate (EER), and Genuine Acceptance Rate (GAR) parameters.

False Acceptance Rate (FAR) is the percentage of unknown samples accepted by identification system. FAR can be calculated using Eq. (15)

$$FAR = \frac{the\ score\ of\ inter\ class\ samples(Impostor)\ >\ threshold}{number\ of\ all\ impostor\ score} \times 100 \tag{15}$$

FRR is the percentage of known samples rejected by the identification system. The FRR was calculated using Eq. (16)

$$FRR = \frac{score\ of\ inter\ class\ samples(Genuine) < threshold}{number\ of\ all\ genuine\ score} \times 100 \tag{16}$$

ERR is the intersection point plotted on the graph where FAR and GAR intersects each other. The EER is try to measure the FAR performance against another FRR (The lowest the ERR the highest the accuracy) by finding the equal (the nearest to equal or have the min distance in case if not equal) between False Acceptance Rate (FAR) and False Rejected Rate. ERR can be calculated using Eq. (17)

$$ERR = \frac{FAR + FRR}{2}. \tag{17}$$

Genuine Acceptance Rate (GAR):

This is defined as a percentage of genuine leaf samples accepted by the system. GAR have been calculated using Eq. (18):

$$GAR = 100 - FRR \tag{18}$$

4.3 Score Level Fusion of Leaf Images

The features produced from each technique i.e. GIST and LBP delivered to matching process subsequently it produces scores for individual method, the scores from each technique was produced with the help of threshold values discussed in decision criteria Sect. 4.1. A single score was obtained from fusion of GIST + LBP scores. For the score level fusion, sum rule was adopted by using Eq. 19 in this study.

$$score_{(i,j)} = \sum_{f=1}^{2} score_{GIST(fi)}, score_{LBP(fi)} \tag{19}$$

Where, $score_{(i,j)}$ retains the sum of scores from GIST and LBP.

5 Experimental Results

The three types of experiments have been done in this study. The GIST and LBP techniques used separately and the score level fusion of GIST + LBP was carried out separately. All the experiments was carried out on *VISLeaf* dataset. The dataset was divided into training and testing set as 70% and 30% respectively.

5.1 Results of GIST Features

The features were extracted from GIST descriptor for each leaf sample. GIST produces 1 * 512(16 * 32) size of feature vector by concatenating the 16 averaged value of all 32 feature maps. These feature vectors for all 60 plant species were extracted and stored as feature matrix for further process of matching. Afterward, the score matching have been done and score matrix was produced for all leaf images in the dataset. Further the intra-class (same plant species or known sample i.e. Genuine in terms of biometric) and inter-class (from different plant species i.e. Impostor in terms of biometric) were generated. Then system performance measures FAR, FRR, EER and GAR was calculated with the help of Threshold value ***T*** using genuine and impostors.

The results obtained from GIST descriptor gave EER 12.77% and with GAR reached to 87.22%. The Table 1 shows the results of GIST descriptor and Fig. 9(a) represents the ROC curve of system performance with EER. Figure 9(b) depict the ROC curve of performance of system GAR against FAR with rate of Threshold value T.

Table 1. Result of GIST features.

FAR	Threshold (T)	FRR (%)	EER (%)	GAR (%)
19.7619	0.6173	6.820016	13.29096	86.70904
17.06349	0.6585	9.039548	13.05152	86.94848
14.04762	**0.6996**	**11.49314**	**12.77038**	**87.22962**
12.61905	0.7408	14.41216	13.5156	86.4844
10.71429	0.7820	17.49395	14.10412	85.89588

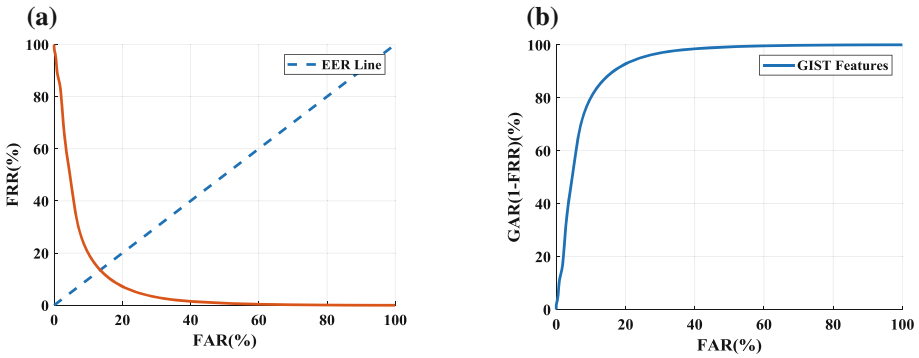


Fig. 9. (a) ROC curve of GIST features (b) GAR Performance of the system.

5.2 Results of LBP Features

The textual information was extracted using LBP technique for each leaf sample of dataset. The extracted feature vectors for all 60 plant species were computed and stored as feature matrix for further process of score matching. Afterward, the score matching have been done and score matrix was obtained for all leaf samples in the dataset. Further, the intra-class and inter-class were generated. Then by using genuine, impostors and threshold ***T*** system performance measures FAR, FRR, EER and GAR were calculated.

The results gained from LBP texture descriptor produces EER 21.60% and with GAR reached to 78.39%. The Table 2 shows the results of LBP descriptor and Fig. 10 (a) represents the ROC curve of system performance with EER. Figure 10(b) represents the ROC curve of performance of system GAR against FAR with rate of Threshold value T.

Table 2. Result of LBP features

FAR	Threshold (T)	FRR (%)	EER (%)	GAR (%)
36.98413	0.1356	16.91552	26.94983	73.05017
31.19048	0.1582	20.93758	26.06403	73.93597
26.19048	**0.1808**	**25.0148**	**21.60264**	**78.39736**
22.06349	0.2035	29.34759	25.70554	74.29446
18.88889	0.2261	33.61851	26.2537	73.7463

5.3 Score Level Fusion of GIST and LBP Features

The scores obtained from GIST and LBP have been fused together by sum rule to achieve single score vector. The resultant score vectors was used to achieve best recognition performance of the plant classification system. The system performance measures like FAR, FRR, EER and GAR have been obtained by applying threshold

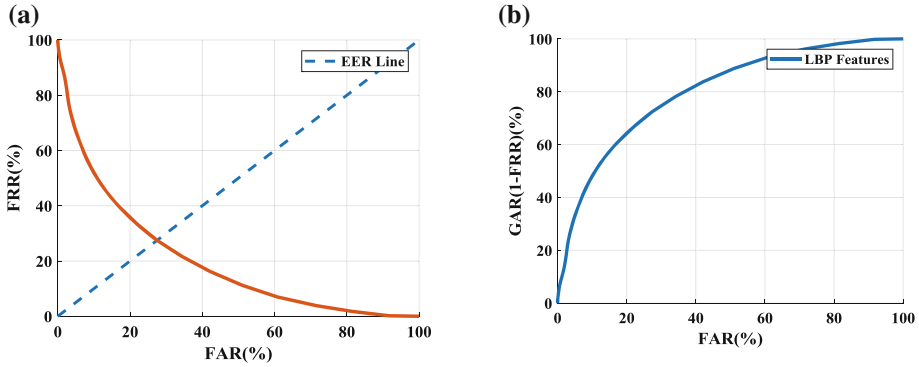


Fig. 10. (a) ROC curve of LBP features. (b) GAR Performance of the system.

value. The results shows that the score level fusion of GIST and LBP features produced best efficiency with minimizing EER to 10.76% and increasing t GAR to 89.23% with threshold value T to 0.6996. Table 3 shows the result of score fusion on GIST + LBP features. Figure 11(a) depict the ROC curve with respect to EER. Figure 11(b) represents the GAR Performance of the system using score level fusion of GIST + LBP features.

Table 3. The result of score fusion technique produced from GIST + LBP features together.

FAR	Threshold (T)	FRR (%)	EER (%)	GAR (%)
26.7619	0.6173	1.084019	13.92296	86.07704
21.06349	0.6585	1.336858	11.20017	88.79983
18.04762	**0.6996**	**3.482378**	**10.765**	**89.235**
16.61905	0.7408	6.386602	11.50282	88.49718
14.71429	0.7820	9.48184	12.09806	87.90194

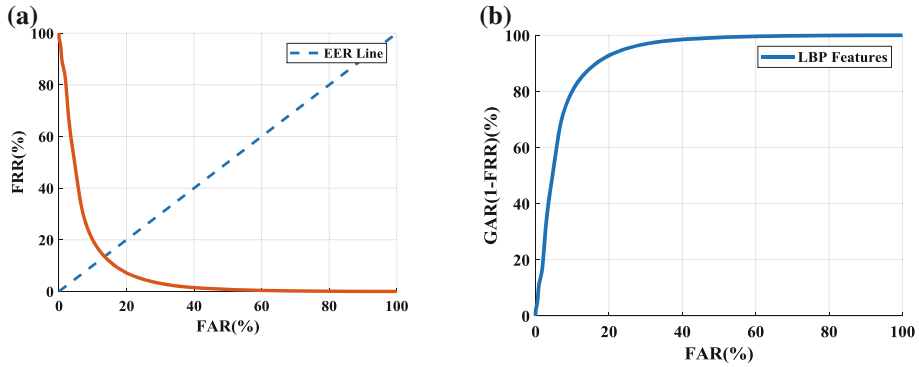


Fig. 11. (a) ROC curve of score level fusion of GIST + LBP features (b) GAR Performance of the system using score level fusion of GIST + LBP features

5.4 Overall Comparison of System Performance

The comparison of three modalities are shown in this section. The highest genuine acceptance rate using GIST, LBP and combined score of both are shown in Table 4.

Table 4. The overall performance of the proposed 3 modalities

	FAR	Threshold (T)	FRR (%)	EER (%)	GAR (%)
GIST	**14.04762**	**0.6996**	**11.49314**	**12.77038**	**87.22962**
LBP	**26.19048**	**0.1808**	**25.0148**	**21.60264**	**78.39736**
GIST + LBP	**18.04762**	**0.6996**	**3.482378**	**10.765**	**89.235**

6 Conclusion

In this piece of research work, The Multimodal plant classification system by using individual GIST, LBP features and fused GIST + LBP feature classification method was explored, and performance of classification were evaluated. All experiments have been carried out on *VISLeaf* dataset Experiments have been done to validate the accuracy of system. The results for single modal and multimodal recognition have been validated using ROC curves and EER values. The dataset is divided into 70:30 ratio for training and testing purpose, respectively. The sum rule based score level fusion were adopted to achieve multimodal plant classification. The system optimal performance were obtained on threshold ***T be 0.6996*** on genuine and impostor scores. The results shows that the highest accuracy using GIST descriptor is 87.22% GAR with minimum EER of 12.77% while threshold value is 0.6996. Similarly system produced highest accuracy with LBP features is 78.39% GAR when threshold value set to 0.1808. Finally, the result of score level fusion of GIST + LBP shows the increased in accuracy with 89.23% GAR when threshold value set to 0.6996. It is observed that the fusion of scores improves the GAR with 2.01%. The work may be extended in variety of ways towards exploring classification of medicinal, wild, cultivated plants.

Acknowledgment. Authors would like to acknowledge UGC-MANF fellowship for financial support and technical supports of GIS & Remote sensing Lab of Department of Computer Science & IT, Dr. Babasaheb Ambedkar Marathwada University, Aurangabad, Maharashtra, India.

References

1. Dhondt, S., et al.: Quantitative analysis of venation patterns of Arabidopsis leaves by supervised image analysis. Plant J. **69**(3), 553–563 (2012)
2. Chai, D., Forstner, W., Lafarge, F.: Recovering line-networks in images by junction-point processes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1894–1901 (2013)

3. Rankothge, W.H., Dissanayake, D.M.S.B., Gunathilaka, U.V.K.T., Gunarathna, S.A.C.M., Mudalige, C.M., Thilakumara, R.P.: Plant recognition system based on Neural Networks. In: 2013 International Conference on Advances in Technology and Engineering (ICATE), pp. 1–4. IEEE (2013)
4. Rahmadhani, M., Herdiyeni, Y.: Shape and vein extraction on plant leaf images using Fourier and B-spline modeling. In: AFITA International Conference, The Quality Information for Competitive Agricultural Based Production System and Commerce, pp. 306–310 (2010)
5. Sun, Z., Lu, S., Guo, X., Tian, Y.: Leaf vein and contour extraction from point cloud data. In: 2011 International Conference on Virtual Reality and Visualization (ICVRV), pp. 11–16. IEEE (2011)
6. Clarke, J., et al.: Venation pattern analysis of leaf images. In: Bebis, G., et al. (eds.) ISVC 2006. LNCS, vol. 4292, pp. 427–436. Springer, Heidelberg (2006). https://doi.org/10.1007/11919629_44
7. Ford, G.E., Estes, R.R., Chen, H.: Space scale analysis for image sampling and interpolation. In: 1992 IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP-92, vol. 3, pp. 165–168. IEEE (1992)
8. Valliammal, N., Geethalakshmi, S.N.: Hybrid image segmentation algorithm for leaf recognition and characterization. In: 2011 International Conference on Process Automation, Control and Computing (PACC), pp. 1–6. IEEE (2011)
9. Salve, P., Sardesai, M., Manza, R., Yannawar, P.: Identification of the plants based on leaf shape descriptors. In: Satapathy, S.C., Raju, K.S., Mandal, J.K., Bhateja, V. (eds.) Proceedings of the Second International Conference on Computer and Communication Technologies. AISC, vol. 379, pp. 85–101. Springer, New Delhi (2016). https://doi.org/10.1007/978-81-322-2517-1_10
10. Deng, W., Huang, Y., Zhao, C., Chen, L., Wang, X.: Bayesian discriminant analysis of plant leaf hyperspectral reflectance for identification of weeds from cabbages. Afr. J. Agric. Res. **11**, 551–562 (2016)
11. Anjomshoae, S.T., Rahim, M.S.M.: Enhancement of template-based method for overlapping rubber tree leaf identification. Comput. Electron. Agric. **122**, 176–184 (2016)
12. De Souzaa, M.M.S., Medeiros, F.N.S., Ramalho, G.L.B., de Paula Jr, I.C., Oliveira, I.N.S.: Evolutionary optimization of a multiscale descriptor for leaf shape analysis. Expert Syst. Appl. **63**, 375–385 (2016)
13. Oliva, A., Torralba, A.: Building the gist of a scene: the role of global image features in recognition. Prog. Brain Res. **155**, 23–36 (2006)
14. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognit. **29**(1), 51–59 (1996)
15. Ali, M.M.H., Yannawar, P.L., Gaikwad, A.T.: Multi-algorithm of palmprint recognition system based on fusion of local binary pattern and two-dimensional locality preserving projection. Procedia Comput. Sci. **115**, 482–492 (2017)

Real-Time Poultry Health Identification Using IoT Test Setup, Optimization and Results

Arun Gnana Raj Alex[1(✉)] and Gnana Jayanthi Joseph[2(✉)]

[1] Department of Computer Science, Bharathiyar University, Coimbatore, Tamilnadu, India
arun.art06@gmail.com
[2] Department of Computer Science, Rajah Serfoji Government College, Thanjavur 613005, Tamilnadu, India
jgnanamtcy@gmail.com

Abstract. Poultry industry needs a system to automate the process of identifying the hen is infected or not. We have proposed a system that uses IoT and sensors to analyze and identify the infected hen. This reduces the cost of labor and increase the accuracy of the identification process. In this paper we discuss about the overall system, audio and video analysis methods and comparing the results using Matlab. The process of sick identification has been optimized using the Matlab results.

Keywords: IoT · Poultry automation · Sick detection · Sound analysis

1 Introduction

Detecting a sick hen in early stage is very important to avoid the rick of discarding all the infected hen. Common diseases like Infectious Bronchitis, Avian Influenza, Infectious Sinusitis, Fowl Pox and Infectious Coryza are affecting the poultry industry [1]. Early stage sick detection also eliminates the spread of disease and to avoid consuming the sick hen. There are various methods can be used to identify the sick for example temperature of a hen, food eating pattern, moving pattern, checking weight of the hen, sound analysis of the hen. There were very few research has been done in the area of poultry to identify the sick [2–5]. There are significant research on traceability of poultry [6–8].

We are proposing two different methods to identify the sick hen. First is using the image analysis including the RGB images (SEN-11745 res 728X488 a RGB camera, with the lens angle 70°) and thermal images (Far Infrared Thermal Sensor Array – AMG8833 Grid-EYE 8X8 RES from Panasonic). AMG8833 can measure temperatures ranging from 0 °C to 80 °C (32 °F to 176 °F) with an accuracy of ±2.5 °C (4.5 °F) and the maximum distance is 7 m. The second (SEN-14262 a Microphone controller) is using the sound of the hen. In this paper we are describing the analysis methods we have done for identifying the infected hen and the performance of the various audio feature extraction and classification methods used for analysis the sick hen. The overview is shown in Fig. 1. The IoT-based real-time poultry monitoring and health status identification has been discussed in our previous research paper [9].

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 30–40, 2019.
https://doi.org/10.1007/978-981-13-5758-9_3

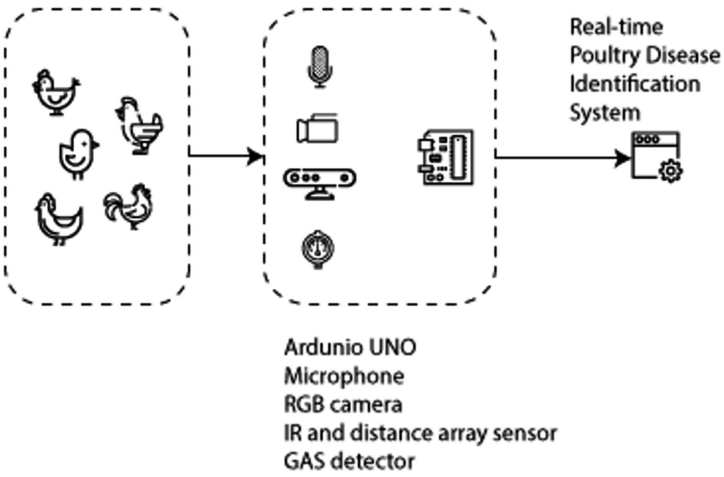


Fig. 1. Proposed IoT enabled poultry disease detection system.

The paper is organized as follows: Sect. 2 discusses about the setup and the dimension of the observable area, sensors tracing ability. The height of the mounted sensors from the ground. The data collection frequency samples has been discussed in the Sect. 3. Image analysis process for the sick identification mapping with image is in the Sect. 4. The Sect. 5 shows the method of audio analysis using the audio feature extraction with KNN machine learning technique. Various performance analysis of extraction techniques has been detailed in Sect. 6. Section 7 concludes the paper with the feature work.

2 Setup Dimensions for Testing

The test environment has been setup using a cardboard fence with the size of 1.25 m × 1.25 m shown in Fig. 2. The total number of hens inside is 4. The sensor module got mounted at the 2 m height. So that we can cover all the 1.56 m^2 area. The thermal array sensor with the narrow field view of 35.6° and can sense the distance up to 7 m. So the thermal array can cover up to 1.7424 m^2. The RGB camera with the view angle of 70°. With the height of 2 m the RGB camera can cover up to 7.84 m^2.

The distance it is mounted known as opposite (opp) = 2 m
To find the area which can be tracked we need to find the adjacent (adj) side value

$$adj = \frac{opp}{\tan\theta}$$

The total width (w) can be covered = 2 * adj
The area (a) it could cover = w^2
AMG8833 - Thermal sensor (TS) - Field view = 36.5°
The θ value for the thermal sensor = 90 – (36.5/2) = 71.75°

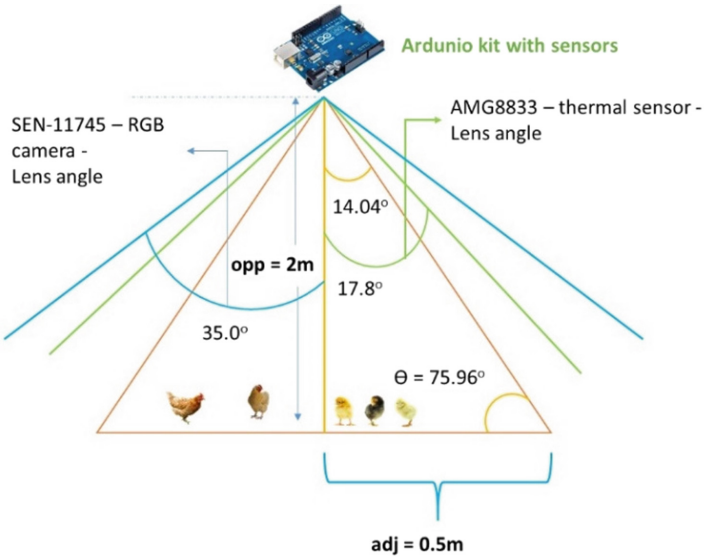


Fig. 2. Proposed IoT enabled poultry disease detection system.

The adjacent value for thermal sensor

$$TSadj = \frac{2}{\tan(71.75)} = \frac{2}{3.03} = 0.66\,\text{m}$$

TS_w = 2 * 0.66 = 1.32 m
Ts_a = 1.7424 m^2
SEN-14262 – RGB camera (RG) - lens angle = 70°
The θ value for the Thermalsensor = 90 – (70/2) = 55.0°
The adjacent value for rgb camera

$$RGadj = \frac{2}{\tan(55.0)} = \frac{2}{1.43} = 1.39\,\text{m}$$

RG_w = 2 * 1.39 = 2.78 m
RG_a = 7.7284 m^2

From the above calculation it has been proved that both sensors can able to track the value 1.56 m^2. The average recommended gauge size for 4 chickens 1.48 m^2.

3 Data Frequency

We have observed the sensor values for 1 h and take the sample of 1 image/second from both rgb camera and the thermal image sensor. So we got sample of 18000 images to analyze. We have collected the audio of the hen in the poultry with the duration of 2

days. In this audio collection there were 2 different test like normal hen with the group and the sick hen with the group. There were around 100000 samples for each category got collected. Each audio sample is about 25 ms. For audio we need to collect more so that we could remove the environment noises. There were audio recorded like human talking, some birds noise etc., needs to be removed. The normalization of removal of environmental sound is can be done manually. The normalization can be automated which are already addressed in various researches [10, 11].

4 Image Analysis

In the image analysis the data from the sensor like RGB (SEN 11745 – 728 × 488) and the thermal sensor value (Far infrared thermal sensor – 8 × 8) needs to be mapped. The mapping can be done in various stages

- Gray scale image conversion
- Denoise using the Gaussian conditional distribution
- Monochrome image conversion
- K-means clustering for image segmentation
- Finding and matching the boundary of RGB and thermal data
- Temperature of the hen can be identified
- Based on the Gi we can calculate the number of the hens (Gn) in a grid. Now can identify the temperature of the hen using the temperature mapping function.

$$f(t, G_n, i) = \begin{cases} t_i, G_n = 1 \\ -1, G_n > 1 \\ -1, G_n = 0 \end{cases}$$

From the thermal sensor the temperature in each pixel will be analyzed. The range of the temperature has been set. If any pixel which is not in the rage has been marked as sick. If the total movement of the hen in a particular time interval is not meeting the defined movement also will be marked as infected. The both results will be correlated to find the infected hen. The infection alert hi will return the Boolean value based on the function gt and gd. The OR operator can be applied to get the result.

$$h_i(g_t, g_d) = \begin{cases} 1, & (g_t + g_d) > 1 \\ 0, & otherwise \end{cases}$$

5 Audio Analysis

The microphone controller sensor (SEN-14262) will gather the sound data and produce the output as 0.01 V to 2.69 V. This sensor can operate on Gain Margin of 12 dB. This signal can be observed from Ardunio and store the digital signal into the SD card. The stored data will be sent to MATLAB for extracting various parameters using the audio extraction techniques. Once the data has been extracted then KNN clustering analysis

used to find the infected hens. Audio signal processing is an efficient way to detecting the infection [12].

The process for audio feature extraction as follows.

- Audio files needs to reduce the noise
- Audio needs to be Normalized
- Audio features extraction technique for feature parameters
- Classification using Machine learning algorithms for identifying normal or sick hen
- The trained classifier test to decide the hen is sick or not using the known audio feature values.

5.1 Study of Mel-Frequency Cepstral Coefficients (MFCC) of Audio Signal

Mel-Frequency Cepstrum Coefficients (MFCCs) have been important to processing the speech recognition. The main point to understand about speech is that the sounds generated by a hen are filtered by vocal size, air flow and pressure. This shape has the characteristics of what sound comes out. MFCCs can accurately represent the shape of the vocal tract manifests the envelope of the short time power spectrum. MFCCs are actually a type of cepstral representation of the signal, where the frequency bands are distributed according to the mel-scale, instead of the linearly spaced approach. The MFCC analysis along with the frequency domain is displayed in Figs. 3 and 4. The audio files used for this analysis has the following parameters

- Time: 10 s
- Codec: WAV PCM
- Sample Rate: 44100 Hz
- Channels: Mono
- Bitdepth: 32

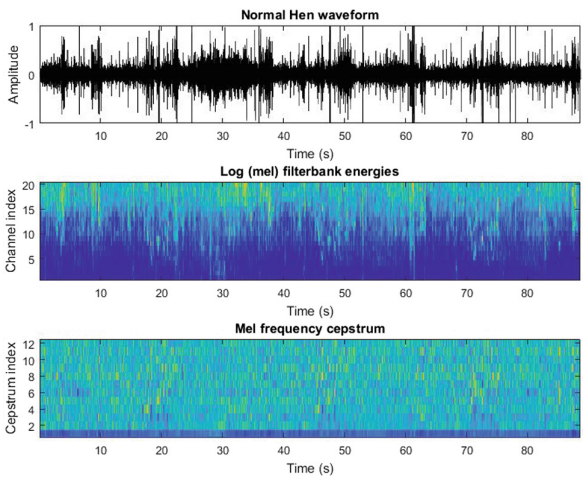


Fig. 3. Audio spectrum of group of hens

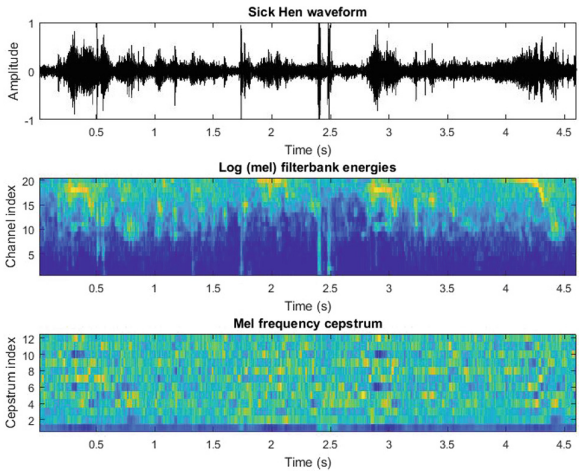


Fig. 4. Audio spectrum of sick hen

5.2 MATLAB Simulation and Results

Matlab has been chosen for testing and developing the proposed algorithm. We have captured the audio from the group of hens. That audio got normalized and saved as a files. These files then sent as an input to the MFCC for extract the features of the audio. The MFCC feature extraction has been done with the following parameters like the analysis frame duration is 25 ms, frame shift of 10 ms, Tw = 25, Ts = 10, alpha = 0.97, M = 20, C = 12, L = 22, LF = 300, HF = 3700. This gives an array of 53 parameters output for each 25 ms audio. The figure, Fig. 4 shows the rainbow chart of the sample signals like normal and the sick hen.

The gathered 200,000 samples are flagged as a normal and sick as 0 and 1. These results has been used as an input for KNN. The KNN trained algorithm has been tested and can classify the sample test audio.

5.3 KNN – Machine Learning for Classification

The classification of sick or not from the MFCC features can be done using is a non-parametric method called k-nearest neighbors algorithm (k-NN). In k-NN classification, an object with the majority vote from neighbors will be marked as classified. If k = 1, then the object is assigned to the class of that single nearest neighbor. The figure Fig. 5 provides the insight of the KNN trained data distribution of 2 variables and figure, Fig. 6 provides the confusion matrix for the KNN performance analysis.

Confusion matrix is a specific table layout that allows visualization of the performance of KNN machine learning classification algorithm for normal and infected

classification [13, 14]. We are doing the supervised learning so for make surer the false positive rates we can use the Confusion matrix. KNN has been trained with the following parameters

- Preset: Fine KNN
- Number of neighbors: 10
- Distance metric: Euclidean
- Distance weight: Equal
- Standardize data: true

The result of the KNN analysis then stored as a MATLAB file. This file later used to test and classify the audio signal.

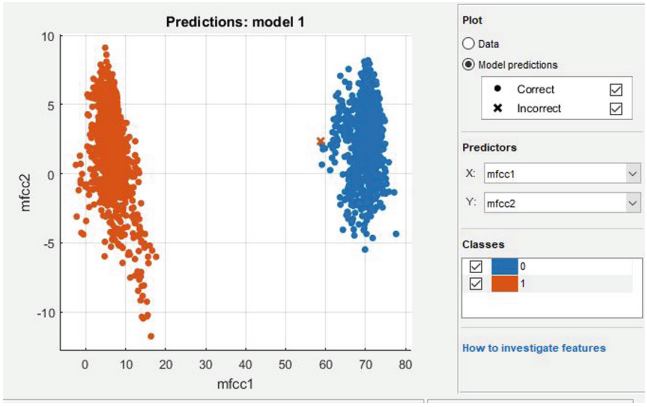


Fig. 5. KNN training

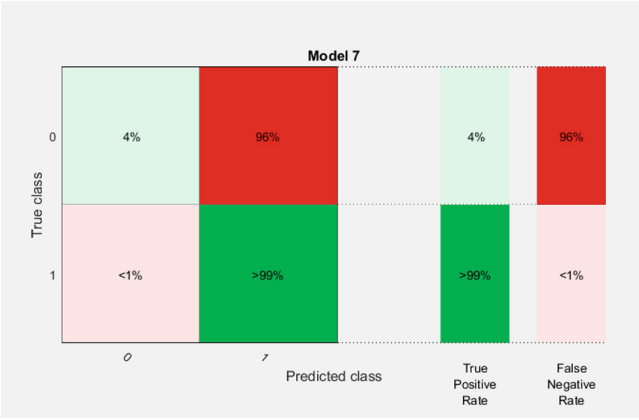


Fig. 6. Confusion matrix

6 Performance Analysis of Extraction Techniques

The process of extracting audio can be done using various techniques. And there are many machine learning languages are there. So we first done an analysis on various feature extraction technique and the machine learning algorithm. Table 1 shows the full list of analysis report of various algorithm against various audio feature extraction methods. We classified the audio feature extraction technique into 3 categories,

- MFCC parameters [13 param]
- Other extraction [20 param - Bias (1), CrestFactor (1), Formant (6), FundamentalFrequency (1), SFF (11)]
- Statics extraction [9 param - Skew, Kurtosis, Slope, Bias, Mean, Min, Range, Std]

All audio feature parameters got extracted. Then these parameters are analyzed using various machine learning techniques. The study has been conducted into the following machine learning algorithms.

- SVM (Coarse Gaussian SVM, Fine Gaussian SVM, Linear SVM, Medium Gaussian SVM, Quadratic SVM)
- KNN (Coarse KNN, Cosine KNN, Cubic KNN, Fine KNN, Medium KNN, Weighted KNN)
- Tree (Complex Tree, Medium Tree, Simple Tree)
- Linear Discriminant
- Logistic regression
- Quadratic Discriminant
- Ensumble (Bagged Trees, Boosted Trees, RUSboosted trees, Subspace Discriminant, Subspace KNN)

The best algorithms has been identified in terms of best prediction accuracy and time. For all these analysis processes, MATLAB 2017a has been used.

The test has been conducted with 100000 normal chicken audio samples and 100000 audio samples with the sick chicken. The first issue in normalization. Once the audio got normalized then the labeling has been done. In the labeling process the identification of various segments like human sound, environmental sounds like bus, car and other hen sounds which comes as a reaction from the hen needs to be marked. Not only that the correct and sick sounds also has been labeled. The process of identifying the other sounds is done by manual process. We can automate the environmental sound identification process using other algorithms and will be done later in the research work. The timelines which has the normal and sick labels are cut from the files and the extracted audio cut has been saved in a separate file. Then these files are sent to the Matlab and the various featured has been extracted.

The table, Table 1 clearly displays all the test result (accuracy % and time in seconds) with various audio feature extraction techniques and the classification studies. From the Table 1 it's very clear that the MFCC audio feature extraction technique with the KNN medium has better prediction accuracy and the performance. So that the further study will be carried out on MFCC with KNN.

Table 1. Various machine learning algorithm performance against various audio feature extraction methods.

Algorithm name	Result % accuracy				Time taken to predict the result (sec)			
	MFCC (%)	Stats (%)	Other params (%)	Average (%)	MFCC-time (s)	Stats-time (s)	Other-time (s)	Average - time (s)
Ensumble - Bagged Trees	95.1	94.9	95	95	15.563	12.224	20.69	16.159
Ensumble - Boosted Trees	95.1	95.1	95.1	95.1	17.691	13.331	25.596	18.87267
Ensumble - RUSboosted trees	70.5	67.2	68	68.56667	25.475	20.442	32.909	26.27533
Ensumble - Subspace Discriminant	95.1	94.6	94.9	94.86667	19.175	15.624	25.239	20.01267
Ensumble - Subspace KNN	95	93.8	95	94.6	36.54	20.009	36.555	31.03467
KNN - Coarse KNN	95.1	95.1	95.1	95.1	14.001	6.9221	20.865	13.92937
KNN - Cosine KNN	95.1	95.1	95	95.06667	15.44	7.9043	19.336	14.22677
KNN - Cubic KNN	95	95.1	95	95.03333	129.86	23.841	194.78	116.1603
KNN - Fine KNN	93	91.2	90.6	91.6	8.0839	2.523	12.465	7.690633
KNN - Medium KNN	**95.1**	**95.1**	**95.1**	**95.1**	**10.53**	**3.9848**	**12.925**	**9.1466**
KNN - Weighted KNN	95	95	95.1	95.03333	20.317	10.308	27.21	19.27833
Linear Discriminant - Linear Discriminant	95.1	94.3	One of the predictor is constant					
Logistic regression - Logistic Regression	95.1	95	94.9	95	4.3031	1.1806	3.6912	3.0583
Quadratic Discriminant - Quadratic Discriminant	94.5		One of the predictor is constant					
SVM - Coarse Gaussian SVM	95.1	95.1	95.1	95.1	50.834	127.86	364.89	181.1947
SVM - Fine Gaussian SVM	95.1	95.1	95.1	95.1	31.846	120.02	362.94	171.602
SVM - Linear SVM	95.1	95.1	95.1	95.1	2.947	12.267	49.308	21.50733
SVM - Medium Gaussian SVM	95.1	95.1	95.1	95.1	41.559	131.42	354.98	175.9863
SVM - Quadratic SVM	95.1		Time exceed > 400 s					
Tree - Complex Tree	93.9	93.8	93.8	93.83333	2.4028	1.3682	5.8609	3.210633
Tree - Medium Tree	94.9	94.8	94.9	94.86667	5.1323	1.1521	4.5904	3.624933
Tree - Simple Tree	95.1	95.1	95	95.06667	3.1579	1.6424	5.5508	3.450367
Average	**93.78**	93.28	93.31	-	-	-	-	-

7 Conclusion

MFCC audio feature extraction along with the KNN medium has better prediction for identifying the sick hen. Image and audio analysis together produce accurate results. The signature thump of sound for various diseases is important point of research could be carried out. The data storage and optimization is another brad area of research to make the small farms to carry out this solution. A complete circuit diagrams for real-time and the design details will be presented in our future research papers.

References

1. Jacob, J.P., Butcher, G.D., Mather, F.B., Miles, R.D.: Common poultry diseases. Animal Sciences Department, UF/IFAS Extension. April 2014. http://edis.ifas.ufl.edu
2. Du, Z., Jeong, M.K., Kong, S.G.: Band selection of hyperspectral images for automatic detection of poultry skin tumors. J. IEEE Trans. Autom. Sci. Eng. **4**(3), 332–339 (2007). https://doi.org/10.1109/tase.2006.888048
3. Fujii, T., Yokoi, H., Tada, T., Suzuki, K., Tsukamoto, K.: Poultry tracking system with camera using particle filters. In: The IEEE International Conference on Robotics and Biomimetics, Bangkok, pp. 1888–1893 (2009). https://doi.org/10.1109/robio.2009.4913289
4. Dogra, S., Chatterjee, S., Ray, R., Ghosh, S., Bhattacharya, D., Sarkar, S.K.: A novel proposal for detection of avian influenza and managing poultry in a cost efficient way implementing RFID. In: 2010 International Conference on Advances in Recent Technologies in Communication and Computing, Kottayam, pp. 111–115 (2010). https://doi.org/10.1109/artcom.2010.48
5. Okada, H., Suzuki, K., Kenji, T., Itoh, T.: Avian influenza surveillance system in poultry farms using wireless sensor network. In: 2010 Symposium on Design Test Integration and Packaging of MEMS/MOEMS (DTIP), Seville, pp. 253–258 (2010)
6. Chansud, W., Wisanmongkol, J., Ketprom, U.: RFID for poultry traceability system at animal checkpoint. In: 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Krabi, pp. 753–756 (2008). https://doi.org/10.1109/ecticon.2008.4600540
7. Laokok, S., Ketprom, U.: Web services with poultry traceability system. In: 2008 5th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, Krabi, 2008, pp. 197–200. https://doi.org/10.1109/ecticon.2008.4600406
8. Hu, S., Tan, J.: BioLogger: a wireless physiological sensing and logging system with applications in poultry science. In: 2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Minneapolis, MN, pp. 4828–4831 (2009). https://doi.org/10.1109/iembs.2009.5332674
9. Raj, A.A.G., Jayanthi, J.G.: IoT-based real-time poultry monitoring and health status identification. In: 2018 11th International Symposium on Mechatronics and Its Applications (ISMA), Sharjah, pp. 1–7 (2018). https://doi.org/10.1109/isma.2018.8330139
10. Chachada, S., Kuo, C.C.J.: Environmental sound recognition: a survey. In: 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kaohsiung, pp. 1–9 (2013). https://doi.org/10.1109/apsipa.2013.6694338

11. Chu, S., Narayanan, S., Kuo, C.C.J.: Environmental sound recognition with time-frequency audio features. IEEE Trans. Audio Speech Lang. Process. **17**(6), 1142–1158 (2009). https://doi.org/10.1109/TASL.2009.2017438
12. Carroll, B.T., Anderson, D.V., Daley, W., Harbert, S., Britton, D.F., Jackwood, M.W.: Detecting symptoms of diseases in poultry through audio signal processing. In: 2014 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Atlanta, GA, pp. 1132–1135 (2014). https://doi.org/10.1109/globalsip.2014.7032298
13. Murugappan, M.: Human emotion classification using wavelet transform and KNN. In: 2011 International Conference on Pattern Analysis and Intelligence Robotics, Putrajaya, pp. 148–153 (2011). https://doi.org/10.1109/icpair.2011.5976886
14. Chinmayi, R., Nair, G.J., Soundarya, M., Poojitha, D.S., Venugopal, G., Vijayan, J.: Extracting the features of emotion from EEG signals and classify using affective computing. In: 2017 International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, pp. 2032–2036 (2017). https://doi.org/10.1109/wispnet.2017.8300118

IoT Based Food Inventory Tracking System

S. P. Lakshmi Narayan, E. Kavinkartik, and E. Prabhu(✉)

Department of Electronics and Communication Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
spnaraen96@gmail.com, kavince28@gmail.com,
e_prabhu@cb.amrita.edu

Abstract. A key component in effective kitchen management is inventory control. Keeping track of the kitchen inventory leads to more informed planning and decision-making. With technology advancing in a fast pace and everything around us becoming automated, people prefer to monitor and perform their day-to-day activities by using the smart devices they carry everywhere rather than manually recording and monitoring things. Maintaining and keeping track of everyday common food inventory is becoming one of the major problems in various households, restaurants and food chains. Replenishing the containers at the right moment and also knowing the expiry of foods is a major concern. Working people and busy restaurants find it difficult to keep track because it requires human intervention at the right time. Through this, it is easy to keep an eye on potential problems related to waste and pilferage. In this project we propose an IOT (Internet of Things) based food inventory tracking system, which ensures real time monitoring of the kitchen inventory. The collected data can be analysed in real time to understand the daily or weekly consumption and also predict usage/consumption patterns. There is also provision to check the real time status, history of consumption through an android application. The system contains a Microcontroller, load cell and wireless Module, MQTT broker, a desktop application and an Android application through which real time inventory tracking is performed. The proposed solution is completely wireless and reliable for both domestic and commercial purposes.

Keywords: Internet of Things · Message queue telemetry transport
Load cell · NodeMCU · Inventory tracking · Android · Arduino

1 Introduction

The Internet of Things (IoT) has risen as a mix of numerous innovations with various applications. Diverse meanings of IoT have risen with the goal of buttoning the components of IoT. This paper leads the way into the future with respect to smart food inventory systems. With the population increasing in a rapid pace, keeping track of things we daily require is needed as well, this project helps to have a better and clearer picture of what is there in the inventory of a restaurant or anywhere food is involved. Advancing technology has made almost everything around us automated such that

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 41–52, 2019.
https://doi.org/10.1007/978-981-13-5758-9_4

people prefer to look up to their smart devices they carry everywhere, for even the slightest of information.

Real-time tracking of the inventory availability has proved to be very methodical and labour-saving as it minimizes human intervention. Also notification of the expiry of inventory stock comes as an added advantage for busy restaurant chains and working professionals to manage their kitchen quite efficiently. The system actually allows the user to predefine the threshold value for the containers and set an identifier (label) for ease of access before the setup. As these parameters can be set for every container, it is more companionate to the user.

This paper proposes an IoT based food inventory tracking system, which ensures real time monitoring of the kitchen inventory. The collected data can be analysed in real time to understand the daily/weekly consumption and also predict usage/consumption patterns. There is also provision to check the real time status, history of consumption through an android application. The system contains a Microcontroller, load cell and ESP8266 Module. The proposed solution is completely wireless and reliable for both domestic and commercial purposes.

2 Related Work

Before the advent of new age technology like IoT, food inventory tracking was either done manually (in household, very small restaurants) or with the use of POS (Point of Scale) machine (in restaurants, supermarkets). These two methods were the most common way to keep track of inventory. As technological advancements were made, many sensors and programming were used to keep track of inventory usage. This has led to the creation of a number of research areas. Extensive research on the sensors used, hardware design, and communication methods is being carried out, to help facilitate accurate and best design of an inventory tracking.

Present generation we are living in, needs lot of sophisticated equipment to make our life easy and simple. Recent development in the field of wireless networks and cloud based analysis helps us to do and maintain our day to day activities easily. Cloud computing delivers services based on Quality of Service [QOS] [4]. To reduce the physical involvement with the systems, there are methods developed to respond to voice commands and eye-gazing [3].

The current suite of 10 kg electronic scale is joined alongside Arduino UNO to make a Bluetooth electronic scale. Framework engineering of Bluetooth Electronic Scale, this framework consists of Arduino board UNO, Bluetooth Module HC-06 and HX711 is an extraordinarily outlined 24-bit A/D converter chip for high-exactness electronic scale. The Bluetooth module uses radio waves which are less hazardous to transmit data [9]. It uses CSR Bluetooth Core chip a 6–8 Mbit programming's storage room which suits Arduino UNO. The scale sensor module has VCC, GND, DOUT, SCK individually to associate with Arduino's energy supply 3.3 V, GND, A1, A0.

Furthermore, VCC, GND, TXD, RXD of the Bluetooth Module HC-06 individually associated with Arduino's 3.3 V, GND, RX, TX. The Arduino is associated by means of USB to connection to the PC [1].

This scale gives users the weight it measures and if users want to throw away the values: "Spare" button can be tapped. Bluetooth electronic scale differs from other electronic scales in a way that it is able to get information from advanced mobile phones by means of transmission of Bluetooth. The Bluetooth electronic scale's product is given to Arduino and cell phone App. Electronic scale can read the weight through the Bluetooth scale and demonstrate the outcome on advanced cells [1].

Low profile RAPG load sensors [5] have been mounted with a stand, along with the HX711 amplifier to process the weight of the inventory. This complete entity is taken as a node and many such nodes are setup. All such nodes are connected to a central node from where data from all the nodes is sent to the cloud using an online Data Stream Network (DSN). Such a setup of nodes connected to a central node is termed as Wireless Sensor Networks (WSN) [2, 7].

The collective data from all the nodes is sent to Thing Speak (DSN), from the central node. Data from the sensors is uploaded to the Thing Speak IoT channel. Time difference between the transmitter and receiver has been taken into account and a timeperiod of 14 s has been set for data points to be uploaded to the channel. However this parameter can be changed based on the user's needs. MATLAB has been used to perform data analysis of the channel data. Visualizations and statistics of the inventory data has been shown as results [2].

Recently, large number of computations involve more data produced and to be stored. Thankfully due to large improvements in wireless technology this has been made easy through cloud computing to publish data and retrieve it whenever we need. Cloud computing is preferred as it is cost efficient and easy to store large data by not depending purely on hardware [8].

The requirement to know the items and its quantity present in our kitchen resulted in development of a tagging system known as RFID was developed. RFID tagging systems are used in food industry to keep track of inventory items, expiry dates [6].

A smart meter is designed in [10] with reduced hardware area. This device operated with low power [10].

3 System Architecture

In this paper, the Inventory Tracking Systems is mainly divided into two parts, the system hardware and system software. The system overview diagram is shown in Fig. 1.

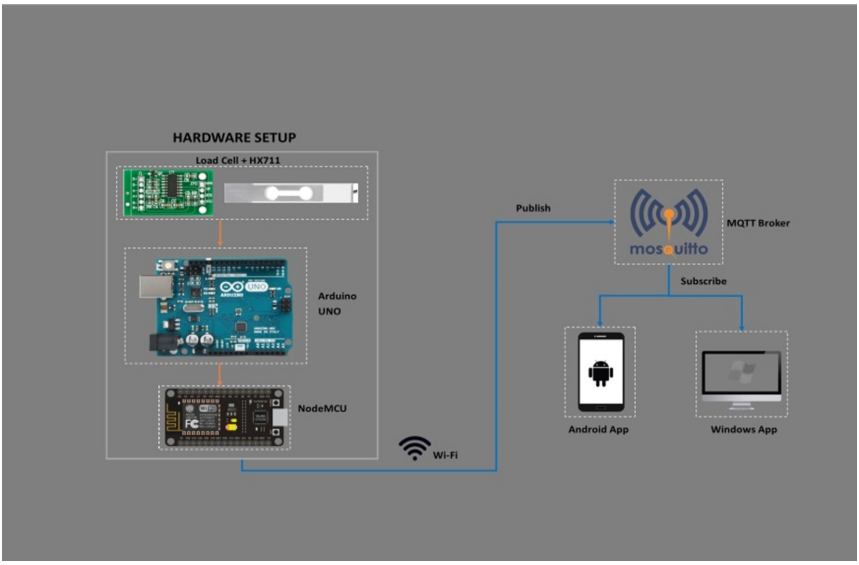


Fig. 1. System overview

The system hardware mainly consists of the Arduino UNO board, 10 kg rectangular Load cell, HX711 amplifier, NodeMCU module and the hardware casing. The wide bar load cell can make an interpretation of up to 10 kg of weight (drive) into an electrical flag. Each load cell can quantify the electrical protection that adjustments in light of, and corresponding of, the strain (e.g. weight or power) connected to the bar [5].

Arduino Uno R3 is a hardware with ATmega328 microcontroller. It has 14 advanced pins which comprises of 6 analog pins and 8 digital pins. It also has 16 MHz stone oscillator, a USB port, power jack, an ICSP header, reset button. 5 V is the operating voltage and it is capable of giving output voltage of 3.3 V for external applications. It is useful to perform advanced operations like managing external interrupts, PWM, Referencing. Arduino software (IDE) is used to code.

NodeMCU (Node Microcontroller Unit) is an open source IoT platform. It runs with System-on-a-Chip (SoC) firmware called the ESP8266. ESP8266 is made by Espressif Systems. It contains CPU, RAM, organizing (Wi-Fi), advanced working framework and SDK. NodeMCU uses the Lua scripting language and also supports MQTT, an IoT connectivity protocol which is used in our project. NodeMCU also supports Arduino software (IDE) providing greater ease.

This Weight Sensor enhancer depends on HX711, which comprise of an intensifier and it has an accurate 24-bit simple to-advanced convertor intended for measure scale and modern control applications to interface specifically with the sensor. Contrasted and different chips, HX711 not just has a couple of essential capacity, additionally contains high reconciliation, quick reaction, insusceptibility, and different highlights.

The steps involved in this project is represented in Figs. 2 and 3 basically is an overview of the MQTT broker which uses Publisher-Subscriber methodology.

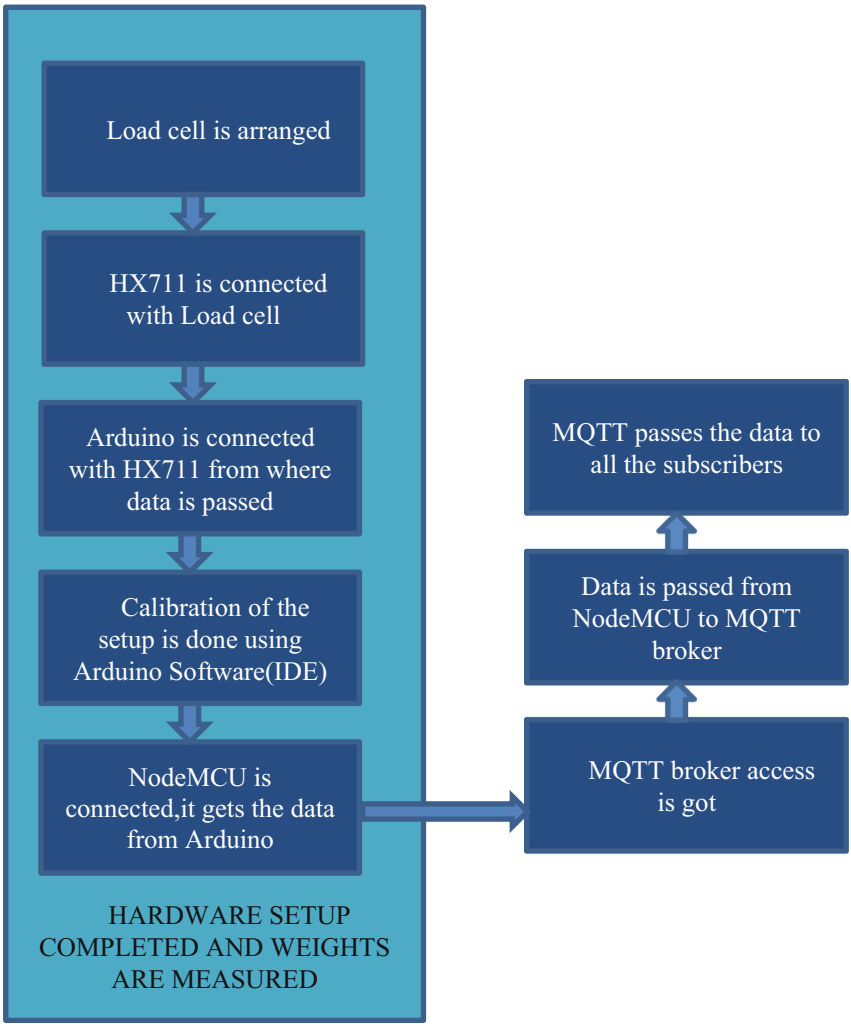


Fig. 2. Block diagram of the system design

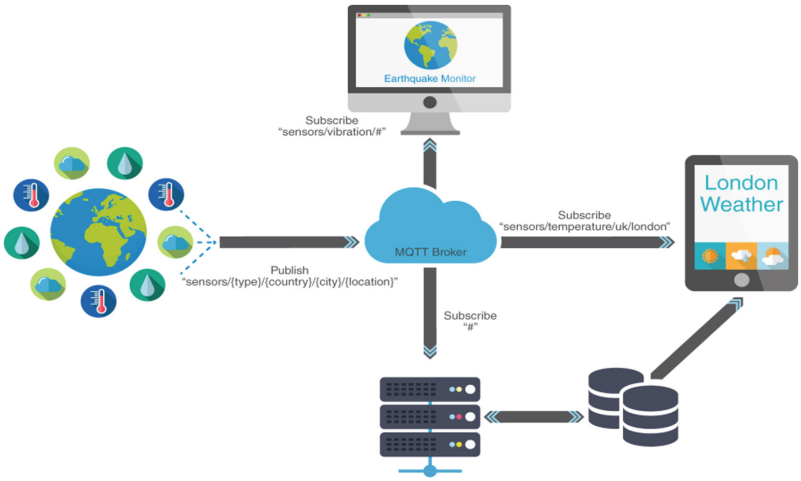


Fig. 3. MQTT framework

4 System Implementation

The prototype used for system implementation mainly involves the development of the system hardware and the system software. The system hardware is designed as follows. The key components required to set up the hardware are, Arduino UNO, rectangular load cells, NodeMCU, HX711 amplifier and the hardware casing. First, the load cell is fixed using screws in the hardware casing. The range of the rectangular load cell used is 10 kg. The schematic diagram showing the project hardware wiring is shown in Fig. 4.

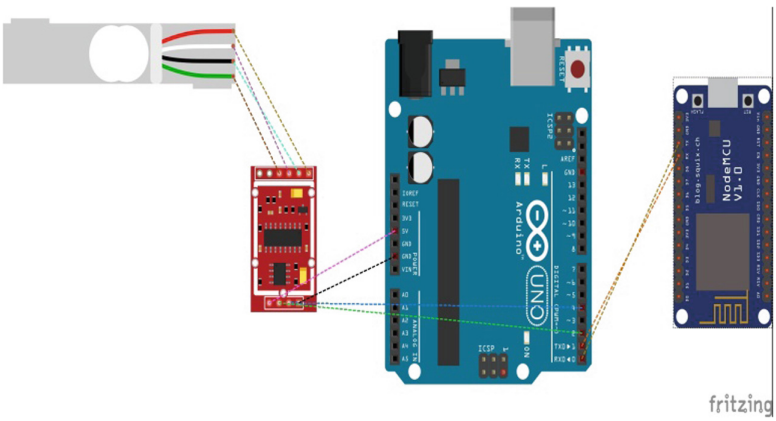


Fig. 4. Schematic diagram (Color figure online)

As shown in Fig. 4, the following connections must be made to implement the system hardware. First, to connect the load cell to HX711; red, white, black, green of the load cell must be connected with the pins E+ , A+ , E-, A- of the HX711 respectively. Next, to connect the HX711 to Arduino UNO board; Vcc, GND, DT, SCK pins of HX711 must be connected with pins 5 V, GND, 4, 2 of the UNO board respectively.

Before we take the weight readings the Load cell must be calibrated using the Arduino IDE. Steps to calibrate are given below:

Step 1: Connect the Load Cell to HX711 and then to Arduino UNO as shown in the schematic.

Step 2: Power up Arduino UNO and open the calibration code in Arduino IDE.

Step 3: Upload the program to the Arduino UNO.

Step 4: Open the Serial Monitor on the Arduino IDE.

Step 5: Wait for sensor to calibrate and set initially to zero weight, then place a known weight of 1 kg on the load cell.

Step 6: Check for the weight to be in grams.

Step 7: If the weight does not match with 1000 g (output through serial console), change the calibration value in the calibration source code and upload the program to the Arduino UNO until the weight attains 1000 g (by repeating Step 3 to Step 7).

Step 8: Once the weight attains 1000 g, note the calibration value, in the source code for the load cell.

Step 9: Update the calibration factor in the source code and upload the final program to the UNO board.

Most importantly, while connecting Arduino UNO and NodeMCU, both microcontrollers must be disconnected from the computer and then given pin connections. Failing to do this will prevent the code from being uploaded into the microcontrollers. To enable information exchange between the server and the sensors, TX and RX of the UNO board must be connected to RX and TX of NodeMCU respectively.

Once the system hardware has been setup, the load cells must be calibrated. This can be done by using standard weights of known measurements. For example, first we place a known weight of 1 kg on the load cell and change the calibration factor accordingly so as to display the proper value of 1000 g on the screen. When the screen shows the correct value, the calibration factor can be fixed.

This calibration factor must be changed in the calibration code and then be uploaded into the UNO board. In order to enable data publishing from the hardware, NodeMCU must be programmed to connect to the Wi-Fi and upload values to the subscribers. Before uploading the code to NodeMCU, an important setup must be done. An access point must be created to enable the board to connect to the Wi-Fi. Also a web server must be provided on the same access point. Note down the IP address of the Wi-Fi access point. Update the noted values in the respective parts of the code and upload the code on to the NodeMCU board.

5 System Results

After successful calibration with known weights, we now test the functionality by connecting to a windows application. However before testing, we must check whether the Wi-Fi and MQTT connections is successful or not. Once the NodeMCU board is successfully connected to the Wi-Fi and the MQTT broker, we can start testing the functionality of the windows and android applications. The functionality test with the windows app is shown below.

When no weights are placed on the cell, the value shown on the app screen is 0. This can be seen from Fig. 5. When an empty container is placed, the weight of the container is displayed as shown in Fig. 6. The weight of the almonds along with the container can be seen in Fig. 7.

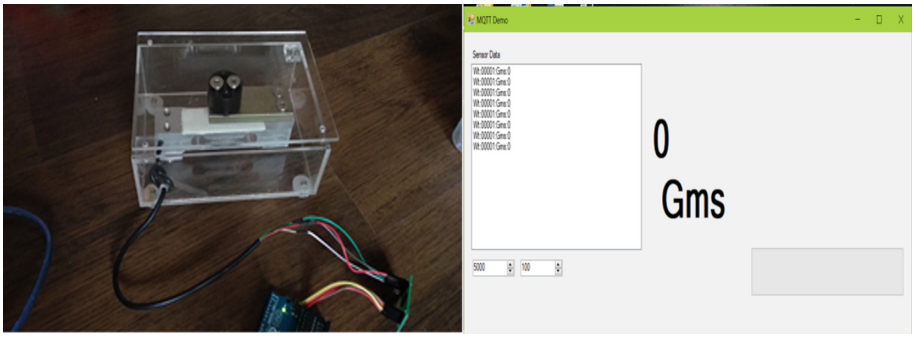


Fig. 5. Desktop app result with no weight

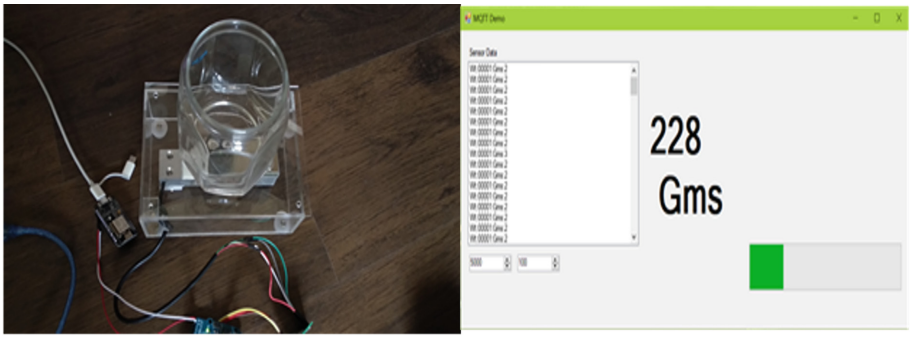


Fig. 6. Desktop app result with empty container

Fig. 7. Desktop app result with almonds

Now we test the android application in a way similar to that that of the desktop application.

First we place an empty container and enter its weight in the Container weight option. After it is entered, we can see from Fig. 8 that the reading in the weight meter is reset to zero.

Fig. 8. Android app test with no weights

In the next case, we test the re-order level or the threshold level. From Fig. 9, it is shown that the re-order level is set to 100. When we add a few amounts of almond in the container, the weight shown is 48 grams and a message in red is displayed indicating less quantity and to refill the container.

As we add more almonds into the container, we can see from Figs. 10 and 11, we can see that the total weight is now 108 grams which is above the threshold level.

Also the re-order level and the container weight can be set as desired by the user. Figure 12 shows this case. These results provided by the above figures gives us the details of the data collected form the sensors and the status of our inventory.

Fig. 9. Android app result with few almonds. (Color figure online)

Fig. 10. Container with more almonds.

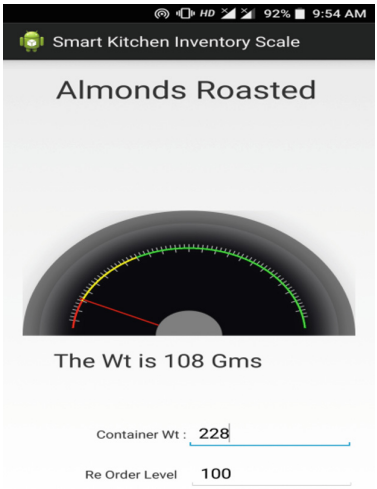


Fig. 11. Android app result with more almonds.

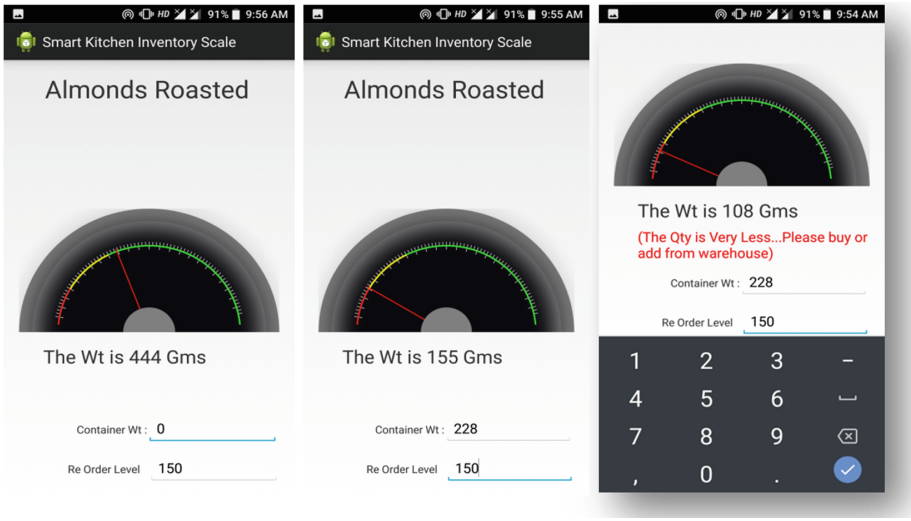


Fig. 12. Android app feature setting.

6 Conclusion and Future Scope

Internet's reconciliation with regular exercises has given us the choice of logging information for future access. It has likewise empowered the observing and breaking down the gained data. This paper illustrates the improvement of a model for checking the basic supply levels. On transferring the staple level over web they gained information can be remotely gotten to. Usage of a system of all the local hubs, can help total the information from an extensive geographic zone and further examine it in view of the application. With monstrous future degree, this framework can be utilized in kitchens, grocery stores and other storage rooms by monitoring the products.

In this project, we have successfully designed and implemented an IoT based food inventory tracking system. The calibration of the load cells was done and the calibration factor for our application use was found to be −457.05. Once the calibration was done, the basic functionality of the prototype was tested using the windows and android applications. The results achieved have also been shown. Further applications and scope includes:

- The information could give bits of knowledge into the normal use of the kitchen wares, and can be utilized to enhance the procedure of basic need acquisition.
- Data accumulated from numerous sources can be dissected to comprehend the utilization designs in light of different geographic territories, financial segments or groups in the general public. This statistic data could be imperative in the advancement of the nourishment business.

- The information could assume an essential part in the arrangement of more beneficial dinners based on the utilization midpoints recorded, contingent on occasional changes in the utilization design. Machine learning calculations can be actualized for prescient investigation. This must be encouraged by intermittent notices given to the client.
- A 'Smart Kitchen' framework can be an inventive execution that consolidates intuitive administrations for checking of nourishment things kept. It helps to improve the shopping patterns and its efficiency. It helps in reducing unnecessary storage and buying of things. This can be improved in such a way that the shopping destination also get notified about the consumer requirements so they can avoid long queues.
- By utilizing the procured information, it is additionally conceivable to deal with the normal utilization of some perishables. This could additionally be used for planning the month to month cost on nourishment and different products.

References

1. Dai, B., Chen, R.-C., Yang, W.-B.: Using arduino to develop a bluetooth electronic scale for water intake. Presented at the International Symposium on Computer, Consumer and Control (2016)
2. Desai, H., Guruvayurappan, D., Merchant, M., Somaiya, S., Mundra, H.: IoT based grocery monitoring system. Presented at the Fourteenth International Conference on Wireless and Optical Communications Networks (WOCN). IEEE, Mumbai, 24–26 February 2017
3. Bradbury, J., Shell, J.: Hands on Cooking: towards an attentive kitchen. In: Extended Abstracts on CHI 2003, pp. 996–997 (2003)
4. Mikulecky, P.: Cloud-based solutions for intelligent environments. In: Proceedings of the 18th International Database Engineering & Applications Symposium, IDEAS 2014, pp. 322–325 (2014)
5. Muller, I., de Brito, R.M., Pereira, C.E., Brusamarello, V.: Load cells in force sensing analysis—theory and a novel application. IEEE Instrum. Meas. Mag. **13**(1), 15–19 (2010)
6. Bravo, B.B., Fernandez, J.C., Barrera, M.M., Sanchez, J.R.: Implementation of RFID tags in food containers in catering business. ITG-Fachbericht 224 - RFID Systech (2010)
7. Surie, D., Laguionie, O., Pederson, T.: Wireless sensor networking of everyday objects in a smart home environment. In: Proceedings of International Conference on Intelligent Sensors, Sensor Networks and Information Processing, pp. 189–194 (2008)
8. Sakr, S., et al.: A survey of large scale data management approaches in cloud environments. IEEE Commun. Surv. Tutor. **13**(3), 311–336 (2011)
9. Reddy, P.P., Suresh, P.V.B., Reddy, P.T., Manitha, P.V., Deepa, K.: Remote control of an electronic device using EOG. In:2017 International Conference On Smart Technologies For Smart Nation (SmartTechCon), Bangalore, pp. 780–783 (2017)
10. Vigneshu, R.I., Dinesh Udhayan, R., Raghav, S., Wilfred Thadeus, P., Anguselvan, S., Prabhu, E.: Design and implementation of digital household energy meter with a flexible billing unit using FPGA. Int. J. Appl. Eng. Res. **10**(11), 28331–28340 (2015)

On Acoustic Monitoring of Farm Environments

Stavros Ntalampiras(✉)

Department of Computer Science, University of Milan,
via Celoria 18, 20133 Milan, Italy
stavros.ntalampiras@unimi.it

Abstract. Green revolution suggests that agriculture systems, such as farms turn into dynamic entities boosting animal production in an eco-friendly way. In this direction, we propose exploiting the acoustic modality for farm monitoring. Such information could be used in a stand-alone or complimentary mode to monitor the farm constantly and provide a great level of detail. To this end, we designed a scheme classifying the vocalizations produced by farm animals. We employed a feature set able to capture diverse characteristics of generalized sound events seen from different domain representations (time, frequency, and wavelet). These are modeled using state of the art generative and discriminative classification schemes. We performed extensive experiments on a publicly available dataset, where we report encouraging recognition rates.

Keywords: Acoustic farm monitoring · Intelligent farming
Audio signal processing · Hidden Markov model · Echo state network
Random forest · Support vector machine
Multidomain acoustic parameters

1 Introduction

The area of Computational Bioacoustic Scene Analysis has received increasing attention by the scientific community in the last decades [4,8,33,34]. Such interest is motivated by the potential benefits that can be acquired towards addressing important issues directly affecting the environment, such species not native to an ecosystem and cause harm, disorders caused by organisms (bacteria, fungi, etc.), climate change, proper land-use, etc. Having in possession, precise information with respect to the amount of species population as well as its tendencies, is crucial for quantifying the conservation status of the species of interest. Such information can be obtained via traditional surveys based on human observation; however these are becoming inadequate since they are (a) expensive, (b) depended on climate status, (c) can provide information for a narrow time window within a constrained space, etc. To this end, autonomous recording units (ARUs) are extensively employed by biologists [12,15,26,29]. This is also motivated by the cost of the involved acoustic sensors which is constantly decreasing due to the advancements in the field of electronics.

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 53–63, 2019.
https://doi.org/10.1007/978-981-13-5758-9_5

One of the first approaches employed for classifying animal vocalizations is described in [18]. The authors extracted Linear predictive coding coefficients, cepstral coefficients based on the Mel and Bark scale, along with time-domain features describing the peaks and silence parts of the waveform. The classifier was a Support Vector Machine, while three kernels were considered, i.e. polynomial, radial basis function, and sigmoid. These were compared with nearest neighbor and linear vector quantization schemes. The specific dataset included sounds of four animal classes, i.e. birds, cats, cows, and dogs. The literature further includes several approaches which concentrate on specific species, classification of Australian anurans [13], interpretation of chicken embryo sounds [9], classification of insects [20], etc. However, a systematic approach addressing the specific case of farm monitoring, is not present in the literature. This work wishes to cover exactly this gap.

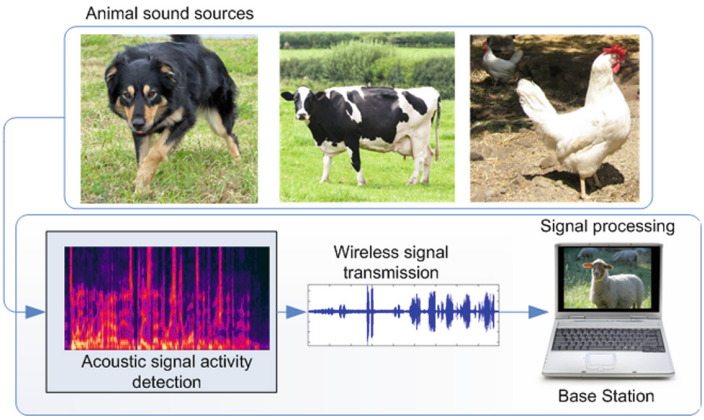


Fig. 1. The main idea of this work regarding acoustic monitoring of farm environments.

Indeed, the acoustic modality could provide complementary information to monitor the health as well as population of animals. For example it could be used in combination with solutions such as [2,16,19] which record physiological parameters of the animals, such as heart rate, temperature, etc. The valuable information that can be obtained via the acoustic modality could assist an overall assessment of the current status of the animals as well as the farm in general. More precisely, acoustic farm environment monitoring could assist in the following applications:

- tracking of similar breed animals and parturitions,
- identification of specific animal(s) for several reasons (vaccination, medication, diseases, diet, etc.),
- animal health monitoring,
- population monitoring,

– detect animals missing from the farm, and
– intruder detection and identification,

Of course, this is a non-exhaustive list of the potential applications, while the overall aim is to optimize animal production.

This work, the main idea of which is depicted in Fig. 1, reports a study of the efficacy of various feature extraction and pattern recognition algorithms on acoustic monitoring of farm environments. To this direction, we used feature sets of heterogeneous propertied, i.e. Mel-Frequency Cepstral Coefficients, MPEG-7 low level descriptors and perceptual wavelet packets, capturing the characteristics of the audio signals as represented in the temporal, frequency and wavelet domains. Subsequently, these are modeled by generative and discriminative classification schemes (echo state network, class-specific and universal hidden Markov models, support vector machines, and random forest). At the same time, this works aims at constructing a comprehensive classification scheme, the operation of which does not follow the black-box logic, i.e. where one is able to "open" the classifier, and by inspecting the misclassifications, obtain clear insights on how the performance can be boosted.

During the experimentations, we used part of the dataset called Environmental Sound Classification-10 described in [30] which includes the animals typically encountered in a farm environment, i.e. *dog*, *rooster*, *pig*, *cow*, *cat*, *hen*, and *sheep*. There, a preliminary classification analysis on the entire dataset provided a recognition rate of 72.7%.

The rest of this article is organized as follows: Sect. 2 formulates the problem, while Sect. 3 details the proposed sound classification framework. Section 4 provides information on the dataset we employed, the system parameterization, and presents and analyzes the experimental results. Finally, Sect. 5 concludes this work.

2 Problem Statement

In this work, we assume availability of a single-channel audio datastream denoted as y_t with unknown duration. Sources included in y may belong to the set $\mathcal{C} = \{C_1, \ldots, C_m\}$, where m is the number of known sound classes. In addition, we suppose that the probability density function $\mathcal{P}$ characterizing each source is stationary over time.

A training sequence $TS = y_t, t \in [1, T_0]$ is available and includes annotated pairs (y_t, C_i), where t is the time instant and $i \in [1, m]$. The ultimate goal of the proposed classification system is to automatically annotate the incoming unknown stream of audio with the lowest possible amount of wrongly classified instances t.

3 Sound Classification Framework

This section describes the feature set used in this work along with the audio pattern recognition schemes.

3.1 The Feature Set

We have used three feature sets which have shown promising performance in genaralized sound recognition tasks, i.e. Mel-Frequency Cepstral Coefficients, MPEG-7, and Perceptual Wavelet Packets set. These are explained in the following paragraphs.

Mel-Frequency Cepstral Coefficients (MFCC). This feature set includes the first thirteen MFCCs. The standard process is followed during the extraction process, i.e. the STFT powers are passed through the Mel-scale. Afterwards, we apply the log operator and discrete cosine transform. The first coefficient, which is directly related to the signal's energy, is kept. Three derivatives of the initial vector are appended resulting to 52 dimensions. This processing stage was based on the openSMILE feature extraction tool [10]. The extraction process is depicted in Fig. 2. Representative mel-scaled spectrograms w.r.t all considered sound classes are shown in Fig. 3.

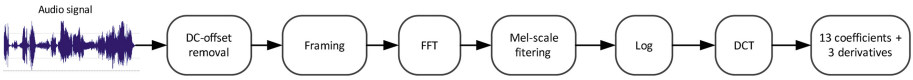


Fig. 2. The extraction process of the MFCC feature set.

MPEG-7 Audio Protocol Descriptors. The specific protocol offers a standardized way to automatically describe audio content, thus includes several interesting features suitable for generalized sound recognition. In the next, we provide a brief description of parameters that we employed, while the interested reader is referred to [5] for a detailed analysis.

- Audio Spectrum Flatness (ASF): This parameter reflects the deviations existing in the power spectrum with respect to a flat curve. Non-overlapping frames are used along with a 1/4-octave spectrum division. ASF is the ratio of the geometric mean to the arithmetic mean of the spectrum power existing in a specific band.
- Audio Waveform: This parameter characterizes the audio waveform by calculating the minimum and maximum samples in subsequent non-overlapping windows.
- Audio Fundamental Frequency: This parameter comprises the fundamental frequency $f0$, which essentially is the pitch estimation according the MPEG-7 audio protocol.

Perceptual Wavelet Packets (PWP). This set was used due to its ability in capturing perceptual properties of audio signals as a result of its initial filtering stage [28]. The specific feature set describes the acoustic signals in a multi-band manner, while each spectral area is analyzed by wavelet packets. Such a critical band division is able to weight each spectral area according to how much

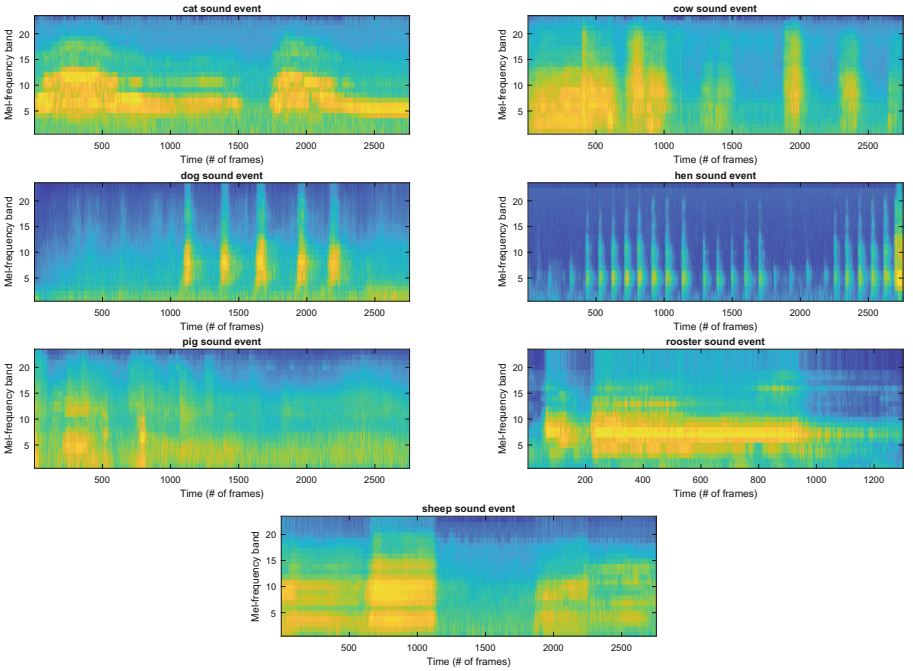


Fig. 3. Mel-scaled spectrograms extracted out audio signals associated with all considered classes.

each one affects human perception. The PWP audio parameters capture the variability of each wavelet coefficient within a critical band, which is important for the generalized sound recognition technology. An implementation of this feature set for research purposes is freely available at https://sites.google.com/site/stavrosntalampiras/home. More information is offered in [28].

3.2 The Classifiers

A series of generative and discriminative classification systems was used to model, and subsequently identify the patterns from by extracted features.

– Class-specific HMMs [14]: HMMs are able to model the temporal evolution followed by the parameters extracted out of an audio signal. Such temporal structures are highly important towards sound identification, and when combined with the corresponding spectral content, provide the ability to characterize sound events in a distinctive way. Here, the HMM states include Gaussian mixture models with diagonal covariance matrices [17]. The Torch machine learning mechanism[1] was employed to learn and evaluate GMMs and HMMs.

[1] Freely available at http://torch.ch/.

Table 1. The recognition rates (in %) achieved by the presented feature sets and classification schemes. The approach providing the highest rate is emboldened.

Classifier	MFCCs (+derivatives)	MPEG-7 audio protocol	PWPs
Class-specific HMMs	**77.1**	70.2	71.8
Universal HMM	68.6	64.9	68.1
SVM	52.3	63.8	61.5
ESN	60	54.8	53.5
RF	54.3	59.6	56.3

- Universal HMM with a Kullback-Leibler divergence (KLD) based data selection scheme [24]: such a classification type constructs one HMM (called universal) using the entire training dataset. Then, each class is represented by adapted version of the universal one. Since the present dataset exhibits high degree of variability, we employed fully-connected HMMs. There, all possible transitions across states are allowed. One of the most important aspects in the creation of the universal HMM is selecting the audio data to learn such a holistic distribution. In other words, all the classes should be adequately represented by this model. Motivated by the findings of work [21], we employed a KLD-based scheme which discovers the most central audio samples with respect to each class.
- Support vector machine (SVM): SVMs aim at discovering the hyperplanes maximizing the margin between support vectors extracted out the feature sets of the classes of interest. Such a feature set representation divides the classes by a clear gap with the goal being to widen it as much as possible. Essentially, a series of 2-class classification tasks is formed. In this paper, we employed a kernel encompassing the radial basis function [7]. It should be noted that the implementation of SVMs was based on the *libsvm* library [6].
- Random Forest (RF) [1]: Such decision tress can be learned with relative easiness without making a-priori assumptions regarding the distributions followed by the feature sets corresponding to each class. Nonetheless, at the same time it is possible that only minor alterations in the training set may output a completely different decision tree structure. RF is essentially an ensemble model since is includes multiple decision trees. There, class predictions are based on combining inferences made by individual trees [1].
- Echo State Network (ESN) [32]: Reservoir Computing (RC) comprises a rather recent trend in audio modeling [23]. In essence, an ESN is a Recurrent Neural Network (RNN) with a-priori determined weights accompanied by an output layer of linear degree of complexity. RC is based on the hypothesis that the computational intensive back propagation-based RNN training alters only slightly the internal weights of the network, thus it may be completely omitted. To this end, training becomes significantly less costly. The ESN implementation is based on the Echo State Network toolbox[2]).

[2] Freely available at https://sourceforge.net/projects/esnbox/.

4 Experimental Protocol and Analysis of the Results

In this section, we analyze the: *(a)* dataset used to acoustically simulate a farm environment, *(b)* parametrization of the feature extraction and pattern recognition algorithms, and *(c)* achieved results.

4.1 Dataset

We collected data associated with the following typical farm animals: *dog*, *rooster*, *pig*, *cow*, *cat*, *hen*, and *sheep*. These are taken from the Environmental Sound Classification-10 described in [30], while they are sampled at 44.1 KHz. Each class includes 40 recordings, each one with a duration of 5 s.

4.2 System Parametrization

After the MPEG-7 audio protocol suggestion, low-level feature extraction is performed on frames of 30 ms with 10 ms overlap between consecutive windows. Moreover, sampled data are windowed using the hamming type smoothing any discontinuities. The FFT size is 512. Standard normalization techniques, i.e. mean removal and variance scaling, were applied.

The parameters of the classification schemes were optimized on TS defined in Sect. 2 by employing a part of it (20%) as validation set.

HMM creation follows the Expectation-Maximization and Baum Welch algorithms [31]. Moreover, their parameters (number of states and Gaussian functions) are optimized based on the maximum recognition rate criterion. Cluster initialization is carried out via k-means, having a limit of 50 iterations. The transition matrix was estimated based on the Baum-Welch algorithm. Here, the number of iterations has an upper bound equal to 25 with a threshold of 0.001 between subsequent iterations. Overall, the explored states range from 3 to 7 while the number of Gaussian components comes from the {2, 4, 8, 16, 32, 64, 128, 256, and 512} set. The final parameters were selected based on the maximum recognition rate criterion.

The RBF kernel was selected for the SVM as early experimentation demonstrated its superiority over other well-known ones (linear, polynomial homogeneous, polynomial inhomogeneous, and hyperbolic tangent). Moving on, RF training was carried out using the standard algorithm described in [11]. Finally, the RN parameters were selected by means of exhaustive search following the logic described in [27].

4.3 Experimental Results

Table 1 includes the results achieved by the approaches presented in Sect. 3. The data division protocol is the ten-fold cross validation one. Identically selected folds were used during the training and testing processes of all approaches, enabling a reliable comparison.

Table 2. The confusion matrix (in %) with respect to the class-specific HMMs trained on the MFCC feature set. The average classification rate is 77.1%.

Presented	Responded						
	Dog	*Rooster*	*Pig*	*Cow*	*Cat*	*Hen*	*Sheep*
Dog	**75.2**	3.6	-	5.4	3.3	2.1	10.4
Rooster	-	**82.1**	-	3.8	4.6	0	9.5
Pig	13.2	3.5	**70.1**	6.1	-	5.8	1.3
Cow	6.3	-	-	**85.1**	5.4	-	3.2
Cat	14.5	-	5.2	-	**70.3**	3.9	6.1
Hen	3.5	7.6	-	-	2.9	**81.9**	4.1
Sheep	4.3	-	3.2	4.9	-	12.3	**75.3**

A first observation is on the difficulty of the task which is relatively high since many classifications schemes fail to provide a satisfactory recognition rate. Then, as we can see, the class-specific HMM solution trained on the MFCC set outperformed the rest of the approaches. The second best approach considers the same pattern recognition algorithm trained on PWPs. As in [14], the superiority of the MFCCs in a generalized sound recognition task, was confirmed.

The general tendency of the experimental results shows that the generative schemes are superior w.r.t the discriminative ones. The UBM logic provides lower rate that the class-specific one showing the high degree of diversity characterizing the common feature space. ESN and SVM solutions are unable to discover reliable boundaries between the classes. Similarly, the RF rules fail to classify the feature space in a reliable manner.

The confusion matrix achieved by the best-performing approach is tabulated in Table 2. We observe that the class recognized with the highest accuracy is *cow* one (85.1%), while the one presenting the worst rate is the *pig* one (70.1%). Several misclassifications occur due to the great inter-class variability since various audio signals are similar despite coming from different classes. This is particularly evident in the cases of *sheep-hen*, *cat-dog*, and *pig-dog* pairs.

We conclude that an important aspect in addressing the specific acoustic monitoring task is taking into account the temporal evolution of the acoustic parameters. Overall, the conducted study highlighted the strengths and limitations of the considered approaches, thus forming the basis for constructing a recognition mechanism able to achieve even better classification rates.

5 Conclusions

This paper presented a classification scheme within the bioacoustics scientific area addressing the problem of acoustic farm monitoring. We systematically evaluated a series of multimodal acoustic parameters modeled by generative and discriminative classifiers. Several observations arose, the most important

of which is the relevance of the temporal evolution in recognizing the involved sound classes.

In the future, we wish to *(a)* create a (hybrid) mechanism able to offer higher classification rates based on the conclusions of the present study, and *(b)* enhance the recognition framework so that it is able to operate in a concept drift environment [22], i.e. being able to evolve itself during operation and recognize altered (noisy or reverberant) versions of the existing classes, increase the dictionary of the animal vocalizations, etc. To this direction we plan to explore the transfer learning technology [3,25].

References

1. Al-Maathidi, M.M., Li, F.F.: Audio content feature selection and classification a random forests and decision tree approach. In: 2015 IEEE International Conference on Progress in Informatics and Computing (PIC), pp. 108–112, December 2015. https://doi.org/10.1109/PIC.2015.7489819
2. Anu, V.M., Deepika, M.I., Gladance, L.M.: Animal identification and data management using RFID technology. In: International Conference on Innovation Information in Computing Technologies, pp. 1–6, February 2015. https://doi.org/10.1109/ICIICT.2015.7396069
3. Banerjee, D., et al.: A deep transfer learning approach for improved post-traumatic stress disorder diagnosis. In: 2017 IEEE International Conference on Data Mining (ICDM), pp. 11–20, November 2017. https://doi.org/10.1109/ICDM.2017.10
4. Blumstein, D., et al.: Acoustic monitoring in terrestrial environments using microphone arrays: applications, technological considerations and prospectus. J. Appl. Ecol. **48**(3), 758–767 (2011). https://doi.org/10.1111/j.1365-2664.2011.01993.x
5. Casey, M.: MPEG-7 sound-recognition tools. IEEE Trans. Circuits Syst. Video Technol. **11**(6), 737–747 (2001)
6. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. **2**, 27:1–27:27 (2011). http://www.csie.ntu.edu.tw/~cjlin/libsvm
7. Chen, L., Gunduz, S., Ozsu, M.T.: Mixed type audio classification with support vector machine. In: 2006 IEEE International Conference on Multimedia and Expo, pp. 781–784, July 2006. https://doi.org/10.1109/ICME.2006.262954
8. Dong, X., Towsey, M., Zhang, J., Roe, P.: Compact features for birdcall retrieval from environmental acoustic recordings. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 762–767, November 2015. https://doi.org/10.1109/ICDMW.2015.153
9. Exadaktylos, V., Silva, M., Berckmans, D.: Automatic identification and interpretation of animal sounds, application to livestock production optimisation (Chap. 4). In: Glotin, H. (ed.) Soundscape Semiotics - Localization and Categorization. InTech, Rijeka (2014). https://doi.org/10.5772/56040
10. Eyben, F., Weninger, F., Gross, F., Schuller, B.: Recent developments in opensmile, the munich open-source multimedia feature extractor. In: Proceedings of the 21st ACM International Conference on Multimedia, MM 2013, pp. 835–838. ACM, New York (2013). https://doi.org/10.1145/2502081.2502224
11. Frank, E., et al.: Weka-a machine learning workbench for data mining. In: Maimon, O., Rokach, L. (eds.) Data Mining and Knowledge Discovery Handbook, pp. 1269–1277. Springer, Boston (2010). https://doi.org/10.1007/978-0-387-09823-4_66

12. Grill, T., Schlüter, J.: Two convolutional neural networks for bird detection in audio signals. In: 2017 25th European Signal Processing Conference (EUSIPCO), pp. 1764–1768, August 2017. https://doi.org/10.23919/EUSIPCO.2017.8081512
13. Han, N.C., Muniandy, S.V., Dayou, J.: Acoustic classification of australian anurans based on hybrid spectral-entropy approach. Appl. Acoust. **72**(9), 639–645 (2011). https://doi.org/10.1016/j.apacoust.2011.02.002
14. Kim, H.G., Sikora, T.: Comparison of MPEG-7 audio spectrum projection features and MFCC applied to speaker recognition, sound classification and audio segmentation. In: 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing. vol. 5, pp. V-925-8, May 2004. https://doi.org/10.1109/ICASSP.2004.1327263
15. Kojima, R., Sugiyama, O., Hoshiba, K., Suzuki, R., Nakadai, K.: A spatial-cue-based probabilistic model for bird song scene analysis. In: 2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA), pp. 395–404, October 2017. https://doi.org/10.1109/DSAA.2017.34
16. Kumar, A., Hancke, G.P.: A zigbee-based animal health monitoring system. IEEE Sens. J. **15**(1), 610–617 (2015). https://doi.org/10.1109/JSEN.2014.2349073
17. Matějka, P., et al.: Full-covariance UBM and heavy-tailed PLDA in I-vector speaker verification. In: 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 4828–4831, May 2011. https://doi.org/10.1109/ICASSP.2011.5947436
18. Mitrovic, D., Zeppelzauer, M., Breiteneder, C.: Discrimination and retrieval of animal sounds. In: 2006 12th International Multi-Media Modelling Conference, 5 p. (2006). https://doi.org/10.1109/MMMC.2006.1651344
19. Nagpal, S.K., Manojkumar, P.: Hardware implementation of intruder recognition in a farm through wireless sensor network. In: 2016 International Conference on Emerging Trends in Engineering, Technology and Science (ICETETS), pp. 1–5, February 2016. https://doi.org/10.1109/ICETETS.2016.7603012
20. Noda, J.J., Travieso, C.M., Sánchez-Rodríguez, D., Dutta, M.K., Singh, A.: Using bioacoustic signals and support vector machine for automatic classification of insects. In: 2016 3rd International Conference on Signal Processing and Integrated Networks, pp. 656–659, February 2016. https://doi.org/10.1109/SPIN.2016.7566778
21. Ntalampiras, S.: A novel holistic modeling approach for generalized sound recognition. IEEE Sig. Process. Lett. **20**(2), 185–188 (2013). https://doi.org/10.1109/LSP.2013.2237902
22. Ntalampiras, S.: Automatic analysis of audiostreams in the concept drift environment. In: 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP), pp. 1–6, September 2016. https://doi.org/10.1109/MLSP.2016.7738905
23. Ntalampiras, S.: Moving vehicle classification using wireless acoustic sensor networks. IEEE Trans. Emerg. Topics Comput. Intell. **2**(2), 129–138 (2018). https://doi.org/10.1109/TETCI.2017.2783340
24. Ntalampiras, S.: Universal background modeling for acoustic surveillance of urban traffic. Digit. Sig. Process. **31**, 69–78 (2014). https://doi.org/10.1016/j.dsp.2014.05.003. http://www.sciencedirect.com/science/article/pii/S1051200414001390
25. Ntalampiras, S.: A transfer learning framework for predicting the emotional content of generalized sound events. J. Acoust. Soc. Am. **141**(3), 1694–1701 (2017). https://doi.org/10.1121/1.4977749
26. Ntalampiras, S.: Bird species identification via transfer learning from music genres. Ecol. Inform. **44**, 76–81 (2018). https://doi.org/10.1016/j.ecoinf.2018.01.006

27. Ntalampiras, S., Potamitis, I.: Transfer learning for improved audio-based human activity recognition. Biosensors **8**(3), 60 (2018). https://doi.org/10.3390/bios8030060
28. Ntalampiras, S., Potamitis, I., Fakotakis, N.: Exploiting temporal feature integration for generalized sound recognition. EURASIP J. Adv. Sig. Process. **2009**, 807162 (2009)
29. Ntalampiras, S., Potamitis, I., Fakotakis, N.: Acoustic detection of human activities in natural environments. J. Audio Eng. Soc. **60**(9), 686–695 (2012). http://www.aes.org/e-lib/browse.cfm?elib=16373
30. Piczak, K.J.: ESC: dataset for environmental sound classification. In: Proceedings of the 23rd ACM International Conference on Multimedia, MM 2015, pp. 1015–1018. ACM, New York (2015).https://doi.org/10.1145/2733373.2806390
31. Rabiner, L.R.: A tutorial on hidden Markov models and selected applications inspeech recognition. Proc. IEEE **77**(2), 257–286 (1989). https://doi.org/10.1109/5.18626
32. Scardapane, S., Uncini, A.: Semi-supervised echo state networks for audio classification. Cogn. Comput. **9**(1), 125–135 (2017). https://doi.org/10.1007/s12559-016-9439-z
33. Stowell, D.: Computational bioacoustic scene analysis. In: Virtanen, T., Plumbley, M., Ellis, D. (eds.) Computational Analysis of Sound Scenes and Events, pp. 303–333. Springer International Publishing, Cham (2018). https://doi.org/10.1007/978-3-319-63450-0_11
34. Towsey, M.W., Truskinger, A.M., Roe, P.: The navigation and visualisation of environmental audio using zooming spectrograms. In: 2015 IEEE International Conference on Data Mining Workshop (ICDMW), pp. 788–797, November 2015. https://doi.org/10.1109/ICDMW.2015.118

Clickbait Detection Using Swarm Intelligence

Deepanshu Pandey, Garimendra Verma(✉), and Sushama Nagpal

Netaji Subhas Institute of Technology, Delhi, India
garimendrav.co@nsit.net.in

Abstract. Clickbaits are the articles containing catchy headlines which lure the reader to explore full content, but do not have any useful information. Detecting clickbaits solely by the headline without opening the link, can serve as a utility for users over internet. This can prevent their time from useless surfing caused by exploring clickbaits. In this paper Ant Colony Optimization, a Swarm Intelligence (SI) based technique has been used to detect clickbaits. In comparison with algorithms used in the past, this SI based technique provided a better accuracy and a human interpretable set of rules to classify clickbaits. A maximum accuracy of 96.93% with a set of 20 classification rules was obtained using the algorithm.

Keywords: Clickbaits · Ant Colony Optimization (ACO) · Classification

1 Introduction

The ease with which it's become convenient to browse through articles over the internet, has introduced the challenges of dealing with links which lead to pages that end up showing little information than what they seem to promise [11]. The goal of a clickbait is to provide a piece of information that the viewer may tend to find highly tempting to read more about. This contributes towards the number of visits on that page or "the number of clicks", which is used by the authors and websites to compete in the world of media campaign for advertisements [3]. Some examples of clickbaits are -

"21 places you should visit before getting old. The 10th will leave you spellbound"

"Can you solve this ancient riddle? 90% gave the wrong answer"

It is the human psychology that is played with and the situation of the Curiosity Gap [11] which is exploited, by the clickbaits. With such journalistic problems around, it is hard for people on the internet to build the ability to stay for long over some more informative content [15]. Thus, they end up switching over from one article to the next news story.

Considering the far-fetched issues posed by clickbaits, Facebook introduced an update to reduce clickbaiting [22]. It considered two factors to combat this issue. First was, how much time people spent reading an article off facebook. This factor implied that less the time, less informative the content. Second was, by considering the correlation between the count of clicks to that of likes, shares or comments. One of the major drawbacks of this approach is that all types of links including the clickbaits need

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 64–76, 2019.
https://doi.org/10.1007/978-981-13-5758-9_6

to be opened first, to analyze the time spans as well as to compute the required ratios. This way, the information learnt from one post is not passed on to another, making the method repetitive for very similar posts.

Along similar lines, a clickbait challenge was organised by the Bauhaus University-Weimar [17]. It aimed at determining the level of clickbaiting for a given tweet into one of the four different classes, indicating the extent of clickbaiting. Various researchers who participated in this challenge suggested techniques for feature engineering and proposed approaches employed in the domain of clickbait detection [25].

Researchers have also put efforts to tackle this problem and used techniques like SVM, Deep Learning and Random Forests. Potthast was among the first researchers who automated clickbait detection by analyzing tweets using Random Forest, Logistic Regression and Naive Bayes [14]. Another automatic clickbait detection engine was developed by Chakraborty et al. [3]. It provided customized settings to block clickbaits from showing up and took into consideration the preferences of user, using SVM [3]. Deep Learning and Neural Network based approaches have also been explored in the domain of clickbait detection [15, 16].

Machine learning based approaches create models which are mathematically complicated and act as a black box when it comes to computing results [28]. Fields like credit risk evaluation, system security in information technology sector, diagnosis related to medical domain etc. involve high accuracy and risk analysis [26, 27]. Such fields cannot trust decisions made by black box models and need human supervision and validation.

This drawback is well attended by Swarm Intelligence (SI) based algorithms like Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), Cuckoo Search etc. These models provide human understandable rules which at the final stage can be well interpreted and modified with human supervision. Among these approaches ACO has shown promising results and has been popularly used in areas like web-spam detection, data-mining, image-edge detection etc. [4, 23, 24]. ACO aims at tackling computational problems with the help of probabilistic heuristics which can be condensed to determining optimal paths using graph based data structures. This paper aims at detecting a clickbait by analysing the headline without opening the link to the web page using ACO.

In order to analyse the performance of our approach, we have used the publicly available dataset released by Chakraborty et al. [3]. The results indicate that the rules generated by the ACO model, classifies clickbaits more accurately than the previous works. We obtained an average accuracy of 96.71%.

The remaining paper is organised as follows. Section 2 describes the Related Work in the field of clickbait detection. Section 3 explains in detail the Proposed Approach that has been implemented to solve the problem. The Experimental Setup & Results are presented in Sect. 4. Lastly, Sect. 5 concludes the paper and discusses Future Work.

2 Related Work

Clickbait detection is not an altogether new problem and has previously been in focus of many researchers who have used different techniques for detecting and classifying headlines [3, 15, 16]. The feeling of deprivation induced by the image that a clickbait headline produces is what urges the user to open the link and know for oneself the truth [11]. In his study Lowenstein puts forward that the notion of gap in understanding or knowledge, initiates a longing for bridging the space in between.

In the domain of clickbait detection, Potthast et al. proposed an approach to analyze the tweets at multiple levels [14]. This included information from the teaser message or the headline of the article, the linked web page containing the article and the meta-information related to the tweets, i.e. time of posting, retweets and attached media. They have used supervised learning based approaches like Random Forests, Naive Bayes and Logistic Regression to detect clickbaits and the results were compared on the corpus of around 3000 tweets. Among these models Random Forest performed best with an ROC-AUC of 0.74.

Chakraborty et al. [3] worked towards automatically detecting clickbaits using a browser extension, in order to warn readers about such links and provided the feature of blocking such pages based on user's interests and preferences. For classification researchers suggested fourteen different features exploiting sentence structure and Parts Of Speech (POS) based features of the headlines. The models they trained for classification were Decision Trees, Random Forests and Support Vector Machine (SVM) with Radial Basis Function kernel. It was observed that SVM performed best among all of their proposed models with an accuracy of 0.93.

Deep Learning has also been explored in order to detect clickbaits. In a very recent study by Agrawal [15], a Convolutional Neural Networks (CNN) based model was proposed which segregated headlines into clickbait and non-clickbait with an accuracy of 0.90. The feature extraction has been incorporated in CNN. On similar lines, another work by Anand et al. [16] focused on RNNs for detecting clickbaits and their approach gave remarkable results with accuracy of 0.98. However there hasn't been much focus on feature extraction and domain knowledge in their work.

SI techniques like ACO work well in exploring a large search space. ACO as compared to the machine learning techniques performs well in forming rules for classification due to its malleable nature [23]. It explores all mix and match of logical conditions during the supervised learning phase and creates transparent and understandable rules. ACO has proved its efficiency in domains like spam classification, image detection, data-mining, etc. Spam classification is one such area where Link-Based Ant Colony Optimization has been suggested by Taweesiriwate which uses TrustRank assumption to generate classification rules for spam detection [4]. The problem of image-edge detection has also been approached with ACO, showing superior results [24]. Tian et al. applied the SI technique, which for every position of a pixel in the image, builds the pheromone matrix recognising image's edge related information. To the best of our knowledge, though ACO received much attention among researchers working in diverse domains, it has not yet been explored in classifying clickbaits. Therefore the aim of this work is to use ACO for clickbait detection.

3 Proposed Approach

In this section we start by first discussing in detail the feature engineering done during data preprocessing phase and then explain the algorithm used. The workflow adopted has been described in Fig. 1.

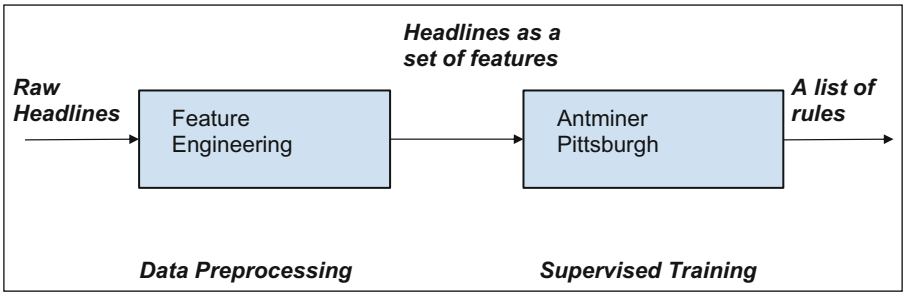


Fig. 1. Workflow diagram

3.1 Feature Engineering

After analysing the structure of various clickbait headlines, a total of 19 features were extracted which can be broadly classified into following categories: -

(A) **Stop Words, Common Phrases and Slangs**

Researchers have observed that clickbaits have a high frequency of various common phrases like "What Happened Next!", "It will leave you awestruck" etc. [3, 9]. Also it has been observed that a large number of stop words and slangs are present in clickbaits. Stop words are the highly abundant words in the dictionary of a language, while slangs include abbreviations like "WTF", "AF", "WTH", "LOL", "LMAO" etc.

In this work we analysed the sentiment values of words in headlines using Stanford Sentiment Analysis Tool and the clickbaits had high occurrence of 'Very Positive' sentiments [8].

Some regexes released by downworthy have also been used to match the typical punctuation patterns used in clickbaits [9]. Clickbaits also have a high tendency to start with a number, like "11 Places To Visit Before Dying", "21 Signs That You Are Confident!" etc.

(B) **N-Grams and Parts Of Speech (POS) Based Features**

There is a list of repetitive two-grams, three-grams and four-grams which occur in clickbait headlines and their count has been exploited to differentiate between headlines. Using the Stanford Named Entity Recognizer Tool, POS based features like count of proper nouns, adverbs and determiners, personal and possessive pronouns, verbs etc. has also been calculated in this work [7].

(C) **Sentence Structure**

The Flesch–Kincaid readability score, number of words, average word length and maximum word length also contribute towards the set of features exploiting the sentence structure. The maximum length of syntactic dependency proposed by Chakraborty et al. which is computed as the maximum distance in terms of the number of words isolating the *dependent* and *governing words* is also used as one of the features for classification of clickbaits [3].

3.2 Algorithm Used

In the following part of this section we have described the ACO technique and then explained how Antminer Pittsburgh algorithm (an application of ACO) has been used for clickbait classification.

Ant Colony Optimization

Ant colony optimization (ACO) is a swarm intelligence technique which mimics behaviour of ants to achieve several computational tasks [2]. The ants while walking secrete a substance called pheromone which is used by other ants to track the path created. Pheromone evaporates with time, hence if the same path is travelled upon frequently the pheromone density increases and the paths which are less travelled, their pheromone density decreases. This creates a feedback mechanism which is utilised by ant colonies to make optimised decisions for simple tasks like travelling the best path to destination as shown in Fig. 2.

```
procedure ACO
    while(not_termination)
        GenerateSolutions()
        ProblemSpecificActions()
        PheromoneUpdate()
    end while
end procedure
```

Fig. 2. ACO high level pseudocode

Antminer-Pittsburgh Algorithm

Parpinelli et al. proposed Antminer algorithm which used ACO for classification task [23]. Antminer works by finding an enumeration of rules for classification using a sequential covering strategy. The strategy constructs the rules one by one till each and every training example gets covered. Various variations of antminer [18–21] have been proposed till now but we have used the Pittsburgh [1, 2] approach implemented by Otero et al. [5] for clickbait detection. The major difference between naive antminer and the one which has been used in our task lies in the difference of sequential covering

strategy. Earlier the focus was on greedily enumerating independently best rules, while on the other hand best list of rules altogether is constructed by Pittsburgh version.

In a supervised manner, headlines with their corresponding class i.e. Clickbait or Non Clickbait are fed into the Antminer Model. In the process of Antminer-Pittsburgh, the ant keeps on adding terms probabilistically to the current rule which is under construction. The probability is decided using pheromone values and a problem specific heuristic value. A minimum value is decided which until exceeded, terms are added to the current rule so as to avoid too specific rules.

Further the constructed rule is pruned to avoid redundant terms which don't affect the quality of the rule formed. The headlines in the training set covered by the constructed list are expelled from the dataset and the subsequent of rule is decided as the

```
Pseudo-Code
Input: headlines with their corresponding class
Output: best list of rules to classify headlines
1. InitialisePheromones();
2. list_gb ← φ;
3. m ← 0;
4. while m < maximum iterations and not stagnation do
5.    list_ib ← φ;
6.    for n ← 1 to colony-size do
7.          headlines ← all training headlines;
8.          list_n ← φ;
9.          while |headlines| > maximum uncovered do
10.               ComputeHeuristicInformation(headlines);
11.               rule ← CreateRule(headlines);
12.               Prune(rule);
13.               headlines ← headlines - Covered(rule, headlines);
14.               list_n ← list_n + rule;
15.         end while
16.         if Quality(list_n) > Quality(list_ib) then
17.               list_ib ← list_n;
18.         end if
19.   end for
20.   UpdatePheromones(list_ib );
21.   if Quality(list_ib) > Quality(list_gb) then
22.         list_gb ← list_ib;
23.   end if
24.   m ← m + 1;
25. end while
26. return list_gb ;
```

Fig. 3. Algorithm used for clickbait detection [5]

class value among the satisfying examples. All ants of the colony form rules iteratively and the finest list of rules updates the pheromone matrix. The target headline is hence classified as the class decided by the rules. Figure 3 describes the high level pseudocode of the approach discussed. In our work we have used the open sourced implementation for rule creation using antminer PB model [10].

4 Experimental Setup and Results

In this section we discuss about the dataset used, parameter settings, performance measures and finally the results obtained.

4.1 Experimental Setup

Dataset

We ran antminer pittsburgh on the dataset released by Chakraborty et al. comprising of 15,000 news headlines, which has well distributed 7,500 clickbait and 7,500 non-clickbait headlines. The clickbait headlines are taken from ViralNova, ViralStories, BuzzFeed, Upworthy and Scoopwhoop. The non-clickbait headlines are collected from Wikinews [3].

A total of 19 features were extracted from the given headlines, which exploited domains like Sentence Structure, finding hyperbolic and common phrases, stop words and exploring POS tag based features. These features were extracted using stanford core-nlp tool and are inspired from various resources [3, 17].

Parameter Setting

We used the default values calculated by racing procedure in F-race by Otero et al. in all parameters (Table 1) other than ant colony size and number of iterations, as these default values had generated promising results previously [5].

Table 1. Parameter settings for Antminer-Pittsburgh

S. No	Parameter	Value
1	Colony size	{5, 10, 15, 20}
2	Number of iterations	{10, 50, 100}
3	Minimum covered examples per rule	10
4	% of allowed uncovered examples	0.01
5	Iterations for convergence	40
6	Evaporation factor	0.9
7	Rule pruner method	Backtrack
8	Heuristic method	Gain
9	Discretisation	mdl
10	Rule quality function	sen_spe

Performance Measures
To evaluate the efficiency of our work we have used the following performance metrics:

(a) *Accuracy*: Accuracy is the ratio of total true predictions and overall predictions made by the model.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \tag{1}$$

(b) *Precision*: Precision is defined as the ratio of true positives to the sum of true and false positives.

$$Precision = \frac{TP}{(TP + FN)} \tag{2}$$

(c) *Recall*: Recall is defined as the ratio of true positives to the sum of true positives and false negatives.

$$Recall = \frac{TP}{(TP + FN)} \tag{3}$$

(d) *F1 Score*: The harmonic mean of precision and recall is defined as F1 score.

$$F1\ Score = 2\frac{(Recall\ *\ Precision)}{(Recall\ +\ Precision)} \tag{4}$$

In above definitions TP represents true positives, FP represents false positives, TN is for True negatives and FN denotes false negatives. To calculate above parameters we did a 67:33 train test split on 15,000 headlines dataset.

4.2 Results

We varied the colony size from 5 to 20 in steps of 5 and the number of iterations were kept {10, 50, 100}. The reason for limiting iterations at 100 (Table 2) lies behind the observation that increasing number of iterations above this led to increased computation time and caused little change in accuracy.

The evaluation metrics achieved by running Antminer PB on the open sourced dataset [3] are mentioned in Table 4. The Antminer PB produced a model (Fig. 4) which consists of 20 rules and 23 terms that successfully classify the headlines into clickbaits and non clickbaits with maximum accuracy of 96.93% (Table 2). The confusion matrix obtained is shown in Table 3. It can be observed in Table 4 that our model outperformed SVM, Decision Tree and Random Forest.

Table 2. Accuracy and computation time (by varying colony size and iterations)

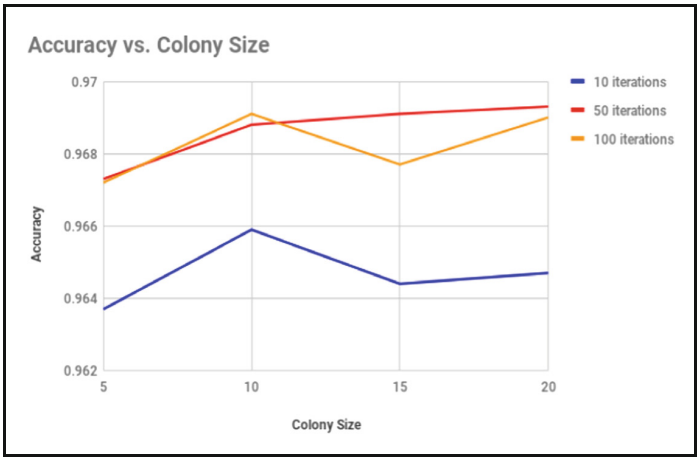
Ant colony size	Iterations	Accuracy	Computation time (in seconds)
5	10	96.37	3829.22
5	50	96.73	14509.59
5	100	96.72	22586.64
10	10	96.59	7623.44
10	50	96.88	29274.37
10	100	96.91	37645.76
15	10	96.44	9416.26
15	50	96.91	36210.36
15	100	96.77	52099.96
20	10	96.47	14811.62
20	**50**	**96.93**	**46652.46**
20	100	96.90	72206.39

```
=== Discovered Model ===

IF stopword_to_word_ratio > 0.114379 THEN Non_Clickbait
IF count_3gram > 0.5 THEN Clickbait
IF count_4gram > 0.5 THEN Clickbait
IF stopword_to_word_ratio > 0.069048 THEN Non_Clickbait
IF count_abbrvtns > 0.5 THEN Clickbait
IF count_pronouns > 0.5 THEN Clickbait
IF count_2gram > 0.5 THEN Clickbait
IF starts_with_number = True THEN Clickbait
IF count_adv+det > 1.5 THEN Clickbait
IF average_word_length > 6.279221 AND count_hyperbolic <= 0.5 THEN Non_Clickbait
IF count_clickbait_phrases > 0.5 THEN Clickbait
IF number_of_words > 10.5 THEN Clickbait
IF count_proper_noun > 0.5 AND max_word_length <= 10.5 AND number_of_words <= 8.5
AND count_adv+det <= 0.5 THEN Non_Clickbait
IF average_word_length <= 4.3875 THEN Clickbait
IF count_proper_noun > 3.5 THEN Non_Clickbait
IF count_hyperbolic > 0.5 THEN Clickbait
IF number_of_words > 8.5 THEN Clickbait
IF count_adv+det <= 0.5 THEN Non_Clickbait
IF average_word_length <= 5.133929 THEN Clickbait
IF <empty> THEN Clickbait

Number of rules: 20
Total number of terms: 23
Average number of terms: 1.15
List quality: 0.967024
```

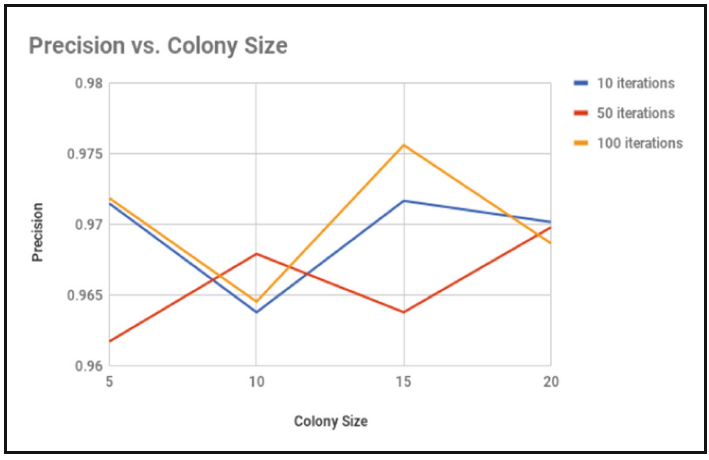
Fig. 4. Discovered model

Table 3. Confusion matrix

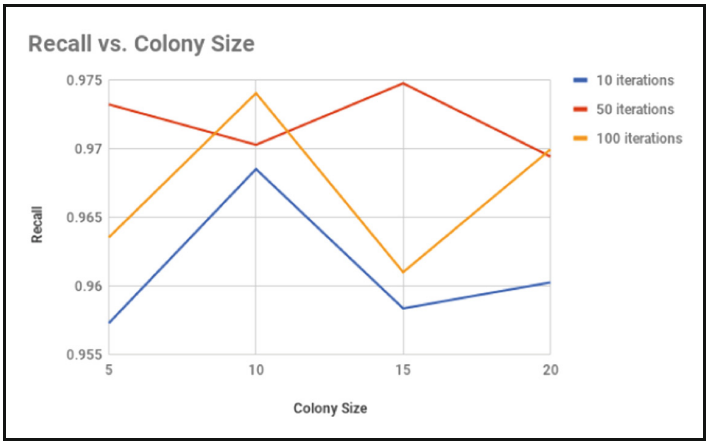
Clickbait	Non-clickbait	← Classified as
5179	150	Clickbait
196	5035	Non-clickbait



(i)



(ii)



(iii)

Fig. 5. Plots of (i) Accuracy vs colony size (ii) Precision vs colony size (iii) Recall vs colony size

Table 4. Results

Model	F1 score	Accuracy	Precision	Recall
Antminer (Pittsburgh University)	0.96	0.96	0.97	0.96
Chakraborty et al. [3] (decision tree)	0.90	0.90	0.91	0.89
Chakraborty et al. [3] (support vector machine)	0.93	0.93	0.95	0.90
Chakraborty et al. [3] (random forest)	0.92	0.92	0.94	0.91

The plots (Fig. 5) shown here are quite critical to the observation that how increasing the number of iterations doesn't lead to an increase in accuracy or recall but increases the precision. The maximum accuracy reached is at 50 iterations for colony size of 20 while precision is maximum for 15 colony size and 100 iterations.

5 Conclusion and Future Work

In this paper, we discussed the application of ACO for clickbait detection. The Research concluded with generation of 20 rules, which classified the headlines into clickbaits and non-clickbaits. Our work achieved a maximum accuracy of 96.91% and outperformed the existing models like SVM, Decision Trees, Random Forest, Deep Learning [3, 15].

The future scope lies in exploring more heuristics affecting the list quality, keeping in mind the coherence between the rules. Apart from amalgamation of models, some domain specific work can be experimented too which includes a more focused feature extraction. The model can also be improved by giving a negative clue whenever discovering of a rule leads into a wrong direction. This would further improve the classification accuracy.

References

1. Smith, S.F.: A Learning System Based on Genetic Adaptive Algorithms, pp. 1–214 (1980)
2. Smith, S.: Flexible learning of problem solving heuristics through adaptive search. In: Proceedings 8th International Joint Conference on Artificial Intelligence, pp. 422–425 (1983)
3. Chakraborty, A., Paranjape, B., Kakarla, S., Ganguly, N.: Stop clickbait: detecting and preventing clickbaits in online news media. In: 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 9–16 (2016)
4. Taweesiriwate, A., Manaskasemsak, B., Rungsawang, A.: Web spam detection using link-based ant colony optimization. In: 2012 IEEE 26th International Conference on Advanced Information Networking and Applications, pp. 868–873 (2012)
5. Otero, F.E.B., Freitas, A.A., Johnson, C.G.: A new sequential covering strategy for inducing classification rules with ant colony algorithms. IEEE Trans. Evol. Comput. **17**, 64–76 (2013)
6. Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., Mcclosky, D.: The Stanford CoreNLP natural language processing toolkit. In: Proceedings of 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations, pp. 55–60 (2014)

7. Finkel, J.R., Grenager, T., Manning, C.: Incorporating non-local information into information extraction systems by Gibbs sampling. In: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics - ACL 2005, pp. 363–370 (2005)
8. Socher, R., et al.: Recursive deep models for semantic compositionality over a sentiment treebank. In: EMNLP (2013)
9. Downworthy: A browser plugin to turn hyperbolic viral headlines into what they really mean. http://downworthy.snipe.net/. Accessed 10 June 2018
10. Myra: A collection of ant colony optimization (ACO) algorithms for the data mining classification and regression tasks. https://github.com/febo/myra. Accessed 10 June 2018
11. Lowenstein, G.: The psychology of curiosity: a review and reinterpretation. Psychol. Bull. **116**, 75 (1994)
12. Becchetti, L., Castillo, C., Donato, D., Leonardi, S., Baeza-Yates, R.: Link-based characterization and detection of web spam. In: Proceedings of the 2nd International Workshop on Adversarial Information Retrieval on the Web (2006)
13. Ahmed, S., Monzur, R., Palit, R.: Development of a Rumor and spam reporting and removal tool for social media. In: 2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE) (2016)
14. Potthast, M., Köpsel, S., Stein, B., Hagen, M.: Clickbait detection. In: Ferro, N., et al. (eds.) ECIR 2016. LNCS, vol. 9626, pp. 810–817. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-30671-1_72
15. Agrawal, A.: Clickbait detection using deep learning. In: 2016 2nd International Conference on Next Generation Computing Technologies (NGCT), pp. 268–272 (2016)
16. Anand, A., Chakraborty, T., Park, N.: We used neural networks to detect clickbaits: you won't believe what happened next! In: Jose, J.M., et al. (eds.) ECIR 2017. LNCS, vol. 10193, pp. 541–547. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-56608-5_46
17. Clickbait Challenge 2017: Challenge on detection of clickbait posts in social media. http://www.clickbait-challenge.org. Accessed 10 June 2018
18. Otero, F.E.B., Freitas, A.A., Johnson, C.G.: *c*Ant-Miner: an ant colony classification algorithm to cope with continuous attributes. In: Dorigo, M., Birattari, M., Blum, C., Clerc, M., Stützle, T., Winfield, A.F.T. (eds.) ANTS 2008. LNCS, vol. 5217, pp. 48–59. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-87527-7_5
19. Otero, F.E.B., Freitas, A.A., Johnson, C.G.: Handling continuous attributes in ant colony classification algorithms. In: 2009 IEEE Symposium on Computational Intelligence and Data Mining, pp. 225–231 (2009)
20. Liu, B., Abbas, H., Mckay, B.: Classification rule discovery with ant colony optimization. In: IEEE/WIC International Conference on Intelligent Agent Technology, IAT 2003, pp. 83–88 (2003)
21. Chan, A., Freitas, A.: A new classification-rule pruning procedure for an ant colony algorithm. In: Talbi, E.-G., Liardet, P., Collet, P., Lutton, E., Schoenauer, M. (eds.) EA 2005. LNCS, vol. 3871, pp. 25–36. Springer, Heidelberg (2006). https://doi.org/10.1007/11740698_3
22. El-Arini, K., Tang, J.: Click-Baiting: Facebook Newsroom. https://newsroom.fb.com/news/2014/08/news-feed-fyi-click-baiting. Accessed 10 June 2018
23. Parpinelli, R.S., Lopes, H.S., Frietas, A.A.: Data mining with an ant colony optimization algorithm. IEEE Trans. Evol. Comput. **6**, 321–332 (2002)
24. Tian, J., Yu, W., Xie, S.: An ant colony optimization algorithm for image edge detection. In: 2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence), pp. 751–756 (2008)

25. Cao, X., Le, T., Zhang, J., Lee, D.: Machine learning based detection of clickbait posts in social media. In: Submission for the Clickbait Challenge. arXiv preprint arXiv:1710.01977 (2017)
26. Baesens, B., Setiono, R., Mues, C., Vanthienen, J.: Using neural network rule extraction and decision tables for credit-risk evaluation. Manage. Sci. **49**, 312–329 (2003)
27. Mani, S., Shankle, W.R., Pazzani, M.J.: Acceptance of rules generated by machine learning among medical experts. Methods Inf. Med. **40**, 380–385 (2001)
28. Martens, D., Backer, M.D., Haesen, R., Vanthienen, J., Snoeck, M., Baesens, B.: Classification with ant colony optimization. IEEE Trans. Evol. Comput. **11**, 651–665 (2007)

IoT-Enabled Distributed Detection of a Nuclear Radioactive Source via Generalized Score Tests

Giampaolo Bovenzi[1], Domenico Ciuonzo[2(✉)], Valerio Persico[1,2], Antonio Pescapè[1,2], and Pierluigi Salvo Rossi[3]

[1] University of Naples "Federico II", Naples, Italy
giampaolo.bovenzi@gmail.com, {valerio.persico,pescape}@unina.it
[2] Network Measurement and Monitoring (NM2), Naples, Italy
domenico.ciuonzo@ieee.org
[3] Kongsberg Digital AS, Trondheim, Norway
salvorossi@ieee.org

Abstract. A decentralized detection method is proposed for revealing a radioactive nuclear source with unknown intensity and at unknown location, using a number of cheap radiation counters, to ensure public safety in smart cities. In the source present case, sensors nodes record an (unknown) emitted Poisson-distributed radiation count with a rate decreasing with the sensor-source distance (which is unknown), buried in a known Poisson background and Gaussian measurement noise. To model energy-constrained operations usually encountered in an Internet of Things (IoT) scenario, local one-bit quantizations are made at each sensor over a period of time. The sensor bits are collected via error-prone binary symmetric channels by the Fusion Center (FC), which has the task of achieving a better global inference. The considered model leads to a one-sided test with parameters of nuisance (i.e., the source position) observable solely in the case of $\mathcal{H}_1$ hypothesis. Aiming at reducing the higher complexity requirements induced by the generalized likelihood ratio test, Davies' framework is exploited to design a generalized form of the locally optimum detection test and an optimization of sensor thresholds (resorting to a heuristic principle) is proposed. Simulation results verify the proposed approach.

Keywords: CBRN sensors · Data fusion · Distributed detection
IoT · Public safety · Smart cities · Wireless Sensor Networks

1 Introduction

Almost 70% of the population of the world (≈ six billion people) is expected to live in cities (and, also, neighboring regions) in next thirty years. So, the need for smart(er) cities, using information and communication technologies, is becoming an imperative to make their services and monitoring more efficient, interactive and aware, and keep them on the track to flourish as platforms enabling

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 77–91, 2019.
https://doi.org/10.1007/978-981-13-5758-9_7

well-being from economic, environmental and social viewpoints. The smart city concept is fostered (also from a technological standpoint) by the Internet of Things (IoT) paradigm, representing an unprecedented Internet evolution into a pervasive network of interconnected entities that (a) collects data from the environment, (b) allows interaction with the physical world and (c) uses Internet infrastructure to furnish services for data analytics, information transfer, and applications usage [1].

Wireless Sensor Networks (WSNs) constitute the actuation/sensing arm of the IoT and are able to naturally permeate the urban infrastructure, thanks to their flexibility, reduced costs, and applicability to several domains [2]. Their collected information is expected to be shared across diverse applications and tools to build-up a shared operating city "picture", representing the milestone to enable different smart city applications. Among those, *public safety and security* is of critical and paramount importance, and innovations in the IoT will likely convert into safety improvement of citizens. Indeed, nowadays crowded places are subject to increasingly high risk and traditional security approaches, such as searches and checkpoints, have become obsolete while altering public experience and people lifestyle. Then, the development of an environment that presents counter-terrorism to both *cyber* and *physical* aspects of urban systems as a "built-in" feature is the stepping stone toward safety-resilient cities, and represents the hope of ensuring improved security in both effective and acceptable ways [3].

For the mentioned reasons, one key task is represented by detection of radioactive sources using a WSN made of massively-deployed and low-cost sensors, usually Geiger–Muller counters, in densely-populated areas. Indeed, detection of radioactive emissions from nuclear materials constitutes an important objective, given the increasing hazards from potential terrorist activities, e.g. the recognition of *dirty bomb*s, apparently-common explosive devices with release of radioactive isotopes upon explosion (such as Caesium-137, obtainable with considerable ease). Therefore, a relevant scenario is to *proactively* detect radiations having a low-level from this type of sources when they are carried or stored and before they are employed, or, *reactively* detect traces of radiation and quantify their extent after ordinary explosions, ensuring awareness and protection of first responders against the low-level (yet highly hazardous) radiation [4]. Indeed, in both cases, the radiation levels are as low as to be confused with "ordinary" fluctuations of the background radiation and therefore sophisticated algorithmic solutions are required. The benefits (as well as the costs) of using a set of radiation detectors (instead of a single one) for revealing a radioactive source have been investigated and evaluated in [5,6].

For the mentioned reasons, several approaches have been devised in the literature for WSN-based detection of nuclear sources. For example, in [7] the authors consider detection of a single source modeled as time-inhomogeneous process following a Poisson law, embedded in the background radiation using full-precision measurements from a network of nodes, assuming all the parameters *known*. The same setup is studied in [8] to obtain Chernoff bounds for the missed detection and false alarm probabilities, which are employed to design an

optimal motion control approach to drive the sensors, and to obtain the optimal values of the threshold for the decision test, versus the sensor and source trajectories. On the other hand, in [9] the problem of jointly estimating an unknown number of sources and the corresponding locations and intensities is tackled, and two different approaches are proposed therein. The first is based on the generalized maximum likelihood rule, whereas the second approach resorts to parameter estimation in the Bayesian framework (via Monte Carlo integration). Similarly, in [10] the same task is approached via a Bayesian framework and, specifically, source detection is tackled as a model order selection problem where partial Bayes factors (whose efficient computation is accomplished via importance sampling using progressive correction) are employed to weight each potential befitting model. Later, in [11] a search lead by the information gain, including a Bayesian estimator in sequential form coupled with a sensor/observer control unit, is developed and successfully compared to a uniform search along a predefined path. The control unit leads the observers to migrate to new locations and acquires measurements that maximize the information gain in the Renyi divergence sense. Unfortunately, all the above works *require full-precision reporting*, which may be unsuitable in a realistic IoT scenario where inexpensive nodes are usually employed.

Indeed, stringent bandwidth and energy constraints hinder full-precision reporting and, as a consequence, each sensor usually provides one bit to the Fusion Center (FC) regarding the inferred hypothesis. In such a case, the optimal sensor-individual decision procedure (from both Bayesian and Neyman-Pearson standpoints) corresponds to the local Likelihood-Ratio Test (LRT) being quantized into one-bit [12,13]. Unfortunately, the design of quantization thresholds has an exponential complexity and, equally important, sensor LRT evaluation is precluded by ignorance of the parameters of the source to be detected [14]. Hence, the bit reported is either the outcome of a quantization made via a "dumb" rationale [15,16] or exemplifies the inferred binary-valued event (obtained via a sub-optimal detection statistic [17]). In both situations, FC gathers sensors bits and fuses them via a wisely-designed rule to improve (single-)sensor detection capability. Under the assumption of conditional independence, the optimum decision statistic is a weighted sum of decisions, *with weights being function of all the parameters specifying the unknown radiation source* [2]. Then, simple fusion strategies, such as the well-known counting rule or based on simplified design assumptions (for the sensing model), have been initially devised to circumvent such unavailability [18–21]. For example, referring to the radioactive source detection problem, in [22] a two-step approach is proposed, based on (*i*) geometric sensor localization methods employed to estimate the location and strength of the source and (*ii*) a sensor-specific test via the sequential probability ratio based on estimated parameters and new measurements to decide individually the source presence. Finally, decisions from different sensors are finally combined through a majority voting rationale. Similarly, in [23] one- and two-sided Wald-statistics at each sensor are adopted, whereas at the FC the counting rule is applied to combine the sensors' hard decisions.

On the other hand, in [24], the authors derive a test for fusing of correlated decisions and obtaining the sensor (optimal) thresholds for the case of two sensors via *copula-theory*, while a counterpart for the general case of N sensors is provided in [25]. Differently, in [26] a hierarchical Poisson-Gamma law is used to model the probability mass function of the received sensors' counts, and the nodes are assumed to adopt threshold-based (viz. deterministic) binary quantizers sending a decision vector during time to the FC for improved decision-making. Resorting to the hypothesized sensing model, the authors propose a Generalized Likelihood Ratio Test (GLRT) based on a Maximum Likelihood Estimator (MLE) in *constrained form* as the relevant FC decision statistic. Indeed, when the model is parametrically specified, the FC is in charge to tackle a composite test of hypotheses and the GLRT is usually taken as the natural design solution [27]. Indeed, GLRT-based fusion of quantized data has been extensively studied in WSN literature [16,28,29] for target detection in the following cases: (*i*) a cooperative target with unknown location, (*ii*) an cooperative target modelled by observation coefficients assumed known, and (*iii*) an unknown source at unknown position. The last scenario represents the most interesting and suitable for nuclear source detection, as it requires the *least source knowledge*. A different approach is instead pursued in [30], where a Locally Optimum Detection (LOD) scheme is proposed for detecting a radioactive weak source embedded in background clutter. Based on the latter rationale, a decentralized approach, exploiting the Alternating Direction Method of Multipliers (ADMM), is developed for a totally-distributed WSN (no FC), and a (low-overhead) ADMM algorithm robust to attacks, consisting in data falsification, demonstrated. Unfortunately, although appealing (as they attempt to solve the resulting composite hypothesis testing), the aforementioned two methods do not take explicitly into account source location when formulating the hypothesis testing.

Accordingly, in this paper we study detection, in decentralized fashion, of a nuclear radioactive source (*i*) having a spatially-dependent signature, (*ii*) with unknown location and (*iii*) with emitted intensity modeled as deterministic (non-random) but not known [31]. We investigate a system consisting of a network of low-cost radiation counters (such as Geiger–Müller) which collaboratively operate to detect the presence of the radioactive source. Specifically, when the source is present, each sensor records an (unknown) emitted Poisson-distributed radiation count with a rate decreasing with the (unknown) sensor-source distance, according to a known Intensity Attenuation Function (IAF), embedded in a known Poisson background and Gaussian measurement noise. Each node transmits a single bit version to a FC, over noisy reporting channels (modelled as Binary Symmetric Channels, BSCs, and emulating low-energy communications), having the task of a global (more accurate) decision output. The FC employs the Generalized LOD (G-LOD) test [31,32] as a lower-complexity alternative to GLRT, and a *novel* quantizer threshold design is proposed herein, based on a *heuristic rationale* developed resorting to the performance of Position-Clairvoyant (PC) LOD in asymptotic form. The resulting design is *sensor-individual*, considers the channel status between each sensor

and the FC, and depends upon neither the source intensity nor its position, thus allowing *offline* computation. Finally, numerical results provide a comparison of the aforementioned rules (while determining the performance drop with respect to PC LOD) vs. the sensor thresholds and test our design in a radioactive source scenario.

The paper organization is the following: Sect. 2 states the considered problem; Sect. 3 develops GLR and G-LOD tests for the considered problem; then, in Sect. 4 we focus on optimization of the quantizer; numerical results are reported and discussed in Sect. 5; finally, concluding remarks (with a mention to further avenues of research) are provided in Sect. 6.

List of commonly-employed notations - bold letters in lower-case indicate vectors, with a_n representing the nth component of $\boldsymbol{a}$; $\mathbb{E}\{\cdot\}$ and $(\cdot)^T$ are the expectation and transpose operators, respectively; the unit (Heaviside) step function is denoted with; $p(\cdot)$ and $P(\cdot)$ differentiate probability density functions (pdf) and probability mass functions (pmf), respectively; we denote a Gaussian pdf having mean μ and variance σ^2 with $\mathcal{N}(\mu, \sigma^2)$ is used to; $\mathcal{Q}(\cdot)$ (resp. $p_{\mathcal{N}}(\cdot)$) denotes the complement of the cumulative distribution function (resp. the pdf) of a normal random variable in its standard form, i.e. $\mathcal{N}(0, 1)$; finally, the symbol $\sim$ (resp. $\overset{a}{\sim}$) corresponds to "distributed as" (resp. to "asymptotically distributed as").

2 Problem Statement

We focus on a binary test of hypotheses in which a set of nodes $k \in \mathcal{K} \triangleq \{1, \ldots, K\}$ is displaced to monitor a given area to decide the absence ($\mathcal{H}_0$) or presence ($\mathcal{H}_1$) of a radioactive source with an isotropic radiation pattern, incompletely-specified spatial signature, and intensity attenuation depending on the sensor-source distance, namely:

$$\begin{cases} \mathcal{H}_0 : & z_k = b_k + w_k \\ \mathcal{H}_1 : & z_k = c_k + b_k + w_k \end{cases}, \tag{1}$$

where $z_k \in \mathbb{R}$ is the observation of kth sensor and $w_k \sim \mathcal{N}(0, \sigma^2_{w,k})$ is the measurement noise. In view of the spatial separation of the sensors, we hypothesize that the contribution due to noise terms w_ks are statistically independent. On the other hand, the terms b_k and c_k denote the background and the source radiation counts, respectively. More specifically, the radiation count of the background is assumed to obey a Poisson distribution with a rate λ_b (*known*), independent of w_k. Similarly, the count of source radiation at kth sensor is assumed to follow a Poisson law with parameter $\lambda_{c,k}$, and can be expressed through the parametric form

$$\lambda_{c,k}(\theta, \boldsymbol{x}_T) = \theta\, g^2(\boldsymbol{x}_T, \boldsymbol{x}_k) \triangleq \frac{\theta}{1 + \|\boldsymbol{x}_T - \boldsymbol{x}_k\|^2}, \tag{2}$$

where θ denotes the source intensity, here assumed *unknown* (which well suits a realistic scenario) and *deterministic*, and $g^2(\boldsymbol{x}_T, \boldsymbol{x}_k)$ has the meaning of an

IAF. In this work, $\boldsymbol{x}_T \in \mathbb{R}^d$ is employed to denote the *unknown* source position, whereas $\boldsymbol{x}_k \in \mathbb{R}^d$ refers to the *known* position of kth sensor, with the couple $(\boldsymbol{x}_T, \boldsymbol{x}_k)$ giving the value of $g(\boldsymbol{x}_T, \boldsymbol{x}_k)$, being the IAF modelling the considered radioactive source.

Literature has shown that the considered signal model is well approximated with a Gaussian pdf [24], based on the Central Limit Theorem. Accordingly, we have:

$$\begin{cases} z_k|\mathcal{H}_0 \sim & \mathcal{N}(\lambda_b,\, \lambda_b + \sigma_{w,k}^2) \\ z_k|\mathcal{H}_1 \sim & \mathcal{N}(\lambda_{c,k}(\boldsymbol{x}_T) + \lambda_b,\, \lambda_{c,k}(\boldsymbol{x}_T) + \lambda_b + \sigma_{w,k}^2) \end{cases}, \tag{3}$$

Then, to cope with stringent energy and bandwidth budgets in realistic IoT scenarios, the kth sensor quantizes z_k within one bit of information, i.e. $d_k \triangleq u\,(z_k - \tau_k)$, $k \in \mathcal{K}$, where τ_k represents the quantizer threshold. For the sake of simplicity, we confine the focus of this paper to deterministic quantizers; the use and the analysis of the more general stochastic quantizers falls out the scope of our paper and left for future studies. Additionally, with the aim of modeling a reporting phase with constrained energy, we assume that kth sensor bit d_k is transmitted over a BSC and the FC, due to non-ideal transmission, observes an error-prone form $\hat{d}_k = d_k$ (resp. $\hat{d}_k = (1 - d_k)$) with probability $(1 - P_{\mathrm{e},k})$ (resp. $P_{\mathrm{e},k}$), which are here collected within $\widehat{\boldsymbol{d}} \triangleq \left[\hat{d}_1 \cdots \hat{d}_K\right]^T$. In this paper $P_{\mathrm{e},k}$ represents the bit-error probability of kth link, assumed known.

In view of the aforementioned assumptions, the bit probability under $\mathcal{H}_1$ is given by

$$\alpha_k(\theta, \boldsymbol{x}_T) \triangleq (1 - P_{e,k})\beta_k(\theta, \boldsymbol{x}_T) + P_{e,k}(1 - \beta_k(\theta, \boldsymbol{x}_T)), \tag{4}$$

where $\beta_k(\theta, \boldsymbol{x}_T) \triangleq \mathcal{Q}([\tau_k - \lambda_b - \theta g(\boldsymbol{x}_T, \boldsymbol{x}_k)^2]/\sqrt{\sigma_{w,k}^2 + \lambda_b + \theta g(\boldsymbol{x}_T, \boldsymbol{x}_k)^2})$. On the other hand, the bit probability under $\mathcal{H}_0$ is obtained as $\alpha_{k,0} \triangleq \alpha_k(\theta = 0, \boldsymbol{x}_T)$ (see Eq. (4)), thus giving:

$$\alpha_{k,0} = (1 - P_{e,k})\beta_{k,0} + P_{e,k}(1 - \beta_{k,0}), \tag{5}$$

where $\beta_{k,0} \triangleq \mathcal{Q}([\tau_k - \lambda_b]/\sqrt{\lambda_b + \sigma_{w,k}^2})$.

We highlight that the *unknown* source position $\boldsymbol{x}_T$ can be *observed* at FC *only* when the expected intensity $\theta > \theta_0$ ($\theta_0 \triangleq 0$). Thus, we cast the problem as *a one-sided test where parameters of nuisance ($\boldsymbol{x}_T$) are observable only under* $\mathcal{H}_1$ [32], where $\{\mathcal{H}_0, \mathcal{H}_1\}$ corresponds to $\{\theta = \theta_0, \theta > \theta_0\}$. The objective of our study is tantamount to a simple test derivation (from a computational standpoint) deciding in favour of $\mathcal{H}_0$ (resp. $\mathcal{H}_1$) when the statistic $\Lambda(\hat{\boldsymbol{d}})$ is below (resp. above) the threshold γ_{fc}, and the design of the quantizer (i.e. an optimized τ_k, $k \in \mathcal{K}$) for each sensor.

We evaluate the system performance of the FC adopting the generic decision statistic Λ through the detection ($P_{\mathrm{D}} \triangleq \Pr\{\Lambda > \gamma_{\mathrm{fc}}|\mathcal{H}_1\}$) and false alarm ($P_{\mathrm{F}} \triangleq \Pr\{\Lambda > \gamma_{\mathrm{fc}}|\mathcal{H}_0\}$) probabilities, respectively.

3 Fusion Rules

The GLR represents a widespread technique for composite hypothesis testing [29], and its implicit form is given by

$$\Lambda_{\mathrm{GLR}}(\hat{\boldsymbol{d}}) \triangleq 2 \ln \left[\frac{P(\hat{\boldsymbol{d}}; \hat{\theta}_1, \widehat{\boldsymbol{x}}_T)}{P(\hat{\boldsymbol{d}}; \theta_0)} \right], \tag{6}$$

where $P(\hat{\boldsymbol{b}}; \theta, \boldsymbol{x}_T)$ represents the likelihood as a function of $(\theta, \boldsymbol{x}_T)$. On the other hand, $(\hat{\theta}_1, \widehat{\boldsymbol{x}}_T)$ are the *Maximum Likelihood* (ML) *estimates* under $\mathcal{H}_1$, i.e.

$$(\hat{\theta}_1, \widehat{\boldsymbol{x}}_T) \triangleq \arg \max_{(\theta, \boldsymbol{x}_T)} P(\hat{\boldsymbol{d}}; \theta, \boldsymbol{x}_T), \tag{7}$$

with $\ln P(\hat{\boldsymbol{d}}; \theta, \boldsymbol{x}_T)$ being the logarithm of the likelihood function as a function of $(\theta, \boldsymbol{x}_T)$, whose explicit form is [29,31]

$$\sum_{k=1}^{K} \left\{ \widehat{d}_k \ln \left[\alpha_k(\theta, \boldsymbol{x}_T) \right] + (1 - \widehat{d}_k) \ln \left[1 - \alpha_k(\theta, \boldsymbol{x}_T) \right] \right\}, \tag{8}$$

and an analogous expression holds for $\ln P(\hat{\boldsymbol{d}}; \theta_0)$ by replacing the term $\alpha_k(\theta, \boldsymbol{x}_T))$ with $\alpha_{k,0}$.

It is clear from Eq. (6) that Λ_{GLR} requires an optimization problem to be tackled. Sadly, an explicit expression for the couple $(\hat{\theta}_1, \widehat{\boldsymbol{x}}_T)$ is not available. This increases GLR computational complexity, usually involving an *approach based on grid discretization* on $(\theta, \boldsymbol{x}_T)$, see e.g. [29].

On the other hand, Davies' work represents an alternative approach for capitalizing the one-sided nature of the considered hypothesis test [32], allowing to generalize score-based tests to the more challenging scenario of nuisance parameters observed only under $\mathcal{H}_1$. In fact, score tests are based on ML estimates of nuisances under $\mathcal{H}_0$ [27], that sadly cannot be obtained, because they are not *observable* in our case. In detail, if $\boldsymbol{x}_T$ were available, the LOD would represent an effective, yet simple, fusion statistic for the corresponding problem testing a one-sided hypothesis [27]. Unfortunately, since $\boldsymbol{x}_T$ is not known in the present setup, we rather obtain a *LOD statistics family*by varying such parameter. Such technical difficulty is overcome by Davies through the use of the *maximum of the family* as the decision statistic, that is:

$$\Lambda_{\mathrm{GLOD}}\left(\widehat{\boldsymbol{d}}\right) \triangleq \max_{\boldsymbol{x}_T} \frac{\left. \frac{\partial \ln \left[P(\widehat{\boldsymbol{d}}; \theta, \boldsymbol{x}_T) \right]}{\partial \theta} \right|_{\theta=\theta_0}}{\sqrt{I(\theta_0, \boldsymbol{x}_T)}}, \tag{9}$$

where $I(\theta, \boldsymbol{x}_T) \triangleq \mathbb{E}\left\{ \left(\partial \ln \left[P(\widehat{\boldsymbol{d}}; \theta, \boldsymbol{x}_T) \right] / \partial \theta \right)^2 \right\}$ represents the Fisher Information (FI) assuming $\boldsymbol{x}_T$ *known*. Henceforth, the above decision test will be referred to as *Generalized LOD* (G-LOD), to highlight the usage of LOD as the basic

statistic within the umbrella proposed by Davies [31]. The closed form of Λ_{GLOD} is drawn resorting to the explicit forms of the score function and the FI, as stated via the following corollaries, whose proof is omitted for brevity.

Corollary 1. *The score function* $\partial \ln \left[P(\widehat{\boldsymbol{d}}; \theta, \boldsymbol{x}_T)\right] / \partial\theta$ *for the considered radioactive source model is given explicitly as:*

$$\frac{\partial \ln \left[P\left(\widehat{\boldsymbol{d}}; \theta, \boldsymbol{x}_T\right)\right]}{\partial \theta} = \sum_{k=1}^{K} \left\{ (1 - 2P_{e,k})\, g^2(\boldsymbol{x}_T, \boldsymbol{x}_k)\, p_{\mathcal{N}} \left(\frac{\tau_k - \lambda_b - \theta g^2(\boldsymbol{x}_T, \boldsymbol{x}_k)}{\sqrt{\sigma_{w,k}^2 + \lambda_b + \theta g^2(\boldsymbol{x}_T, \boldsymbol{x}_k)}} \right) \times \frac{2\sigma_{w,k}^2 + \lambda_b + \theta g^2(\boldsymbol{x}_T, \boldsymbol{x}_k) + \tau_k}{2\left(\sigma_{w,k}^2 + \lambda_b + \theta g^2(\boldsymbol{x}_T, \boldsymbol{x}_k)\right)^{3/2}} \right\} \tag{10}$$

Proof. The proof can be obtained analogously as [31] by exploiting in the derivative calculation the separable form expressed by Eq. (8).

Corollary 2. *The FI* $I(\theta, \boldsymbol{x}_T)$ *for the considered radioactive source model has the following closed form:*

$$I(\theta, \boldsymbol{x}_T) = \sum_{k=1}^{K} \psi_k(\theta, \boldsymbol{x}_T)\, g^4(\boldsymbol{x}_T, \boldsymbol{x}_k), \tag{11}$$

where the following auxiliary notation has been employed

$$\psi_k(\theta, \boldsymbol{x}_T) \triangleq \frac{(1 - 2P_{\mathrm{e},k})^2}{\alpha_k(\theta, \boldsymbol{x}_T)\left[1 - \alpha_k(\theta, \boldsymbol{x}_T)\right]} \times \frac{\left\{2\sigma_{w,k}^2 + \lambda_b + \theta g^2(\boldsymbol{x}_T, \boldsymbol{x}_k) + \tau_k\right\}^2}{4\left(\sigma_{w,k}^2 + \lambda_b + \theta g^2(\boldsymbol{x}_T, \boldsymbol{x}_k)\right)^3}\, p_{\mathcal{N}}^2 \left(\frac{\tau_k - \lambda_b - \theta g(\boldsymbol{x}_T, \boldsymbol{x}_k)^2}{\sqrt{\sigma_{w,k}^2 + \lambda_b + \theta g(\boldsymbol{x}_T, \boldsymbol{x}_k)^2}} \right). \tag{12}$$

Proof. The proof can be obtained analogously as [31], exploiting conditional independence of decisions (which implies an additive form for the FI) and similar derivation results as $\frac{\partial \ln\left[P\left(\widehat{\boldsymbol{d}}; \theta, \boldsymbol{x}_T\right)\right]}{\partial \theta}$.

Then, combining the results in (10) and (11), the G-LOD statistic is obtained in the final form as

$$\Lambda_{\mathrm{GLOD}}(\widehat{\boldsymbol{d}}) \triangleq \max_{\boldsymbol{x}_T} \Lambda_{\mathrm{LOD}}(\widehat{\boldsymbol{d}}, \boldsymbol{x}_T), \tag{13}$$

where

$$\Lambda_{\mathrm{LOD}}\left(\widehat{\boldsymbol{d}}, \boldsymbol{x}_T\right) = \frac{\sum_{k=1}^{K} \widehat{\nu}_k(\widehat{d}_k)\, g^2(\boldsymbol{x}_T, \boldsymbol{x}_k)}{\sqrt{\sum_{k=1}^{K} \psi_{k,0}\, g^4(\boldsymbol{x}_T, \boldsymbol{x}_k)}}, \tag{14}$$

is the LOD statistic when $\boldsymbol{x}_T$ is assumed known, and we have employed $\widehat{\nu}_k(\widehat{d}_k) \triangleq (\widehat{d}_k - \alpha_{k,0})\, \Xi_k$, $\psi_{k,0} \triangleq \alpha_{k,0}\,(1 - \alpha_{k,0})\, \Xi_k^2$ and

$$\Xi_k \triangleq \frac{(1 - 2P_{\mathrm{e},k})\left[2\sigma_{w,k}^2 + \lambda_b + \tau_k\right]}{\alpha_{k,0}\,(1 - \alpha_{k,0})\; 2\left(\sigma_{w,k}^2 + \lambda_b\right)^{3/2}}\, p_{\mathcal{N}} \left(\frac{\tau_k - \lambda_b}{\sqrt{\sigma_{w,k}^2 + \lambda_b}} \right), \tag{15}$$

as compact auxiliary definitions. We motivate the attractiveness of G-LOD with a *lower (*resp. a *simpler) complexity (resp. implementation),* as we do not require $\hat{\theta}_1$, and only a grid search with respect to $\boldsymbol{x}_T$ is imposed, that is

$$\Lambda_{\mathrm{GLOD}}(\widehat{\boldsymbol{d}}) \approx \max_{i=1,\ldots N_{\boldsymbol{x}_T}} \Lambda_{\mathrm{LOD}}(\widehat{\boldsymbol{d}}, \boldsymbol{x}_T[i]). \tag{16}$$

Hence, the complexity of its implementation scales as $\mathcal{O}(K\, N_{\boldsymbol{x}_T})$, which implies a *significant reduction of complexity* with respect to the GLR (corresponding to $\mathcal{O}(K\, N_{\boldsymbol{x}_T}\, N_\theta)$).

It is evident that Λ_{GLOD} (the same applies to Λ_{GLR}, see Eqs. (6) and (14)) depends on τ_k (via the terms $\widehat{\nu}_k(\hat{d}_k)$ and $\psi_{k,0}$), $k \in \mathcal{K}$, (the threshold set is gathered within $\boldsymbol{\tau} \triangleq \left[\tau_1 \cdots \tau_K\right]^T$) which can be *designed* to optimize performance. Accordingly, in Sect. 4 we accomplish this purpose.

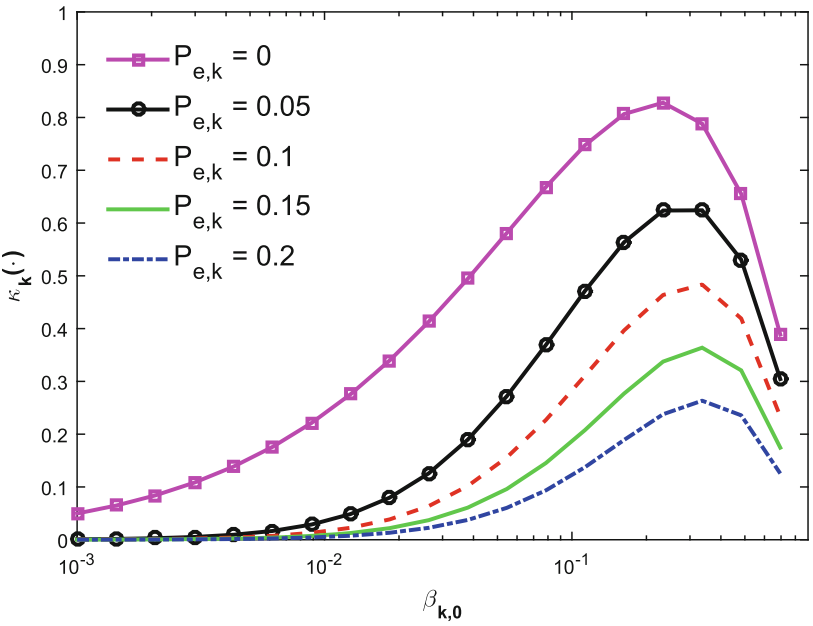


Fig. 1. Sensor threshold optimization: objective function $\kappa_k(\beta_{k,0})$ is illustrated for sensor parameters $(\sigma^2_{w,k}, \lambda_b) = (1, 1)$ and different $P_{\mathrm{e},k}$ values.

4 Design of Quantizers

We point out that the rationale in [16,33] cannot be applied to design (asymptotically-) optimal deterministic quantizers, since no closed-form performance expressions exist for tests built upon Davies approach [32]. In view of this

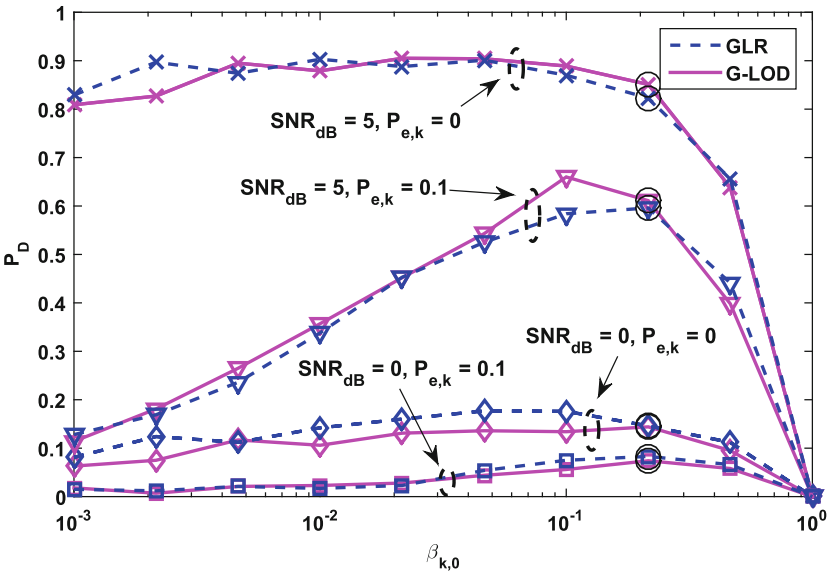


Fig. 2. $P_{\rm D}$ vs. $\beta_{k,0} = \beta_0$, when the FC false-alarm probability is set to $P_{\rm F} = 0.01$. A WSN with $K = 100$ sensors, having sensing SNR $\in \{0, 5\}$ dB, is considered. Corresponding decisions are sent over BSCs with $P_{{\rm e},k} \in \{0, 0.1\}$. Circled markers correspond to (optimized) $P_{\rm D}(\boldsymbol{\tau}^\star)$.

reason, we use a *modified rationale with respect to* [16,33] (that is resorting to a heuristic, yet intuitive, basis) and demonstrate its effectiveness in Sect. 5 through simulations, as done for uncooperative target detection problems in [15,34]. In detail, it is well known that the (position $\boldsymbol{x}_T$) clairvoyant LOD statistic $\Lambda_{\rm LOD}$ is distributed (under an asymptotic, weak-signal, assumption[1]) as [27]

$$\Lambda_{\rm LOD}(\boldsymbol{x}_T, \boldsymbol{\tau}) \overset{a}{\sim} \begin{cases} \mathcal{N}(0,1) & \text{under} \quad \mathcal{H}_0 \\ \mathcal{N}\left(\delta_Q(\boldsymbol{x}_T, \boldsymbol{\tau}), 1\right) & \text{under} \quad \mathcal{H}_1 \end{cases}, \tag{17}$$

where the deflection[2] $\delta_Q(\boldsymbol{x}_T, \boldsymbol{\tau}) \triangleq (\theta_1 - \theta_0)\sqrt{\mathrm{I}(\theta_0, \boldsymbol{x}_T, \boldsymbol{\tau})}$ (underlining dependence on $\boldsymbol{x}_T$ and $\boldsymbol{\tau}$) is given as:

$$\delta_Q(\boldsymbol{x}_T, \boldsymbol{\tau}) = \theta_1 \sqrt{\sum_{k=1}^{K} \psi_{k,0}(\tau_k)\, g^4(\boldsymbol{x}_T, \boldsymbol{x}_k)}, \tag{18}$$

with θ_1 representing the *true value* when $\mathcal{H}_1$ holds true. Obviously, the higher $\delta_Q(\boldsymbol{x}_T, \boldsymbol{\tau})$, the better the performance for the $\boldsymbol{x}_T-$clairvoyant LOD test is

[1] That is $|\theta_1 - \theta_0| = c/\sqrt{K}$ for a certain value $c > 0$ [27].

[2] By doing a slight notation abuse, we adopt the notation $\mathrm{I}(\theta, \boldsymbol{x}_T, \boldsymbol{\tau})$ (resp. $\psi_{k,0}(\tau_k)$ and $\beta_{k,0}(\tau_k)$) in the place of $\mathrm{I}(\theta, \boldsymbol{x}_T)$ (resp. $\psi_{k,0}$ and $\beta_{k,0}$) to highlight their parametric dependence on the local thresholds τ_k's.

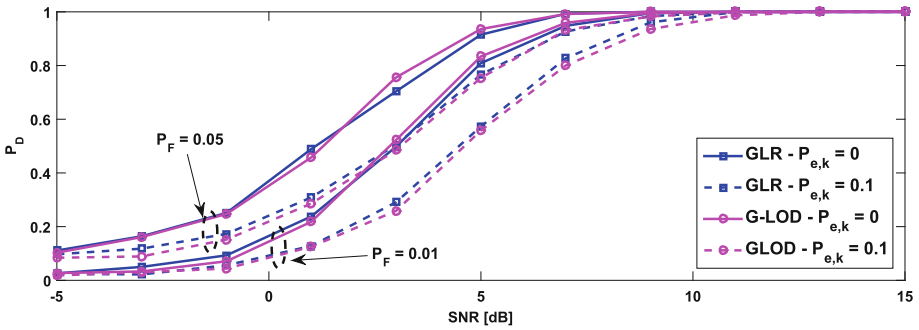


Fig. 3. P_D vs. sensing SNR (dB), when the FC false-alarm probability is set to $P_{\mathrm{F}} \in \{0.05, 0.01\}$. A WSN with $K = 100$ sensors is considered, with sensor thresholds set as $\tau_k = \tau_k^\star$ and whose decisions are sent over BSCs with $P_{e,k} = P_e \in \{0, 0.1\}$

expected to perform in a one-sided test which assumes that the source to be revealed is placed at $\boldsymbol{x}_T$. We highlight that, for one-sided testing, no immediate expressions in the literature are found for GLRT in the asymptotic case [27]. Nonetheless, for the sake of a complete comparison, our proposed design will be also applied to the GLRT in the numerical results later reported in Sect. 5.

For the mentioned reasons, we design vector of thresholds $\boldsymbol{\tau}$ to maximize $\delta_Q(\boldsymbol{x}_T, \boldsymbol{\tau})$, namely

$$\boldsymbol{\tau}^\star \triangleq \arg\max_{\boldsymbol{\tau}} \ \delta_Q(\boldsymbol{x}_T, \boldsymbol{\tau}) \tag{19}$$

In the general case such approach would imply a $\boldsymbol{\tau}^\star$ which depends on $\boldsymbol{x}_T$ (hence not practical). Still, for this particular problem, the optimization requires only the solution of K *decoupled* threshold designs (hence the optimization complexity presents a linear scale with the number of sensors K), being also *independent of* $\boldsymbol{x}_T$ (cf. Eq. (18)), that is:

$$\max_{\tau_k} \left\{ \psi_{k,0}(\tau_k) = \frac{\left(2\sigma_{w,k}^2 + \lambda_b + \tau_k\right)^2}{4\left[\sigma_{w,k}^2 + \lambda_b\right]^3} \frac{p_{\mathcal{N}}^2\left(\left[\tau_k - \lambda_b\right] / \sqrt{\sigma_{w,k}^2 + \lambda_b}\right)}{\Delta_k + \beta_{k,0}(\tau_k)\left[1 - \beta_{k,0}(\tau_k)\right]} \right\}, \tag{20}$$

where $\Delta_k \triangleq [P_{\mathrm{e},k}\,(1 - P_{\mathrm{e},k})]/(1 - 2P_{\mathrm{e},k})^2$. Examples of the objective $\psi_{k,0}(\tau_k)$ for a number of $P_{\mathrm{e},k}$ (or, equivalently, Δ_k) values are illustrated in Fig. 1. We reformulate such maximization (being in *bijective* correspondence with τ_k through the relationship $\tau_k = \mathcal{Q}^{-1}(\beta_{k,0})\sqrt{\lambda_b + \sigma_{w,k}^2} + \lambda_b$), as

$$\kappa_k(\beta_{k,0}) = \frac{p_{\mathcal{N}}^2\left(\mathcal{Q}^{-1}(\beta_{k,0})\right)\left[1 + \frac{1}{2}\mathcal{Q}^{-1}(\beta_{k,0}) / \sqrt{\sigma_{w,k}^2 + \lambda_b}\right]^2}{\Delta_k + \beta_{k,0}\,(1 - \beta_{k,0})} \tag{21}$$

Examples of the objective $\kappa_k(\beta_{k,0})$ for a number of $P_{\mathrm{e},k}$ (viz. Δ_k) values are illustrated in Fig. 1. Therefore, we can be evaluate the optimized $\tau_k^\star$ (viz. $\beta_{k,0}^\star$) via a simple numerical search on a 1-D grid.

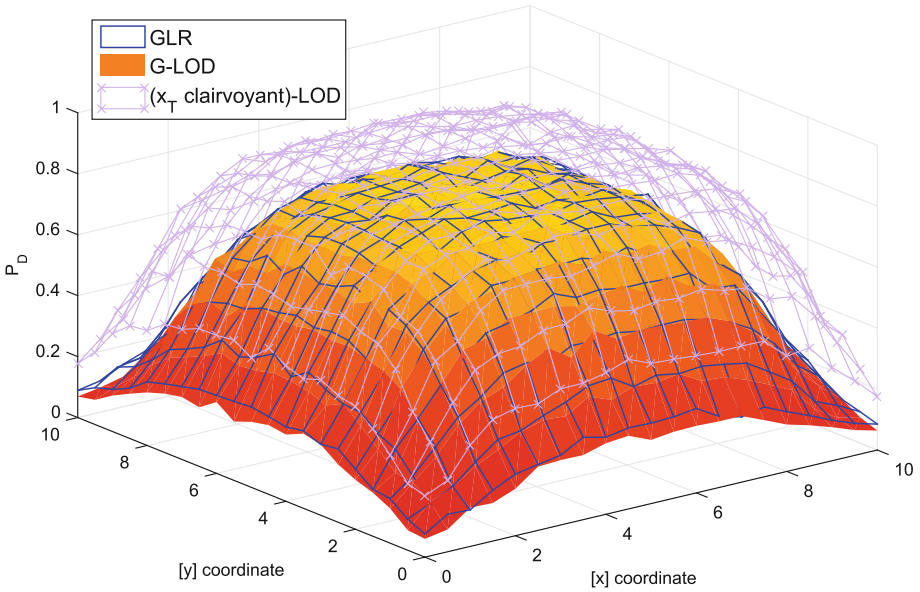


Fig. 4. Profile of P_{D} vs. source position $\boldsymbol{x}_T$, when the FC false-alarm probability is set to $P_{\mathrm{F}} = 0.01$. A WSN with $K = 100$ sensors, having sensing SNR $= 5\,\mathrm{dB}$, is considered. Corresponding decisions are sent over BSCs with $P_{\mathrm{e},k} = 0.1$. Finally, $\beta_{k,0}$'s are selected to maximize Eq. (21).

5 Numerical Results

Henceforth we delve into performance investigation of both tests and we assess the threshold-optimization developed in Sect. 4. With this aim, a 2-D area ($\boldsymbol{x}_T \in \mathbb{R}^2$) is considered, in which the presence of a radioactive source in the surveillance region $\mathcal{A} \triangleq [0, 10]^2$ (i.e. a square) is monitored by a WSN composed of $K = 100$ sensor nodes. For the sake of a simplicity, we arrange sensors on a uniform square grid which covers the whole $\mathcal{A}$. Regarding the model assumptions for the sensing phase, we hypothesize $w_k \sim \mathcal{N}(0, \sigma_w^2)$, $k \in \mathcal{K}$ and, for simplicity, we set $(\sigma_w^2, \lambda_b) = (1, 1)$. Lastly, we define $\mathrm{SNR} \triangleq 10\,\log_{10}[\theta/(\sigma_w^2 + \lambda_b)]$ the source Signal-To-Noise Ratio (SNR), to measure the strength of the radioactive source to be detected.

According to Sect. 3, the implementation of Λ_{GLR} and Λ_{GLOD} resorts to search on grids for θ and $\boldsymbol{x}_T$. First, the search space of θ is constrained within $S_\theta \triangleq \left[0, \bar{\theta}\right]$, where $\bar{\theta} > 0$ corresponds to $\mathrm{SNR} = 20\,\mathrm{dB}$. The vector collecting the points on the grid is then defined as $\left[0\ \boldsymbol{g}_\theta^T\right]^T$, with $\boldsymbol{g}_\theta$ including the intensity values which correspond to the sampling of the SNR (dB) $-10 : 2.5 : 20$ (leading to $N_\theta = 12$). Secondly, the search support of $\boldsymbol{x}_T$ is naturally assumed to be coincident with the monitored area, i.e. $\mathrm{S}_{\boldsymbol{x}_T} = \mathcal{A}$. Accordingly, the 2-D grid is the result of sampling $\mathcal{A}$ uniformly with $N_{\boldsymbol{x}_T} = N_c^2$ points, where $N_c = 100$.

In this setup, the evaluation of G-LOD requires $N_c^2 = 10^4$ grid points, as opposed to $N_c^2 N_\theta = 1.2 \times 10^5$ points for GLR, i.e. a *more than tenfold complexity decrease*.

Initially, in Fig. 2 we report P_{D} (under $P_{\mathrm{F}} = 0.01$, SNR $\in \{0, 5\}$ dB and $P_{\mathrm{e},k} \in \{0, 0.1\}$) vs. the same *local (*sensor*)* bit-activation probability $\beta_{k,0} = \beta_0$, $k \in \mathcal{K}$ (enforced via a common threshold choice) so as to assess sensitivity w.r.t. sensor quantizer threshold. At each run, the source position $\boldsymbol{x}_T$ is drawn randomly based on a *uniformly-distributed pdf* within the region $\mathcal{A}$. The optimized threshold $\tau_k^\star$ (viz. $\beta_{k,0}^\star$) proposed in Sect. 4 apparently represents a solution of appealing (circled markers highlight the corresponding P_{D}), since the optimal P_{D} of each curve is a function on the *(unknown)* source SNR measured, and since a naively chosen β_0 may imply a high performance loss. This reasoning *also* applies to the GLR.

Then, in Fig. 3 we provide a P_D comparison (for $P_F \in \{0.05, 0.01\}$) of considered rules (for a source whose position is randomly generated within $\mathcal{A}$ at each run, similarly as Fig. 2) versus SNR (dB), in order to assess their detection sensitivity as a function of the dispersion strength of the radioactive source. From inspection of the figure, we concluded that both rules perform very similarly for a different quality of the BSC ($P_{e,k} = P_e \in \{0, 0.1\}$) and over the whole SNR range.

Thirdly, in Fig. 4, we report P_{D} (under $P_{\mathrm{F}} = 0.01$) profile vs. source location $\boldsymbol{x}_T$ (for SNR $= 5$ dB, $P_{\mathrm{e},k} = 0.1$ and optimized $\beta_{k,0}^\star$), to draw a detailed overview of detection capabilities over the whole monitored area $\mathcal{A}$ and underline possibly *blind spots*. Remarkably, G-LOD test performs only *negligibly worse* than the GLRT, and moderately worse in comparison to a test based on PC LOD (given in Eq. (14), having an $\mathcal{O}(K)$ complexity). Unfortunately, the latter assumes *the unrealistic* knowledge of $\boldsymbol{x}_T$ and thus merely constitutes an *upper-limiting bound* on performance achievable. Finally, we notice that both rules have a similar $P_{\mathrm{D}}(\boldsymbol{x}_T)$ profile, and its "shape" highlights a lower detection rate at the edge of the monitored area. Such result can be ascribed to the *regularity* of the WSN arrangement in the monitored area $\mathcal{A}$.

6 Concluding Remarks and Further Directions

In this paper, a distributed scheme using a WSN for detection of a nuclear radioactive source was developed. More specifically, we considered a source emitting an unknown intensity (θ) at unknown position ($\boldsymbol{x}_T$), as an alternative enjoying low-complexity as opposed to the GLRT (which, instead, requires a grid search within the space of $(\theta, \boldsymbol{x}_T)$). Specifically, we generalized the LOD test, based on the rationale proposed by [32]. Furthermore, we developed an effective criterion (originating from performance expressions having a semi-theoretical background) to design sensor thresholds of G-LOD in an optimized fashion, resulting in *a sensor-individual (simple) numerical search* on a *1-D grid*. Remarkably, we capitalized the obtained result so as to optimize both GLR and G-LOD tests performance. Numerical results underlined the close performance of G-LOD test to the GLRT in the scenarios investigated, and a small (yet reasonable) loss of GLOD compared to a test based on PC LOD. Our future work

will consist of investigating design of fusion rules in more challenging radioactive source scenarios, such as inhomogeneous radiation background, multiple moving sources and burstiness (time-correlation) of reporting channels.

References

1. Jin, J., Gubbi, J., Marusic, S., Palaniswami, M.: An information framework for creating a smart city through Internet of Things. IEEE Internet Things J. **1**(2), 112–121 (2014)
2. Varshney, P.K.: Distributed Detection and Data Fusion, 1st edn. Springer, New York (1996). https://doi.org/10.1007/978-1-4612-1904-0
3. Coaffee, J., Moore, C., Fletcher, D., Bosher, L.: Resilient design for community safety and terror-resistant cities. In: Proceedings of the Institution of Civil Engineers-Municipal Engineer, vol. 161, pp. 103–110. Thomas Telford Ltd. (2008)
4. Brennan, S.M., Mielke, A.M., Torney, D.C., MacCabe, A.B.: Radiation detection with distributed sensor networks. IEEE Comput. **37**(8), 57–59 (2004)
5. Brennan, S.M., Mielke, A.M., Torney, D.C.: Radioactive source detection by sensor networks. IEEE Trans. Nucl. Sci. **52**(3), 813–819 (2005)
6. Stephens, D.L., Peurrung, A.J.: Detection of moving radioactive sources using sensor networks. IEEE Trans. Nucl. Sci. **51**(5), 2273–2278 (2004)
7. Pahlajani, C.D., Poulakakis, I., Tanner, H.G.: Networked decision making for Poisson processes with applications to nuclear detection. IEEE Trans. Autom. Control **59**(1), 193–198 (2014)
8. Pahlajani, C.D., Sun, J., Poulakakis, I., Tanner, H.G.: Error probability bounds for nuclear detection: improving accuracy through controlled mobility. Automatica **50**(10), 2470–2481 (2014)
9. Morelande, M., Ristic, B., Gunatilaka, A.: Detection and parameter estimation of multiple radioactive sources. In: 10th International Conference on Information Fusion (FUSION), pp. 1–7 (2007)
10. Morelande, M.R., Ristic, B.: Radiological source detection and localisation using Bayesian techniques. IEEE Trans. Signal Process. **57**(11), 4220–4231 (2009)
11. Ristic, B., Morelande, M., Gunatilaka, A.: Information driven search for point sources of gamma radiation. Signal Process. **90**(4), 1225–1239 (2010)
12. Hoballah, I.Y., Varshney, P.K.: Distributed Bayesian signal detection. IEEE Trans. Inf. Theory **35**(5), 995–1000 (1989)
13. Reibman, A.R., Nolte, L.W.: Optimal detection and performance of distributed sensor systems. IEEE Trans. Aerosp. Electron Syst. **1**, 24–30 (1987)
14. Viswanathan, R., Varshney, P.K.: Distributed detection with multiple sensors - part I: fundamentals. Proc. IEEE **85**(1), 54–63 (1997)
15. Ciuonzo, D., Salvo Rossi, P., Willett, P.: Generalized Rao test for decentralized detection of an uncooperative target. IEEE Signal Process. Lett. **24**(5), 678–682 (2017)
16. Fang, J., Liu, Y., Li, H., Li, S.: One-bit quantizer design for multisensor GLRT fusion. IEEE Signal Process. Lett. **20**(3), 257–260 (2013)
17. Ciuonzo, D., Salvo Rossi, P.: Decision fusion with unknown sensor detection probability. IEEE Signal Process. Lett. **21**(2), 208–212 (2014)
18. Aalo, V.A., Viswanathan, R.: Multilevel quantisation and fusion scheme for the decentralised detection of an unknown signal. IEE Proc. Radar Sonar Navig. **141**(1), 37–44 (1994)

19. Chen, B., Jiang, R., Kasetkasem, T., Varshney, P.K.: Channel aware decision fusion in wireless sensor networks. IEEE Trans. Signal Process. **52**(12), 3454–3458 (2004)
20. Ciuonzo, D., Romano, G., Salvo Rossi, P.: Channel-aware decision fusion in distributed MIMO wireless sensor networks, decode-and-fuse vs. decode-then-fuse. IEEE Trans. Wireless Commun. **11**(8), 2976–2985 (2012)
21. Niu, R., Varshney, P.K.: Performance analysis of distributed detection in a random sensor field. IEEE Trans. Signal Process. **56**(1), 339–349 (2008)
22. Chin, J.R., et al.: Identification of low-level point radioactive sources using a sensor network. ACM Trans. Sens. Netw. (TOSN) **7**(3), 21 (2010)
23. Sen, S., et al.: Performance analysis of Wald-statistic based network detection methods for radiation sources. In: 19th International Conference on Information Fusion (FUSION), pp. 820–827 (2016)
24. Sundaresan, A., Varshney, P.K., Rao, N.S.V.: Distributed detection of a nuclear radioactive source using fusion of correlated decisions. In: 10th IEEE International Conference on Information Fusion (FUSION), pp. 1–7 (2007)
25. Sundaresan, A., Varshney, P.K., Rao, N.S.V.: Copula-based fusion of correlated decisions. IEEE Trans. Aerosp. Electron. Syst. **47**(1), 454–471 (2011)
26. Sundaresan, A., Varshney, P.K., Rao, N.S.V.: Distributed detection of a nuclear radioactive source based on a hierarchical source model. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 2901–2904 (2009)
27. Kay, S.M.: Fundamentals of Statistical Signal Processing, Volume 2: Detection Theory. Prentice Hall PTR, Upper Saddle River (1998)
28. Niu, R., Varshney, P.K.: Joint detection and localization in sensor networks based on local decisions. In: 40th Asilomar Conference on Signals, Systems and Computers, pp. 525–529 (2006)
29. Shoari, A., Seyedi, A.: Detection of a non-cooperative transmitter in Rayleigh fading with binary observations. In: IEEE Military Communications Conference (MILCOM), pp. 1–5 (2012)
30. Kailkhura, B., Ray, P., Rajan, D., Yen, A., Barnes, P., Goldhahn, R.: Byzantine-resilient locally optimum detection using collaborative autonomous networks. In: IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP) (2018)
31. Ciuonzo, D., Salvo Rossi, P.: Distributed detection of a non-cooperative target via generalized locally-optimum approaches. Inf. Fusion **36**, 261–274 (2017)
32. Davies, R.D.: Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika **74**(1), 33–43 (1987)
33. Ciuonzo, D., Papa, G., Romano, G., Salvo Rossi, P., Willett, P.: One-bit decentralized detection with a Rao test for multisensor fusion. IEEE Signal Process. Lett. **20**(9), 861–864 (2013)
34. Ciuonzo, D., Salvo Rossi, P.: Quantizer design for generalized locally optimum detectors in wireless sensor networks. IEEE Wirel. Commun. Lett. **7**(2), 162–165 (2018)

Wemotion: A System to Detect Emotion Using Wristbands and Smartphones

Bao-Lan Le-Quang[1(✉)], Minh-Son Dao[2], and Mohamed Saleem Haja Nazmudeen[2]

[1] University of Information Technology Ho Chi Minh City, Ho Chi Minh City, Vietnam
baolan2005@gmail.com
[2] Universiti Teknologi Brunei, Gadong, Brunei Darussalam
{minh.son,mohamed.saleem}@utb.edu.bn

Abstract. Understanding students' emotion, especially during the classroom time, can help to improve the positive classroom emotional climate towards promoting academic achievement. Unfortunately, most of the exisiting studies that try to understand the emotion of students have just utilized a questionnaire method to discover the link between the classroom emotional climate and academic achievement. Such methods do not reflect exactly the emotion of students in the real-time mode. There are also other studies that leverage hi-tech technologies (e.g. video camera, sensors, smartphones) to capture data generated by people themselves (e.g. physiological data, facial expression, body postures, human-smartphone interaction) to recognize emotion. Nonetheless, these methods build either a general model for all users or an individual model for a selected user leading to having a weak adaptive ability. In this paper, we introduce Wemotion, a smart-sensing system built by smartphones and wristbands that can not only detect students' emotion in real-time mode and also evolve to continuously improve the accuracy during the life cycle. The system is evaluated by real data collected from volunteers and compared to several existing methods. The results show that the system works well and satisfies the purpose of our research.

Keywords: Emotion detection · Wearable sensor · Smartphone
Ubiquitous ambient · Machine learning · Adaptive learning

1 Introduction

Many studies have been carried out to understand the influence of emotion in education during the last decade [1–6]. Along with traditional research methods that make use of questionnaire, methods that pay attention in understanding human activities (e.g. facial expression [7], posture [9], human-smartphone interaction [10,12]) towards detecting their emotion are also utilized to find the link between the classroom emotional climate and academic achievement.

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 92–103, 2019.
https://doi.org/10.1007/978-981-13-5758-9_8

Nevertheless, using physiological data to detect emotion could be considered as highly accurate method since these data reflects exactly the biological reaction of emotions. To collect physiological data from human, wearable sensors need to be attached to a human body. Many systems using wired wearable sensors to detect emotion have been proposed [11,19,24]. Unfortunately, wired sensors do not let people move freely; and with a web of wires, it is difficult to apply these systems into reality. Other systems use wireless wearable sensors but lack of online data transmission protocols and inconvenience for people to wear such sensors for a whole day [13,28]. Recently, wristbands are known as flexible wearable sensors that can measure physiological data, transfer data to a cloud, monitor health status online, even give a good recommendation for healthcare [25,26]. Light, touchable, accurate, cheap, fashionable, and easy to use are significant benefits of these wristbands towards making them popular. Besides, smartphones are also integrated with a lot of sensors. Among these sensors gyroscope and accelerometer sensors totally can help to understand physical activities of human [12,14,25].

Leveraging on these advantages of smartphones and wristbands, we build *Wemotion*, a smart-sensing system using smartphones and wristbands to detect students' emotion towards improving learning-teaching experience in education. A smartphone is used to collect emotional tags and transmit data to a cloud. A wristband is utilized to collect physiological data (e.g. heartbeat, walking step) and send such data to the smartphone. The *Wemotion* has two types of model for detecting emotion: (1) general model that can work for any user who just joins our system, (2) personal model that can personalize the general model to a specific user. The general model is first generated by using a training data collected from a certain number of users. Then, via the interactive learning process, this model will be personalized to be a personal model that fits for an individual user. Periodically, the general model is re-trained again by updating training data collected from personal models. Thus, the accuracy of recognition rate, the diversity and the imbalance of training data will be improved over a period of time. The main contribution in this paper is that our method can personalize a model to increase the accuracy of the system. Moreover, our model continues to evolve for its whole life cycle: the longer people use our system, the more accurate the results are. Whereas, other methods that use wristbands and/or smartphones for detecting, build either a general model for all users or an individual model for a selected user.

The paper is organized as follows: Sect. 2 gives a brief review of existing methods which pay attention to detecting emotion using wristband and/or smartphones; Sect. 3 describes our system; Sect. 4 reports the evaluation of our system as well as the comparison with others, and the last Section concludes our paper and raises the future direction of our research.

2 Related Works

In last decade, we have witnessed many studies on emotion/mood recognition [15–18]. In [8] the authors propose a formula that can recognize students' mood

during an online self-assessment process. The system gets feedback from each student to build a personal formula that is applied to recognize his/her mood. Unfortunately, due to heavy dependence on the individual data, this personal formula can only be applied to an individual student and cannot be extended to a general case. In [9], the authors introduce a framework for recognizing mood by analyzing upper body postures and movements of a human. Nonetheless, this framework is lacking testing with real cases. Moreover, the accuracy of this framework could be doubted due to the upper body postures of human can lead to misclassifying the mood. For example, a staff who is working in an office and listening to music. He/she can have some different mood depending on which music he is listening to, but his/her upper body postures may remain the same. In [28], a system, namely PSYCHE, is introduced. This system uses Heart Rate Variability (HRV) data and data mining algorithms to recognize mood. One of the disadvantages of this system is that users must wear a T-shirt with embedded wearable sensors to collect physiological data. In [12], the system, namely MoodScope, is proposed to recognize mood using data collected from a smartphone such as contact, text message email, application usage, phone call history, web browsing history and GPS data. Unfortunately, the privacy issue and lack of data due to objective reasons can limit this system.

Stress is a specific type of mood and attracts a lot of attention of people due to its effect on human life. In [27] the authors use physiological data to recognize stress in office-like situations. Electrocardiogram (ECG), respiration, skin conductance (SC), and surface electromyogram (sEMG) data collected by wearable sensors are used to recognize stress. The accuracy rate of this method is up to 74.5%. In [25], the authors propose a new method using wearable sensor and smartphone to recognize stress. Accelerometer (ACC), call history, SMS, location history, and screen on/off log data are collected from an Android phone; and skin conductance (SC) data are collected using wearable sensors. Besides, user's profiles are created by gathering data from survey forms. All these data are utilized to train a model that can help to detect stress with acceptance rate. Unfortunately, the accuracy of this method depends on the volume of data the system can collect. In [20], the authors measure stress by using SVM running on HRV data collected from sensors and facial expression detected by the camera. In general, existing approaches for emotion/mood/stress recognition using data collected from smartphones and/or wearable sensors have gained significant results and brought convenience and comfort for users. Nonetheless, models used for recognizing the mood are weak in adapting itself since they have been designed either as general type or individual type leading to the fact that these models limit the scope of application.

3 Wemotion

In this section, we discuss the overview of the *Wemotion* framework including the working environment setup, algorithms and methods used to detect emotion.

3.1 The Framework's Overview

Figure 1a illustrates the overview of the Wemotion system. First, a student is required to wear a wristband (e.g. Xiaomi Miband 1 S illustrated in Fig. 1b [21]), to activate an app in their smartphone, and to start Bluetooth Low Energy (BLE) connection to let the device and the smartphone talk together. The app is built to display and receive tags from a student whenever his/her emotion changes. Second, the *heart-beat* and *step-count* data captured by the wristband will be sent to the smartphone and forwarded to the backend. Third, at the backend, models being trained before are used to detect emotion. The result is sent back to smartphones that are already registered before for displaying. At this stage, the student can re-tag the emotion if he/she feels that the result does not match his/her emotion. This cycle will be looped unlimited times. Occasionally, new versions of models are re-trained again based on the number of re-tags generated by students.

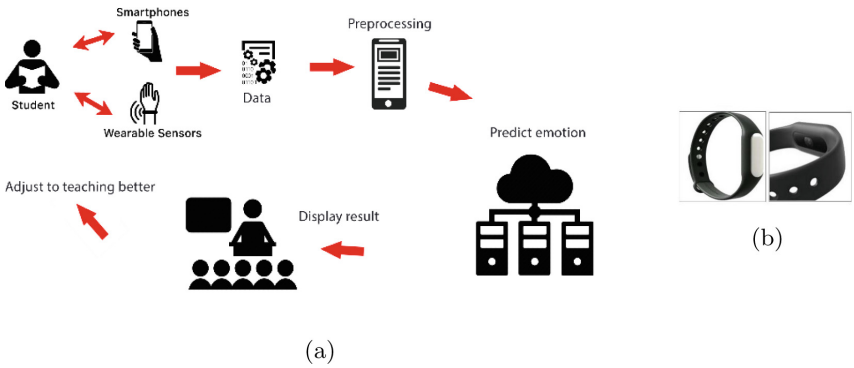


Fig. 1. (a) The framework's overview; (b) Xiaomi Miband 1S

3.2 Data Cleaning

The data received from wristband do not usually in a perfect status. There are a lot of noises due to the quality of the wristband and the Bluetooth connection. Thus, the data needs to be cleaned before sending to the next stage. We use the method proposed in [23] to remove outlier. This method uses Median Absolute Deviation (MAD) described in the Eq. 1. After removing outliers, we sample the polished data by 500 ms and fill missing data by linear interpolation. We run this filter in a smartphone and send the polished data to the backend.

$$M - 3 * MAD < x_i < M + 3 * MAD \tag{1}$$

$$MAD = bM_i(|x_i - M_j(x_j)|) \tag{2}$$

where M and MAD denotes a Median and Median Absolute Deviation, respectively.

3.3 Features Extraction

In [22], Heart Rate Variability (HRV) and Electrodermal Activity (EDA) are used to create features. We leverage this method to extract features for our system usages. Nevertheless, most of the cheap wristbands in the current market can only measure *heart-beat* and *step-count*. In order to adapt this method to our system, we modify the original one as follows:

– Calculating the Inter-beat Interval (IBI) in milliseconds from a *heart-beat* value (HBV) by using Eq. 3

$$IBI = 60000/HBV \tag{3}$$

– Applying the sliding-window whose size is set as W seconds along the time axis. With each sliding step, following values are calculated:
  - *meanIBI*: the mean of IBIs inside the slide window
  - *SDNN*: the standard deviation of NN interval where NN interval is the interval between normal heart-beats (beat-to-beat). In our case, it is IBI.

$$SDNN = \sqrt{\frac{1}{N-1}\sum_{n=2}^{N}[I(n) - \overline{I}]^2} \tag{4}$$

  - *RMSSD*: Root Mean Square of successive differences between two adjacent NNs

$$RMSSD = \sqrt{\frac{1}{N-2}\sum_{n=3}^{N}[I(n) - I(n-1)]^2} \tag{5}$$

  - *pNN50*: NN50/(total of NNs) where NN50 is the number of pairs of successive NNs whose successive differences are larger than 50 ms.
  - *meanStepcount*: Mean of step-counts being counted within the slide window

where I, N and $\overline{I}$ denote NN interval, total heart-beats, and meanIBI, respectively.

The feature vector is formed by the Eq. (6)

$$fv = (meanIBI, SDNN, RMSSD, pNN50, meanStepcount) \tag{6}$$

Table 1. Examples of feature vectors

MeanIBI (ms)	SDNN	RMSSD	pNN50	MeanStep
755.90	23.56	11.63	0.022	0.0
808.02	63.98	14.62	0.033	0.0
774.98	36.16	14.13	0.032	0.0

In other words, after every sliding step, one fv is created by using the equation (6) and sent to the next stage for processing. Table 1 shows a sample of feature vectors in training set.

3.4 Building Classifiers

In order to detect emotion, we create general and personal models that receive the features generated in the previous step as the input and classify such an input to a suitable label. To display the results of emotion detection (i.e. for the testing stage) as well as to manually label the unknown/misclassified emotion (i.e. for training and feedback stage), we develop an app running in a smartphone (Android phone in our case).

General Model: We use Support Vector Machine (SVM) to build the general model (Algorithm 1). Feature vectors {fv} are fed to SVM for training with full labels manually annotated by volunteers. In the testing stage, {fv} with no label is fed to SVM to get the right label.

```
while a user doesn't exit do
    read data from the user;
    send data to system and predict;
    display results to the user;
end
```

Algorithm 1. Algorithm for General Model

Personal Model: Emotion, in general, differs from person to person and situation to situation. Thus, the accuracy of the general model might depend on how big and heterogeneous the collected samples are. Nevertheless, collecting such samples is time-consuming and there is no guarantee that we can collect all necessary samples. Hence, we propose a personal model that is based on the general model and feedback learning schema. First, when a new user joins our system, he/she is assigned his/her personal model, a copy of the general model. Second, the user will join the feedback learning process by correcting misclassified emotion displayed on his/her smartphone. This feedback process will continue until the user satisfies with the results of the system. Figure 2a and Algorithm 2 illustrate how the personal model works.

```
while a user doesn't exit do
    read data from the user;
    send data to the system and predict;
    send result to the user;
    if user has feedback then
        update feedback into the training set for re-training;
    else
        update the result into the training set for re-training;
    end
end
```

Algorithm 2. Algorithm for Personal Model

Fusion Model: Practically, the personal model helps to increase the accuracy of emotional detection. Nevertheless, continuing training for a long time can lead to the overfitting of the model. Since the personal model does not use samples come from other people, some common senses can be misclassified. Thus, we

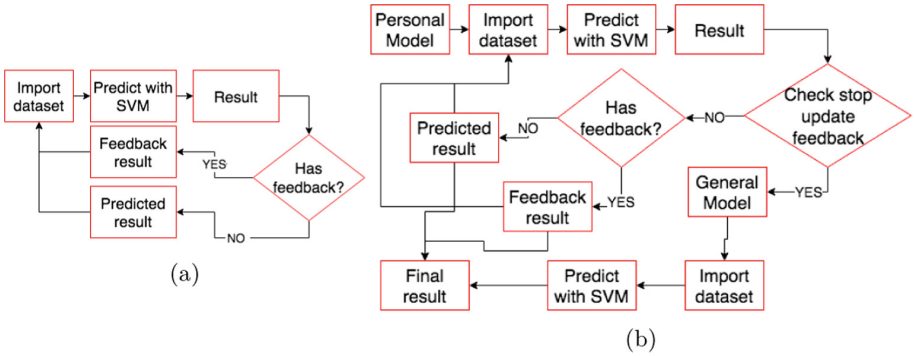


Fig. 2. (a) The workflow of the personal model; (b) The workflow of the fusion model

propose a fusion model that combines both the general model and the personal model. Figure 2a and Algorithm 3 describe how the fusion model works.

```
while user doesn't exit do
    read data from a user;
    send data to the system and predict;
    if the system stops receiving feedback then
        send the result to user;
    else
        if the user has feedback then
            update feedback into the training set for re-training;
        else
            update the result into the training set for re-training;
        end
        send the result to the user;
    end
end
```

Algorithm 3. Algorithm for Fusion Model

The idea behind this model is to capture emotional common senses from people and emotion from an individual. The former could be understood that most people have the same reaction to certain situations (e.g. people can laugh - positive emotion - when listening to a comedy show, people can cry - negative emotion - when attending a funeral). The latter may be considered that the level of emotion can be varied from person to person (e.g. fans of folk songs can have a positive emotion while fans of rock songs can only have a neutral emotion).

3.5 Understanding Emotion and Improving Learning-Teaching Process

In [29], the authors point out the link between classroom emotional climate and academic achievement. Thus, we design a sub-system that can help a teacher

monitoring students' emotion during a class time. By understanding students' emotion, the teacher could adjust his/her classroom emotional climate (e.g. the teaching methods, classmate interaction, quiz, teacher-student interaction) to decrease the portion of negative and neutral emotion inside his/her class.

We build a website that can be accessed by a teacher with his/her account. Each class is assigned the classID. Teachers and students are required to login to a virtual classroom identified by the classID with their ID (i.e. teacherID and studentID). The virtual classroom classID displays the percentage of positive, negative, and neutral emotion inside his/her class during the class time.

4 Experimental Results

To evaluate the *Wemotion*, four volunteers, three male students (user 1, user 2, and user 3) and one female pupil (user 4), are invited to test the system. The Xiaomi Miband 1S (Fig. 1b) and different Android smartphones are used. User 1 joins the system for one and a half month; user 2 uses the system for one day; user 3 tests the system for two months; and user 4 plays with the system for one month.

We investigate in three types of emotion: positive (e.g. fun, desire, happiness, joy, interest, and compassion), negative (e.g. depression, sadness, anger, stress, and anxiety), and neutral. Each user is asked to manually label "positive", "negative", and "neutral" whenever he/she feels his/her emotion changes. This task will repeat for feedback process (i.e. the Wemotion will display a label when the user's emotion changes, and ask the user to correct the label).

We utilize Google Cloud with Intel Haswell CPU, 3.75 GB RAM for setting the working environment. The parameters of SVM are set as default, using the *scikit-learn* library. The sliding window is set by $W = 60$ s.

Tables 2, 3, and 4 describes the results of emotion detection by using the general model, the personal model, and the fusion model, respectively. Table 5 denotes the comparison among these three models.

Table 2. Results of using the general model

	User 1	User 2	User 3	User 4
	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)
Neutral	661 - 4 (99.40%)	23 - 0 (100%)	303 - 41 (88.08%)	121 - 14 (89.63%)
Positive	219 - 39 (84.88%)	63 - 09 (87.50%)	66 - 04 (94.28%)	96 - 12 (88.89%)
Negative	12 - 01 (92.31%)	0 - 0 (0%)	07 - 01 (87.50%)	05 - 03 (62.50%)

Table 3. Results of using the personal model

	User 1	User 2	User 3	User 4
	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)
Neutral	606 - 53 (91.96%)	23 - 0 (100%)	296 - 40 (80.09%)	112 - 14 (88.89%)
Positive	225 - 41 (85.59%)	63 - 09 (87.50%)	74 - 06 (92.5%)	107 - 12 (89.91%)
Negative	10 - 01 (90.91%)	0 - 0 (0%)	05 - 01 (83.33%)	04 - 02 (66.66%)

Table 4. Results of using the fusion model

	User 1	User 2	User 3	User 4
	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)	Correct - Incorrect (Accuracy)
Neutral	586 - 33 (94.67%)	26 - 0 (100%)	284 - 25 (91.91%)	119 - 3 (97.54%)
Positive	262 - 40 (86.75%)	63 - 05 (92.64%)	101 - 02 (98.06%)	117 - 05 (95.90%)
Negative	15 - 0 (100%)	01 - 0 (100%)	09 - 01 (90%)	07 - 0 (100%)

In general, the fusion model generates the best results comparing with others. This confirms our hypothesis of common senses and individual reaction for a certain situation mentioned in Subsect. 3.4.

We use the methods introduced in [26] and in [12] to compare with our prposed method. In [26], Heart Rate Variability (HRV) and Electrodermal Activity (EDA) are used to create features. These data come from expensive devices that are not affordable by pupils. On the other hand, we use Xiaomi Miband 1S, a cheap device and only measure heart-beat and step-count which are very popular for most wristband sold in the current market. In [26], the k-NN, Decision Tree (DT), and Bagged Ensembles of Decision Trees (BE-DT) are leveraged to classify emotion, while we use SVM to classify emotion. In [12], data from smartphone are used to build a "sensor" with personalized, all-user and hybrid models. The personalized model requires some training data for each person, this led a problem if a new user join into their system and it can impact to the result of their hybrid model. So we test our data just with their all-user model.

Table 5. General model vs. personal model and fusion model

	General model	Personal model	Fusion model
User 1	89.95%	89.85%	92.02%
User 1	90.50%	90.53%	94.07%
User 3	89.10%	88.86%	93.36%
User 4	88.45%	88.84%	96.81%

Table 6 denotes the comparison of our method and method introduced in [26]. Clearly, our method can detect emotion with higher accuracy rate comparing to the method introduced in [12,26] though we use a cheap device with poorer information.

Table 6. Comparison with method introduced in [26]

	General model	Personal model	Fusion model	[26]	[12]
Neutral	90.65%	90.65%	94.33%	91.4%	88.1%
Positive	87.40%	87.34%	91.26%	91.61%	93.1%
Negative	82.76%	82.6%	96.97%	78.85%	82.4%

To evaluate the satisfaction of teachers and students when using the *Wemotion* for educational purpose, we conduct a survey to collect idea from them after using our system for a while. Generally, we received a positive feedback of how useful our system can be in education. Figure 3 illustrates the feedback results from teachers and students.

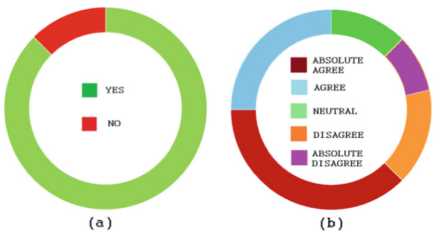


Fig. 3. Feedback from teachers and students when using *Wemotion*. (a) Is Wemotion useful in education? (b) does *Wemotion* help to promote academic achievement?

5 Conclusions

In this paper, we introduce the *Wemotion*, the emotion detection system using smartphones and wristbands. The general, personal, and fusion models are introduced, evaluated, and compared to other methods. We also conducted a survey

to see how good our system is in education. Our system differs from other methods that either build a general model for all or build several personal models for individuals. We build a fusion model that integrates both the general and personal models to overcome the overfitting, common senses and emotion level differences among people. Moreover, our system has the ability to evolve itself during their life cycle towards increasing the accuracy rate occasionally. In the future, we will develop the sub-system of education to understand better the link between monitoring students' emotion during class time, as well as student-student and teacher-student interaction towards proposing a good policy for emotional classroom climate.

References

1. Pekrun, R.: The impact of emotions on learning and achievement: towards a theory of cognitive/motivational mediators. Appl. Psychol. Int. Rev. **41**(4), 359–376 (1992)
2. Bryan, T., Mathur, S., Sullivan, K.: The impact of positive mood on learning. Learn. Disabil. Q. **19**(3), 153–162 (1996)
3. Febrilia, I., Warokka, A.: The effects of positive and negative mood on university students' learning and academic performance: evidence from indonesia. In: The 3rd International Conference on Humanities and Social Sciences, pp. 1–12 (2011)
4. Sottilare, R.A., Proctor, M.: Passively classifying student mood and performance within intelligent tutors. J. Educ. Technol. Soc. **15**(2), 101 (2012)
5. Lewine, R., Sommers, A., Waford, R., Robertson, C.: Setting the mood for critical thinking in the classroom. Int. J. Sch. Teach. Learn. **9**(2), 1–4 (2015)
6. Liew, T.W., Tan, S.-M.: The effects of positive and negative mood on cognition and motivation in multimedia learning environment. Educ. Technol. Soci. **19**(2), 104–115 (2016)
7. Kulkarni, S.S., Reddy, N.P., Hariharan, S.I.: Facial expression (mood) recognition from facial images using committee neural networks. BioMed. Eng. OnLine **8**, 16 (2009)
8. Moridis, C.N., Economides, A.A.: Mood recognition during online self-assessment tests. IEEE Trans. Learn. Technol. **2**(1), 50–61 (2009)
9. Thrasher, M., Van der Zwaag, M.D., Bianchi-Berthouze, N., Westerink, J.H.D.M.: Mood recognition based on upper body posture and movement features. In: D'Mello, S., Graesser, A., Schuller, B., Martin, J.-C. (eds.) ACII 2011. LNCS, vol. 6974, pp. 377–386. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-24600-5_41
10. Gjoreski, M., Gjoreski, H., Lutrek, M., Gams, M.: Automatic detection of perceived stress in campus students using smartphones. In: 2015 International Conference on Intelligent Environments (IE), pp. 132–135. IEEE (2015)
11. Maaoui, C., Pruski, A.: Emotion recognition through physiological signals for human-machine communication. In: Cutting Edge Robotics, pp. 317–333. INTECH Open Access Publisher (2010)
12. LiKamWa, R., Liu, Y., Lane, N.D., Zhong, L.: MoodScope: building a mood sensor from smartphone usage patterns. In: Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services, pp. 389–402. ACM (2013)

13. Odinaka, I., et al.: ECG biometric recognition: a comparative analysis. IEEE Trans. Inf. Forensics Secur. **7**(6), 1812–1824 (2012)
14. Tapia, E.M., et al.: Real-time recognition of physical activities and their intensities using wireless accelerometers and a heart rate monitor. In: 11th IEEE International Symposium on Wearable Computers. IEEE (2007)
15. Sebe, N., et al.: Multimodal approaches for emotion recognition: a survey. In: Internet Imaging VI, vol. 5670. International Society for Optics and Photonics (2005)
16. Vinola, C., Vimaladevi, K.: A survey on human emotion recognition approaches, databases and applications. ELCVIA Electr. Lett. Comput. Vis. Image Anal. **14**(2), 24–44 (2015)
17. Sharma, N., Tom, G.: Objective measures, sensors and computational techniques for stress recognition and classification: a survey. Comput. Methods Programs Biomed. **108**(3), 1287–1301 (2012)
18. Basu, S., et al.: A review on emotion recognition using speech. In: 2017 International Conference on Inventive Communication and Computational Technologies (ICICCT). IEEE (2017)
19. Wioleta, S.: Using physiological signals for emotion recognition. In: 6th International Conference on Human System Interactions (HSI), pp. 556–561. IEEE (2013)
20. McDuff, D., Gontarek, S., Picard, R.: Remote measurement of cognitive stress via heart rate variability. In: 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) (2014)
21. https://www.amazon.com/Original-Xiaomi-Monitor-Wristband-Display/dp/B01GMQ4Y3O
22. Alexandratos, V.: Mobile Real-Time Stress Detection (2014)
23. Leys, C., et al.: Detecting outliers: do not use standard deviation around the mean, use absolute deviation around the median. J. Exp. Soc. Psychol. **49**(4), 764–766 (2013)
24. Healey, J.A.: Detecting stress during real-world driving tasks using physiological sensors. IEEE Trans. Intell. Transp. Syst. **6**(2), 156–166 (2005)
25. Sano, A., Picard, R.W.: Stress recognition using wearable sensors and mobile phones. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII). IEEE (2013)
26. Zenonos, A., et al.: HealthyOffice: mood recognition at work using smartphones and wearable sensors. In: 2016 IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops). IEEE (2016)
27. Wijsman, J., et al.: Wearable physiological sensors reflect mental stress state in office-like situations. In: 2013 Humaine Association Conference on Affective Computing and Intelligent Interaction (ACII). IEEE (2013)
28. Valenza, G., et al.: Wearable monitoring for mood recognition in bipolar disorder based on history-dependent long-term heart rate variability analysis. IEEE J. Biomed. Health Inf. **18**(5), 1625–1635 (2014)
29. Reyes, C.R., et al.: Classroom emotional climate, student engagement, and academic achievement. J. Educ. Psychol. **104**(3), 700–712 (2012)

Obstacle Detection in Drones Using Computer Vision Algorithm

N. Aswini[1(✉)] and Satyanarayana Visweswaraiya Uma[2]

[1] MVJ College of Engineering,
Visveswaraya Technological University, Bangalore, India
shijiaswini@gmail.com
[2] RNS Institute of Technology, Visveswaraya Technological University,
Bangalore, India

Abstract. Obstacle detection and collision avoidance are complicated particularly in drones where accuracy matters a lot to avoid collision between a vehicle and an object. This complication arises due to restricted number of heavy sensors like radar. To overcome the drawbacks of heavy sensor, light weight monocular cameras can be employed. Monocular cameras are capable of analyzing and computing depth by giving the three-dimensional representation. In the proposed method, key point features are extracted from each video frame using Computer vision algorithms like Harris corner detector and Scale Invariant Feature Transform algorithm (SIFT). Then by using Brute Force Matching (BFM), key points of consecutive frames are matched. As drone move towards obstacle, size of obstacle increases i.e. convex hull size around key point increases which shows obstacle is detected.

Keywords: Unmanned Aerial Vehicle (UAV) · Harris corner detector
Scale Invariant Feature Transform · Brute Force Matching · Obstacle detection

1 Introduction

The ability to detect and avoid obstacles autonomously can be achieved by incorporating distance sensors to Unmanned Aerial Vehicles (UAV). Addition of sensors like laser scanners, stereo cameras or combination of sensors is expensive. Time of Flight (TOF) cameras are faster than lasers, but again it's costly. Due to recent advancements in solid state electronics, low price and smaller size monocular cameras are now able to sense, localize and detect obstacles efficiently [12].

Collision avoidance techniques basically depend upon the sensors used in detecting obstacles and the algorithms used. When it comes to flying objects, the techniques used to avoid collision are more complex. The complexity is due to the motion of objects like birds or aircrafts and the vehicle on which the avoidance system need to be implemented. We can consider visual sensors rather than radar or lidar, because they are too bulky and too expensive to fit on a small or medium sized UAV. For micro and small aerial vehicles, the constraints of size, weight and power makes them choose digital cameras as the best choice for obstacle detection. Camera based sensing can depend on Simultaneous Localization and Mapping (SLAM) techniques, Optical Flow

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 104–114, 2019.
https://doi.org/10.1007/978-981-13-5758-9_9

(OF) methods, stereo vision or monocular (single camera) vision-based techniques. The various existing monocular vision-based object detection techniques include estimation of relative size of obstacles, relative clarity, texture gradient, linear perception, inter-position, and relative motion.

Unmanned aircrafts exploit many applications in military system, surveillance system, rescue operations, traffic monitoring and in agricultural systems to provide the geographical information [13]. Increase in digitization technology and complications in the system lead to high level automatic aerial vehicle development, while maintaining constant stabilization. For UAV, detection of objects, avoiding collision and achieving good accuracy is a challenging problem [14].

There are several methods proposed to extract features of input image captured from camera. However, they are not capable of extracting accurate features and not capable of measuring distance of moving vehicle. For indoor environments, OF assumes object as stationary. It will not work properly for outdoor surroundings where both camera and object are not stationary (i.e. vehicles are moving freely).

Advancement in aerial vehicles involves various issues which are required to be considered, like directing vehicle, providing security, identifying and evading obstacle, providing guidance, considering safety etc. The basic operation of sensing and detecting obstacle is depicted in Fig. 1. Aircraft sensors, sense and detect the obstacle. Once detected, collision is avoided by taking action to direct vehicle in a different and safe path.

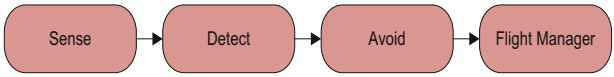


Fig. 1. Basic diagram for sensing and avoiding object

In the proposed work, video frames generated are initially filtered using median filter. Harris detector detects the corners and then key points are extracted from each frame using SIFT [15]. The extracted key points of current frame are matched with its previous frame. This matching process continues for all the frames. The key points which exists in both current and previous frames are only matched. A rectangular box is created around the keypoints, and the points beyond this rectangular box are neglected. To detect presence of obstacle, a convex hull is created around keypoints. When drone move towards obstacle, size of convex hull increases i.e. size of obstacle increases which shows obstacle is detected.

2 Literature Survey

Al-Kaff et al. [1] proposed a similar real-time collision avoidance and object detection algorithm for UAV. A singular monocular camera is mounted on vehicle to capture the image. SIFT and Brute Force algorithm is used for generating and matching key points. In this work they can detect objects within 90 to 120 cm range. Kovacs et al. [2]

proposed a deconvolution technique to discover the object region to take out features of that object and to create feature map which usually called as D-map. Monocular camera is employed to capture obstacle with low collision ratio and can be frequently used in various environments. Methodology used in this work is helpful in navigation system, surveillance, military, mapping and odometry.

Esrafilian et al. [3] proposed a collision avoidance scheme where video streams recorded using front camera is transmitted to ground placed laptop through wireless network connection. The navigation data received is processed by Oriented Fast and Rotated Brief (ORB) and SLAM to compute 3D maps and three-dimensional position. The scaling parameter of monocular SLAM is figured out using linear filtering. Kalman Filter (KF) is used for fusing sensor in monocular camera of quadrotor. For controlling three-dimensional position of obstacle, Proportional Integral Derivative (PID) controller is designed. To discover free path between precise point and robot position, a path is created by Rapidly Exploring Random Tree (RRT) algorithm to avoid obstacle. Engel et al. [4] introduced an UAV attached to a ground based laptop for directing and navigating in Global Positioning System (GPS) enabled surroundings. The Quadcopter based system includes SLAM Algorithm, Extended Kalman Filter (EKF) for fusing the sensor. PID offer steering command, to control and direct the drone. Shankar et al. [5] proposed an obstacle avoidance algorithm with Support Vector Machine (SVM). The generated video frames are given to OF algorithm to detect and describe key point features, and then the extracted key points are given as input to SVM algorithm to evaluate presence of obstacle. This system is used for both indoor and outdoor. SVM classifier is chosen to offer high accuracy and robustness against noise. Wu et al. [6], lists the advantages and disadvantages of different autonomous navigation algorithms and discuss the prospect of monocular obstacle detection technology. They cover feature extraction and matching, focusing and defocusing model, background subtraction, frame differencing and optical flow.

Mori et al. [7] introduced an object detection technique for objects arriving near and exactly in front of vehicle. Image is divided into patches to limit drawbacks and to find object by estimating the change in size of image patches. They employ Speeded up Robust Features (SURF) algorithm for detecting keypoints and template matching algorithm to compute size of object. However, obstacles must comprise adequate texture to create SURF keypoints. The future, work was expected to enhance detection accuracy with a good camera scheme. Since high resolution image indicate more processing exploitation, on-board processing may be difficult. Magree et al. [8] presented an effective monocular camera vision based system to capture object based on fast terrain mapping. The feature detector is a modified Harris corner detector and EKF is used to estimate the location of image features in the environment.

Sapkota et al. [9] proposed detection and estimation algorithm to track and locate target position using a single moving camera. On line detection of micro aircraft, measuring speed and position in 3D locations is elaborated in this work. To avoid misdetections and to address the requirement of computationally limited payload, AdaBoost is integrated with on line visual tracking technique. On line visual based tracking algorithm extract position and size of aerial vehicle from the video frames. AdaBoost detects objects by Histogram of Oriented Gradients (HOG) features, where the image is alienated into patches and each image slides on other image. Images are

further divided into object containing patches and non- object containing patches. Future enhancement includes similar progression in outdoor environment. Boroujeni et al. [10] introduced an obstacle detection method using two types of optical flow techniques. After computing OF, K-means clustering is used for obstacle reconstruction. Huang et al. [11] uses a RGB-D camera to provide color image and to find out rotation and position of UAV. By examining the color images of RGB-D camera with visual odometry process, it became possible for them to approximate the depth of the obstacles. To the image captured, Gaussian pyramid is applied and then features of Gaussian pyramid are taken out using FAST feature detector. Features having corresponding depth are extracted and others are discarded.

3 Methodology

This section briefly illustrates the proposed system designed for obstacle detection. Obstacle is static while Drone is in motion. The work done in [1] has been extended by the inclusion of median filter and corner detection before extracting SIFT features. Input frames are filtered using median filtering to reduce noise and to maintain image edges. Harris corner detector is used to detect corner points and then SIFT descriptor is used to extract key points which are invariant to rotation and scale. The extracted key points of current frame are matched with extracted key points of previous frame using brute force algorithm and this process continues until all current frame key points are matched with previous frame key points. The key points which are there in both frames are only matched and appeared. Convex hull is created around key points of each frame to detect obstacle. Generally, as drone move towards obstacle, obstacle size increases and hence the convex hull size increases around key points that show obstacle is detected. Figure 2 depicts the architecture of the detecting obstacle. Harris-SIFT detector improves the salience of the SIFT feature and at the same time reduces the time complexity.

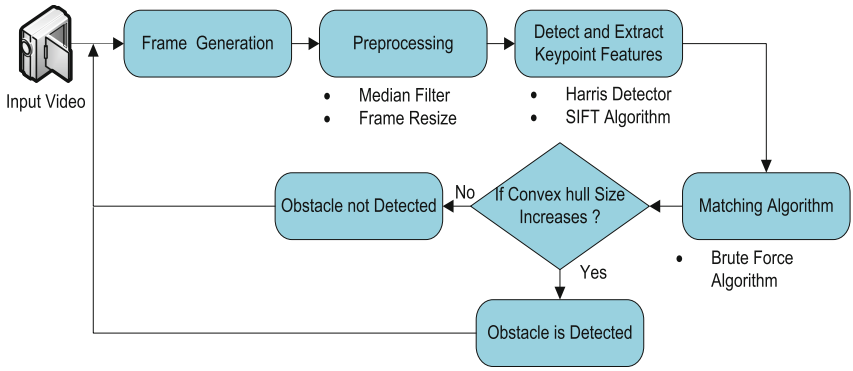


Fig. 2. Architecture of proposed system to detect obstacle in UAV

3.1 Median Filtering

Filtering of images is usually performed for removing noise, sharpening, and detecting the edges. In a median filter, a window slides along the image, and the median intensity value of the pixels within the window becomes the output intensity of the pixel being processed. It is a nonlinear filtering technique. Initially, all pixels in the median filtering technique are sorted in the ascending order and then a middle value replaces all these pixel values.

3.2 Harris Corner Detector

Harris Corner detector mathematically finds the corners in an image. It is one of the popular methods of finding corners of an image. Corner detection is applied in obstacle recognition, detection and in video tracking applications. In proposed method combination of Harris and SIFT descriptor is incorporated for extracting key points as Harris detector detects the corner edges and SIFT provide invariance to rotation and scale. Harris detector is based on auto-correlation matrix shown in Eq. (1) to detect image corners.

$$M(x,y) = \sum_{u,v} w(u,v) * \begin{bmatrix} I_x^2(x,y) & I_xI_y(x,y) \\ I_xI_y(x,y) & I_y^2(x,y) \end{bmatrix} \tag{1}$$

I_x and I_y are image derivatives in (x, y) directions. $W(u, v)$ denotes weighting window over (u, v) area. Harris detector cornerness measure $C(x, y)$ for each (x, y) pixel as given in Eq. (2).

$$C(x,y) = \det(M) - K(trace(M))^2 \tag{2}$$

Where $\det(M) = \lambda_1 * \lambda_2$ and $trace(M) = \lambda_1 + \lambda_2$, K is an adjusting parameter and λ_1, λ_2 are Eigen values of the auto-correlation matrix M.

3.3 SIFT Descriptor

SIFT algorithm is proposed by Lowe in 2004 [15]. It has a classic approach, and is the original inspiration for most of the descriptors proposed later. It is accurate, and invariant to rotation and scaling. But it is mathematically complicated and computationally heavy. SIFT uses difference of Gaussian (DOG), which is an approximation of Laplacian of Gaussian (LoG). To extract the feature vector, SIFT algorithm require four steps.

Create Scale Space

At first, position and dimension space is determined for input filtered image using a dimension space method as in Eq. (3). Invariance to scale is achieved by representing internal structure of filtered image using Scale Space function.

$$S(q,r,\sigma) = G(q,r,\sigma) * I(q,r) \tag{3}$$

$$\sigma = \sigma_0\sqrt{k^2 - 1} \tag{4}$$

$S(q, r, \sigma)$ shows scale space function and $*$ indicate the convolution operation. $G(q, r, \sigma)$ indicates Gaussian scale as in Eq. (5), $I(q, r)$ represents filtered input image. σ indicates scale parameter. Convolution of image and Gaussian operator provide blurred image. q and r are location coordinators.

$$G(q, r, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left\{ -\frac{q^2 + r^2}{2\sigma^2} \right\} \tag{5}$$

Now DOG function is adopted to determine the maximum stable key point position. DOG between two images is represented by Eq. (6).

$$D(q, r, \sigma) = S(q, r, K\sigma) - S(q, r, \sigma) \tag{6}$$

Where K is a constant multiplicative factor.

Key Points Localization

Maxima and minima of the difference-of-Gaussian images are detected by comparing a pixel to its 26 neighbours in 3 × 3 regions at the current and adjacent scales. Position of extrema (max and min) of key point is represented by Eq. (7).

$$L = -(\frac{\partial^2 D}{\partial Q^2})^{-1} \frac{\partial D}{\partial q} \tag{7}$$

Equation (8) represents the value of Keypoints in DOG,

$$D(L) = D + \frac{1}{2} \frac{\partial D^{-1}}{\partial q} L \tag{8}$$

If L value is less then threshold, then extreme points are eliminated having very less illumination. Eliminating regions, edges with low contrast provide more efficiency and robustness. However, we compute the SIFT descriptors for computed Harris interest points which has drastically reduced the computation time.

Assign Orientation

In this step, each key point is assigned with orientation for rotation invariance. Steps evolved in assigning keypoints orientation are given below.

1. Initialize scale space Gaussian image *S*.
2. Calculate gradient magnitude

$$m(q, r) = \sqrt{(s(q+1, r) - s(q-1, r))^2 + (\begin{matrix} s(q, r+1) \\ -s(q, r-1) \end{matrix})^2} \tag{9}$$

3. Calculate orientation, θ.

$$\theta(q, r) = \frac{tan^{-1}(s(q, r+1) - s(q, r-1))}{s(q+1, r) - s(q-1, r)} \tag{10}$$

4. Create orientation histogram from the gradient orientations of sample points within a region around the key point.
5. Each sample is weighted by its gradient magnitude and a Gaussian window having σ value 1.5 times greater than scale of the key point.
6. Determine highest peak in histogram, use 80% height of this peak to generate keypoints having similar orientation.

Fit a parabola to the histogram values close to each peak.

Key Point Descriptor

A key point descriptor is created. A 16 × 16 neighborhood around the key point is taken. It is divided into 16 sub-blocks of 4 × 4 size. For each sub-block, 8 bin orientation histograms are created. So, a total of 128 bin values are available. It is represented as a vector to form key point descriptor. Figure 3 shows the flowchart of SIFT detector.

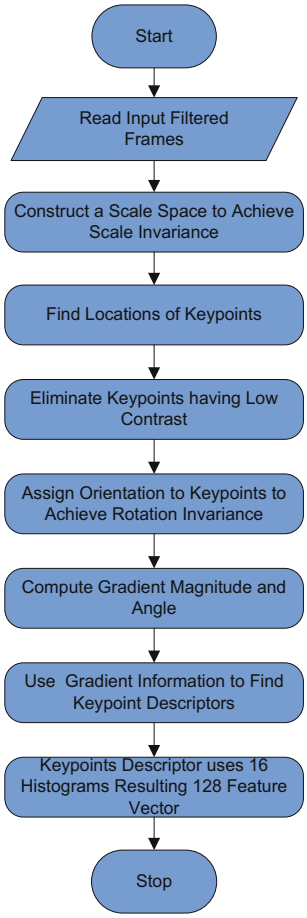


Fig. 3. Flow chart of SIFT descriptor

3.4 Brute Force Algorithm

Brute force is simple and uncomplicated method generally used to find similarity and matching pairs. Brute force or template matching algorithm recognize object in tracking system. It takes the descriptor of one feature in first set and is matched with all other features in second set using distance calculation. In this paper brute force is employed to match key point features of current and previous frames. After computing the distance of all features, a threshold value is set using the minimum distance value. Key points of each frame having distance value less than threshold value is selected as matched key points since minimum distance value shows more similarity and nearest corresponding features.

The obstacle detection algorithm is as follows: -

Algorithm 1. Object Detection Algorithm

- **Input:** Image frames F. **Output**: Match keypoints
- Generate video frames.
- Apply median filleting to eliminate noise.
- Harris corner detection.
- Extract keypoints using SIFT descriptor; keypoints having less contrast are eliminated.
- Determine *(x, y)* vector position of keypoints in each frame.
- Match keypoints using brute force algorithm.
- For better accuracy eliminate matched keypoints having distance greater than threshold value.
- If keypoints are matched, it signifies presence of object.

4 Experimental Result

Videos were taken at a resolution of 720 × 1280 at 30 fps. Video frames are extracted resized to 360 × 640, and initially filtered using median filter. The corner points are found using Harris Corner detector. SIFT descriptor operation is performed only to those features which are extracted by Harris algorithm. This operation drastically reduces the computation time taken. The time taken to compute is 0.2–0.3 s. When we performed the simulation by extracting SIFT features without Harris corner detection, it takes a huge amount of time. As in Table 2, for 80 × 80 resolution itself, it is taking more than 2 s to compute. So, with SIFT alone, we cannot detect obstacles at a faster rate especially in real time video processing for drones.

The entire algorithm is simulated in MATLAB. Figures 4 and 5 shows the generated key points for current and previous frame. Figure 6 shows the key point features with in rectangular box. The keypoints beyond this rectangular box are assumed as noise and are avoided. Next, keypoints of current and previous frame are matched using brute force. The increase in size of convex hull represents an approaching obstacle.

Table 1. Performance results of Harris + SIFT

Parameters	Current frame	Previous frame
Resolution	360*640	360*640
Time taken for feature extraction	0.23 s	0.35 s
Total keypoints extracted using Harris + SIFT	10752	6784
Matched keypoints	31	31

Table 2. Performance results of SIFT algorithm

Parameters	Current frame	Previous frame
Resolution	80*80	80*80
Time taken for feature extraction	2.01 s	2.50 s
Total keypoints extracted using SIFT	21502	24111
Matched keypoints	337	337

Table 1 illustrate the performance of the system, using SIFT and Harris detector. Time consumed to generate keypoints of current and previous frame is 0.23 s and 0.35 s. Out of 10752 and 6784 extracted keypoints only strongest 31 keypoints are matched which are found in both frames.

Fig. 4. Previous frame

Fig. 5. Current frame

Fig. 6. Convex hull applied on key points in rectangular box

5 Conclusion and Future Enhancement

The proposed work was conducted as an initial experimental study on object detection for autonomous navigation of drones using advanced computer vision algorithms. SIFT is one of the basic feature extracting algorithms which is very accurate. But due to its complexity and time consumption, in future work, we will focus to enhance the speed performance by adopting improved algorithms like FAST algorithm for corner detection, Speeded Up Robust Features (SURF), Binary Robust Independent Elementary Features (BRIEF) etc. The inclusion of neural networks for classification of objects will also be explored.

References

1. Al-Kaff, A., Meng, Q., Martin, D., De La Escalera, A., Armingol, J.M.: Monocular vision-based obstacle detection/avoidance for unmanned aerial vehicles. In: Intelligent Vehicles Symposium (IV). IEEE (2016)
2. Kovács, L.: Visual monocular obstacle avoidance for small unmanned vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop, pp. 59–66 (2016)
3. Esrafilian, O., Taghirad, H.D.: Autonomous flight and obstacle avoidance of a quadrotor by monocular SLAM. In: Robotics and Mechatronics (ICROM), 4th International Conference on IEEE, pp. 240–245 (2016)
4. Engel, J., Sturm. J., Cremers, D.: Camera-based navigation of a low-cost quadrocopter. In: Intelligent Robots and Systems (IROS), IEEE/RSJ International Conference on IEEE (2012)
5. Shankar, A., Vatsa, M., Sujit, P.B.: Collision avoidance for a low-cost robot using SVM based monocular vision. In: Robotics and Biomimetics (ROBIO), International Conference on IEEE (2014)
6. Wu, Q., Wei, J., Li, X.: Research progress of obstacle detection based on monocular vision. In: Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Tenth International Conference on IEEE (2014)
7. Mori, T., Scherer, S.: First results in detecting and avoiding frontal obstacles from a monocular camera for micro unmanned aerial vehicles. In: Robotics and Automation (ICRA), International Conference on IEEE (2013)
8. Magree, D., Mooney, J.G., Johnson, E.N.: Monocular visual mapping for obstacle avoidance on UAV. J. Intell. Robot. Syst. **74**(1-2), 17–26 (2014)
9. Sapkota, K.R., et al.: Vision based unmanned aerial vehicle detection and tracking for sense and avoid systems. In: Intelligent Robots and Systems (IROS), International Conference on IEEE, pp. 1556–1561 (2016)
10. Boroujeni, N.S., Etemadi, S.A., Whitehead, A.: Fast obstacle detection using targeted optical flow. In: Image Processing (ICIP), 19th International Conference on IEEE, pp. 65–68 (2012)
11. Huang, A.S., et al.: Visual odometry and mapping for autonomous flight using an RGB-D camera. In: Christensen, Henrik I., Khatib, O. (eds.) Robotics Research. STAR, vol. 100, pp. 235–252. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-29363-9_14
12. Alvarez, H., Paz, L.M., Sturm, J., Cremers, D.: Collision avoidance for quadrotors with a monocular camera. In: Hsieh, M.A., Khatib, O., Kumar, V. (eds.) Experimental Robotics. STAR, vol. 109, pp. 195–209. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-23778-7_14
13. Huh, S., Shim, D.H., Kim, J.: Integrated navigation system using camera and gimbaled laser scanner for indoor and outdoor autonomous flight of UAVs, pp. 3158–3163. IEEE (2013)
14. Eresen, A., Imamoglu, N., Efe, M.O.: Autonomous quadrotor flight with vision-based obstacle avoidance in virtual environment. Expert Syst. Appl. **39**(1), 894–905 (2012)
15. Lowe, D.G.: Distinctive image features from scale invariant KeyPoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)

Intelligent Systems Applied to the Classification of Multiple Faults in Inverter Fed Induction Motors

Wagner Fontes Godoy[1(✉)], Alessandro Goedtel[1], Ivan Nunes da Silva[2], Rodrigo Henrique Cunha Palácios[1], and Alexandre L'Erario[1]

[1] Department of Electrical and Computer Engineering, Federal Technological University of Paraná, Cornélio Procópio, PR 86300-000, Brazil
{wagnergodoy,agoedtel,rodrigopalacios,alerario}@utfpr.edu.br
[2] São Carlos School of Engineering, University of São Paulo, São Carlos, SP 13566-590, Brazil
insilva@sc.usp.br

Abstract. The monitoring condition of electrical machine is an important parameter for maintenance of industrial process operation levels. In this paper, an investigation based on learning machine classifiers to proper classify machine multiple faults i.e. stator short-circuits, broken rotor bars and bearings in three phase induction motors driven by different inverters models is proposed. Experimental tests were performed in 2 different motors, running at steady state, operating under variable speed and torque variation resulting in 2967 samples. The main concept of proposed approach is to apply the three phase current amplitudes to immediately detect motor operating conditions. The high dimensionality of the input vectors in the algorithms was solved through the discretization of the current data, which allows the reduction the classification complexity providing a optimized waveform in comparison with the original one. The results show that it is possible to classify accurately these faults.

Keywords: Intelligent systems · Three-phase induction motor
Inverters · Fault diagnosis

1 Introduction

This paper deals with three-phase induction motor as this type of machine is commonly found in various industrial applications. After working in the most harshest environments and conditions some parts of this machine are subjected to present multiples faults which are inevitable throughout its useful life. In the modern industrial applications, the employment of efficient control strategies based on various speed operation levels are request in several process applications.

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 115–125, 2019.
https://doi.org/10.1007/978-981-13-5758-9_10

Continuous observation of the machinery operating conditions is an important role to diagnose the fault type and then proper analyse and act in the problems still in their incipient phase of evolution. This procedure is primarily employed to increment machine operation time, minimize breakages during production and basically prevent economic losses [1].

Machine learning techniques are studied and employed in the area of rotating machinery diagnosis [1–4]. According to [5], there are crescent efforts to develop monitoring models that can distinguish and provide proper machinery operation conditions. Such intelligent fault diagnosis approaches are mainly based in artificial intelligence techniques, such as multilayer perceptron network (MLP) [1], support vector machine/Sequential minimal optimization (SVM/SMO) [2], fuzzy Artmap network (FAM) [6] and k-Nearest Neighbour (k-NN) [7].

From the diagnosis produced by these techniques, operators can take appropriate actions to guarantee machinery availability. Hence, the development of expert monitoring units for these machines has become a strategic component in industrial drives [2].

Predictive maintenance methods are essential for industry. Advances in data processing associated with the application of intelligent monitoring tools allow everyone involved in the process of repair and fault prevention to have an effective and comprehensive understanding of the problems faced while minimizing breakdowns and unexpected equipment stops [8].

These strategies have shown a movement in the modern industry, where machinery efficiency and focus on increase operational performance and reducing costs are the goal [9].

In this way, equipment inspection frequency is no longer defined and begin to be closely followed and programmed based on the machine operation status, where decision is taken based on results gathered form the predictive inspection techniques which aims to improve overall process reliability, since it is possible to optimize the allocation of human resources necessary to maintenance procedures and machine parts are explored at maximum [8].

Such machines parts however, are susceptible to common defects that can affect their correct operation. In general, the faults can be divided in two groups: the electrical faults and the mechanical faults. Electrical faults are usually corresponds to stator faults (21%) and rotor (7%), whereas the mechanical one are mainly associated with bearings (69%) and shaft/coupling (3%) [10]. These numbers states that bearings, rotor and stator account for 97% of the fault sources connected with three-phase induction motors.

In view of described scenario, this work presents the classification performance of MLP, SVM/SMO, FAM and k-NN learning machine techniques employed in the diagnosis of bearing, stator short-circuit and broken rotor bars defects which are the most common faults in tree phase induction motors. Proposed technique uses the current amplitudes to monitor anomalies that may characterize evolving faults. A total of 2967 tests gathered from two 1 HP motors driven by three different frequency inverter models were collected. As for the

testes, the frequency range were varied from 12 Hz to 60 Hz and load conditions consider a range of 10% to 110% of the nominal torque.

The paper is structured into the following sections: Sect. 2 presents the proposed methodology, data processing and configuration parameters used in the classifiers. In Sect. 3 the results and discussions are presented. Finally, conclusions are drawn in Sect. 4.

2 Methodology

Proposed methodology considers the use of the currents amplitudes to be presented as input to the classifiers, aiming to diagnosis motor faults allowing correct decision concerning to machinery maintenance.

In this work, two 3-phase induction motors of 1 CV, 60 Hz, 220/380 V, 3,02/1,75 A, 4 poles, 4.1 N.m and 1730 rpm were used, one of which has customized windings to emulate stator short-circuit faults. These motors were powered by the following inverters: Schneider ATV12H075M2, Sinamics G110 and Micromaster 440 (both from Siemens) with the objective of validating the proposed strategy in some commercial inverters.

For each experiment, frequency were varied in the range of 12 Hz to 60 Hz, in steps of 6 Hz, while load were varied from 0.5 N.m to 5.0 N.m, in steps of 0.5 N.m. Exploring speed and load variations conditions allows to evaluate the method performance in conditions similar to those found in the industrial environment. Figure 1, shows the experimental workbench employed to perform tests under evaluation in this work.

Fig. 1. Experimental workbench

The employment of artificial procedures to create defects at controlled premisses are commonly found in the literature, as such procedure enables to control proper fault severity level besides its adequate reproduction associated with the investigation goal [11].

Acquired data is stored as spreadsheets using Matlab software. The acquisition rate was 25 kS/s, with an acquisition time of 6 s for the tests simulating broken rotor bars and bearing faults and 3 s for the tests considering stator short-circuit faults.

As shown in Fig. 2, as for the tests simulating stator faults a customized motor where derivations from several coils allows the emulation of short circuit of 1%, 3%, 5% and 10% was artificially assembly.

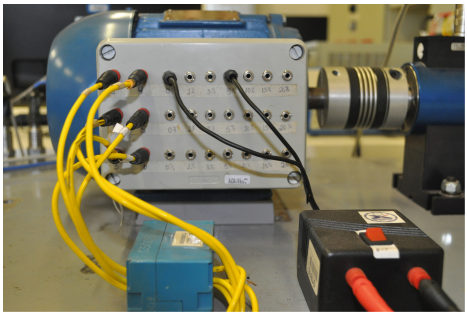
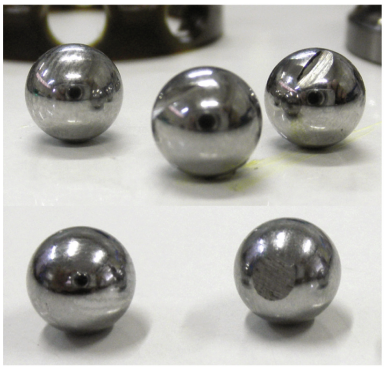


Fig. 2. Motor 1 - tests considering stator short-circuit in one single phase.

Also, 5 different rotors are used to emulate broken rotor bar faults. The rotor fault is artificially generated by drilling holes to bars as shown in Fig. 3. The following broken bar cases were created: one, two, and four consecutive bars, as well as two consecutive broken bar in opposite side of the rotor circumference.

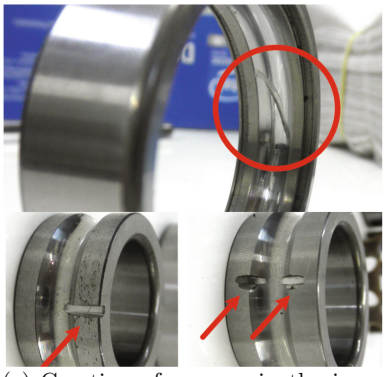
Fig. 3. Artificial generation of broken rotor bars.

As per Figs. 4(a), (b) and (c), the following defects were artificially created for the tests considering bearing defects: balls wear, excessive wear and the creation of grooves in the inner and outer.



(a) Creation of balls faults.

(b) Creation of excessive wear on bearings.



(c) Creation of grooves in the inner and outer bearings tracks.

Fig. 4. Creation of bearing faults.

2.1 Data Preprocessing

The methods used to reduce the size of input data are essential for the performance machine learning techniques. Preprocessing of high data matrices by an appropriate tool can lead to a lower computational cost in addition to an improvement in the interpretation of the characteristics of interest associated with the signal under analysis.

Thus, the proposed approach directly employs the amplitude current signal to diagnose the occurrence of faults and in the sequence properly distinguish which class of fault is occurring in the machine.

Considering that a motor under fault operation conditions has a characteristic signal of its wave cycle over time, in order to reduce the input matrix dimensionality, 3 positive half-cycles representing each current phase per experiment are selected at random form to create each input vector. As the next step, these vectors are normalized based on the peak current. This procedure guarantee that data can be properly applied regardless the machine load or speed.

As per [12], in this paper 50 points of I_a, I_b and I_c from a single positive cycle representing each adopted test condition were collected, and vectors considering these normalized amplitudes values are presented to the selected machine learning classifiers. Proposed approach employs the linear discretization, which is considered as method capable of recreate the waveform without changing its original format.

As the method deals with current amplitude values, it is mandatory to assemble sequential column vectors considering each specific phase order, to create a matrix with 150 points size, as per Fig. 5.

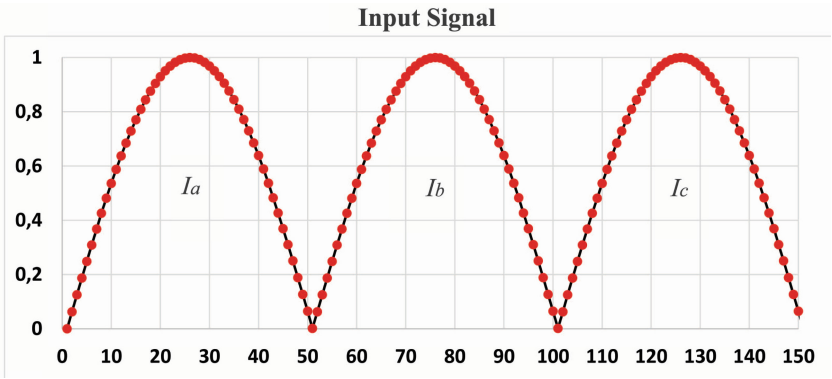


Fig. 5. Input data modeling: current discretization.

Figure 6 presents in a simplified form the general diagnostic process flowchart considered in this work.

2.2 Classification Structures

In this work, the pattern classification methods MLP, SVM/SMO, FAM and k-NN are compared in order to proper identify which intelligent algorithms presents better robustness for the task of multiple fault classification in inverter fed induction motors.

Performed tests combines the use of 2 motors, 3 different inverter models, variation of the load torque and also explore a wide range of frequency. Stator current are directly used as inputs to the intelligent structures properly indicate each fault class accordingly. Same sets of samples are used for all the tests performed in the learning machine algorithms under evaluation in this paper.

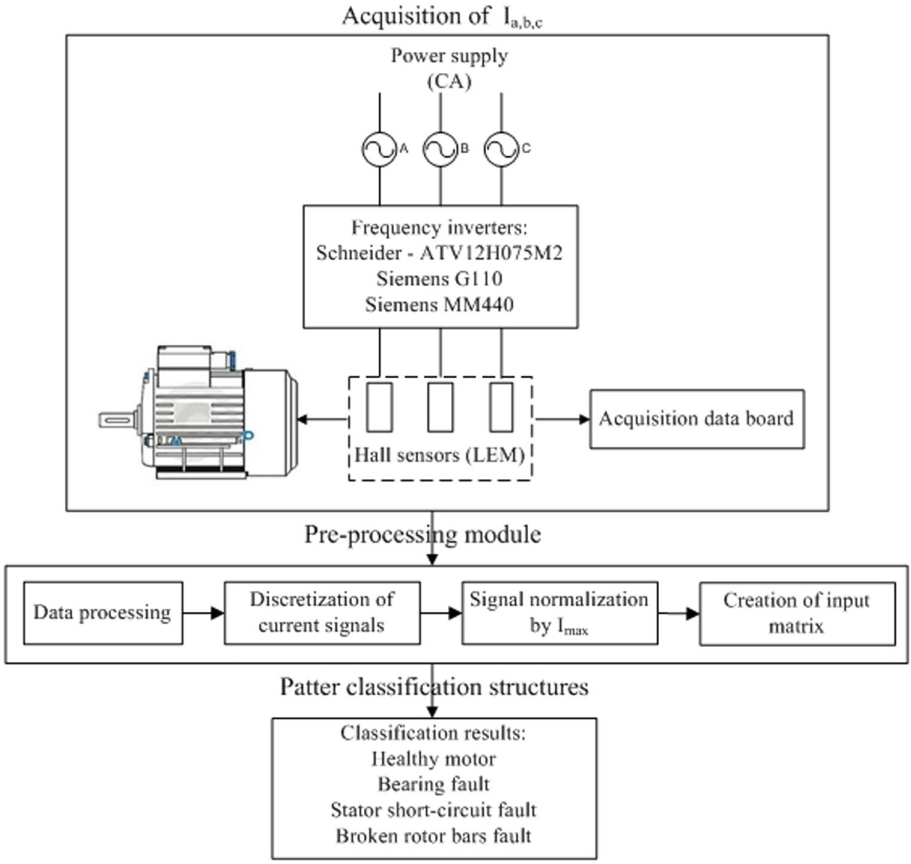


Fig. 6. General overview of proposed fault diagnosis system.

The literature related to multiple class classifiers points out advantages of learning machine approaches in comparison with traditional methods such as statistics and vibration analysis commonly applied in the diagnosis and classification of faults in electrical machines [4,13,14].

Thus, the machine learning algorithms used in this work were tuned for supervised training, where all parameters and configurations employed specifically for each classifier remained unchanged for all the test performed in this investigation. Also, tests were carry out in the WEKA tool (Waikato Environment for Knowledge Analysis), considering for training and validation the use of cross validation with $k = 10$.

The deep detailing and the scientific basis of the concepts about the intelligent classifiers used in this work can be found in [15–17].

3 Results and Discussions

Section experimental results and discussions presents the performance of the algorithms FAM, SVM/SMO, k-NN and MLP, trained from the amplitudes of three-phase stator current signals, in the time domain, obtained by means of experimental tests using motors driven from inverters models commercially available.

The results consider diagnosis and classification of healthy motor, motor operating with stator short-circuit faults, motor with broken rotor bar and bearing defects. Results consider binary classification (Healthy and Faulty motor), as well as evaluation of classifiers configured to proper classify multiple faults. Same set of samples were used for both situations.

Next, comparative classification results obtained by the different algorithms evaluated in this study are presented.

Tables 1, 2, 3 and 4 bring the results concerning the set consisting of 2967 tests performed at laboratory conditions, referring to the data obtained through the tests performed in this study. Since each test comprises 3 currents signals, the experiments resulted in 8901 samples which were properly divided into training and validation. Tables 1 and 2 present the results obtained with the classifiers configured for the binary diagnosis.

Table 1. Experimental results - binary classification

Attributes	FAM	SVM/SMO	k-NN	MLP
Correctly classified instances (%)	77,58	85,10	91,24	89,40
Incorrectly classified instances (%)	22,42	14,90	8,75	10,60
Kappa statistic	0,36	0,29	0,72	0,64
Mean absolute error	0,22	0,14	0,08	0,12
Root mean squared error	0,47	0,38	0,29	0,28
Relative absolute error (%)	74,12	49,27	28,98	39,68
Root relative squared error (%)	121,77	99,27	76,08	72,52
Time taken to build model (s)	1,22	34,3	0,02	904,08

Based on the results presented on Table 1 it is noticed that the algorithm k-NN reached 91.24%, which can be associated with the best overall accuracy. In the sequence the MLP, SVM/SMO and FAM algorithms, respectively, with 89.40%, 85.10% and 77.58%. The classifiers k-NN and MLP obtained a Kappa index greater than 0.60 indicating substantial concordance with the obtained results. As for the experiments performed with SVM/SMO and FAM classifiers, obtained Kappa indices indicate poor agreement.

As per Table 2, the algorithm k-NN obtained 81.5% accuracy for healthy motor diagnostics and 93.3% for faulty motor classification. The MLP algorithm reached 70% of precision for the detection of healthy motor and a precision of 93.8% for the estimation of faulty motor. The FAM classifier obtained

Table 2. Confusion matrix - healthy × defective motor

Predicted classes								
FAM			SVM/SMO		k-NN		MLP	
Classes	Healthy	Defective	Healthy	Defective	Healthy	Defective	Healthy	Defective
Healthy	1015	637	340	1312	1346	306	1157	495
Defective	1358	5891	14	7235	473	6776	448	6801

respectively an accuracy of 61.4% and 81.3% for indication of healthy motor and defective motor. Already the SVM/SMO algorithm obtained 20% accuracy for indication of motor operating at normal conditions and 99.8% of precision in case of motor operating with defective status.

Tables 3 and 4 present the classification results for motors operating with multiple faults.

Table 3. Experimental results - multiple fault classification

Attributes	FAM	SVM/SMO	k-NN	MLP
Correctly classified instances (%)	74,05	76,22	87,56	89,37
Incorrectly classifed instances (%)	25,95	23,78	12,44	10,63
Kappa statistic	0,64	0,67	0,83	0,85
Mean absolute error	0,12	0,27	0,06	0,06
Root mean squared error	0,36	0,35	0,24	0,20
Relative absolute error (%)	35,49	76,10	17,06	17,54
Root relative squared error (%)	84,25	82,54	58,32	47,30
Time taken to build model (s)	1,67	28,69	0,02	917,33

Based on the results presented on Table 1 it is possible to observe that in the case of the classification of multiple defects, in general, the MLP algorithm reached the best classification accuracy with 89.37% of global precision. In the sequence, the algorithms k-NN, SVM/SMO and FAM, obtained 87.56%, 76.22% and 74.05% of overall accuracy, respectively. The classifiers k-NN and PMC obtained a Kappa index higher than 0.80 indicating perfect concordance. Both SVM/SMO and FAM algorithms, obtained Kappa coefficients indicates substantial agreement.

As per Table 4, the MLP algorithm reach an accuracy of 75.3% for the classification of healthy machine, 92.1% for diagnostics of motor operating with stator short-circuit fault and 98.3% of precision for the diagnostics of broken rotor bars, and 82.8% for the diagnosis of bearing fault. In this same scenario, the algorithm k-NN has the highest precision index for the diagnosis of healthy motor, reaching 82.6%. As for the diagnosis of motor operating with stator short-circuit fault an accuracy of 86.6% was obtained, followed by 96.2% accuracy for motor diagnosis

Table 4. Confusion matrix - multiple fault classification

Predicted classes								
FAM					SVM/SMO			
Classes	Healthy	Stator	Rotor	Bearing	Healthy	Stator	Rotor	Bearing
Healthy	1129	224	91	208	780	145	86	641
Stator	258	2037	169	164	182	2064	186	196
Rotor	124	198	2544	158	56	127	2793	48
Bearing	389	169	157	882	325	64	61	1147
k-NN					MLP			
Classes	Healthy	Stator	Rotor	Bearing	Healthy	Stator	Rotor	Bearing
Healthy	1364	176	20	92	1240	159	17	236
Stator	216	2276	40	36	146	2420	24	38
Rotor	35	36	2910	43	6	31	2973	14
Bearing	239	73	41	1244	210	37	28	1322

of motor operating with broken rotor bars, and 77.9% for diagnosis of motor operating with bearing defect. Also, Table 4 shows that the lowest classification rates were obtained by the SVM/SMO algorithm reaching only 47.2% of precision.

4 Conclusion

In this work, different classification strategies based on learning machines algorithms were employed for the primary fault identification followed by the configuration of these same classifiers for the diagnose of multiple faults in machines driven by different inverter models running at a wide speed and load levels.

The main aspect related with this work is the investigation performed from the direct analysis of the amplitude of stator current by means of the employment of learning machine techniques. Also, this work present the comparison of 4 different intelligent strategies for the task of fault diagnosis.

k-NN and MLP algorithms achieved classification rates over to 87%.

Additionally, this work presents the concept of multiple faults classification in inverter fed three-phase induction motors where bearings, stator short-circuit and broken rotor bars faults were properly addressed. The results indicate that this technique allows the employment of this strategy in real-time monitoring conditions as a support tool for the creation of indicators that can be used for proper equipment shut-down planning.

Acknowledgement. The authors would like to thank the support and motivation provided by the Federal Technological University of Paraná.

References

1. Seera, M., Lim, C.P., Nahavandi, S., Loo, C.K.: Condition monitoring of induction motors: a review and an application of an ensemble of hybrid intelligent models. Expert. Syst. Appl. **41**(10), 4891–4903 (2014)
2. Seshadrinath, J., Singh, B., Panigrahi, B.: Investigation of vibration signatures for multiple fault diagnosis in variable frequency drives using complex wavelets. IEEE Trans. Power Electron. **29**(2), 936–945 (2014)
3. Zarei, J., Tajeddini, M.A., Karimi, H.R.: Vibration analysis for bearing fault detection and classification using an intelligent filter. Mechatronics **24**(2), 151–157 (2014)
4. Das, S., Purkait, P., Koley, C., Chakravorti, S.: Performance of a load-immune classifier for robust identification of minor faults in induction motor stator winding. IEEE Trans. Dielectr. Electr. Insul. **21**(1), 33–44 (2014)
5. Jia, F., Lei, Y., Lin, J., Zhou, X., Lu, N.: Deep neural networks: a promising tool for fault characteristic mining and intelligent diagnosis of rotating machinery with massive data. Mech. Syst. Signal Process. **72–73**, 303–315 (2016)
6. Xu, Z., Li, Y., Wang, Z., Xuan, J.: A selective fuzzy artmap ensemble and its application to the fault diagnosis of rolling element bearing. Neurocomputing **182**, 25–35 (2016)
7. Baraldi, P., Cannarile, F., Maio, F.D., Zio, E.: Hierarchical k-nearest neighbours classification and binary differential evolution for fault diagnostics of automotive bearings operating under variable conditions. Eng. Appl. Artif. Intell. **56**, 1–13 (2016)
8. Riera-Guasp, M., Antonino-Daviu, J.A., Capolino, G.A.: Advances in electrical machine, power electronic, and drive condition monitoring and fault detection: state of the art. IEEE Trans. Ind. Electron. **62**(3), 1746–1759 (2015)
9. Lee, J., Wu, F., Zhao, W., Ghaffari, M., Liao, L., Siegel, D.: Prognostics and health management design for rotary machinery systems–reviews, methodology and applications. Mech. Syst. Signal Process. **42**(1–2), 314–334 (2014)
10. Bellini, A., Filippetti, F., Tassoni, C., Capolino, G.A.: Advances in diagnostic techniques for induction machines. IEEE Trans. Ind. Electron. **55**(12), 4109–4126 (2008)
11. Maruthi, G.S., Hegde, V.: Application of MEMS accelerometer for detection and diagnosis of multiple faults in the roller element bearings of three phase induction motor. IEEE Sens. J. **16**(1), 145–152 (2016)
12. do Nascimento, C.F., de Oliveira Jr., A.A., Goedtel, A., Serni, P.J.A.: Harmonic identification using parallel neural networks in single-phase systems. Appl. Soft Comput. **11**(2), 2178–2185 (2011)
13. Lughofer, E., Buchtala, O.: Reliable all-pairs evolving fuzzy classifiers. IEEE Trans. Fuzzy Syst. **21**(4), 625–641 (2013)
14. Lughofer, E., Weigl, E., Heidl, W., Eitzinger, C., Radauer, T.: Integrating new classes on the fly in evolving fuzzy classifier designs and their application in visual inspection. Appl. Soft Comput. **35**, 558–582 (2015)
15. Haykin, S.O.: Neural Networks and Learning Machines, 3 edn. (2008). Hardcover
16. Han, J., Kamber, M., Pei, J.: Data Mining: Concepts and Techniques, 3rd edn. Elsevier, Morgan Kaufmann, Amsterdam (2011)
17. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification, 2nd edn. Willey, New York (2001)

A v-Hog Tensor Based Discriminant Analysis for Small Size Face Recognition

Belavadi Bhaskar(✉) and K. V. Mahendra Prashanth

SJB Institute of Technology, Affiliated to VTU, Bengaluru, Karnataka, India
bhaskar.brv@gmail.com, kvmprashanth@yahoo.com

Abstract. Apart from Illumination, Pose and Expression variations, low dimension is also a primary concern that spiflicate the performance of face recognition system. This work distils to applying v-Hog Tensor discriminant analysis on small sized face image to yield good result in terms of correctness rate. Firstly the face image is mapped on to w-quintuple Colorspace to effectively interpret information existing in the image. Further discriminant features are extracted out of Tensor plane to bore on the confounded image due to reduction of image size. To exhibit the beauty of the feature, v-Hog [22] is adopted. The obtained features are further mapped to a lower dimension space for efficient face recognition. In this work the effect of fSVD [17] with bias is also considered to fortify the recognition system. Finally, for classification five different similarity measures are used to obtain an average correctness rate.

Keywords: w-Quintuple · Biased-fSVD · v-Hog · Tensor

1 Introduction

The colourful cosmos of god's creation has some singular lineament in every entity created by him, among which color provides useful visual information. In recent years, considerable research has been conducted on face recognition for color images. Image gradients with schur decomposition are invariant to changes in illumination and pose variations [1]. Face images can be represented in terms of Eigen faces [2], which facilitates gradual changes in face and also simple to compute. A rather more unquestionable 2DPCA algorithm was developed in [20], in which covariance matrix is directly constructed using original image matrices. In [11] volume measure was used as classifier with 2DPCA, which outperformed classical distance measure classifiers. For effectual face representation [3], (2D) 2PCA was coined which works on row and column directions of face images with much diluted coefficients for face representation. Multiple Kernel Local Fisher Discriminant analysis (MKLFDA) was proposed [12] to delineate the nonlinearity of face images. Independent component neighborhood preserving analysis [16] (IC-NPA)was proposed to retain the inviolable incisive potential of LDA and at the same time upholding the intrinsic geometry of the within-class information of face images. As stated earlier, combining the features always aids

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 126–136, 2019.
https://doi.org/10.1007/978-981-13-5758-9_11

in improved efficiency, in [4,5] Eigen and Fisher face are combined using wavelet fusion and 2FNN (Two- Feature Neural Network) is used for classification to improve the recognition rate. To surmount the problem of small sample size, an unsupervised nonlinear Spectral Feature Analysis (SFA) was minted, which extracts discriminative features out of a small sample sized face image. In [15] the author proposed a method for the recognition of facial expressions in face images using landmark points in linear subspace domain.

Nonlinear kernel sparse representations [6] aids in robustness of face recognition against illumination and occlusion by developing pixel-level and region-level kernels hiring local features. Immix of DWT, DFT and DCT, expeditiously educe pose, translation and illumination invariant features [7]. Decimated redundant DWT was proposed in [10] to cauterize the effect of translation of face image. Gabor jets [8] feature extractor improves accuracy rate and also reduces the computational time with the facilitation of Borda count classifier. Grayscale Arranging Pairs (GAP) [9] delineates the holistic data in the whole face image and demonstrates mellowed raciness against illumination changes. Combination of two effectual local descriptors Gabor wavelet and enhanced local binary pattern with generalized neural network [13] is insensitive to small changes in input data and is robust-to-slight variation of imaging conditions and pose variations. A more versatile improved Quarternion LPP [14] was purported to exlploit the color information in face images. Multiple features improve the recognition accuracy than having single features in face recognition system [23]. Eigen space based sparse algorithm was proposed by Hou et al [24] which adopted HOG features to give good results in terms of accuracy, but was not suited for multiple face recognition problems. A low dimensional feature with the essence of discriminant factor will always improve the efficiency of the face recognition system [25]. Keeping this in mind we have come up with the idea of discriminant analysis with our earlier work to assist in good recognition system.

This paper is indited with three sections leaving out the current section, starting with proposed methodology. In Sect. 2 we elucidate our idea of overcoming the different perturbs that prevails in the face recognition system. Section 3 focuses on the result component. The last section speaks about the conclusion of this paper spotlighting the kernel of the entire paper.

2 Proposed Methodology

In this section we delineate the proposed face image recognition technique in particular. We have considered the baseline problem of any face image, which is its size and have come up with an intriguing method to tackle small sized images. Initially face image is mapped onto w-quintuple Colorspace to improvise on highlighting significant information existing in the image. Further fSVD with variable weights is applied to make face image ready for better feature extraction. Biased fSVD is applied to low rank Tensor discriminant analysis on to a low dimension face image to yield good result in terms of correctness rate. Firstly the face image is mapped on to w-quintuple Colorspace to effectively interpret

information existing in the image. Further discriminant features are extracted out of Tensor plane with a low rank to bore on the confounded image due to reduction of image size. To inhibit the feature size further, the low rank features are further mapped to a lower dimension space for efficient face recognition. Finally, for classification five different similarity measures are used to obtain an average correctness rate.

2.1 w-Quintuple Colorspace

Different Color models possess different discriminative index. Compared to greyscale images, a color image embeds in it much more utile information for improving face recognition performance [21]. Making use of these properties, we derive an image from the original image by mapping it on quintuple Colorspace. The color face image M (x, y, z) of size (l x m x 3) is mapped as given below,

$$M_{cj}(x,y,z) = M(x,y,z) \rightarrow \{K_1{*}Lab, K2{*}LUV, K3{*}HSV, K4{*}YIQ, K5{*}XYZ\} \tag{1}$$

where i = 1, 2, 3, 4, 5 and K_1, K_2, K_3, K_4, K_5 are weights for colorspace

The weights for the Quartette Colorspace are set based on the behaviour of the Colorspace with fSVD.

$$I_{fi}(x,y,z) = w * \frac{(F_{ci}(x,y,z) + \xi_1 * D_i(x,y,z))}{(1 + (\xi_1, \xi_2, \xi_3, \xi_4))} \tag{2}$$

where $D_i(x,y,z) = U * \Sigma^{\gamma_i} * V^T$ and w is the bias for fSVD

$$[U, \Sigma, V] = SingularValuedecomp(Inorm(x,y))$$

$$\xi_1 = 0.2, \xi_2 = 0.9, \xi_3 = 0.05, \xi_4 = 0.07, \xi_5 = 0.1, \gamma_1 = 0.695, \gamma_{2,3,4,5} = 0.895$$

2.2 Tensor Based Discriminant Analysis

A third order tensor Linear Discriminant Analysis is adopted to perform structured dimensionality reduction. To achieve this reduction, a three Eigen vector problem is solved each for one color plane. This analysis finds a map from $R^{a_1}XR^{a_2}$ to $R^{b_1}XR^{b_2}$ $(a_1 < b_1$ & $a_2 < b_2)$. The matrices in the eigen- problems are of size b_1Xb_1 or b_2Xb_2, which are much smaller than the matrices of size b X b ($b = b_1Xb_2$) obtained by other linear subspace methods like PCA. Hence Tensor based Discriminant Analysis is much more computationally efficient in time and storage.

Instead of performing the feature size reduction by upholding the class discriminative information, we try to reduce the dimension of the face image at the image level itself by singular value decomposition and to maintain the information content intact the resized actual image is also taken forwarded. Now in finding the projections that maximises the ratio of inter space face and intra space face spread, both tensor face and original resized images are taken into account. The projections C of length k from I image of dimension a_1 X a_2 X 3

are obtained by means of projection vectors e_k. The equation for the projection from I to C is given below

$$C = e_k^T * I \tag{3}$$

The projections e_k that we are interested in here is obtained in such a way that it maximises the ratio of inter space face to intra space face spread. As the projection is a vector of size K , we find the vectors $\{e_1, e_2, ..., e_k\}$ whose columns are the eigen vectors gibing to the tumid eigen vector i.e.,

$$E^* = \{e_1|e_2|e_3|\cdots|e_k\} = max\_arg\left\{|(E^T S_{It} E)|/|(E^T S_{Ie} E)|\right\} \tag{4}$$

where $S_I t$ is inter space face spread matrix and $S_I e$ is intra space face spread matrix.

S_{It} is given by the equation,

$$S_{It} = \sum (I - mu_i)(I - mu_i)^T \tag{5}$$

where $mu_i = \frac{1}{N_i} \sum(I))$ and $i = 1 \cdots K$

And S_{Ie} is given by the equation,

$$S_{Ie} = \sum (mu_i - mu)(mu_i - mu)^T \tag{6}$$

where $m = \frac{1}{N} \sum(N_i mu_i)$

The final projection vectors are actually found in our case by singular value decomposition of the spread matrices found just now. These vectors are used to project our face image onto a lower dimensional subspace which has proved very effective in terms of recognition accuracy after conducting experiments. A pictorial view of Tensor discriminant analysis is shown in Fig. 1. Examples of tensor discriminant images pertaining to other databases are show in Fig. 2.

2.3 Variant Hoglets

HOG features were first introduced by Dalal and Triggs [18]. The essence of the Histogram of Oriented Gradient descriptor or signifier is that local object appearance and shape within an image can be described by the distribution of local intensity gradients or adjoin directions. The effectuation of these signifiers can be achieved by dividing the image into small colligated regions, called cells, and for each cell amassing a histogram of gradient directions or adjoin orientations for the pixels within the cell. The combination of these histograms then represents the signifiers. The HOG signifiers has few crucial advantages like invariance to geometric and photometric transformation, which are helpful in face recognition applications. Gradient orientation preserves the important structure of the image and is much useful in face recognition [19]. To improve the face recognition accuracy, in our work we propose variant Hoglets by adopting HOG with various gradient filters which is adopted based on the colorspace. In an attempt to imbibe the behaviour of wavelets with HOGs under different

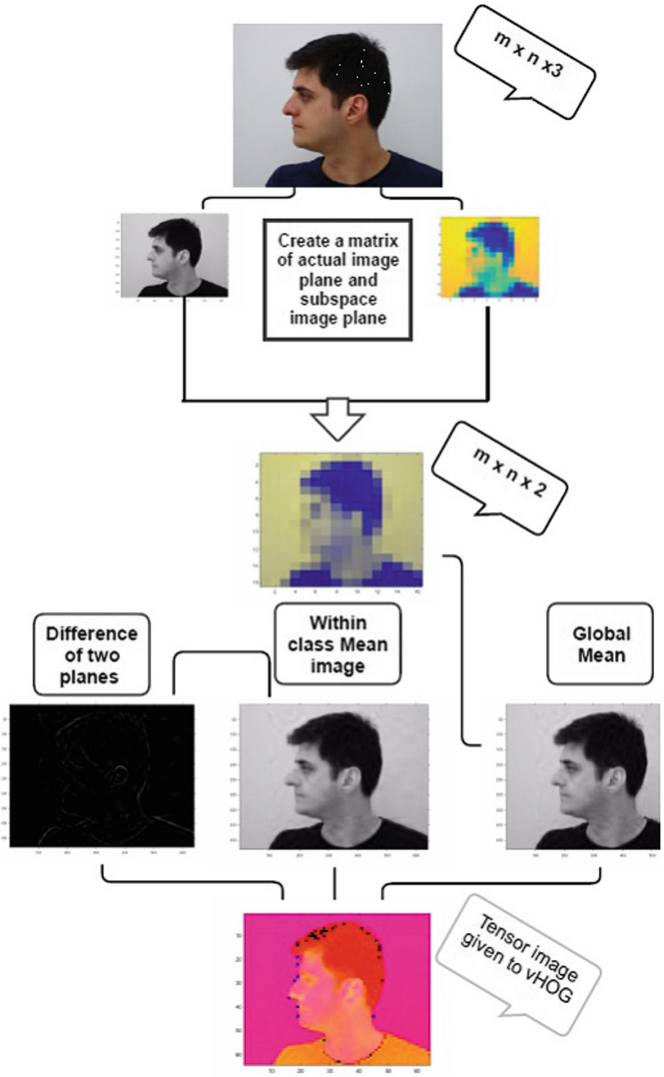


Fig. 1. Flow of Tensor discriminant analysis applied to a single image

colorspace, we have incubated daubechies as filter kernels for HOG which are defined below.

$$db_n = Coeff_1, Coeff_2, .., Coeff_n$$

where n = 7, 9, 11 and 7 are used in combination with LAB, LUV, HSV and XYZ colorspace respectively.

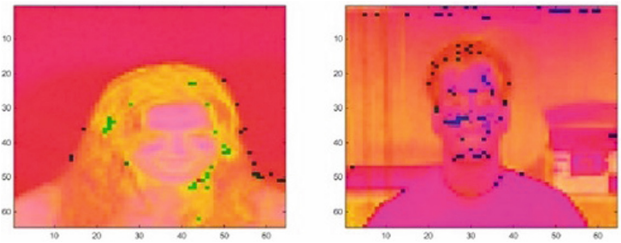


Fig. 2. Examples images of Tensor discriminant

Further more the below filter kernels are also adopted with YIQ alone unlike our previous work, which adopts two different kernels for two different colorspace.

$$[0.707, 0, -0.707] \text{ and } [0.707, 0, -0.707]^T$$

The features obtained from the above proposed method is further compressed by projecting them on to the lower dimensional subspace and classified using an aggregate of five different distance measures.

3 Result

This segment of the paper limns on the results with the help of six different databases which were used in our previous work [22]. We embark into the recognition system by mapping face image onto w-quintuple Colorspace and biased fSVD is then applied to normalize the mapped image. Further to extract the shape descriptors vHOG is applied to the derived face image and is projected onto the reduced dimensionality Eigen subspace. We have used five different distance measures to formalize the inquiry face image.

To assess our proposed technique, we have expended six different database viz., Georgia Tech, Dr. Libor Spacek Faces94, CVL, FEI, Indian face and SJB face database. GTech has 50 persons × 15 images/person. This database shows frontal face with variations in expressions, lighting conditions, cluttered background and scale. Libor has 152 × 20 images with tilt of head and considerable expression changes. CVL consists of 114×7 with pose and expression variations. FEI has 200×14 images in total with change in pose, expression and illumination. Indian face database has 671 images out of which we have used 61×7 images and it has considerable variations in pose and expression. SJB face database is our own face database which consists of 10 persons × 10 images/person. This also has substantial variations in expression, pose and illumination.

In all our experiments we have considered -

- Single image per person for training phase and rest all images are used as query images in every case.
- fSVD coefficients are - $\xi_1 = 0.2, \xi_2 = 0.9, \xi_3 = 0.05, \xi_4 = 0.07, \xi_5 = 0.1, \gamma_1 = 0.695, \gamma_{2,3,4,5} = 0.895$.

- The bias w used in fSVD is 0.9, 0.15 and 0.1 for first 3, 4th and 5th respectively.
- v-Hoglets kernels used are $\text{db}_n = Coeff_1, Coeff_2, .., Coeff_n$, where n = 7, 9, 11 and 7 and also $[0.707, 0, -0.707]$ and $[0.707, 0, -0.707]^T$.

3.1 Experiment - I - Effect of Image Size on Recognition Accuracy

This segment of the paper limns on the results with the help of six different databases to bridle fortification of our proposed algorithm. We embark into the recognition system by mapping face image onto w-quintuple Colorspace and biased fSVD is then applied to normalize the mapped image. Further to extract the shape descriptors vHOG is applied to the derived face image and is projected onto the reduced dimensionality Eigen subspace. We have used five different distance measures to formalize the inquiry face image.

3.2 Experiment - II - Effect of fSVD on the Extracted Features

In this experiment we have found the correlation coefficients for the features of images with and without fSVD. Here we are comparing the proposed method with the earlier work on v-Hog. For experimentation purpose only two databases have been considered.

Below table suggests that the correlation coefficient is high for features extracted using fSVD than without fSVD. And at the same time, our proposed method is able to discriminate within and between class correlation coefficients far better than the previous work. The highlight of this experiment is that it evinces the presence of fSVD changes the way features extracted out of the same face image. Even though the tensor discriminant is widens the spread distance between the class, fSVD is able to further improve the within and between class discrimination (Table 1).

3.3 Experiment - III - Effect of v-Hog Tensor on Feature Vector Size and Execution Time

In the first part of the experiment we have compared the feature vector size required to obtain highest efficiency in terms of recognition accuracy. The result suggests that proposed algorithm gains good accuracy at lower feature dimension compared to the v-Hog alone. V-Hog on the other side requires more features to get good results in terms of accuracy. The results are tabulated in Table 3 and A1 corresponds to present context while A2 is our previous work (Table 2).

The second half of the experiment aims at witnessing a very less execution time as compared to other existing state of the art algorithms. Different size face images are considered in obtaining the result. Table 4 suggests that v-Hog requires highest execution time and the least is our proposed methodology.

Table 1. Effect of image size on accuracy

Database	Algorithm	Image size	Accuracy
GT	v-Hog Tensor	8	80.86
		16	74.22
		32	74.22
	v-Hog	8	74.22
		16	74
		32	75.56
CVL	v-Hog Tensor	8	62.67
		16	60
		32	54.33
	v-Hog	8	51
		16	60
		32	51
SJB	v-Hog Tensor	8	84.85
		16	85.86
		32	85.85
	v-Hog	8	85.86
		16	88.89
		32	84.85
Indian	v-Hog Tensor	8	76.22
		16	82.44
		32	80.22
	v-Hog	8	62.79
		16	66.87
		32	72.79
FEI	v-Hog Tensor	8	92.89
		16	93.11
		32	84.22
	v-Hog	8	67
		16	72.44
		32	79.11
DrLibor	v-Hog Tensor	8	99.78
		16	99.56
		32	99.33
	v-Hog	8	98.89
		16	99.33
		32	99.33

Table 2. Average correlation coefficient

Algorithm	Database	With fSVD		Without fSVD	
		Within	Between	Within	Between
v-Hog Tensor	CVL	0.9785	0.4881	0.9546	0.6895
	GT	0.9994	0.3562	0.9794	0.7652
v-Hog	CVL	0.9719	0.9426	0.9616	0.9485
	GT	0.9584	0.9213	0.9490	0.9229

Table 3. Feature vector size versus accuracy for different database

Database	Size	Feature vector size (accuracy)	
		A1	A2
GT	8	7 (80.86)	33 (74.22)
	16	31 (74.22)	27 (74)
	32	43 (74.22)	32 (75.56)
CVL	8	43 (62.67)	45 (51)
	16	46 (60)	38 (60)
	32	47 (54.33)	41 (51)
SJB	8	2 (84.8485)	3 (85.86)
	16	3 (85.86)	3 (88.89)
	32	3 (85.85)	4 (84.85)
Indian	8	50 (76.2222)	44 (62.7879)
	16	37 (82.4444)	48 (66.8687)
	32	48 (80.2222)	48 (72.7879)
FEI	8	48 (92.8889)	45 (68.67)
	16	50 (93.1111)	48 (72.4444)
	32	50 (84.2222)	50 (79.1111)
Dr Libor	8	21 (99.78)	33 (98.89)
	16	37 (99.56)	33 (99.33)
	32	15 (99.33)	39 (99.33)

Table 4. Comparision of execution time with different algorithms

Algorithm	Time/image (sec)	Image size
v-Hog Tensor	0.377291	16 * 16
	0.315095	8 * 8
Ridgelet PCA	0.884674	592 * 896
	0.419039	32 * 32
	0.330754	8 * 8
PCA-GMM	0.491761	8 * 8
v-Hog	0.530699	8 * 8

4 Conclusion

The work suggested here has not only taken care of PIE problem but also has attended to single sample per person and small sample size. The adumbrated algorithm has redacted the result in a very considerable manner by highlighting the use of v-Hog on Tensor discriminant subspace for different state of the art databases. The purported method is proficient in terms of accuracy, feature vector size, execution time and most importantly the single train image and small sample size which are haunting problems in the recent years of face recognition research problem.

References

1. Ghinea, G., Kannan, R., Kannaiyan, S.: Gradient-orientation-based PCA subspace for novel face recognition. IEEE Access **2**, 914–920 (2014)
2. Turk, M., Pentland, A.: Eigenfaces for recognition. J. Cogn. Neurosci. **3**(1), 71–86 (1991)
3. Zhang, D., Zhou, Z.-H.: (2D) 2PCA: two-directional two-dimensional PCA for efficient face representation and recognition. Neurocomputing **69**(1), 224–231 (2005)
4. Devi, B.J., Veeranjaneyulu, N., Kishore, K.V.K.: A novel face recognition system based on combining eigenfaces with fisher faces using wavelets. Procedia Comput. Sci. **2**, 44–51 (2010)
5. Wang, F., Wang, J., Zhang, C., Kwok, J.: Face recognition using spectral features. Pattern Recognit. **40**(10), 2786–2797 (2007)
6. Kang, C., Liao, S., Xiang, S., Pan, C.: Kernel sparse representation with pixel-level and region-level local feature kernels for face recognition. Neurocomputing **133**, 141–152 (2014)
7. Krisshna, N.L.A., Kadetotad Deepak, V., Manikantan, K., Ramachandran, S.: Face recognition using transform domain feature extraction and PSO-based feature selection. Appl. Soft Comput. **22**, 141–161 (2014)
8. Perez, C.A., Cament, L.A., Castillo, L.E.: Methodological improvement on local Gabor face recognition based on feature selection and enhanced Borda count. Pattern Recognit. **44**(4), 951–963 (2011)
9. Zhao, X., He, Z., Zhang, S., Kaneko, S., Satoh, Y.: Robust face recognition using the GAP feature. Pattern Recognit. **46**(10), 2647–2657 (2013)
10. Li, D., Tang, X., Pedrycz, W.: Face recognition using decimated redundant discrete wavelet transforms. Mach. Vis. Appl. **23**(2), 391–401 (2012)
11. Meng, J., Zhang, W.: Volume measure in 2DPCA-based face recognition. Pattern Recognit. Lett. **28**(10), 1203–1208 (2007)
12. Wang, Z., Sun, X.: Multiple kernel local Fisher discriminant analysis for face recognition. Signal Process. **93**(6), 1496–1509 (2013)
13. Sharma, P., Arya, K.V., Yadav, R.N.: Efficient face recognition using wavelet-based generalized neural network. Signal Process. **93**(6), 1557–1565 (2013)
14. Wu, S.: Quaternion-based improved LPP method for color face recognition. Opt. Int. J. Light. Electron Opt. **125**(10), 2344–2349 (2014)
15. Aifanti, N., Delopoulos, A.: Linear subspaces for facial expression recognition. Signal Process. Image Commun. **29**(1), 177–188 (2014)
16. Hu, H.: ICA-based neighborhood preserving analysis for face recognition. Comput. Vis. Image Underst. **112**(3), 286–295 (2008)

17. Bhaskar, B., Mahantesh, K., Geetha, G.P.: An investigation of fSVD and ridgelet transform for illumination and expression invariant face recognition. In: El-Alfy, E.-S.M., Thampi, S.M., Takagi, H., Piramuthu, S., Hanne, T. (eds.) Advances in Intelligent Informatics. AISC, vol. 320, pp. 31–38. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-11218-3_4
18. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893 (2005)
19. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Principal component analysis of image gradient orientations for face recognition. In: IEEE International Conference on Automatic Face and Gesture Recognition and Workshops (FG 2011), pp. 553–558, March 2011
20. Yang, J., et al.: Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Trans. Pattern Anal. Mach. Intell. **26**(1), 131–137 (2004)
21. Torres, L., Reutter, J.-Y., Lorente, L.: The importance of the color information in face recognition. In: Proceedings of the 1999 International Conference on Image Processing, ICIP 1999, vol. 3. IEEE (1999)
22. Belavadi, B., Mahendra Prashanth, K.V.: An exploration of v-HOG on w-Quartette space for multi face recognition issues. In: Sa, P.K., Sahoo, M.N., Murugappan, M., Wu, Y., Majhi, B. (eds.) Progress in Intelligent Computing Techniques: Theory, Practice, and Applications. AISC, vol. 518, pp. 219–226. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-3373-5_21
23. Wang, D., Wang, X., Kong, S.: Integration of multi-feature fusion and dictionary learning for face recognition. Image Vis. Comput. **31**(12), 895–904 (2013)
24. Hou, Y.-F., Pei, W.-J., Chong, Y.-W., Zheng, C.-H.: Eigenface-based sparse representation for face recognition. In: Huang, D.-S., Jo, K.-H., Zhou, Y.-Q., Han, K. (eds.) ICIC 2013. LNCS (LNAI), vol. 7996, pp. 457–465. Springer, Heidelberg (2013). https://doi.org/10.1007/978-3-642-39482-9_53
25. Bhaskar, B., Anushree, P.S., Divya Shree, S., Mahendra Prashanth, K.V.: Quantitative analysis of kernel principal components and kernel fishers based face recognition algorithms using hybrid gaborlets. Procedia Comput. Sci. **58**, 342–347 (2015)

A Novel Method for Stroke Prediction from Retinal Images Using HoG Approach

R. S. Jeena[1(✉)], A. Sukesh Kumar[1], and K. Mahadevan[2]

[1] College of Engineering Trivandrum, Trivandrum, Kerala, India
jeena_rs@yahoo.com, drsukeshkumar@yahoo.in
[2] Sree Gokulam Medical College and Research Foundation, Trivandrum, Kerala, India
eyemahadevan@rediffmail.com

Abstract. Stroke is one of the principal reasons for adult impairment worldwide. Retinal fundus images are analyzed for the detection of various cardiovascular diseases like Stroke. Stroke is mainly characterized by soft and hard exudates, artery or vein occlusion and alterations in retinal vasculature. In this research work, Histogram of Oriented Gradients (HoG) has been implemented to extract features from the region of interest of retinal fundus images. This innovative method is assessed for the computer aided diagnosis of normal healthy and abnormal images of stroke patients. A comparative analysis has been made between the extracted HoG features and Haralick features. HoG features extracted from the region of interest, when given to a Naïve Bayes classifier provides an accuracy of 93% and a Receiver Operating Characteristic (ROC) curve area of 0.979.

Keywords: Stroke · Histogram of Oriented Gradients (HoG) Haralick features · Naïve Bayes classifier

1 Introduction

Stroke or Cerebrovascular Accident (CVA) is the one of the critical reason for death worldwide [1]. Stroke is a medical condition that happens because of lack of supply of blood to the cerebrum cells. This damages the brain cells eventually leading to their death. Coagulation or a break in the blood vessel can interrupt the blood flow to cerebral nervous system. Coagulation in the vessel supplying blood to brain causes Ischemic stroke [2] and blood vessel rupture causes Hemorrhagic stroke. About 80% of the stroke cases belong to ischemic category. The major reason for stroke is contributed by unhealthy lifestyles [3]. Early identification of stroke is still a challenge and is really significant in the area of biomedical research.

The ability to picture the retina and develop various methods for analyzing the images is of great interest. Investigations show that microvasculature of retina and cerebrum is closely related in terms of structure and function [4]. The retina can be considered as a biomarker, as any variation in the cerebral blood vessel will be reflected in the retinal blood vessel. Cardiovascular ailments like stroke show itself in retina in different ways. High Blood pressure and plaque deposition cause variations in the ratio

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 137–146, 2019.
https://doi.org/10.1007/978-981-13-5758-9_12

between the diameters of retinal blood vessels. Contraction of arteries and expansion of veins is linked with an enhanced danger of stroke. Research shows that changes such as micro aneurysms or narrowing of arterioles seen in the retina present hazard factors for cerebrovascular disease such as stroke [4]. Morphological variations in blood vessels like branching pattern, width, presence of lesions, branching angle and coefficient are some of the abnormalities in retinal vasculature associated with cardiovascular diseases like stroke. This work is an extension of author's earlier works for stroke prediction [5]. Literature survey shows that retinal imaging has a strong potential in predicting cardiovascular diseases like stroke.

Texture and higher order spectral features were employed in [12] for detection of eye diseases. Felzenszwalb et al. [11] demonstrated that directed gradient was specified in accordance with the contrast direction with a measurement of twice the number of orientations. Haralick texture features based glaucoma detection is presented in [13]. Thirteen Haralick features were extracted and are fed to K-Nearest Neighbor classifier which gave an accuracy of 98%. HOG eye detectors were used in combination with Hough transform [14] and used with meta classifiers for better execution [15]. An algorithm for precise eye detection using meta classifiers trained with Haar features and HoG features were implemented in this work. This work gave outstanding results when applied on certain standard databases.

Andreas et al. [17] demonstrated an efficient eye detector by applying Principal Component Analysis on HoG gradients. Dimensionality reduction was achieved and Support Vector Machine was used for testing on standard databases. Research by Sadak et al. [18] demonstrates SURF, HOG and LBP texture descriptors were used for the extraction for the features in automatic diagnosis of diabetic retinopathy. The histogram representation was given to SVM for classification. Linear kernel gave promising results by providing accuracies of 97.20% and 99.8% in single and multiple-based dictionaries. A method for glaucoma detection using GLCM features and logistic regression classifier from ocular thermal images is discussed in [19]. Four features were extracted in GLCM and were used to train a logistic regression classifier for glaucoma detection. Shreya et al. [20] proposed an approach to develop an automated system for classifying extra ocular diseases.

Nowadays, machine learning has been modernized by deep learning, surpassing human prediction on an extensive range of problems. Among the various deep learning frameworks, Convolutional Neural Networks (CNN) can intrinsically perform substantial texture analysis. But it requires a large amount of data and is computationally expensive to train, as it requires machines equipped with extensive GPUs. Since the number of training samples is limited in our work due to data privacy, texture descriptors combined with machine learning were preferred to deep learning.

Early prognosis of stroke is vital for timely prevention and cure, as it can lead to long term disability. Various traditional numerical methods and computational intellectual models have been introduced for predicting the risk and outcome of stroke. Risk prediction is mainly based on physiological factors, behavioral factors, demographic and socio-economic factors [1]. Even though Framingham model has been introduced to provide valuable insights into cardiovascular risk prediction, the accuracy of the model is limited because of the exclusion of certain vital risk factors and its dependence on demography. This novel approach aims in developing a sophisticated diagnostic

strategy for non invasive stroke prediction from retinal fundus images. Texture analysis seems to be appropriate for analyzing retinal images because human retina has differently textured regions and its texture varies in regions of pathological signs. In this work, retinal fundus images of both healthy and stroke patients have been analyzed and the features are extracted using Histogram of Oriented Gradients (HoG) from the region of interest. Haralick features have also been computed from the ROI and are compared with HoG. Experimental results show that HoG outperforms Haralick features by giving promising results. Naive Bayes classifier is used for the classification of healthy and stroke affected images and performance metrics are evaluated. This novel approach can surely aid the physicians in stroke prognosis.

2 Methodology

Retinal images of healthy group and stroke patients have been collected from Sree Gokulam Medical Centre and Research Foundation, Kerala (control n = 80, Stroke n = 50).

The basic block diagram of the proposed method is given in Fig. 1.

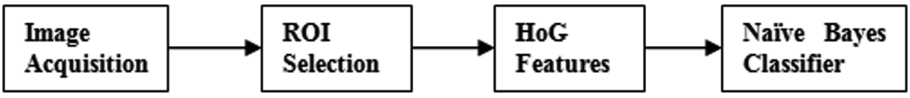


Fig. 1. Block diagram of proposed method

Image acquisition is done by a fundus camera which produces fundus images that document the present ophthalmoscopic view of a retina. Selection of the region of interest to measure HoG features is a critical phase in retinal image analysis. In zonal representation of fundus image, region between optic disc margin to 0.5 optic disc diameters (ODD) is Zone A, Zone B corresponds to 0.5–1.0 ODD and Zone C corresponds to 0.5–2.0 ODD away from the optic disc boundary. As Zone C covers most of the variations in retinal vasculature, the region of interest selected in our study is Zone C. Zonal representation is depicted in Fig. 2.

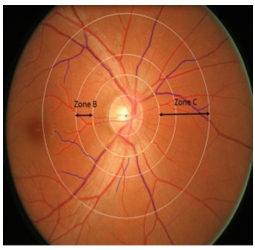
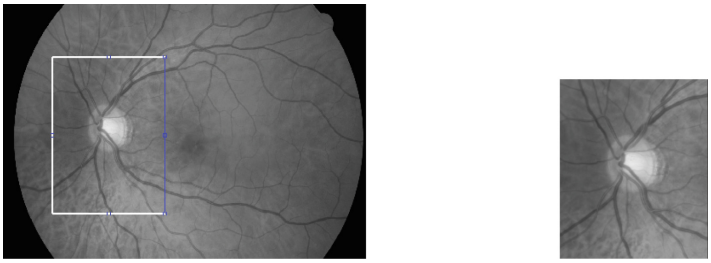


Fig. 2. Division of zones

To obtain Zone C, Centre of the optic disc is located manually and a region of size 256 × 256 has been cropped from the original image for feature extraction. ROI selection and the cropped image are given in Figs. 3(a) and (b) respectively.



(a) (b)

Fig. 3. (a) Gray scaled original image (b) Cropped image

2.1 Histogram of Oriented Gradients

The Histogram of oriented gradients (HOG) was developed by Dalal et al. [6] which finds application in object recognition and other computer vision tasks. This descriptor is capable of capturing fine details in images. Research show that HoG performs better than other descriptors and the implementation is simple. The essential notion behind the implementation of HoG descriptor in this work is that shape and local object appearance within a retinal fundus image can be revealed by the distribution of intensity derivatives. Normalization of local histograms results in preferable invariance to modification in illumination and shadowing [15]. Figure 4 shows the block diagram of HoG approach.

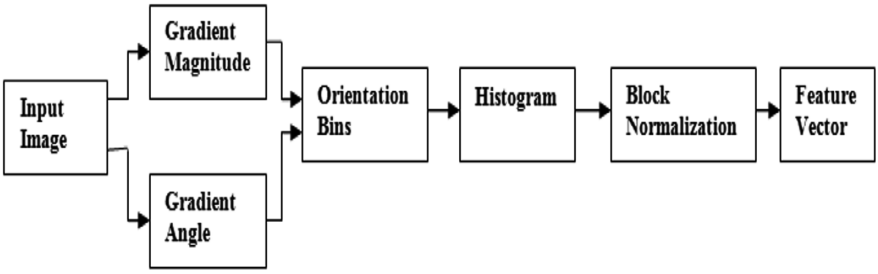


Fig. 4. Block diagram of HoG approach

Detailed explanation of these steps is given below.

(i) Gradient Calculation

Since larger masks seems to decrease the performance, [−1, 0, 1] and its transpose have been used in this work for the calculation of vertical and horizontal gradients.

$$V_x = [-1 \quad 0 \quad 1] \text{ and } V_y = [-1 \quad 0 \quad 1]^{\mathrm{T}}$$

For a cropped retinal fundus image R, derivatives along x and y directions are obtained by convolution as given by Eqs. (1) and (2) respectively.

$$R_x = R * V_x \tag{1}$$

$$R_y = R * V_y \tag{2}$$

Magnitude (M) and orientation are computed using the Eqs. (3) and (4) respectively.

$$|\mathrm{M}| = (R_x^2 + R_y^2)^{0.5} \tag{3}$$

$$\theta = \tan^{-1}\left[R_y / R_x\right] \tag{4}$$

(i) Computation of HoG Gradients in 8 × 8 cells

Here, the image is partitioned into 8 × 8 cells and a HoG is computed for each 8 × 8 cell. Each pixel calculated a weighted vote for an edge orientation histogram channel in accordance with the orientation of the gradient element centered on it. The votes are accumulated into orientation bins over local spatial regions.

(ii) Block Normalization

L2 normalization is applied to each block for providing better invariance to edge contrast, illumination and shadowing. Let ν be the non-normalized vector representing all histograms in a given block, $\|\nu\|_{\mathrm{K}}$ be its k-norm for k = 1, 2 and e is a small constant value. Maximum value of ν is limited to 0.2 [6]. The L2 normalization factor is given by Eq. (5).

$$\mathrm{f} = \nu / \left(\|\nu\|_2^2 + \mathrm{e}^2\right)^{0.5} \tag{5}$$

(ii) HoG feature vector computation

The final HoG feature vector will be the concatenation of all the 36 × 1 vectors.

2.2 Haralick Features

Co-occurrence matrices are joint probability distributions between pairs of pixels at a definite distance and direction. Gray level co-occurrence matrices proposed by Haralick [16] has become one of the widely used texture measures. For each cropped image, features like Angular Second Moment, Contrast, Correlation, Sum of Squares or Variance, Inverse Difference Moment, Sum Average, Sum Variance, Sum Entropy, Entropy, Difference Variance, Difference Entropy, Information Measure of Correlation I, II and Cluster Tendency are computed.

2.3 Classifier

HoG features are given as input to a Naïve Bayes classifier and the performance metrics are evaluated. Naive Bayes [10] require a few samples of training data to estimate parameters which are essential for classification [16]. Bayes Rule is used to compute the probabilities in Naïve Bayes algorithm. The probability of the chance of some conclusion, given some observation E, where an association between R and E, denoted as *P (R/E)* can be computed based on Eq. 6.

$$P(R/E) = P(E/R).P(R)/P(E) \tag{6}$$

3 Evaluation Metrics

The performance of the classifier is often computed in terms of its Predictive Accuracy, Kappa statistic, Root Mean Square Error (RMSE) and Area under Receiver Operating Characteristic curve (AUC) given by Eqs. (7), (8) and (9) respectively. Accuracy of the classifier is evaluated in terms of true positive, true negative, false positive and false negative. It is relatively a good metric when the output variable classes in the data are nearly balanced.

$$Accuracy = (TP + TN)/(TP + TN + FN + FP) \tag{7}$$

Kappa statistic (K) is a non specific term for several similar measures of agreement used with categorical data. The formula is given by Eq. (8).

$$K = (O_o - Oe)/(1 - Oe) \tag{8}$$

where O_o gives the relative observed agreement and Oe represents the hypothetical probability of chance agreement.

Root Mean Square Error (RMSE) is defined as the square root of the average of the sum of squares of residuals.

$$\text{RMSE} = \left\{\left[(t1 - o1)^2 + \ldots\ldots + (tn - on)^2\right]/n\right\}^{0.5} \tag{9}$$

where t represents predicted target values $\{t1, t2, \ldots tn\}$ and o represents original value $\{o1, o2, \ldots on\}$.

Receiver Operator Characteristic (ROC) curves signifies the outcome for binary decision problems in data mining. It shows the association between True Positive Rate and False Positive Rate.

4 Results

4.1 Dataset

Retinal fundus images had been collected from Sree Gokulam Medical College and Research Foundation, Trivandrum. (Ethical Committee Approval: SGMC-IEC No. 25/293/01/2017) Database contains 130 images of which 80 are normal and 50 are those of stroke patients, each having a dimension of 2336 × 3504 pixels.

4.2 Methods

Gray scale conversion of the image is carried out and the centre of the optic disc is located manually. Centre of the optic disc is taken as the seed point. ROI is selected from the fundus image by cropping a sub image of size 256 × 256 with the seed point as centre. The HoG Features are extracted from the ROI with 8 × 8 cell sizes with 2 × 2 cells/block and 9 orientation histogram bins. There are 16 horizontal and 16 vertical positions giving a total of 15 × 15 = 225 positions. Each 16 × 16 block corresponds to a 36 × 1 vector. Concatenation of all these into one single vector gives a size of 36 × 225 = 8100. Therefore, each image yields a feature vector of size 1 × 8100. Experiments show that the extracted HoG features, when given to a Naïve Bayes classifier gave a predictive accuracy of 93%.

4.3 Validation

K-fold cross-validation method with K = 10 has been implemented in this work to generate the output. In 10-fold cross-validation, 9 subsets were used for model training and a single subset was taken for model testing. Performance of the classifier was evaluated in terms of Accuracy, Kappa statistic, Root Mean Square Error (RMSE) and Area under the ROC Curve (ROC AUC). Table 1 shows the comparison of performance metrics evaluated for both cases.

Table 1. Comparative analysis of performance metrics

Performance metrics	HoG	Haralick features
Accuracy	93%	87%
Kappa statistic	0.88	0.73
RMSE	0.212	0.31
ROC AUC	0.979	0.939

ROC for both HoG features and Haralick features are given in Figs. 5(a) and (b) respectively.

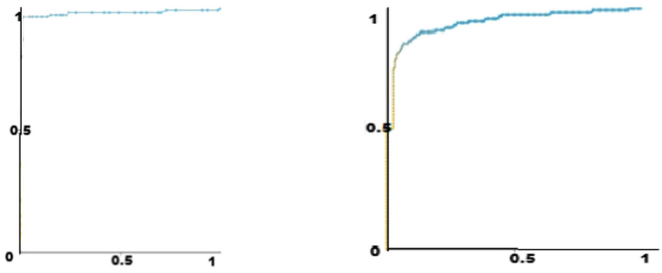


Fig. 5. (a) ROC for HoG features (b) ROC for Haralick features

ROC curve shows the plot of true positive rate versus false positive rate for various cut-off points.ROC curve passes through the top left corner for a test with no overlap in the two distributions. Therefore, closer the ROC curve to the top left corner, better the accuracy. From Fig. (5), we can analyze that ROC curve for HoG features show better results than Haralick features.

5 Discussion

The feature vector generated by HoG with 8 × 8 cell sizes and 2 × 2 cells per block gave a size of 1 × 8100. The texture descriptive feature vector is fed to three different classifiers like Support Vector Machine (SVM), Multilayer Perceptron (MLP) and Naïve Bayes (NB) for analyzing the predictive accuracy. Classifiers trained by different algorithms will perform in different ways. SVM and MLP gave a predictive accuracy of 68.9% and 76.4% respectively while Naïve Bayes gave a promising accuracy of 93%. Hence Naïve Bayes is considered in this work.

Various other texture descriptors like haralick features had been reported in literature. A comparative analysis has been made between HoG and Haralick features. Regression analysis has been performed to find the significant haralick features. After statistical analysis, it has been found that among the fourteen haralick features, only Contrast, Correlation, Angular second moment, Inverse difference moment and Entropy has a p value greater than 0.05. Those variables in conformity to a principal that its p-value is not greater than 0.05 are considered as the selected features of stroke in our study. By this standard, only 5 of the total 14 features can be chosen as input attributes. These selected haralick features are then given to Naïve Bayes for automatic detection of stroke which gave an accuracy of only 87%.

To the best of our knowledge, none of the previous works have used the same dataset. All the reported works in literature have used their own private dataset collected from reputed medical institutes, which are not publicly available. Hence, we could not compare our novel approach with existing state of art methods.

6 Conclusion

In this work, a novel method for computer aided stroke prognosis from retinal images has been implemented using Histogram of Oriented Gradients. HoG features outperform Haralick features by giving an accuracy of 93%. Prior changes necessary to understand the pathophysiology of stroke is easily demonstrated in the retina by analyzing the retinal vasculature. This method can surely help the physicians to plan for better medication that were not possible with conventional assessment systems which will aid in Stroke diagnosis. Future work involves improvement of the proposed algorithm by incorporating a large database and extracting more features which can ultimately increase the predictive accuracy of the classifier.

Acknowledgments. Our sincere thanks to eminent Neurologist, Dr Manoj P., Sree Gokulam Medical College & Research Foundation for providing us valuable suggestions in the progress of this work.

References

1. Sridharan, S.E., et al.: Incidence, types, risk factors, and outcome of stroke in a developing country, the Trivandrum stroke registry. Stroke **40**(4), 1212–1218 (2009)
2. Zaret, B.L., Marvin Moser, M.D., Cohen, L.S.: Chapter 18 Stroke - Lawrence M. Brass, M. D. (pp. 215–234)
3. Subha, P.P., Pillai Geethakumari, S.M., Athira, M., Nujum, Z.T.: Pattern and risk factors of stroke in the young among stroke patients admitted in medical college hospital, Thiruvananthapuram. Ann. Indian Acad. Neurol. **18**, 20–3 (2015)
4. Baker, M.L., Wang, J.J., Liew, G., et al.: Differential associations of cortical and subcortical cerebral atropy with retinal vascular signs in patients with acute stroke. Stroke **41**, 2143–2150
5. Jeena, R.S., Sukesh Kumar, A.: Stroke prediction using SVM. In: ICCCITT 2016, Tamil Nadu (2016)
6. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893, June 2005
7. Pearson, K.: On lines and planes of closest fit to systems of points in space. Philos. Mag. **2**(11), 559–572 (1901)
8. Gunteman, G.H.: Principal Components Analysis. Sage Publications, CA (1989)
9. Gonzales, R.C., Woods, R.E.: Digital Image Processing, 2nd edn, p. 795. Prentice Hall, Upper Saddle River (2002). ISBN 0-201-18075-8
10. Bhardwaj, B.K., Pal, S.: Data mining: a prediction for performance improvement using classification. Int. J. Comput. Sci. Inf. Secur. (IJCSIS) **9**(4), 1–5 (2011)

11. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893, 25–25 June 2005
12. Acharya, U.R., Dua, S., Du, X., Chua, C.K.: Automated diagnosis of glaucoma using texture and higher order spectra features. IEEE Trans. Inf Technol. Biomed. **15**(3), 449–455 (2011)
13. Simonthomas, S., Thulasi, N., Asharaf, P.: Automated diagnosis of glaucoma using haralick texture features. In: IEEE Conference on Information Communication and Embedded Systems, pp. 1–6 (2014)
14. Ito, Y., Ohyama, W., Wakabayashi, T., Kimura, F.: Detection of eyes by circular hough transform and histogram of gradient. In: International Conference on Pattern Recognition, ICPR (2012)
15. Monzo, D., Albiol, A., Sastre, J., Albiol, A.: Precise eye localization using HOG descriptors. Mach. Vis. Appl. **22**, 471–480 (2011)
16. Haralick, R., Shanmugam, K., Dinstein, I.: Textural features for image classification. IEEE Trans. Syst. Man Cybern. SMC **3**(6), 610–621 (1973)
17. Savakis, A.: Efficient eye detection using HOG-PCA. In: Proceedings of SPIE - The International Society for Optical Engineering, February 2014. https://doi.org/10.1117/12.2036824
18. Ibrahim, S., Sidibé, D., Meriaudeau, F.: Automatic Discrimination of Color Retinal Images using the Bag of Words Approach (2015). https://doi.org/10.1117/12.2075824
19. Harshvardhan, G., Venkateswaran, N., Padmapriya, N.: Assessment of glaucoma with ocular thermal images using GLCM techniques and logistic regression classifier. In: IEEE Conference on Wireless Communications, Signal Processing and Networking (2016)
20. Shah, S., Shloka, S., Maurer, Z., Sidrane, C., Shah, M.: Automated diagnostic classifier for extra ocular diseases. Biomed. J. Sci. Tech. Res. (BJSTR) **4**(1) (2018). https://doi.org/10.26717/BJSTR.2018.04.000995

Addressing Relay Attacks Without Distance-Bounding in RFID Tag Inclusion/Exclusion Scenarios

Selwyn Piramuthu(✉)

Information Systems and Operations Management, University of Florida, Gainesville, FL 32611-7169, USA
selwyn@ufl.edu

Abstract. With the widespread adoption and use of RFID tags, a valid scenario is one in which an RFID-tagged object includes several components that each have their own individual RFID tags. Under such a context, each of the components are bound to be included in or excluded from the main object over its lifetime. In order for only the tags that are a part of the main object to be authenticated by the main object, there is a need for a secure protocol that ensures that no other tag has access to the shared secrets among the main object and the component objects. Moreover, there is also a need to address relay attacks by adversaries under such scenarios. Existing authentication protocols address relay attacks through round-trip distance measurements in such inclusion/exclusion scenarios. While this works in principle, distance-bounding approaches are not always reliable. We consider another approach for inclusion/exclusion scenarios and develop a protocol sketch for this context. We also provide related security analysis.

Keywords: RFID · Tag inclusion/exclusion
Authentication protocol · Non distance-bounding

1 Introduction

An object with its own RFID (Radio-frequency IDentification) tag is not uncommon. For example, in a retailing context, item-level tags are used (e.g., Levi's jeans at Kohls, shoes at Macy's, most items at American Apparel stores) to identify and to individually consider each item at the store. The benefits of item-level RFID tags are many, and include tracking and tracing of each such tagged item, customized handling of each item (e.g., perishables, based on experienced ambient conditions), among others. The use case for item-level RFID tags is clear. In addition to objects with a single RFID tag, there are cases where an object can possibly be associated with several RFID tags. For example, any object with multiple components has this possibility where each of the components have their own RFID tags.

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 147–156, 2019.
https://doi.org/10.1007/978-981-13-5758-9_13

Under such a scenario, it is necessary for all the components' tags to have something in common (e.g., a shared secret) that allows them to communicate with the main object. Such a commonality would facilitate coordination among the components and the main object with the exclusion of tags that do not belong to this constellation. Moreover, such a main object-components scenario generally does not remain static. Over the lifetime of the main object, each of the components could possibly be a part of the main object only during a fraction of the time. This necessitates the possibility of components joining as well as leaving the main object of interest. For example, a computer system can have the monitor replaced by a new monitor; an automobile can have its tires replaced by new tires; a printer can have its almost empty ink cartridge replaced by a new ink cartridge.

As the name suggests, relay attacks are operationalized as messages are simply relayed between multiple entities. In the RFID system context, one or more adversaries relay the messages without any modification between RFID tag and reader. To the reader, it is as if the RFID tag is in close physical proximity and the RFID tag assumes that the reader is in close physical proximity. This is because almost all extant RFID authentication protocols do not specifically check or account for any delay in response from the recipient of a message. Since the messages themselves are not modified, the authentication protocols proceed to successful completion. The extent of the harm done through relay attacks depends on the specific application of interest. For example, it has been shown that RFID-based car keys are vulnerable to relay attacks [1] and allow adversaries to gain entry to a car that is physically parked farther away from the car key.

Although there are dozens of protocols that have been proposed to provide defense against relay attacks for RFID-based systems in general, almost all of them are based on the distance-bounding idea where the time taken for a bit to travel between two entities (usually the tag and reader) is taken as proxy for the distance traveled by that bit. While this works perfectly in principle, there is one major concern with these authentication protocols. The issue is that when tag and reader are expected to be at the most a few meters apart, it is rather difficult to tell reader and tag that are meters apart from those that are miles apart since bits don't just travel between two entities. The receiving entity needs to receive the bit, identify it as something that the receiving entity should send back to the sending entity, and send it back. This process involves time. When the time involved is somehow more than an order or two in magnitude over nanoseconds for whatever reason, distance measurement based on the time taken for the bit to travel between tag and reader becomes unreliable and inaccurate.

We develop a sketch of a non-distance-bounding authentication protocol that is resistant against relay attacks for inclusion/exclusion scenarios. As with similar authentication protocols, we also consider the use of ambient conditions for the proposed authentication protocol that is designed to be resistant against relay attacks, but for the inclusion/exclusion scenario.

We include brief discussion on relay attacks in Sect. 2. We provide an overview of inclusion/exclusion protocols in Sect. 3. We then include a sketch of the

proposed protocol in Sect. 4. We consider the security properties of the proposed protocol in Sect. 5. We conclude the paper in Sect. 6.

2 Relay Attacks

Almost all extant RFID system authentication protocols generally operate with the assumption that the tag and reader are next to each other. When this requirement is satisfied, it allows for minimal opportunities for an adversary to wedge in-between the tag and reader and accomplish a relay attack. Even if an adversary manages to squeeze in-between the tag and reader within this short separation, it is more difficult for the adversary to get away without being noticed. However, unfortunately, for relay attacks to be accomplished, the reader and tag can physically be miles apart. In addition, such relay attacks are oftentimes mounted without the knowledge of the reader or tag.

Hancke and Kuhn's (2005) (Fig. 1) distance-bounding protocol is divided into two phases. The first phase does not include a time component and the second phase includes a time component [2] that is accomplished through the clock at the reader's end. The second phase is done iteratively where a single bit is sent from the reader to the tag and the round-trip taken by this bit is measured. The time taken by several such single bits are measured. The primary purpose of the first phase is to generate $R^0||R^1$, which is then used in the second (timed) phase where either R^0 or R^1 is randomly chosen and its $(i-1)^{th}$ position bit is sent to the tag at the i^{th} iteration. Before completion of the protocol, the following are ensured: (a) each of the n round trip times is at most (Δt_{max}) and (b) the

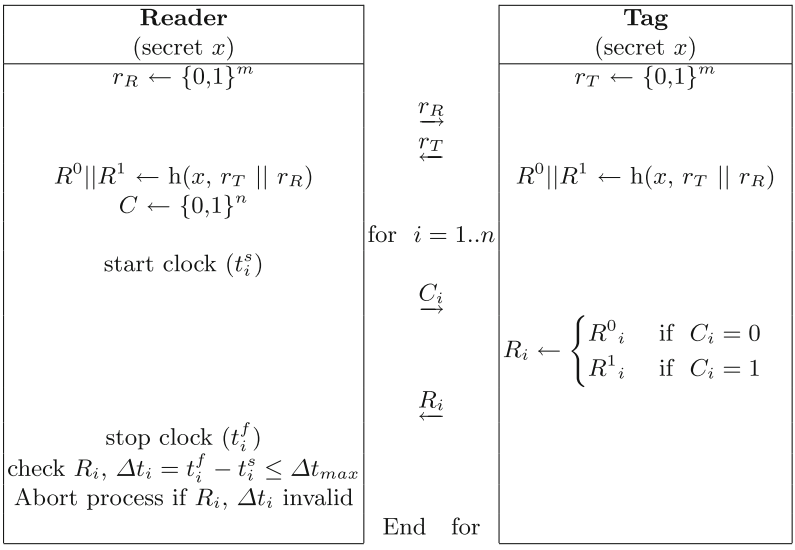


Fig. 1. Hancke and Kuhn's (2005) protocol

R (R^0 or R^1) values are valid. Both of these tests need to be successfully passed for ensuring the absence of a relay attack.

This protocol has several vulnerabilities as was identified in later publications (e.g., Tu and Piramuthu 2017). Reid et al. (2007) avoid the scenario where the dishonest tag shares its secrets with the adversary [10]. Other researchers have developed variants of the distance-bounding protocol such as a different outer loop [11], mixed challenges [3], and pre-computing response [4]. A few other protocols indirectly measure distance (e.g., Rasmussen and Čapkun 2008, which is vulnerable to mafia fraud attack [5,9]).

3 Protocol for Multi-tagged Object

The notations used in this paper include:

$N_t, N_p, N_r, N_u, N_T, N_R, N_R^{'}, N_P, N_T, r_A, r_B$: random n-bit nonce
s_c, s_{c+1}: Tag's current and subsequent keys
$f_k^{'}$, f_k, $\{\}_k$: keyed (with key k) encryption function
H_k: keyed (with key k) one-way hash function
t_j: shared secret between tag$_i$ and TTP, Reader
r_i: shared secret between Reader R_i and TTP
id_{t_j}: tag t_j identifier
T_{AC}, R_{AC}: Measured ambient condition at tag and reader respectively

Piramuthu [8] develops a inclusion/exclusion protocol with no consideration for the possibility of relay attacks (Fig. 2). Clearly, then, this protocol is vulnerable to relay attacks.

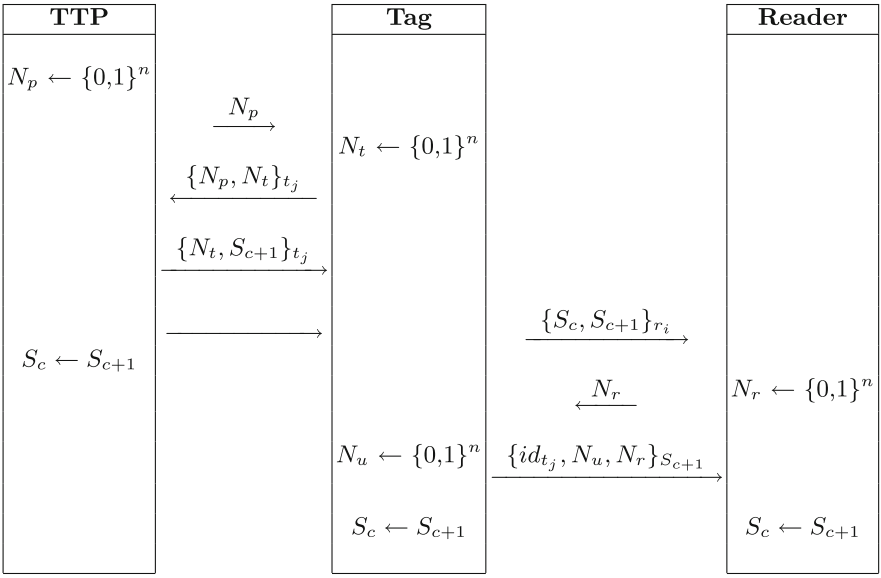


Fig. 2. The inclusion/exclusion protocol of Piramuthu (2018)

Although there are a few studies that have considered the inclusion/exclusion scenario, only one of the existing protocols explicitly considers the issue of relay attacks for the inclusion/exclusion aspect of component RFID tags as a part of a larger ensemble. Piramuthu [6,7] developeds a protocol (Fig. 3) that involves a reader, a trusted third party and the RFID tags. These protocols use a variant of the distance-bounding method to address relay attacks.

This protocol (Fig. 3) is for a scenario where RFID tags are allowed to join as well as leave an ensemble of RFID tagged items (components). A trusted third party (TTP) is assumed to be present in order to mediate communication between the reader and tags, specifically with respect to secret key updates as and when a tag either enters or leaves the ensemble. This protocol assumes that all component RFID tags share the same common secret, which is simultaneously updated for every tag that belongs to the ensemble whenever a change in the ensemble composition occurs. The purpose of the key update is to ensure that all

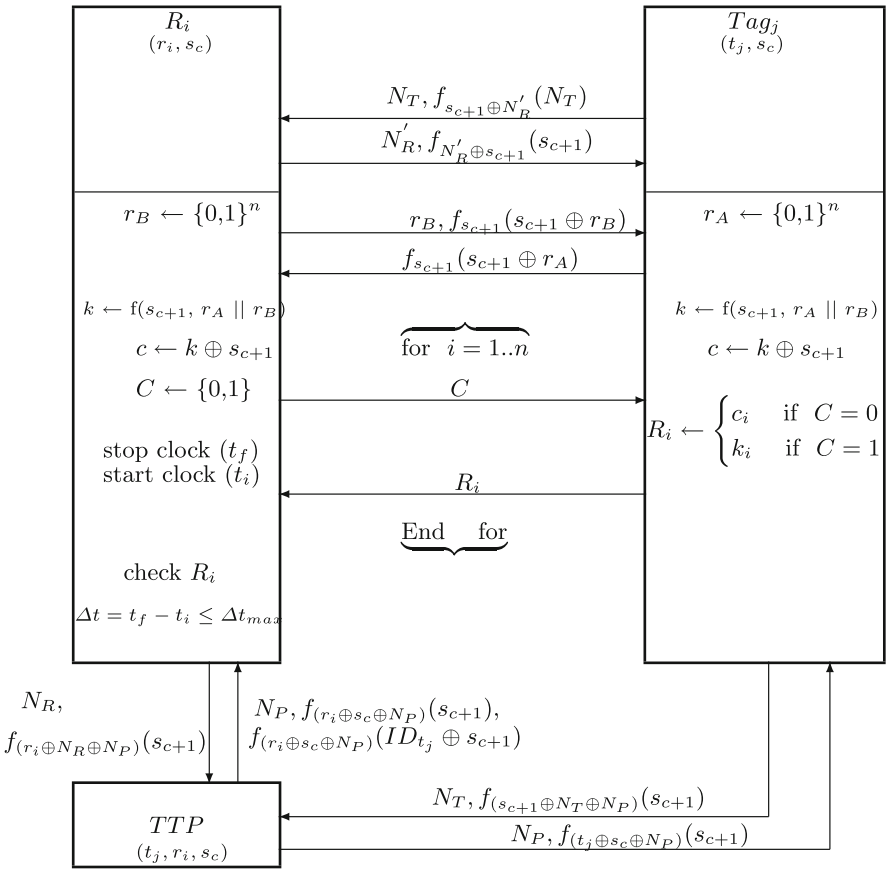


Fig. 3. Tag inclusion/exclusion protocol (Piramuthu 2011)

tags that belong to the ensemble of tags share a common secret while excluding any tag that does not belong to this ensemble from knowledge of this shared secret. A component that leaves this ensemble is not allowed to have knowledge of the updated shared secret as soon as it leaves.

4 The Proposed Protocol

We modify the protocol presented in Piramuthu (2018) to include resistance against relay attacks. The same three set of entities (TTP, tags, reader) are involved in the process. To provide resistance to relay attacks, we make use of ambient condition information [12]. There are several ambient conditions that can be used in this context. However, some of the ambient conditions suffer from directionality issues. The directionality issue results in different ambient condition readings. Other ambient condition facets include pressure, temperature, magnetic field, and location (GPS). Some of these are context specific. For example, GPS readings are not appropriate/accurate for use in locations that block GPS signals (e.g., inside buildings, under tree cover). Moreover, it is not difficult to mess with GPS signals. Given these reasons, we do not specify a specific ambient condition for the proposed protocol. We assume that an appropriate ambient condition sensor is chosen for the application of interest, taking the context into full consideration. The proposed protocol (Fig. 4) is a variant of the protocol developed in Piramuthu (2018).

The TTP begins with a nonce (N_r) that is sent to each of the tags. In response, each of the tags generate their own nonce (N_t) that is sent along with the TTP's tag encrypted with the shared secret (t_j). The TTP sends the encrypted (with t_j) new group shared key (S_{c+1}) along with N_t. This concludes new group key sharing by the TTP with the tags. The next part of the protocol is to ensure that the tag and reader are next to each other and to ensure that both the tag and reader have the new group key. The TTP then sends the previous and new keys encrypted with the shared key (r_i) to the reader and then updates the new key as its current key. It keeps the previous key in memory just in case an adversary blocks any of the messages and attempts to cause desynchronization. The reader sends nonce (N_r) to the tag. The tag generates another nonce (N_u). It then measures its ambient condition (T_{AC}). The tag sends the encrypted (with S_{c+1}) message with $id_{t_j}, N_u, N_r, T_{AC}$ to the reader. The reader measures its ambient condition (R_{AC}) and verifies if its ambient condition measurement is close to the tag's ambient condition measurement. When they are next to each other, these two measurements should not be significantly different from each other. When this is not the case, the reader aborts the protocol. If not, the reader encrypts id_{t_j}, N_u, R_{AC} (with S_{c+1}) and sends it to the tag as evidence that its measured ambient condition value is close to that of the tag's measured ambient condition value. The reader then updates its group shared secret while storing the previous key. The tag does the same once it receives and confirms the last message from the reader. This ends one round of the protocol and all

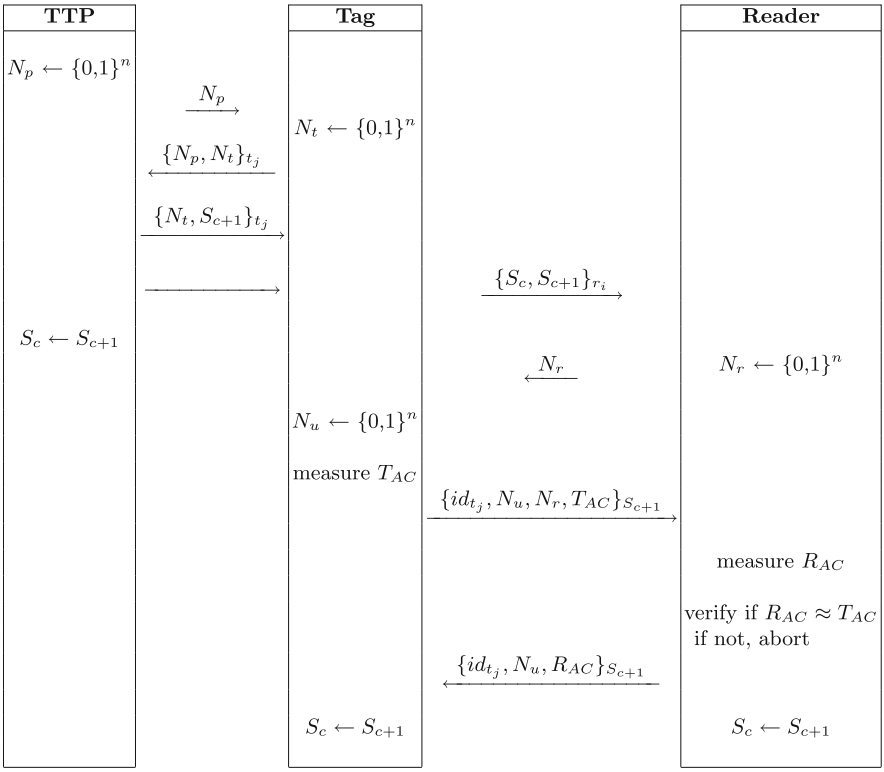


Fig. 4. The proposed inclusion/exclusion protocol

the entities (reader, TTP, and tags) have the same updated shared group secret (S_{c+1}). Both the reader and the TTP wait for response from the recipient of its message. When the response does not arrive on time, the protocol is aborted.

5 Security Analysis

Any pair of the entities are assumed to communicate with each other in an environment that is not assumed to be secure. The proposed protocol is executed when there is a change in the group constellation - when at least one entity (component tag) enters or leaves the ensemble regardless of the reason. We assume the existence of resourceful active adversaries who are able to block, modify, resend any of the messages that are passed between any two entities. Freshly-generated nonce (N_p, N_r, N_u, N_t) are used during each run of the protocol. The adversary cannot learn anything that is usable with knowing one of (t_j, r_i) since both the secrets are required for successful authentication.

Other possible attacks include the following.

Tag/Reader Anonymity: The information that is unique to each tag and reader include their shared secret keys (t_j, r_i) and the group shared secret (S_c). With knowledge of these, it is possible for an adversary to impersonate as well as clone a tag. To protect the tags and reader from such attacks, we ensure that none of this information is made available to adversaries regardless of their attempt(s) to retrieve this information.

Forward Security: The proposed protocol is not forward-secure since knowledge of all shared secrets (i.e., both t_j and r_i) will allow an adversary to decrypt all earlier messages that are sent between any two entities. The group keys are also vulnerable to such attacks.

Secrecy/Data Integrity and Authenticity: None of the identifiers are sent in the open.

DoS/Desynchronization: The purpose of this protocol is to update the shared secret key whenever a change in the constellation of the ensemble of tags occurs. An adversary could potentially block messages after one entity has changed the key to the updated key and cause desynchronization. Since the shared group key is updated during every round of the protocol, both the reader and the TTP store the previous group key just in case a desynchronization attack is mounted in the future. The group key update is accomplished by the reader and tag and in the middle of the protocol by the TTP after it has sent out its last message in the protocol.

Reader/Tag Impersonation Attack: For successful impersonation attack to occur, the adversary should be able to generate messages that are successfully validated by the recipient of those messages. Since the messages that are sent later are dependent on those that are sent earlier (e.g., through use of freshly generated nonce), simply replaying captured messages from an earlier run of the protocol will not work.

Mafia Attack: Since the ambient condition selected and used in the protocol needs to be the same at the reader and the tag end and all the messages (other than just nonce) that are sent between any two entities are encrypted, it is not easy to mount this type of attack.

Terrorist Attack: A tag that colludes with an adversary can share the group secret key with the adversary. However, it cannot share its shared key with the reader or TTP (r_i, t_j) with the adversary since doing so will compromise its identity - the adversary can clone the tag and this is not a desired outcome for the tag of interest.

Replay Attack: With the presence of freshly-generated nonce that affects messages that are passed between any two entities, a replay attack will not succeed in the proposed protocol.

6 Discussion

We considered the inclusion/exclusion scenario for an object with a dynamic ensemble of RFID tags that enter and leave the main object over time. Specifically, we considered the scenario where relay attacks could be mounted on such a system. Relay attacks are especially dangerous since these are rather difficult to detect and address. Almost all of the RFID protocols in existence fall prey to relay attacks unless they are specifically designed to be resistant against such attacks. These attacks are difficult to address since (a) they are passive attacks with no change to the messages that are passed between two entities, and (b) therefore do not depend on cryptography for their success. Among the RFID-based protocols that specifically consider the existence of relay attacks, almost all of them use some variant of distance-bounding means to determine the distance between tag and reader. We did not use the more common distance-bounding means to determine the physical separation between tag and reader since it has serious measurement issues due to (unexpected) delays that could occur at the tag's end which would render the precise measurement of distance between tag and reader useless. We considered the use of ambient conditions for this purpose. We did not consider a specific ambient condition in this protocol since that could be context-specific. The purpose of this paper include creating awareness of the inclusion/exclusion setup in RFID systems and the use of non-distance-bounding means to approach relay attacks in such systems. The hope is that other researchers would study this scenario and eventually develop more secure protocols for this system setup.

References

1. Greenberg, A.: Radio Attack lets Hackers Steal 24 Different Car Models. Wired, 21 March 2016. https://www.wired.com/2016/03/study-finds-24-car-models-open-unlocking-ignition-hack/
2. Hancke, G.P., Kuhn, M.G.: An RFID distance bounding protocol. In: Proceedings of the IEEE/Create-Net SecureComm, pp. 67–73 (2005)
3. Kim, C.H., Avoine, G.: RFID distance bounding protocol with mixed challenges to prevent relay attacks. In: Garay, J.A., Miyaji, A., Otsuka, A. (eds.) CANS 2009. LNCS, vol. 5888, pp. 119–133. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-10433-6_9
4. Mauw, S., Toro-Pozo, J., Trujillo-Rasua, R.: A class of precomputation-based distance-bounding protocols. In: Proceedings of the IEEE European Symposium on Security and Privacy (EuroS&P), pp. 97–111 (2016)
5. Mitrokotsa, A., Onete, C., Vaudenay, S.: Mafia fraud attack against the RC distance-bounding protocol. In: Proceedings of the IEEE International Conference on RFID -Technologies and Applications (RFID-TA), pp. 74–79 (2012)
6. Piramuthu, S.: Relay attack-resisting inclusion/exclusion protocol for RFID. In: 2nd International Workshop on DYnamic Networks: Algorithms and Security (DYNAS) (2010)
7. Piramuthu, S.: Inclusion/exclusion protocol for RFID tags. In: Meghanathan, N., Kaushik, B.K., Nagamalai, D. (eds.) CCSIT 2011. CCIS, vol. 133, pp. 431–437. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-17881-8_41

8. Piramuthu, S.: Authentication protocols for an object with dynamic RFID tags. In: Doss, R., Piramuthu, S., Zhou, W. (eds.) FNSS 2018. CCIS, vol. 878, pp. 93–101. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-94421-0_7
9. Rasmussen, K., Čapkun, S.: Location privacy of distance bounding. In: Proceedings of the Annual Conference on Computer and Communications Security (CCS), pp. 149–160 (2008)
10. Reid, J., Gonzalez Nieto, J.M., Tang, T., Senadji, B.: Detecting Relay Attacks with Timing-Based Protocols. Queensland University of Technology ePrint (2006). http://eprints.qut.edu.au/view/year/2006.html
11. Tu, Y.-J., Piramuthu, S.: RFID distance bounding protocols. In: 1st International EURASIP Workshop on RFID Technology, pp. 67–68 (2007)
12. Tu, Y.-J., Piramuthu, S.: Non-distance-bounding means to address RFID relay attacks. Decis. Support Syst. **102**, 12–21 (2017)

Digital Display Control Using Automation

Deven M. Gupta and Soumya S. Patil(✉)

School of Electronics and Communication Engineering,
KLE Technological University, Hubli 580031, Karnataka, India
devengupta19@gmail.com, soumya@bvb.edu

Abstract. Display system finds vast application in most of the fields. The digital display discussed in this paper uses Raspberry pi, Led dot matrix, shift registers (74HC595) has their main components. The digital display system uses IOT platform (i.e. IBM WATSON) as a communication interface between raspberry pi and the led dot matrix. The user types the message that needs to be displayed into the display (led dot matrix) using web application/android application which can be accessed by the user, only if the authentication process is completed. Then this message becomes input to the raspberry pi (controller) which acts as a gateway in this IOT based application. Generally all micro-controller development boards have 20–30 General Purpose Input Output (GPIO) pins which are insufficient, because it cannot be interfaced with the larger displays. So to solve this problem, the Serial In Parallel Out (SIPO) technique is used. The shift register IC uses this technique and takes one serial input of 8 bits and converts it into 8 different outputs of one bit, which can be connected to the led dot matrix. So the overall system built is user friendly, cost effective and consumes low power as it uses LED display. This digital display is a perfect replacement for posters and trivial noticeboards.

1 Introduction

Digital devices are the most commonly and widely used devices in today's world and have become a new trend in the market. So digitization is the main criteria for the progress of any nation. Among digital devices digital display is playing an important role in many purposes like advertisement, digital notice boards in schools and colleges, in railway stations etc. Previously, there were display that were not portable but with the advancement of technology display system became more flexible, portable etc. Earlier the display was fixed but with different technologies it became so enhanced that display could be changed according to user's needs. The banners used for advertisement purpose requires lot of maintenance and replacement again and again. So we can overcome this problem by using digital display which is controlled by a micro-controller which can change the theme time to time. In the market there are many digital displays but LED

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 157–164, 2019.
https://doi.org/10.1007/978-981-13-5758-9_14

displays are affordable for all classes of people and also they consumes less power and require less maintenance.

Using micro-controllers as a solution there comes a problem of having limited number of GPIO pins in it, which cannot be directly connected to the larger displays which becomes an important concern. For example, to display a single English alphabet we need atleast a 4 * 4 led matrix which uses 8 GPIO pins. Hence imagine the GPIO pins required for a larger display like 1000 * 1000 led matrix. One of the common solution will be connecting multiple micro-controllers, but this is a impractical solution as same micro-controllers with same specifications can have different processing times and would be an expensive setup to build. And the other solution is using shift registers which uses Serial In Parallel Out technique. Shift registers uses a single GPIO pin as a input (which is of 8-bit) and converts it into 8 different outputs of single bit each. So this reduces the usage of GPIO pins by 8 times.

The digital display system built should satisfy some objectives and constraints like: it should be portable, cost effective, Eco-friendly, durable, have less maintenance and should not be affected by environmental conditions (like humidity etc).

Some of the available literature [1–5] uses led dot matrix as their display unit. The authors in [6] uses GSM MODEM (Global System for Mobile Communication) as an interface between the mobile and micro-controller to display message onto the LED display and thus creating a authentic wireless communication between the mobile and the micro-controller. In paper [7], the authors uses Led display control system which can improve the cost performance of the system, shorten the development cycle, and enhance the system scalability and maintainability. And the authors in [8] concentrate on designing portable real time display system which can be accessed by using Ethernet network.

1.1 System Design

(1) *Functional Block Diagram:* The functional block diagram shown in Fig. 1 consists of raspberry pi acting as a controlling unit and the led matrix as a display device. The message entered by the user in the android application is sent to the cloud and further raspberry pi displays the message on to the led dot matrix.
(2) *Morphological Chart:* Morphological chart is used to generate many number of possible solutions for a given problem in terms of different architectures. For example, there are five possible architectures for the controlling unit that can be used in the digital display, which are 8051 micro-controller, PIC controller, ATMEGA328P controller, Raspberry pi controller and arm controller.
(3) *Design Alternatives:* Table 2 shows the design alternatives based on the morphological chart shown in Table 1.

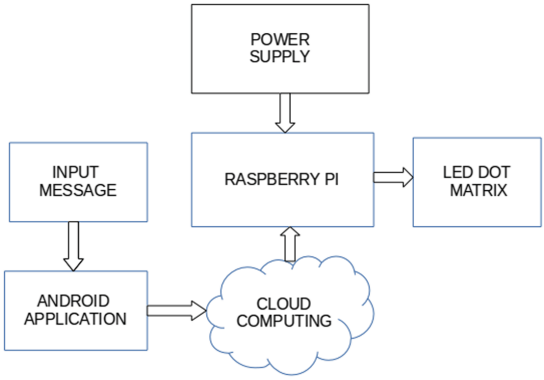


Fig. 1. Functional block diagram of Digital Display

Table 1. Morphological chart

	Option 1	Option 2	Option 3	Option 4	Option 5
Controlling unit	8051	PIC	ATMEGA328P	RASPBERRY-PI	ARM
Communicating device	DTMF decoder	RF Module	Bluetooth module	WIFI module	GSM MODEM
Display system	LED	LCD	PDP	OLED	LCD
Access to display system	Fingerprint recognition	Android APP	Using IOT	Android APP	Web APP

1.2 Final Design

One optimal solution is selected based on its working and ease of the implementation.

Using morphological chart as shown in Table 1, multiple designs for the project is obtained. And further by applying concept screening and scoring techniques a particular design, which is more feasible and optimal is selected.

So the final design mainly consists of raspberry pi and led dot matrix. And the user can get access to the display system by using the android application. Also to communicate between the controller and the application, IOT platform i.e. IBM WATSON is used.

2 Implementation Details

2.1 Specifications and Final System Architecture

The Fig. 2 illustrates the overall detailed block diagram of Display system with their functions. The System has two inputs and a output. The two inputs are the input signal which is the message given by the user and the power supply

Table 2. Design alternatives

Design alternatives	DESIGN-1(A)	DESIGN-2(B)	DESIGN-3(C)	DESIGN-4(D)
Controller used	8051	PIC	RASPBERRY-PI	RASPBERRY-PI
Communicating device	RF Module	Bluetooth Module	WIFI Module	GSM MODEM
Display system used	PDP	LED Matrix	PDP	LCD
Method used to access display system	Finerprint Recognition	Android APP	Using IOT	Android APP

is the second input. The output is the pattern or the text displayed onto the led dot matrix. The intermediate blocks between the input and output are the communicating device or module and the controller. Here the controller used is Raspberry pi and IBM WATSON platform is used to communicate between the android application and raspberry pi.

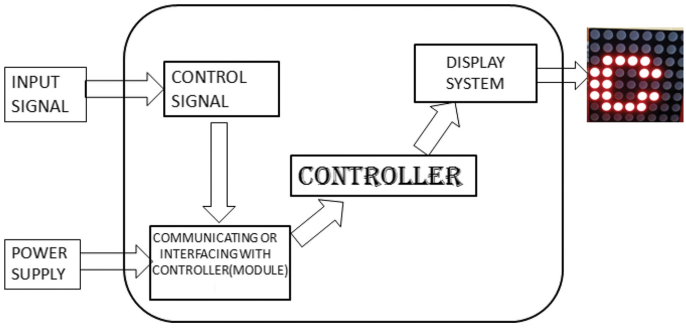


Fig. 2. Detailed block diagram

2.2 Design Specifications

Raspberry pi is a micro-processor which is also called as mini-computer because it is capable of doing all the work done by a general computer.

The Fig. 3 illustrates the functional block schematic of raspberry pi. It consists of the main controlling unit i.e. the CPU or the GPU, the input/output pins, the memory (RAM) and the USB HUB which consists of Ethernet and 2x- USB.

RASPBERRY PI 2 MODEL B:

- 1.2 GHz QUAD-CORE CORTEX A53
  Raspberry pi is a System On Chip (SOC) and made up of ARM cortex A53.
- SOC-BROADCOM BCM2837.

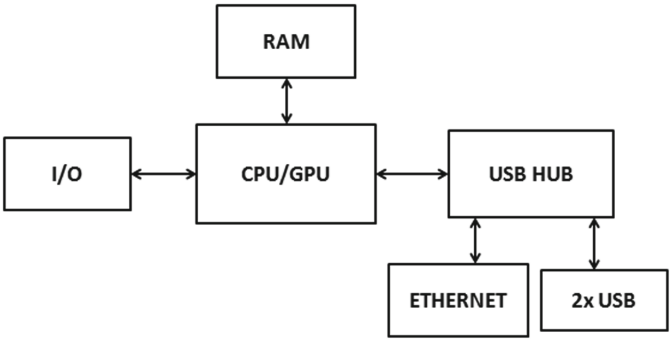


Fig. 3. Functional block schematic of Raspberry pi

- GPU-BROADCOM VIDEOCARE IV @400 MHz
  It has a Graphics Processing Unit of Broadcom videocare which work at 400 MHz.
- Memory-1 GB LPDDR2-900 SDRAM.
- Network-10/100 MBPS Ethernet, 802.11N Wireless LAN, Bluetooth 4.0 These are the connectivity options in Raspberry pi 2 model B.
- Storage-micro-SD.

LED DOT MATRIX USED HAS: LED dot matrix is built using an array of LED's. To build an 8 * 8 led matrix 16 LED's are needed. LED dot matrix are available in two categories i.e. common cathode and common anode. Here, the common cathode LED dot matrix is used with the following specifications:

- Dimension - 2.3 in (58 mm).
- Power Dissipation-125 mW.
- Forward Current-30 mA.
- Reverse Voltage-3 V.
- Operating Temperature: (−45 to +85 °C) [9].

SHIFT REGISTER (74HC595): Shift registers are high speed advanced CMOS circuit fabricated on silicon. It has 2 input AND gates for the transition of clock which is used to enter serial data. It is a high speed Serial In Parallel Out shift register. A supply voltage of 0 V to 7 V can be applied to it as mentioned below.

- Supply voltage range- 0 V to 7 V.
- Output current- +35 mA/−35 mA [10].

2.3 Algorithm

The steps of the algorithm are listed below:

(1) Import RPi.GPIO library for using GPIO pins of raspberry pi.
(2) Assign GPIO pins to the cathodes of the led dot matrix.
(3) Assign GPIO pins to the anodes of the led dot matrix.
(4) Declare the GPIO pins as output pins.
(5) Set the pins 'HIGH' or 'LOW' according to your pattern.
(6) Use delay function 'time.sleep' accordingly.
(7) The GPIO pins of raspberry pi can be cleared by using keyboard interrupt.

2.4 Flowchart

Flowchart is defined as the geometrical representation of the sequence of steps involved in the program or the algorithm. Here, Fig. 4 represents the flowchart for the Digital display algorithm. The first step involved importing rpi.GPIO library as shown in the flowchart. The second step is to assign cathode and anode pins of led dot matrix to the output pins of the shift register (74HC595). And if the cathodes are assigned with '1' and anodes with '0', the LED's of the matrix will glow.

2.5 Result Analysis

Figure 5 shows one of the implementation result of the digital display system where a single alphabet 'A' is displayed onto the 8 * 8 led dot matrix using

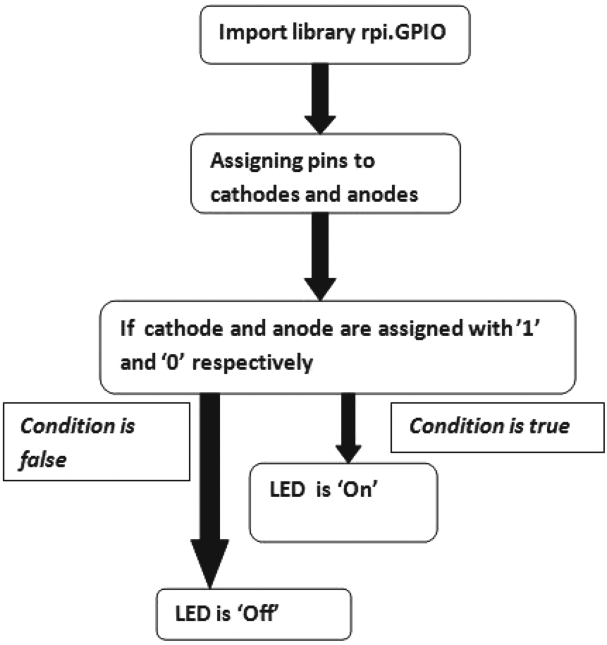


Fig. 4. Flow chart

android application. This can further be used to implement on a larger display so that a pattern can be displayed on it.

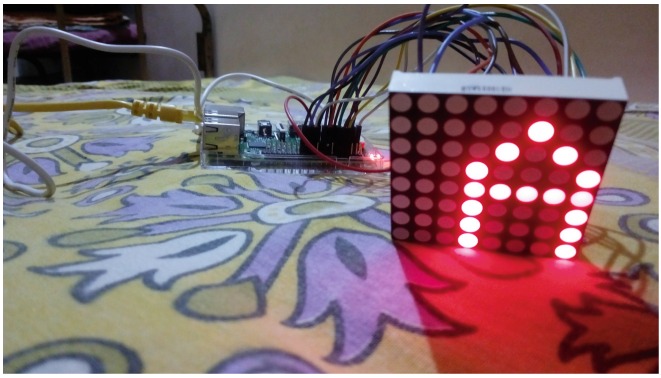


Fig. 5. Implementation image of the project

2.6 Application in the Societal Context

It can be widely used in all the fields where display is required. It finds majority applications in the field of entertainment, market industry etc. Since LED matrix is used, the power consumption is very less. The main advantage of it over other conventional technique is that the message displayed can be changed according to user's requirements. It is pretty economical compared to other displays and Eco-friendly too. And also by using digital displays, replacing the posters and banners would eventually make the nation or the society to step further in the world of digitization. And make the country to be one of the developed countries in the field of technology.

3 Conclusion

The project DIGITAL DISPLAY is finally implemented on a 8 * 8 led dot matrix using raspberry pi and IBM WATSON (IOT PLATFORM) through which we are sending our input signal. The GPIO pin of raspberry pi is connected to the shift register which converts 1 byte of information to 8 different bits of output and thus reducing the usage of GPIO pins. Then the 8 bit outputs are connected to led dot matrix which displays the message fed by the user into the android application. Thus, the project Digital display solves two problems, that is connecting many GPIO pins for larger led matrix and also the problem of maintenance and time to time replacement of banners and posters.

3.1 Future Scope

The project DIGITAL DISPLAY can be expanded by using N * N led dot matrix using limited number of GPIO pins of raspberry pi by interfacing shift registers. And also the problem of flickering of message characters can be controlled by using appropriate refresh intervals. And also for more brightness of the characters current drivers (for example: ULN2804) can be used.

References

1. Pervez, M.F., Al Mortuza, A., Ali, M.H., Haque, M.S., Ghosh, H., Hoq, M.: Novel scanning technique for LED dot-matrix display to reduce flicker. Electr. Electron. Eng. **6**(2), 19–24 (2016)
2. Huang, W.-F.: Three approaches to light an 8 × 8 LED dot-matrix display. In: 2013 IEEE International Symposium on Next-Generation Electronics (ISNE), pp. 37–40. IEEE (2013)
3. Jiang, X., Chen, M., Li, Z., Chen, X.: Wireless real-time LED display control system based on single chip microcontroller. In: 2012 7th International Conference on Computer Science & Education (ICCSE), pp. 951–952. IEEE (2012)
4. Kapoor, D., Pandey, R., Vaid, R., Patil, M.V.: Int. J. Eng. Sci. Res. Technol. Electron. Dice Using Dot Matrix Display, 315–321 (2016). https://doi.org/10.5281/zenodo.51429
5. Zungeru, A.M., Obikoya, G.D., Uche, O.F., Eli, T.: Design and implementation of a GSM-based scrolling message display board. Int. J. Comput. Sci. Inf. Technol. Control Eng. (IJCSITCE) **1**, 21–31 (2014)
6. Darshil, J., Vora, S., et al.: Advertising display system using led and graphical LED. Int. J. Sci. Res. Dev. **1**, 153–157 (2013)
7. Zhao-you, Z., Sheng-hui, D.: Embedded LED lighting control system research and implementation. In: 2nd IEEE International Conference on Computer Science and Information Technology, ICCSIT 2009, pp. 553–557. IEEE (2009)
8. Ali, S., Sardar, A., Patra, K.K., Barua, R.: Implementation of secure portable real time display system over Ethernet network. In: 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), pp. 359–364. IEEE (2016)
9. LC LED Corporation. http://www.lc-led.com/products/lc2088tur1a.html
10. Sn54hc595, sn74hc595 8-Bit Shift Registers with 3-State Output Registers, December 1982. https://www.sparkfun.com/datasheets/IC/SN74HC595.pdf

Fuzzy Based Energy Efficient Protocol for Monitoring of Nuclear Radiation Using Wireless Sensor Network

Ravi Yadav(✉) and A. K. Daniel

Madan Mohan Malaviya University of Technology, Gorakhpur, India
ravics85@gmail.com, danielak@rediffmail.com

Abstract. Wireless sensor network swiftly emerges as one of the most prominent technology for industrial and human security application. The paper presents hybrid routing protocol for monitoring of radiation at nuclear plant using Wireless Sensor Network technology. The proposed protocol uses hybrid clustering for deployment of sensor nodes in the given area. There are two types of heterogeneous sensor nodes deployed in different region based on energy level for efficient monitoring of radiation. The super node sense the radiation level in their region continuously and transmit the sensed data to base station using fuzzy based clustering technique along with node scheduling. The normal nodes are used for routing purpose along with super nodes. The proposed protocol yields the optimum energy consumption and improve the performance and network lifetime for monitoring the nuclear plant surrounding.

Keywords: Nuclear radiation · Heterogeneous nodes · Fuzzy technique Clustering

1 Introduction

Energy is the most vital need of human beings now a days. Without Energy we can't think life is possible. There are many sources of generating energy such as mechanical energy, thermal energy, wind energy, etc. Among these sources, nuclear energy source has large production of energy and known as cleanest form of energy generation. This motivates the Countries to set up nuclear plant for generation of energy. The production of energy at nuclear plant is carried out by process called nuclear fission. In nuclear fission process, the atom of uranium is bombarded with fast moving neutron, it splits into one or more small atoms resulting in generation of energy. The heat produced during nuclear fission is used into the boiler to boil water to generate steam and these steam is used to rotate turbine blades. Rotation of turbine blades, makes generator to produce electricity. The nuclear power plant generates harmful nuclear radiation as alpha, beta, gamma, X-ray and neutrons during nuclear fission process [1, 2]. Gamma radiation has high penetration level than beta and alpha particles. Neutrons are charged particles produced during fission and are very dangerous because they convert non-radioactive material to radioactive material. Radioactive decay of product like radon also result in generation of radiation. The FUKUSHIMA-DAIICHI nuclear power plant

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 165–177, 2019.
https://doi.org/10.1007/978-981-13-5758-9_15

disaster on 11 March 2011 due to tsunami result in major radiation leakage in Japan [3]. The fear of accidental leakage of nuclear radiation in surrounding area of nuclear plant cause many damages to human survival. The radiation can change the balance of human cell and can change gene (DNA) when exposed for long time. So there is need to monitor the nuclear power plant into safe operated mode. International Commission on Radiological Protection (ICRP) has define a limit for exposure to radiation for safe working inside nuclear power plant. Therefore, it is increasingly necessity to build a reliable and efficient nuclear monitoring system using Wireless Sensor Network.

The objective of the paper is to provide a hybrid routing protocol using fuzzy logic for efficient monitoring of radiation at nuclear plant. The paper presented is divided in following section. First section contain introduction to nuclear plant. Second section describe related work to radiation monitoring system and describe various energy efficient routing protocol that are used to transmit sensed data (radiation). Third Section contain the proposed architecture and protocol for data transmission process using fuzzy technique. Fourth section contain simulation and result analysis.

2 Related Work

Energy efficient routing is the important factor during the deployment of sensor nodes. There are different approaches for data gathering and data transmission. The Direct transmission approach for data transmission uses the concept of each node directly transmit their sensed data to base station. The approach of multi-hop transmission is used for data transmitting using intermediate nodes (relay nodes). The approach of clustering is used to divide area into small cluster. Each cluster has a cluster head which is responsible for data collection and data transmission to base station. Clustering helps us to maintain balancing of energy in the sensor network [4].

Low Energy Adaptive Clustering Hierarchy (LEACH) was the first hierarchical clustering protocol designed for homogenous sensor network and every node have same capacity and equal probability for selecting as cluster head [5]. Stable Election Protocol (SEP) is used for heterogeneous nodes. The protocol has different types sensor (normal and super nodes). The weighted probability is used for selecting the cluster heads based on the residual energy [6]. Threshold sensitive Energy Efficient sensor Network (TEEN) Protocol is used for time critical application. TEEN has two type of threshold as hard and soft threshold for data transmission [7]. The hybrid routing protocol divides the whole area into region and uses clustering technique for improving energy utilization and network lifetime [8–10].

Dutta et al. [11] describe E-xtreme scale mote for carrying out radiation monitoring unit for nuclear plant. Mote consists of microcontroller, radio unit, photo, acoustic, infrared and temperature sensor. The deployment of mote is aerial in nature.

Ding et al. [12] uses light weight sensor with Zigbee communication for radiation monitoring at nuclear plant. All nodes are enable with GPS facility for localization among nodes. Sleep-Wake Up Schedule is used for power saving among nodes.

Ranjith Kumar [13] describe Zigbee as a suitable technology for radiation monitoring. Monitoring system consists of Zigbee node have core–board unit with nuclear radiation sensor and uses T1CC2530f256 as chip. Zigbee node embedded with CC2591 to increase link budget for range over. The sensor is embedded with Geiger Muller (GM) Tube. The basic principle of GM Tube is when radiation pass through air,/special gas, it ionizes the air inside the gas tube. This ionization result in generation of signal indicating the level of radiation. More is the radiation, more is ionization of gas and more is signal value [13].

3 Proposed Model

The proposed model consists of following system.

- Radiation model
- Architectural model
- Cluster Head Selection Model using Fuzzy Logic
- Data Transmission and Monitoring Model

3.1 Radiation Model

The proposed system is used to monitor the radiation level at the nuclear plant. The monitored area is divided into regions. The base station (nuclear plant) is at the centre of monitored area. The system consider two types of sensor nodes as normal nodes, and super nodes. The region close to base station deployed with normal nodes while region far from base station are deployed with super node. The normal nodes are used for routing purpose while the super nodes uses multi-hop communication. The super nodes have (α) times higher energy level than normal nodes.

The radiation emitted at the nuclear plant transmit in all direction. The intensity of the radiation is given by following equation:

$$I = E/4\pi D^2 \tag{1}$$

Where E is Energy radiated per unit by source, and D is distance from source.

The coverage area of the system is NXN having m number of cluster. The range of cluster is calculated as follow [14]:

$$\begin{aligned} \mathrm{A} &= \frac{N}{(\sqrt{m}\Pi)} \\ \mathrm{A} &= \frac{1}{\sqrt{\pi}} * \frac{N}{\sqrt{m}} \\ \mathrm{A} &= \mathrm{K} * \frac{N}{\sqrt{m}} \\ \text{Where } \mathrm{K} &= 0.5643 \end{aligned} \tag{2}$$

Data Transmission from cluster head to base station occurs in following ways.

If (A= =R),
  then one-hop transmission is allowed.
Else
  multi-hop transmission is allowed.

where R is communication region of sensor node.

Communication between node have minimum hop-count [15]. For calculating hop count, distance between cluster head and base station must be known. The distance from cluster head and base station is calculated as follow [16]:

$$S = \left(\int \sqrt{(x^2 + y^2)} dA/A \right)$$
$$S = (0.765 * N)/2 \tag{3}$$

For calculating of number of hop from cluster head to base station is given by formula:

$$\text{Number of hop} = \lfloor S/R \rfloor \tag{4}$$

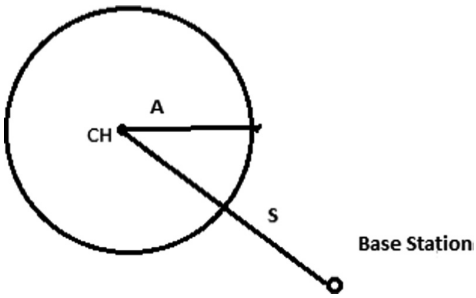


Fig. 1. Calculation model for simple clustering

Where S is average distance between cluster head and base station (Fig. 1).

The range of radiation intensity inside a cluster head (x_1, y_1) to (x_2, y_2) is defined as

$$I_{range} = E\left\{ \frac{1}{\left|(x - x_1)^2\right|} - \frac{1}{(x - x_2)^2} \right\} \tag{5}$$

Where (x, y) is coordinate of base station. So after simplification the intensity variation within cluster is found of 0.0007 unit.

3.2 Architectural Model

The model consists of heterogeneous sensor nodes (normal and super nodes) they are distributed in (N × N) unit^2 network area. The Base Station is located at the centre of the sensing field. The area is divided into region and each region having number of clusters. For efficient coverage of monitoring area. The proposed region 100 × 100 is divided into five region (Fig. 2).

Region 1: Normal nodes are randomly deployed in Region 1 having respective x and y coordinate .20<X<=80 and 20<Y<=80.

Region 2:Region 2 lies between 0<X<=100 and 80<Y<=100 and in this region super nodes are randomly deployed.

Region 3:Region 3 lies between 0<X<=100 and 0<Y<=20 and in this region super nodes are randomly deployed.

Region 4:Region 4 lies between 0<X<=20 and 20<Y<=80 and in this region super nodes are randomly deployed.

Region 5:Region 5 lies between 80<X<=100 and 20<Y<=80 and in this region super nodes are randomly deployed.

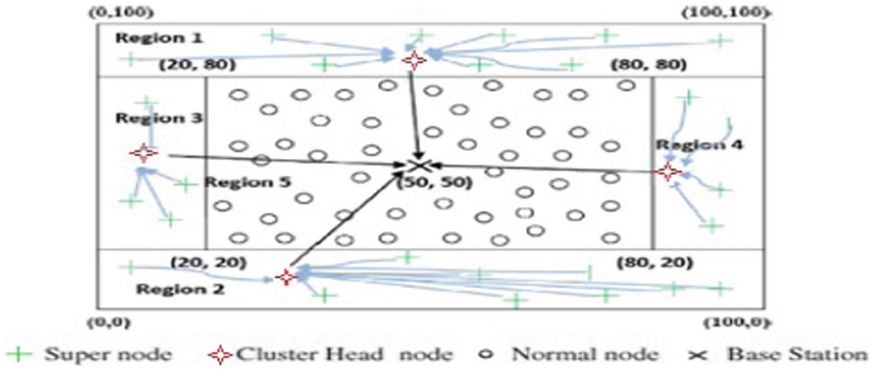


Fig. 2. Deployment of nodes in sensing area.

3.3 Cluster Head Selection Model Using Fuzzy Logics

The protocol is designed on Fuzzy logics for selection of cluster head among super nodes. The system consider three parameter Distance, Residual Energy and Noise factor for selection of cluster head. The parameter can be defined in fuzzy set as follows:

$$X = \{(D, \mu(D))\},\ D \in D\,\text{is}$$

$$Y = \{(RE, \mu(RE))\},\ RE \in \text{Res\_En.}$$

$$Z = \{(N, \mu(N))\},\ N \in Ns$$

Where, Dis represent as a universe of discourse for Distance, Res_En. represent universe of discourse for Residual Energy and Ns represent universe of discourse for Noise. D, RE and N are the elements of Dis, Res_En. And Ns respectively. μ(D), μ (RE) and μ(N) are the membership function of distance, residual energy and noise factor.

Table 1. Degree of input variable of distance, residual energy and noise factor using membership function.

Input	Membership		
Distance	Close	Considerable	Far
Residual energy	Low	Adequate	High
Noise factor	less	Medium	High

Precedence order for input parameter are as follow:

Distance > Residual Energy > Noise Factor.

The Noise, Distance and noise input membership produce following combination for selecting the cluster head.

Table 2. 27 probable output membership function.

Noise (N)	Distance (D)	Residual Energy (RE)	Output membership
Less	Close	High	R1
Less	Close	Adequate	R3
High	Far	Adequate	R26
High	Far	Low	R27

The Output membership represent the criteria for selecting of cluster head based on IF THEN RULE Set and Precedence order for output membership is R1 > R2 > …… > R26 > R27.

3.3.1 Validation of Model

The selection of cluster head using fuzzy technique is as follows. The proposed system uses following membership function for calculating degree of input variables.

$$\mu(N) = \begin{cases} 1 & \text{if } N < Th_1 \\ (Th_2 - N)/(Th_2 - Th_2) & \text{if } Th_1 \le N \le Th_2 \\ 0 & \text{if } N > Th_2 \\ Th_1 = 0.4 \text{ and } Th_2 = 0.8 \end{cases}$$

$$\mu(D) = \begin{cases} 1 & \text{if } N < Th_1 \\ (Th_1 - D)/(Th_2 - Th_1) & \text{if } Th_1 < D \le Th_2 \\ 0 & \text{if } D > Th_1 \\ Th_1 = 40 \text{ and } Th_2 = 80 \end{cases}$$

$$\mu(RE) = \begin{cases} 0 & \text{if } RE \le Th_1 \\ (RE_1 - Th_1)/(Th_2 - Th_1) & \text{if } Th_1 < RE < Th_2 \\ 1 & \text{if } RE \ge Th_2 \\ Th_1 = 0.15 \text{ and } Th_2 = 0.5 \end{cases}$$

where Th_i represent threshold (minimum/maximum) for N, D, RE.

Let us consider a network of 5 super node used for selection of cluster head with following parameter.

The Distance set for super node:

D = {25, 45, 60, 85, 95}

The Residual Energy at Each node are as:

RE = {0.1, 0.22, 0.325, 0.5, 0.8} and

Noise factor at each node as:

N = {0.1, 0.3, 0.5, 0.7, 0.9}.

The degree of membership for distance, noise and residual energy are calculated based on above membership function and are as {1, 0.8, 0.5, 0, 0} {1, 1, .75, .25, 0} and {0, 0.2, 0.5, 1, 1}.

Based on fuzzy decision rule from Table 1, the membership function for Distance are {25|close, 45|considerable, 60|considerable, 81|far, 95|far}, the membership function for Noise are {0.1|less, 0.3|less, 0.5|significant, 0.7|significant, 0.9|high} and the membership function for Residual Energy are {0.1|low, 0.22|adequate, 0.325|adequate, 0.5|high, 0.8|high} (Table 2).

The proposed system uses fuzzy logic (MIN-MAX Rule) for selection of cluster head using AND rule operation between Distance, Noise and Residual Energy

$$\mu(D) \wedge \mu(N) \wedge \mu(RE) = \min(\mu(D), \mu(N), \mu(RE)).$$

Table 3. Fuzzy relation for membership value of distance and noise

Distance/Noise	25	45	60	85	95
0.1	1^1	0.8∧1	0.5∧1	0∧1	0∧1
0.3	1^1	0.8∧1	0.5∧1	0∧1	0∧1
0.5	1^.75	0.8∧.75	0.5∧.75	0∧.75	0∧.75
0.7	1∧.25	0.8∧.25	0.5∧.25	0∧.25	0∧.25
0.9	1∧0	0.8∧0	0.5∧0	0∧0	0∧0

The model consider the threshold for selecting the cluster head is Th > 0.6 for Distance, Noise and Residual Energy (Tables 3 and 4).

Table 4. Output table of distance and noise

(Distance/Noise)	25	45
0.1	1	0.8
0.3	1	0.8
0.5	.75	.75

The cluster head depend on energy also. Now Consider other factor (Distance/ Noise) to compute for cluster head Selection (Table 5).

Table 5. Output table of residual energy, distance and noise

$\frac{\text{Residual Energy}}{(\frac{\text{Distance}}{\text{Noise}})}$	0.5	0.8
25/0.1	1	1
25/0.3	1	1
25/0.5	.75	.75
45/0.1	0.8	0.8
45/0.3	0.8	0.8
45/0.5	.75	.75

Table 6. Output membership with degree of membership

S No.	Distance	Membership function	Noise	Membership function	Residual Energy	Members-hip function
1	25	Close	0.1	Less	0.5	High
2	25	Close	0.1	Less	0.8	High
3	25	Close	0.3	Less	0.5	High
4	25	Close	0.3	Less	0.8	High
5	25	Close	0.5	Medium	0.5	High
6	25	Close	0.5	Medium	0.8	High
7	45	Considerable	0.1	Less	0.5	High
8	45	Considerable	0.1	Less	0.8	High
9	45	Considerable	0.3	Less	0.5	High
10	45	Considerable	0.3	Less	0.8	High
11	45	Considerable	0.5	Medium	0.5	High
12	45	Considerable	0.5	Medium	0.8	High

There are 12 Rule set generated for selecting the cluster head from Table 6. The cluster head is selected on the basis of the above output membership depending upon Distance, Noise and Residual Energy. The first cluster head is selected on basis of output membership when Noise = 0.1, Distance = 25 and Residual Energy = 0.8.

3.4 Data Transmission and Monitoring Model

The Transmission model describe how the data is sensed and transmitted from sensor node to base station. The model consider region-wise deployment of the sensor node. Cluster formation is used for energy balancing of network. Mamdani Inference System is used the selection of cluster head [17]. For Data transmission, the proposed algorithm is described.

Data_ Transmission(node, base station)

Data Transmission(node, base station)
Predefine:
Tn ← Number_sensor nodes
m ← fraction_super nodes
Th ← Threshold value for Data transmission
x.dimension ← N
y.dimension ← N
SN(i) ← sensor node
CM : cluster Member
BS : Base Station
CH : Cluster head
MIS : Mamdani Inference System(AND _OR Operation)

Begin
1: Region wise deployment of sensor node
/* Distribution of node */
No_ Super node ← m * Tn
No_ Normal node ← (1-m) * Tn
2: If(node.type == super)
3: Formation of Cluster
4: Dist ←Distance_SN(i)
5: R.E ←Residual Energy_SN(i)
6: Ns ← Noise_SN(i)
7: Selection of Cluster Head
8: CH_ Selection ←MIS(Dist, R.E, Ns)
9: CH←Data_CM
10: If (Energy.CH >Th)
BS ← Data_CH
11: Else
Data Transmission via Normal node (multi-hop communication)
12: Else(node= normal)
Remain active (node) for routing
End

The proposed model monitor the system depending upon distance as a parameter from base station. The radiation received, indicates that within the distance the radiation strength is high. The radiation strength I and distance is related $I \alpha \frac{1}{D^2}$.

$$I = a*1/D^2$$

Where a is system parameter where $0 < a < 1$. So, Base station will generate alarm to control radiation level.

4 Simulation and Performance Evaluation

The Simulation is carried out on heterogeneous sensor network on area of 100×100 m^2 in MATLAB framework. The network consists of 200 total nodes with region based deployment of nodes based on their energy level. Region 1 contain only normal node while in rest of region super nodes are deployed (Fig. 3).

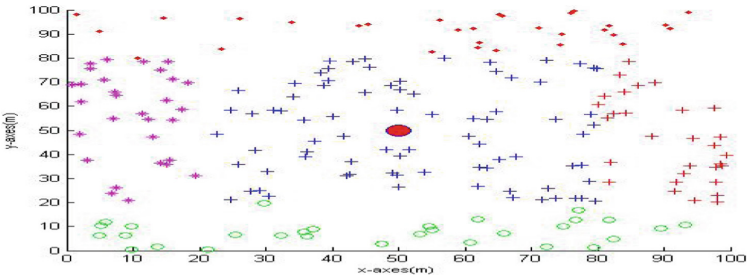


Fig. 3. Deployment of nodes

The simulation parameter are shown in Table 7.

Table 7. Simulation parameter

Parameters	Value
Initial energy	0.5 J
Energy factor (α)	1
Energy consumed in transmission E_{elec}	50 nJ/bit
Energy consumed by amplifier to transmit at shorter distance E_{fs}	10 pJ/bit/m^2
Energy consumed by amplifier to transmit at larger distance E_{amp}	0.0013 pJ/bit/m^4
Energy consumed in data aggregation	5 nJ/bit/report
Packet size	500 byte

We consider 60% and 30% of super node deployment while rest of nodes are deployed as normal nodes. The region-wise deployment of super nodes and normal nodes are shown in Table 8.

Table 8. Region based distribution of nodes

Node types	Total number of nodes (30% super nodes)	Total number of nodes (60% super nodes)
Super node in Region 2	15	30
Super node in Region 3	15	30
Super node in Region 4	15	30
Super node in Region 5	15	30
Normal node in Region 1	140	80

Table 9. Comparison table for different number of super nodes

Performance metrics	Proposed (30% super nodes)	Proposed (60% super nodes)
Stability period (Number of rounds after first node dead)	1116	1163
Network lifetime (Number of rounds after all nodes are dead)	2829	3861
Normal node lifetime	2315	2005
Super node lifetime	2829	3861
Throughput (Total number of packet transmitted to Base Station)	73798	79394

The Simulation result shows that the performance of network lifetime improves as the number of super nodes increases as shown in Fig. 4(a) and (b) (Table 9).

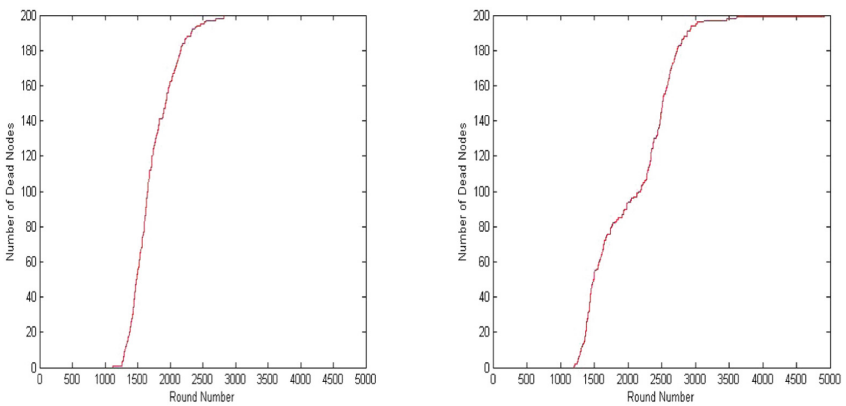


Fig. 4. (a) Number of dead nodes vs Number of round (30% super nodes) (b) Number of dead nodes vs Number of round (60% super nodes)

5 Conclusion

In this paper we proposed a protocol based on Fuzzy technique for heterogeneous sensor node for monitoring of nuclear radiation at nuclear plant. The algorithm mostly selects the optimal path for transmission of data packets from node (source) to base station (destination). Fuzzy technique is used for cluster head selection and energy efficient hybrid routing along with scheduling of nodes. The protocol shows that using clustering of super node and normal node improve the network lifetime.

References

1. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: A survey on sensor networks. IEEE Commun. Mag. **40**, 102–114 (2002)
2. Wu, L., Wang, L., Su, S., et al.: Wireless sensor network based solution for nuclear radiation detection. IEEE Conference 2014, vol 6, 978-1-4799-4808-6 (2014)
3. Ranjith Kumar, R., Macwin, L., Rathna, R.: Nuclear radiation detection using wireless sensor network. In: IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems ICIIECS 2015, 2015
4. Yadav, R., Daniel, A.K.: Performance and analysis of MAC protocols based on application. In: International Conference on Electrical, Electronics, Materials and Applied Science (2018)
5. Mhatre, V., Rosenberg, C.: Homogeneous vs. heterogeneous clustered sensor networks: a comparative study. In: Proceedings of 2004 IEEE International Conference on Communications (ICC 2004), June 2004
6. Smaragdakis, G., Matta, I., Bestavros, A.: SEP: a stable election protocol for clustered heterogeneous wireless sensor networks. In: Proceedings of 2nd International Workshop on Sensor and Actor Network Protocols and Applications (SANPA 2004) (2004) 6. Soro, S., Henizelman, W.B.: Prolonging the lifetime of wireless sensor networks
7. Manjeshwar, A., Agrawal, D.P.: TEEN: a routing protocol for enhanced efficiency in wireless sensor networks. In: 1st International Workshop on Parallel and Distributed Computing Issues in Wireless Networks and Mobile Computing, San Francisco (2001)
8. Baghouri, M., Chakkor, S., Hajraoui, A.: Fuzzy logic approach to improving stable election protocol for clustered heterogeneous wireless sensor networks. J. Theor. Appl. Inf. Technol. **53**(13), July 2013
9. Maurya, S., Daniel, A.K.: Hybrid routing approach for heterogeneous wireless sensor networks using fuzzy logic technique. In: Published in Proceedings of 4th IEEE Computer Society International Conference on Advanced Computing & Communication Technologies (2014)
10. Chaturvedi, P., Daniel, A.K.: An energy efficient node scheduling protocol for target coverage in wireless sensor networks. In: The 5th International Conference on Communication System and Network Technologies (CSNT-2015), pp. 138–142, April 2015. https://doi.org/10.1109/csnt.2015.10
11. Dutta, P., et al.: Design of a wireless sensor network platform for detecting rare, random, and ephemeral events. In: Proceedings of the 4th International Symposium on Information Processing in Sensor Networks. IEEE Press (2005)
12. Ding, F., et al.: A GPS-enabled wireless sensor network for monitoring radioactive materials. Sens. Actuators A Phys. **155**(1), 210–215 (2009)

13. Li, J., Zhu, X., Tang, N., et al.: Study on ZigBee network architecture and routing algorithm. In: 2010 2nd International Conference on Signal Processing Systems (ICSPS), vol. 2, pp. V2-389–V2-393. IEEE (2010)
14. Nam, C.S., Han, Y.S., Shin, D.R.: Multi-hop routing-based optimization of the number of cluster-heads in wireless sensor networks. MDPI J. Sens. **11**, 2875–2884 (2011)
15. Arioua, M., el Assari, Y., Ez-zazi, I., el Oualkadi, A.: Multi-hop cluster based routing approach for wireless sensor networks. The 7th International Conference on Ambient Systems, Networks and Technologies (ANT 2016) (2016)
16. Raghuvanshi, A.S., Tiwari, S., Tripathi, R., Kishor, N.: Optimal number of clusters in wireless sensor networks: an FCM approach. In: International Conference on Computer and Communication Technology (2010)
17. Singh, S., Chand, S., Kumar, B.: Energy efficient clustering protocol using fuzzy logic for heterogeneous WSNs. Wireless Pers. Commun. **86**, 451–475 (2016)

A Framework for Lane Prediction on Unstructured Roads

Rohan Raju Dhanakshirur, Preeti Pillai(✉), Ramesh Ashok Tabib, Ujwala Patil, and Uma Mudenagudi

KLE Technological University, Hubballi, India
preeti_pillai@bvb.edu

Abstract. In this paper, we propose to address the issue of lane prediction on unstructured roads, i.e. roads where lane markings are not available. Lane prediction has received considerable attention in the last decade towards the development of ADAS (Advanced driver assistance systems) system. We consider lane prediction as a vision based problem and propose a learning based framework for lane prediction. We preprocess the data using adaptive thresholding to estimate ROI (Region of Interest) in an image. We develop a variant of Bayesian learning using the evidence based Cascaded Dempster Scafer Combination Rule to categorize the road and non-road sectors of the region of interest. We also propose to post-process the data with improved morphological operations to remove the outliers. Lane prediction finds applications in pothole detection, autonomous driving etc. We demonstrate the results on real datasets captured in different scenarios. We compare the results with different state-of-art techniques of lane prediction to validate the efficiency of proposed algorithm.

Keywords: Lane prediction · Unstructured roads
Cascaded Dempster Scafer Combination Rule · Region of Interest Classification · Improved morphological operations

1 Introduction

Lane detection/prediction is a well-researched area in computer vision with its application in autonomous vehicles and advanced driver assistance systems. Lane detection is a process of identifying best drivable region for any vehicle. The crucial task in lane detection/prediction is identifying the white markings on a dark road (i.e. lane markings), but this simple task can turn out difficult on unstructured roads i.e. roads with no lane marks. The lane prediction system should be able to reliably estimate the lane trajectory even if lane markings are not available.

American study [1] shows that most of the accidents (96.2%) are because of human faults. Crashes reported by the police show that at least 25% of them are due to driver inattention [25]. Due to this, an Advanced Driver Assistance

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 178–189, 2019.
https://doi.org/10.1007/978-981-13-5758-9_16

System (ADAS), in particular, Lane Departure Warning System (LDWS) has received considerable attention in the recent decades. If a system can alert the driver on the lane departure, then many of these accidents can be avoided or atleast be reduced. The safety of the drivers, pedestrians and passengers increases because such systems.

Different methods are proposed in the literature to detect the road area in the images captured from on-board camera mounted on a vehicle and can be classified broadly into detecting lanes on unstructured roads and on structured roads. For structured roads (roads with clear edges and well-painted lane markings), lane detection is performed using image intensity as a key feature. Models that are geometrical [12,30], deformable [10], colour distributional [7,11] etc. are being used for road area detection on structured roads. People use Spline model [9,30] colour cue [26], Hough transform [23] and stereo-vision [3] etc. for the feature extraction and localization. In addition to this, researchers make use of sensors such as laser [24] and radar [15] to detect road. All the above discussed methods perform well only if there is a considerable variation in colours between the surroundings and the road. However, these methods do not perform well on unstructured roads having unpredictable weather, diversified terrain covers and with various lighting conditions [29].

For unstructured roads, researchers prefer learning based methods [19]. Authors in [17] propose a road detection technique based on boundary classifier and multichannel histogram filter. This technique assumes major part of the image to be road, and fails when it is not. Authors in [16] uses modified morphological operations to classify road region with the non-road regions. Authors in [14] tries to identify the road region by blurring the edges. These methods fail to classify the sky region from road. Authors in [2] propose a combination of road boundary detection based on geometric projections and Adaboost-based region segmentation to identify the "drivable" area. However, the training time for this algorithm is very high [8]. Optical flow based techniques as proposed in [13], [4] perform road region segmentation adaptively, but authors in [8] demonstrate, on chaotic roads, the accurracy of the optical flow estimation goes low and the techniques do not work well. The SCARF system as proposed in [5] works for roads with lane marking on degraded surfaces. It uses set of Gaussian colour models for differentiation of road and the non-road sectors. Another method proposed in [6] uses HSV colour representation to learn the road portion's features. The methods based on colour do not work well in scenarios where the intensity distribution between the non-road and the road area is negligible. People generally use texture features to overcome this problem [18,22]. Another way of detecting road is by considering the vanishing point of texture cue [20,21]. In general, vanishing point is estimated by measuring the orientations of texture allied with the road pixels [21]. Based on this fact, Authors in [22] have proposed an algorithm to find the vanishing point by determining the prevailing orientations of texture flow. Disadvantage of this technique is that the shape of the road area is not preserved. It also poses a constraint that the road region must always be in triangular shape. If the edges of the roads do not converge, then major portion of

the road remains uncovered. Moreover, the process of computing the vanishing point is time consuming as it uses Gabor filter in 36 orientations to estimate the principal orientation of each pixel.

In this paper, we propose lane prediction on unstructured roads using evidence based variant of Bayesian learning framework. Towards this,

- We propose to select Region of Interest (RoI) from the test image using the adaptive thresholding to optimize the computational time required for classification.
- We propose evidence based variant of Bayesian learning using the Dempster Scafer Combination Rule to categorize road and the non-road sectors of the image.
- We propose to post-process the result with improved morphological operations to remove the outliers.
- We demonstrate the results on real dataset of unstructured road videos and compare the result with different methods of state-of-art.

In Sect. 2, we discuss the framework of the proposed lane prediction system. In Sect. 3, we demonstrate the results and we conclude the analysis in Sect. 4.

2 Framework for Lane Prediction

The proposed lane prediction approach is shown in Fig. 1. It consists of four phases namely, training a variant of bin based bayes classifier, Selection of Region of Interest (ROI) for test image, Classification of ROI into road & non-road regions and post processing the classified image with improved morphological operations to remove the outliers.

2.1 Training the Classifier

We use fixed number of road and non-road images for training. Every image of the training dataset is divided into patches of fixed size. We separate each channels of image and seven competent texture features namely mean, variance, standard deviation, energy, entropy, smoothness, 3^{rd} moment with PCA (principle component analysis) are extracted for every channel of an image patch to generate a set of 24 competent features. The probability of texture features being similar for non-road and road regions necessitates the use of set of features [28]. We normalize the features using mean normalization technique to represent every feature descriptor as a value between 0 to 1. We generate fixed number of bins for every channel of the image. We model a variant of Naive-bayes classifier as a bin based classifier by estimating the hypothesis for every bin. Bin based classifier necessitates the normalization of each features. Naive bayes hypothesis is distinctively set at every bin of the classifier for every channel independently. These hypothesis are used as training models for classification.

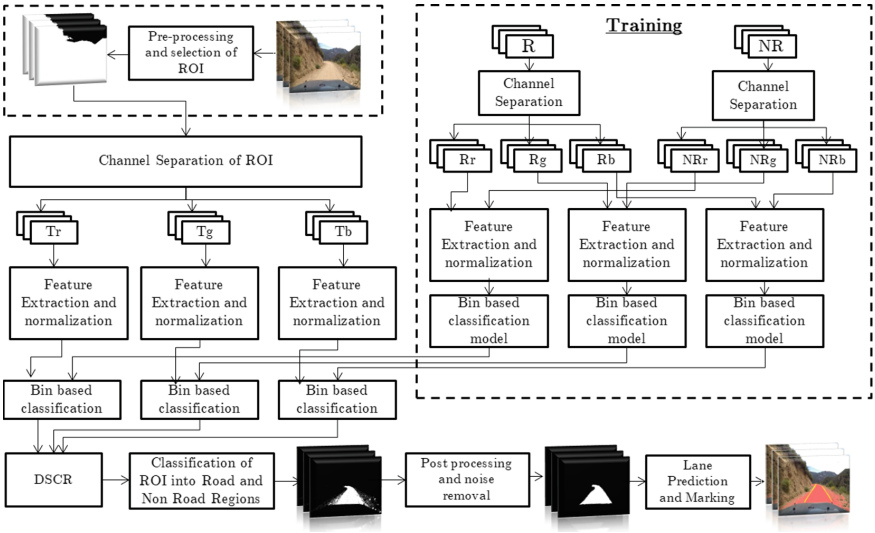


Fig. 1. Framework for lane prediction

2.2 Selection of ROI

We propose to select the ROI by performing adaptive thresholding on the input images. The selection of ROI is necessary because, we propose to classify the ROI rather than classifying the entire image in the further stages. This reduces computational time for classification. Also in many situations, sky is classified as road due to similar feature descriptors for variance and standard deviation. Selecting ROI eliminates the sky region from undergoing classification and thereby improves efficiency.

The input images are pre-processed by performing adaptive thresholding to obtain thresholded images. We use otsu approach [27] to perform adaptive thresholding. The thresholded images decide the ROI. Generally the pixels with high intensity represent the sky region and can be assumed to be non-road. The results of ROI selection are shown in Fig. 2. Here, Fig. 2(a, b and c) shows input frames of 3 different unstructured roads and Fig. 2(d, e and f) shows their corresponding thresholded image obtained from preprocessing. The 'white' region of the thresholded image is considered as ROI.

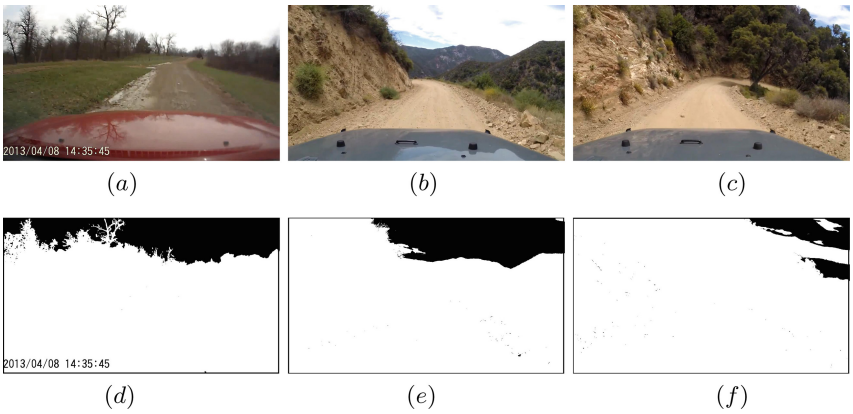


Fig. 2. The pre-processing of input images to select the ROI. (a), (b), (c) show the input frames and (d), (e), (f) the 'White' region show their corresponding ROIs

2.3 Classification of ROI

The Bin based Naive-bayes classifier is applied on the RoI of the image to classify the elements of RoI into the road and non-road regions. The RoI of the test image is divided into patches of fixed size. Same features are extracted for every patch of the test image and normalized independently with the same normalizing factor used in training. The approach estimates the bin of the test patch for every features explicitly. The hypothesis of the corresponding bin is used to classify the test patch into road or non-road regions. This results in 8 classification votes corresponding to 8 features for every patch in every channel. The probability of every vote being correct, is estimated by computing the difference in the number of road and non-road training patches present in the corresponding bin. Weighted fusion of all the votes is performed by considering the probability of a particular vote being correct as its corresponding weight and we obtain the classification probability of every channel independently.

Cascaded Dempster Scafer Combination Rule: We propose Cascaded Dempster Scafer combination rule (CDSCR) to combine the classification probabilities obtained from every channel to generate the final probability for the test patch being road. Proposed CDSCR is shown in Fig. 3. We consider the classification probabilities of red and green channels as masses to be combined using DSCR.

DSCR combines two evidences with their corresponding masses for belief, disbelief and uncertainty to give combined evidence with its masses for belief, disbelief and uncertainty. Let p_1 and p_2 be the probabilities to be combined. We consider these values as masses of belief for the two evidences evd_1 and evd_2 respectively. The masses for disbelief shall be $1 - p_1$ and $1 - p_2$ respectively. We consider the masses for uncertainty as 0.

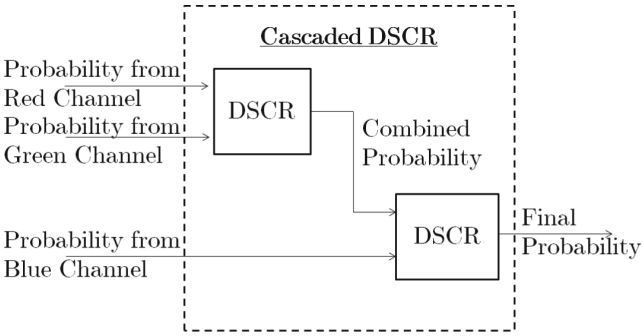


Fig. 3. Framework for cascaded DSCR

The two evidences are combined using the combination rule as demonstrated in Table 1. The product of belief and disbelief gives rise to conflict and is represented by $\emptyset$. The product of belief and belief or the product of belief and uncertainty represents belief and is represented by ψ_1. Similarly ψ_2 represents the disbelief.

Table 1. Evidence combination table

$\cap$	$m(\mathcal{E}_1^{belief})$	$m(\mathcal{E}_1^{disbelief})$	$m(\mathcal{E}_1^{ambiguity})$
$m(\mathcal{E}_2^{belief})$	ψ_1	$\emptyset$	ψ_1
$m(\mathcal{E}_2^{disbelief})$	$\emptyset$	ψ_2	ψ_2
$m(\mathcal{E}_2^{ambiguity})$	ψ_1	ψ_2	Ω

The Combined belief of the two evidences is given by Eq. 1

$$CombinedBelief(CB) = \frac{\sum \psi_1}{1 - \sum \emptyset} \tag{1}$$

The Combined belief of the two evidences is considered as the combined probability CP_1. CP_1 and the probability of the blue channel are considered as masses of belief for evd_1 and evd_2 respectively. We repeat the same process again to obtain the final probability FP. The label to the test patch is decided using the value of FP.

This process is repeated for all the patches in the ROI to obtain all the labels. Obtained labels are used to construct a segmented image classifying the road region from the non-road region.

2.4 Post-processing

Segmented image is post processed by performing improved morphological operations like erosion and dilation sequentially with the help of a small kernel. Erosion will remove the outliers in the road region and dilation will remove the outliers from non-road region. The segmented image is superimposed on the test image and the intensity of the red channel of the pixels whose prediction is road, is made 255. This visualizes the lane present on the unstructured road. For Further visualization, we extact lines from the binary image using hough transform and plot hough lines on the image to demonstrate the road region.

If the formulation of two dimensional hough space is seen as $y = ux + v$, the equation of the line gets transferred to Eq. 2 in the hough transform.

$$xcos(\theta) + ysin(\theta) = \rho \tag{2}$$

where ρ is the vertical length measured from the origin of the co-ordinate system to the corresponding line and θ is the angle measured between the corresponding line and the x axis. Hough transform can be used to amp any point (x, y) to the sine curve of the space (θ, ρ).

3 Results and Discussions

We demonstrate our methodology on 3 different unstructured road datasets. The results shown here are implemented with a patch size of $3X3$. The RoI is divided into patches of size $3X3$ and are considered for Classification. The classification output for 3 different datasets is shown in Fig. 4. The classified output is then used for visualization of the lanes and is shown in Fig. 4. Figure 4(a, b and c) shows the input images. Figure 4(d, e and f) shows the classified images. The white region in the classified images correspond to road region of the input image. The same is visualized in Fig. 4(g, h and i) respectively. We further perform edge detection and Hogg transform on the classified images in order to get the lane markings on the unstructured road. The results of lanes on unstructured roads are shown in Fig. 4(j, k and l) respectively.

The application of lane prediction can be seen in pot hole detection in an image as shown in Fig. 5. The considerable 'black' portion of the road area can be visualized as a pothole.

We compare our results with state of art implementation shown in [14,16,17]. The results of the same are shown in Fig. 6.

In each of the images, White region indicates the road area and the black region indicates the non-road area. It can be seen that our implementation results are better.

We use jaccard score to evaluate the performance of the proposed algorithms and the results are demonstrated in Table 2. We observe, proposed algorithm outperforms the other compared algorithms. We also demonstrate the time taken to process the test datasets on an I5 processor with 8 GB RAM in Table 2. We observe no much variation in the time taken.

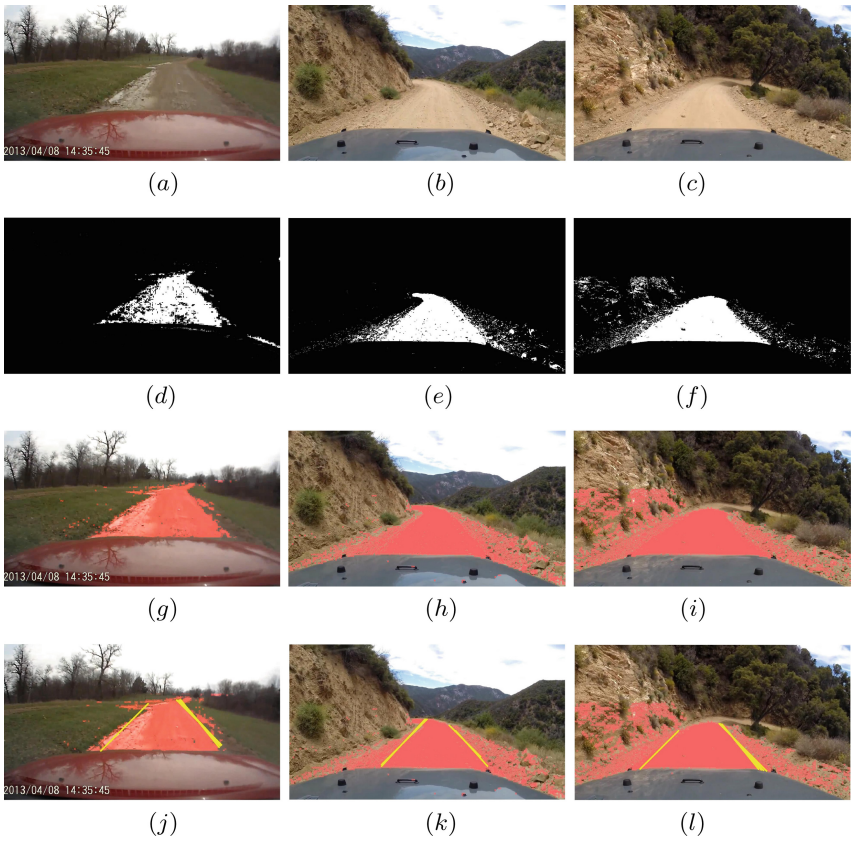


Fig. 4. Classification and lane prediction on unstructured roads (a), (b), (c) show the input frames (d), (e), (f) show the corresponding classified images in which white region is road area (g), (h), (i) show their visualization and (j), (k), (l) show the lane prediction

Fig. 5. Application of lane detection (a) shows the input image and (b) shows the detected potholes

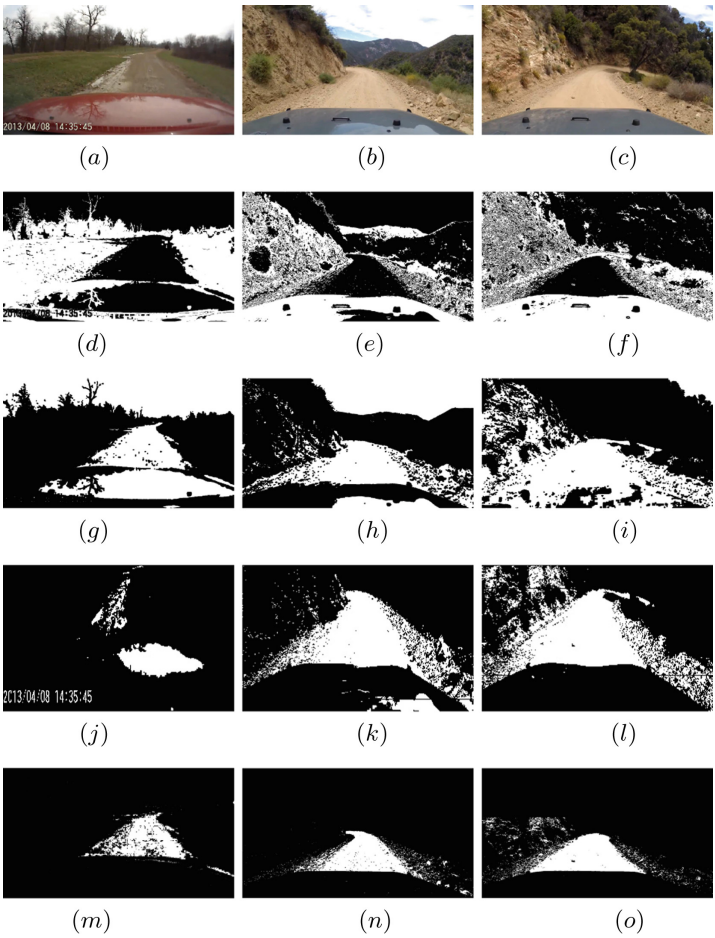


Fig. 6. Comparision results of road detection for lane predection (a), (b), (c) show the input frames, (d), (e), (f) show the corresponding classified images using [17], (g), (h), (i) show the corresponding classified images using [16], (j), (k), (l) show the corresponding classified images using [14] (m), (n), (o) show classification using proposed method. In each of the images, White region is road area and Black is non-road area

Table 2. Jaccard score of segmentation and Time taken to process test data (in sec)

	Jaccard score of segmentation				Time taken to process test data (in sec)			
	Proposed algorithm	[17]	[16]	[14]	Proposed algorithm	[17]	[16]	[14]
Dataset 1	0.83	0.42	0.59	0.65	2.4	3.0	3.2	2.4
Dataset 2	0.92	0.45	0.644	0.50	2.7	3.4	2.6	2.5
Dataset 3	0.89	0.39	0.631	0.51	2.6	3.2	2.7	2.8

4 Conclusion

In this paper, we addressed the problem of lane prediction on unstructured roads. Towards this, we implemented learning based framework for lane prediction on unstructured roads. We considered this as a vision problem and classified road and non road region for lane prediction. We demonstrated the results on real datasets captured in different scenarios. We demonstrated an application of lane prediction on pot-hole detection. We demonstrated the effectiveness of our implementation by comparing it with state-of-art implementation.

References

1. Aidman, E., Chadunow, C., Johnson, K., Reece, J.: Real-time driver drowsiness feedback improves driver alertness and self-reported driving performance. Accid. Anal. Prev. **81**, 8–13 (2015)
2. Alon, Y., Ferencz, A., Shashua, A.: Off-road path following using region classification and geometric projection constraints. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2006), vol. 1, pp. 689–696, June 2006
3. Bertozzi, M., Broggi, A.: GOLD: a parallel real-time stereo vision system for generic obstacle and lane detection. IEEE Trans. Image Process. **7**(1), 62–81 (1998)
4. Cremers, D., Rosenhahn, B., Yuille, A.L., Schmidt, F.R. (eds.): Statistical and Geometrical Approaches to Visual Motion Analysis. LNCS, vol. 5604. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03061-1
5. Crisman, J.D., Thorpe, C.E.: SCARF: a color vision system that tracks roads and intersections. IEEE Trans. Robot. Autom. **9**(1), 49–58 (1993)
6. Gao, Q., Luo, Q., Moli, S.: Rough set based unstructured road detection through feature learning. In: 2007 IEEE International Conference on Automation and Logistics, pp. 101–106, August 2007
7. Hsu, C.M., Lian, F.L., Ting, J.A., Liang, J.A., Chen, B.C.: Road detection based on bread-first search in urban traffic scenes. In: 2011 8th Asian Control Conference (ASCC), pp. 1393–1397, May 2011
8. John, N., Anusha, B., Kutty, K.: A reliable method for detecting road regions from a single image based on color distribution and vanishing point location. Procedia Comput. Sci. **58**(Supplement C), 2–9 (2015). Second International Symposium on Computer Vision and the Internet (VisionNet15)
9. Jung, C.R., Kelber, C.R.: A robust linear-parabolic model for lane following. In: Proceedings of the 17th Brazilian Symposium on Computer Graphics and Image Processing, pp. 72–79, October 2004
10. Kluge, K., Lakshmanan, S.: Deformable-template approach to lane detection, pp. 54–59. IEEE (1995)
11. Lee, J.W., Cho, J.S.: Effective lane detection and tracking method using statistical modeling of color and lane edge-orientation. In: 2009 Fourth International Conference on Computer Sciences and Convergence Information Technology, pp. 1586–1591, November 2009
12. Li, M., Zhao, C., Hou, Y., Ren, M.: A new lane line segmentation and detection method based on inverse perspective mapping. Int. J. Digit. Content Technol. Appl. **5**(4), 230–236 (2011)

13. Lookingbill, A., Rogers, J., Lieb, D., Curry, J., Thrun, S.: Reverse optical flow for self-supervised adaptive autonomous robot navigation. Int. J. Comput. Vis. **74**, 287–302 (2006)
14. Lu, K., Xia, S., Chen, D., Xia, C.: Unstructured road detection from a single image. In: 2015 10th Asian Control Conference (ASCC), pp. 1–6, May 2015
15. Ma, B., Lakshmanan, S., Hero, A.O.: Simultaneous detection of lane and pavement boundaries using model-based multisensor fusion. IEEE Trans. Intell. Transp. Syst. **1**(3), 135–147 (2000)
16. Ming, X., Juan, Z., Zhijun, F.: Research on unstructured road detection algorithm based on improved morphological operations. In: 4th International Conference on Smart and Sustainable City (ICSSC 2017), pp. 1–5, June 2017
17. Munajat, M.D.E., Widyantoro, D.H., Munir, R.: Road detection system based on RGB histogram filterization and boundary classifier. In: 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), pp. 195–200, October 2015
18. Nieto, M., Salgado, L.: Real-time vanishing point estimation in road sequences using adaptive steerable filter banks. In: Blanc-Talon, J., Philips, W., Popescu, D., Scheunders, P. (eds.) ACIVS 2007. LNCS, vol. 4678, pp. 840–848. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-74607-2_76
19. Qian, R., Liu, Q., Yue, Y., Coenen, F., Zhang, B.: Road surface traffic sign detection with hybrid region proposal and fast R-CNN. In: 2016 12th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD), pp. 555–559, August 2016
20. Rasmussen, C.: Grouping dominant orientations for ill-structured road following. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2004, vol. 1, pp. I-470–I-477, June 2004
21. Rasmussen, C., Korah, T.: On-vehicle and aerial texture analysis for vision-based desert road following. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2005) - Workshops, pp. 66–66, June 2005
22. Rasmussen, C.: Texture-based vanishing point voting for road shape estimation (2004)
23. Southall, B., Taylor, C.J.: Stochastic road shape estimation. In: Proceedings of the Eighth IEEE International Conference on Computer Vision, ICCV 2001, vol. 1, pp. 205–212 (2001)
24. Sparbert, J., Dietmayer, K., Streller, D.: Lane detection and street type classification using laser range images. In: Proceedings of the 2001 IEEE Intelligent Transportation Systems, ITSC 2001, (Cat. No.01TH8585), pp. 454–459 (2001)
25. Stutts, J., Reinfurt, D.W., Rodgman, E.: The role of driver distraction in crashes: an analysis of 1995–1999 crashworthiness data system data, vol. 45, pp. 287–301 (2001)
26. Sun, T.-Y., Tsai, S.-J., Chan, V.-H.: HSI color model based lane-marking detection. In: 2006 IEEE Intelligent Transportation Systems Conference, pp. 1168–1172 (2006)
27. Tabib, R.A., Patil, U., Ganihar, S.A., Trivedi, N., Mudenagudi, U.: Decision fusion for robust horizon estimation using dempster shafer combination rule. In: 2013 Fourth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1–4, December 2013
28. Tabib, R.A., Patil, U., Naganandita, T., Gathani, V., Mudenagudi, U.: Dimensionality reduction using decision-based framework for classification: sky and ground. In: Sa, P.K., Sahoo, M.N., Murugappan, M., Wu, Y., Majhi, B. (eds.) Progress in

Intelligent Computing Techniques: Theory, Practice, and Applications. AISC, vol. 519, pp. 289–298. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-3376-6_32
29. Thrun, S., et al.: Stanley: The robot that won the DARPA grand challenge. J. Field Robot. **23**(9), 661–692 (2006)
30. Wang, Y., Teoh, E.K., Shen, D.: Lane detection and tracking using B-Snake. Image Vis. Comput. **22**(4), 269–280 (2004)

Survey on Virtual Assistant: Google Assistant, Siri, Cortana, Alexa

Amrita S. Tulshan(✉) and Sudhir Namdeorao Dhage

Department of Computer Engineering, Sardar Patel Institute of Technology, Mumbai 400058, India
amrita.tulshan@spit.ac.in, sudhir_dhage@gmail.com

Abstract. Virtual assistant is boon for everyone in this new era of 21st century. It has paved way for a new technology where we can ask questions to machine and can interact with IVAs as people do with humans. This new technology attracted almost whole world in many ways like smart phones, laptops, computers etc. Some of the significant VPs are like Siri, Google Assistant, Cortana, and Alexa. Voice recognition, contextual understanding and human interaction are the issues which are not solved yet in this IVAs. So, to solve those issues 100 users participated a survey for this research and shared their experiences. All users' task was to ask questions from the survey to all personal assistants and from their experiences this research paper came up with the actual results. According to that results many services were covered by these assistants but still there are some improvements required in voice recognition, contextual understanding and hand free interaction. After addressing these improvements in IVAs will definitely increased its use is the main goal for this research paper.

Keywords: Human factors · Virtual Personal Assistant · Smart devices

1 Introduction

Intelligent Virtual Assistant (IVA) is "an application that utilization information, for example, the user's voice… and logical data to give help by noting inquiries in normal dialect, making suggestions and performing activities". Inside the writing the term IVA is utilized conversely with terms, for example, Conversational Agents, Virtual Personal Assistants, Personal Digital Assistants, Voice-Enabled Assistants or Voice Activated Personal Assistants, to give a few illustrations. IVAs join talk affirmation lingo understanding, trade organization, tongue age and talk association to respond to clients request and sales. Voice engaged IVAs like Siri, Google Assistant, Microsoft Cortana and Amazon Alexa are by and large open on cutting edge cell phones, and continuously in homes (e.g. Amazon Echo and Google Home) and automobiles (e.g. Google Assistant blend with Hyundai). The market for IPAs is foreseen to reach 4.61 billion by the mid 2020s.

The specialized foundations that empower IVAs have progressed quickly lately and have been the subject of broad research. Be that as it may, investigate concentrated on understanding the users experience of IVAs is more restricted. Obviously, from our point of view, this has had suggestions for the selection and utilization of IVAs. For

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 190–201, 2019.
https://doi.org/10.1007/978-981-13-5758-9_17

instance, in spite of their across the board and generally advanced consideration on cell phones individuals tend to utilize IVAs once in a while or not in the least. A current review demonstrated that 98% of iPhone clients had utilized Siri previously. However just 30% used it routinely, with 70% utilizing it once in a while or just sporadically [8].

IVAs can talk with different great IoT contraptions running a maintained OS. Siri works exclusively with Apple things—iPhone, iPad, iPod Touch, Home Pod, Mac, Apple Watch, and Apple TV devices. Microsoft Cortana works with Windows 10, Android, Xbox One, Skype, iOS, Cyanogen, and Windows Mixed Reality devices. Alexa works with Amazon's Resound, Fire, and Dash thing families and diverse keen contraptions running Android and iOS including PDAs, savvy speakers and headphones, brilliant watches, and advanced cell devices including TVs, radios, lights, indoor controllers, and coolers. Google Assistant in like manner works with Android and iOS contraptions.

IVAs are becoming increasingly popular but still enthusiasm inside the people has been developing for instance, researched the capability of discourse just interfaces (Siri with the screen blocked) for more seasoned users. The writers concentrated on regular errands like requesting the climate, making notes, requesting headings, composing an instant message, requesting an address and setting a schedule and found that members were very positive to utilizing discourse interfaces. While issues were recognized in connection to the voice recognition with four assistants Siri, Google Assistant, Alexa, Cortana. There were still some improvements in voice speech recognition and people are searching best in contextual understanding too. The basic requirements of a user is to feel free and work in relax way. One instance of human free interaction can be taken as pick up a phone call then directly instruct VPs for that.

The pattern is comparative for different IVAs. This paper initiates survey from 100 users; some were frequent users while some were infrequent. They share their experiences regarding intelligent virtual assistant on their smart phones. The inquiry brings out user experiences and finds out some issues on which this paper discussion is held: (1) problem with voice recognition and contextual understanding (2) problem faced in human free interaction. On the off chance that these boundaries can't be tended to, engineers might be best served in tolerating that IPAs will remain a moderately specialty application and spotlight on boosting the experience of users. Eventually, we accept improved users encounters and across the board selection is an achievable target. This paper plans to help this goal.

2 Objective

Currently existing IVA's provide enormous types of functionalities. However, the functionalities may vary depending on the personal assistant, implementation and the person implementing it. The main objective of the survey was to validate the real potential of this personal assistant, which their creators usually brag about. Considering heterogeneity in availability of these personal assistants it may be difficult for user to choose amongst them which one proves to be the best in real life scenario. This survey plays a role of guiding path for the same. In addition to this, after the evaluation of their potentials exploring the possibilities to combine various existing technologies with IVA's for accomplishing some unbelievable feats.

3 Literature Review

A computer primarily based approach for performing a command via a voice consumer interface on a subset of objects. The subset is selected from a fixed of items, each having an object type at least one taggable field is associated with the object type and has a corresponding value. The set of objects is saved in the laptop memory. An utterance is acquired from the person and consists of a command, an object type choice, a tag-gable field selection, and a price for the taggable discipline. Responsive to the utterance, at least one item is retrieved from the set of gadgets, the item of the sort selected through the user and having a price within the taggable area selection that matches the taggable field fee obtained from the user the command is done on the item. The object includes textual content that's converted to voice output [1]. They envisioned that someday computers will recognize natural language and count on what we need, whilst and where we need it, and proactively whole responsibilities on our behalf.

However, speech recognition and machine getting to know have persevered to be refined, and based records served through packages and content providers have emerged. We agree with that as computer systems turn out to be smaller and greater ubiquitous [e.g., wearable's and Internet of Things (IoT) [2]. The recognizer is designed to change a verbal articulation from a individual into an alternate method of data (e.g., text). A hand held individual colleague including a voice-recognizer and a characteristic dialect processor is disclosed. This snippet of data can be a plan for the day, data in the individual's logbook or data from the individual's address book, Such as a telephone number [3].

The Most well known utilization of iPhone is "SIRI" which causes the end client to impart end client versatile with voice and it additionally reacts to the voice charges of the client. It is named as Personal Assistant with Voice Recognition Intelligence, which takes the client contribution to type of voice or content and process it and returns the yield in different structures like activity to be performed or the item is directed to the end client. Furthermore, this proposed framework can change the method for communications between end client and the cell phones [4]. Open Data is currently gathering consideration for imaginative administration creation, predominantly in the zone of government, bio science, and shrewd venture. Be that as it may, to advance its application more for purchaser administrations, a web crawler for Open Data to realize what sort of information is there would be of assistance.

This paper presents a voice colleague which utilizes Open Data as its learning source. It is highlighted by change of precision as per the client criticisms, and obtaining of unregistered information by the client support. We additionally demonstrate an application to help for a field-work and affirm its viability [5]. The paper gives a diagram of the VPA applications, and the normal highlights and future patterns. The paper proposes also a bound together choice model in light of a quantitative appraisal of the significance of the solicitations and the accessibility of the client. Virtual Personal Assistant (VPA) is the up and coming age of bearer administrations for portable clients. VPA is accepted to be the smart advancement of administrations to take care of the regularly expanding demand by the portable experts for portability and network. The VPA controls the phone calls, deals with the individual exercises through logbook,

empowers the client to get to his undertaking administrator by means of voice interfaces, and incorporates every one of the elements of Unified Messaging. The Virtual Personal Assistant (VPA) will empower the client to productively handle expanding interest of phone calls, messages, gatherings and different exercises [6].

In any case, a great many people don't utilize them consistently. Past research has examined the encounters of continuous IPA clients. Utilizing subjective techniques we investigate the experience of rare clients: individuals who have attempted IPAs, yet pick not to utilize them consistently. Obviously occasional clients share a portion of the encounters of successive clients, e.g. dissatisfaction at confinements on completely sans hands collaboration. Critical purposes of difference and beforehand unidentified concerns likewise develop. Humanness of IPAs started correlations with human associates, comparing their restrictions. In particular, critical concerns rose around security, adaptation, information permanency and straight forwardness. Drawing on these discoveries we talk about key difficulties, including: outlining for interrupt ability; reexamination of the human similitude; issues of trust and information proprietorship. Tending to these difficulties may prompt more across the board IPA utilize [7].

As virtual assistants move toward becoming more intelligent and the IVA biological community of administrations and gadgets extends, there's a developing need to comprehend the security and protection dangers from this rising innovation. A few late occurrences feature noteworthy vulnerabilities in IVAs. Better demonstrative testing can uncover such vulnerabilities and prompt more reliable frameworks [8]. It enables the objective clients to connect with PCs and web based administrations with a wide cluster of usefulness in light of different web administrations and social media. There are four standard parts of the system; the voice recognition module, the natural language processing module, conversational agent and the content extraction module. The current screen per client writing computer programs are not fitting for getting to Internet in perspective of the base help they give for web content and the nonattendance of voice affirmation. The Virtual Right hand programming open in the market are not especially given everything and unfit to utilize it similarly. Some may confront issue now too [9].

This paper presents a usability of four Virtual assistant voic-baesd and contextual text (Google assistant, Coratan, Siri, Alexa) [10]. Cortana can likewise read your messages, track your area, watch your perusing history, check your contact list, watch out for your date-book, and set up this information together to propose valuable data, on the off chance that you enable it. You can likewise type your inquiries or solicitations, in the event that you want to not stand up uproarious. It is only desktop based virtual assistant. [11] Siri: Siri has been an integral part of iOS since the dispatch of iOS 5 of every 2011. It began with the nuts and bolts, for example, climate and informing, yet has extended significantly from that point forward to help all the more outsider mix with MacOS. While Siri's jokes are unbelievable, the virtual aide is getting more able consistently. Presently, you can request that it call individuals, send messages, plan gatherings, dispatch applications and recreations, and play music, answer questions, set updates, and give climate conjectures [12]. Google Assistant: Google Assistant (which has consolidated capacities from the more seasoned Google now, as now is being eliminated) is unique in relation to Cortana and Siri.

The significantly conversational VA is capable at interpreting essential vernaculars and understanding the importance behind unobtrusively complex request like, "What should we have for dinner?" It can in like manner see up to six unmistakable voices for couples and families, each voice settling to different logbook events and slants, great position amazing to Assistant and impeccable in a condition where everyone uses the voice helper on a singular gadget [13]. Alexa: While sharing different features similarly as various VAs, Alexa is in its own one of a kind class. Amazon's voice partner isn't centred on portable or PC purposes, but instead for the independent Amazon Echo speaker and a set number of Amazon Fire gadgets, with a more prominent focus on entire house administration and administrations as opposed to PC situated errands [14]. Each business visionary, side trickster and multitasking proficient out there would love to have a virtual assistant right hand to go up against a portion of the dull every day errands that accompany existing in the advanced time.

Similarly as with any developing innovation, in any case, it can be hard to isolate the build up from the certainties. There are four noteworthy players seeking consideration: Amazon (Alexa), Apple (Siri), (Google Assistant) and Microsoft (Cortana). I invested hours testing each of the four assistants by making inquiries and giving charges that numerous business clients would utilize [15]. Amid the testing procedure, I noticed the accomplishment of the AI's reaction to me, and in addition different components a planned users may think about, for example, simplicity of setup, general capacity to perceive my voice and relevant comprehension. About each cell phone and PC available today has a brilliant right hand caught inside, similar to an accommodating phantom—however how might they stack up against each other? While it may seem like Siri, Cortana, and the mysterious Google Assistant are in general just assortments of the same virtual partners, they each have their own specific unconventionality's, imperfections, and characteristics. So which one's best for clients? All things considered, that isn't a basic request to answer, as they're like the point that it's hard to take a gander at them without plunging significant into their capacities. Along these lines, we should start on this virtual right hand connection [16].

4 Methodology

We recruited 100 users (55 Female, 45 Male) among them 30 were frequent users and 70 were non frequent users. Almost all were native English speakers. We focused on best virtual assistants. All the 100 users using four devices to build their experiences the majority of our users addressed that they were not currently active users (Figs. 1 and 2).

Our survey involved only one stage. Users were provided 21 Multiple choice devices and multiple questions which they were supposed to ask to all of these four IVAs. These questions were based on following categories: Travel, Email, Messaging, Sports, Music, Weather, Calendaring, Social, Translation, Basic Tasks, General Knowledge, and Personality. The evaluation of the results of survey was based on which IVA did the used find the best, the devices used to access each personal assistant are mentioned in Table 1.

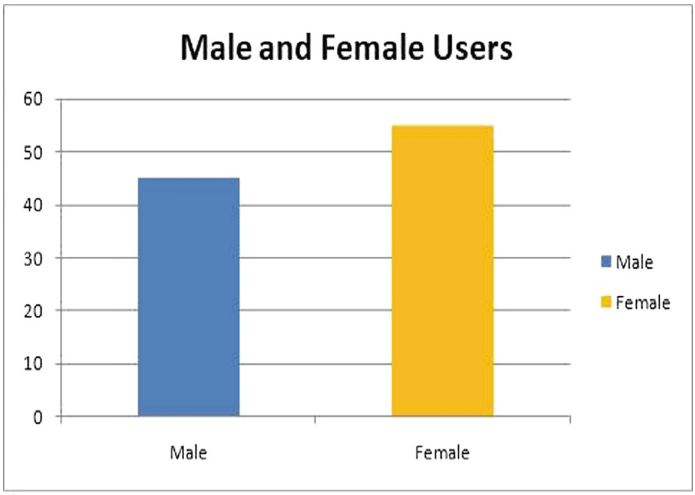


Fig. 1. Numbers of male and female.

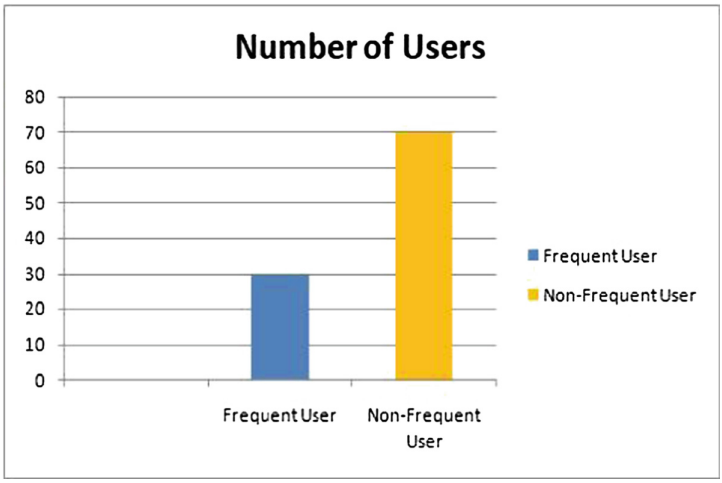


Fig. 2. Numbers of frequent and none frequent users.

Table 1. Devices used to access each personal assistant.

Virtual assistant	Devices
Google assistant	Android smart phone
Alexa	Echo Dot
Cortana	Windows 10 Laptop
Siri	IPhone 6s

5 Results

5.1 Evaluation Results

The aim of this task was to test voice recognition and contextual understanding between users and human free interaction. From above graph, it can be said that question based on travel i.e. "How do I get to the Mumbai Airport" and "Book me a flight to Pune", Google assistant answered more accurately than other virtual assistants. In the next question which was "Call me an Uber" winners were Alexa and Siri. Next three questions asked were based on email functioning, those were "Send an email to Amrita"; "Do I have any email from Amrita" and "Send a text to Amrita". All three assistants responded to these questions but Google assistant answered more accurately. Next question was to message someone using Whatsapp in which only Siri was able to execute. Google assistant did open Whatsapp but was unable to perform the task. Since, there is no user interface in Alexa; no such response could be expected.

The question related to sports was answered by all the assistants except Cortana which opened various links. When the questions related to music category were asked all of them answered correctly. On the question related to weather all the IVAs were able to answer except Alexa. On the question "Will I need umbrella tomorrow?" Cortana and Google Assistant were the winners. When asked to schedule dinner with

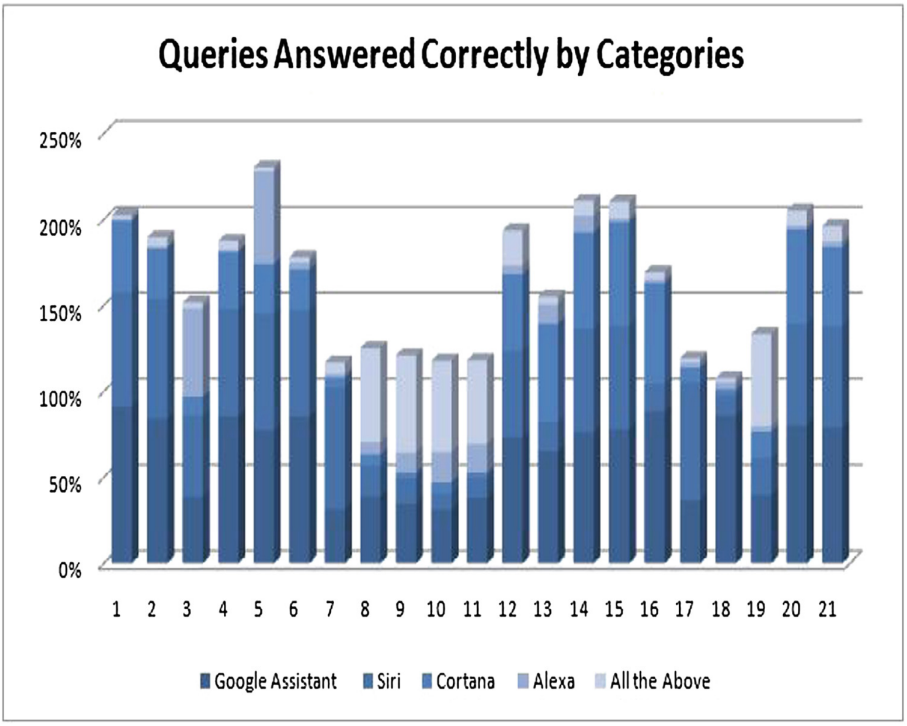


Fig. 3. Queries answered correctly by categories

my Dad, Google assistant, Siri and Cortana did setup reminder on the device. Google assistant successfully answered question related to navigation like searching south Indian restaurants near me whereas only siri was able to reserve a table at Hotel Taj. Google assistant not only answered but also provided accurate pronunciation when asked "How do I say where are my clothes in Gujarati?" When asked tricky questions like Do you know siri? and which virtual assistant is best? all the assistants except Alexa did answer however, sarcastically.

With this the survey came to an end. Here, researcher reached to conclude the survey. According to this results Google assistant's answers were more accurate then all three. Google assistant easily understood voice variations. It also avoids all disturbances and managed to listen the voice. Now, Table 2 has 21 questions and it shows results which are based on three categories Best, Better and Good. Figure 3 shows Queries answered correctly by categories. The bar chart was made based on the 21 questions asked to the users. From Fig. 3 every question was evaluated. After analyzing each and every question results are shown in above paragraph. From that best virtual assistant has been evaluated. After analyzing Google assistant was found to be technically very smooth to use and users didn't need to repeat the statements. Following figure shows the graph bar based on queries correctly answered by the categories.

Table 2. Best, better and good personal assistants by feature. Abbreviations: GA = Google Assistant, S = Siri, C = Cortana A = Alexa, AL = All, N/A = Not applicable

Question asked	Virtual assistants and human free interaction		
Range	Best (100% > 60%)	Better (60% > 25%)	Good (25% ≥ 0%)
Q.1 Asked "How do I get to the Mumbai Airport?"	**GA, S**	**C**	**A, AL**
Q.2 Asked "Book me a flight to Pune"	**GA, S**	**C**	**A, AL**
Q.3 Asked "Call me an Uber"	**A**	**S**	**GA, C, AL**
Q.4 Asked "Send an email to Amrita"	**GA**	**S, C**	**A, AL**
Q.5 "Do I have any email from Amrita"	**GA**	**S, C**	**A, AL**
Q.6 "Send a text to Amrita"	**GA**	**S, C**	**A, AL**
Q.7 "Do I have any new text from Whatsapp?"	**S**	**GA, C**	**A, AL**
Q.8 Who is the best player in IPL 2015?	**N/A**	**GA, S, AL**	**C, A**
Q.9 "Last IPL winner team?"	**GA**	**S, AL**	**C, A**
Q.10 "Play me some old music	**GA**	**C**	**S, C, AL**
Q.11 "Play my favourite song on youtube?"	**GA**	**C**	**S, C, AL**
Q.12 What's the weather is going to be like day after tomorrow?"	**GA**	**S, C, AL**	**A**
Q.13 "Will I need umbrella in next two months?"	**C**	**N/A**	**GA, S, A, AL**

(*continued*)

Table 2. (*continued*)

Question asked	Virtual assistants and human free interaction		
Range	Best (100% > 60%)	Better (60% > 25%)	Good (25% ≥ 0%)
Q.14 "Do I have anything scheduled for coming week?"	**GA**	**C, AL**	**S, A**
Q.15 "Asked each assistant to" Scheduled dinner with Dad on my calendar Sunday night,"	**GA**	**C**	**S, A, AL**
Q.16 What are some South Indian restaurants near me?	**GA**	**C**	**S, A, AL**
Q.17 "I want to make a reservation at Taj hotel in Mumbai"	**S**	**GA**	**C, A, AL**
Q.18 "How do I say 'where are my clothes' in Gujrati?"	**GA**	**N/A**	**S, A, C, AL**
Q.19 Who is the current Prime Minister of India?"	**GA**	**S, C, AL**	**A**
Q.20 Do you know Siri?"	**GA, C**	**N/A**	**S, A, AL**
Q.21 "What do you think who is the best as a virtual assistant?"	**N/A**	**GA, C, S**	**A, AL**

5.2 Analysis on Results

Following Fig. 4 shows overall percentage of quality of Voice recognition and human free interaction exhibited by all four IVAs. Google assistant excelled in VR and HFI by bagging 59.80% while Siri responded 43.98% accuracy in VR and HFI. There were very slight differences between Google Assistant's and Siri's accuracy but next assistant i.e. Cortana decreased gradually. Its accuracy responded up to 28.42%. And last assistant Alexa, responded up to 7.91% which shows very poor result as compared to all the others.

1. Voice Recognition and Contextual Understanding

For voice recognition, all needed to know was the way regularly the virtual assistant could perceive the words users were stating. I didn't consider setting or the estimation of the reaction I was given, simply essential acknowledgment. In this survey users tried voice recognition at different separations from the gadgets and also with changing levels of foundation noise. Google and Siri comprehended us well when according the graphs. There was a couple amusing misconceptions, similar to when users asked Siri, "Will I need umbrella for next two months?" And it gave me the date for one week later, on the grounds that it is unable to understand tricky questions. Google assistant is good at understanding natural language, Alexa is good at sound and music, Alexa is not comfortable with basic questions, Cortana is not that good in basic

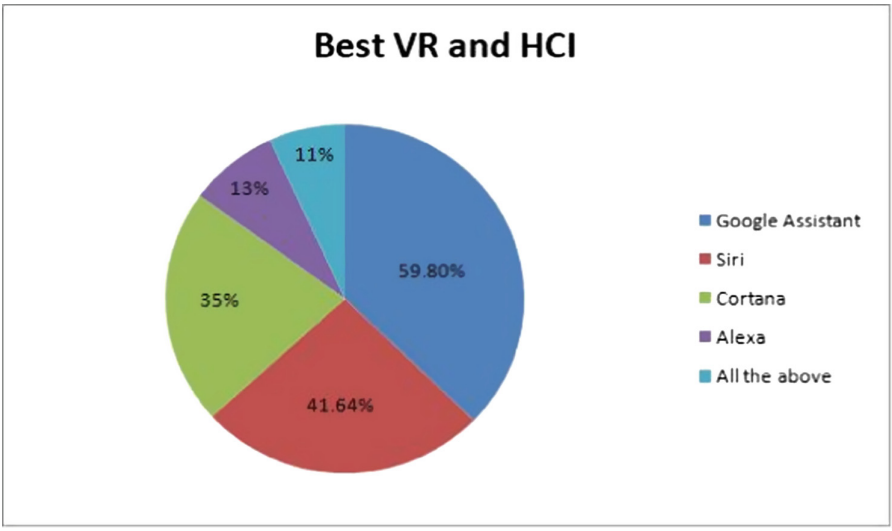


Fig. 4. It shows comparison between four assistants all users' select best virtual assistants on voice based recognition, contextual understanding and hand-free interaction.

voice recognition. Microsoft's assistant had issues understanding user voice even with zero noise interference.

2. Human Free Interaction

This was the first issue our survey found that for rare user, hands free connection was the real utilize case. Our members featured that, in this specific situation, interrupt ability was a huge hindrance. They were baffled when they were requested that by the intelligent virtual assistant draw in outwardly with the screen, or to affirm or select alternatives by taping the touch screen as opposed to through utilizing discourse. This appeared to fundamentally interfere with the without hands involvement of IVAs and was considered especially tricky in circumstances, for example, driving. We propose that keeping up discourse as the fundamental information and yield all through connection should be a need in future plan of IPAs with a specific end goal to guarantee that hands free association is upheld completely and that undertaking are not hindered by a collaboration methodology move.

6 Discussion and Future Work

Virtual assistants are growing to be prevalent in our day to day lives. Due to good outreach of smart phones many of us own at least one IVA may it be Siri on iPhone or Google Assistant on Android phones. Cortana also has a good reach due to wide users of Windows 10 and Alexa as a home speaker. All of the 100 users were provided with the same set of questions which they were supposed to ask to all the four personal assistants. Therefore the IVAs were rated on how many questions they attempted to

answer and how many they got right. From whole scenario it nearly came to conclusion that recognizing voice required number of large different variations such as environment, voice modulation, frequency etc. Main challenge for voice recognition is that people's voices vary and they speak in different ways and in different number of languages.

All the IVAs are evolving over time. While all the four IVAs were capable of answering nearly 17.35% of daily based questions, only Google Assistant was found to be most efficient by answering 59.80% of the question. Contrary to the expectations, Siri brought a bit of disappointment by answering only 45%. Whereas Cortana could only cover one third of the questions. Alexa loss the race badly by only answering 7.91%. That was surely a terrible defeat. According to our survey, IVAs have quite a poor retention rate, with only 25% of frequent day to day life users. Alexa, Google Assistant, Siri, and Cortana are at good pace for betterment over years. It's quite likely that one day they will actually meet our expectations.

IVA's are of broader benefits for future then they appeared to be. There is a large portion of people from disabled community mostly with cognitive disabilities who may have hard time in forming full sentences and communication, for such people personal assistant can be a deal breaker. Though there is the place for enhancement in all tested devices new studies can be conducted and the potential of IVAs can be tested in various untouched areas like education, banking, business, counseling, sales etc. And also a merger of these devices with various machine learning technologies and algorithms may give birth to different new possibilities.

References

1. Gong, L.: San Francisco, CA (US) United States US 2003.01671.67A1 (12) Patent Application Publication c (10) Pub. No.: US 2003/0167167 A1 Gong (43) Pub. Date: 4 September 2003 for Intelligent Virtual Assistant
2. Sarikaya, R.: The technology behind personal digital assistants. IEEE Signal Process. Mag. **34**, 67–81 (2017). https://doi.org/10.1109/msp.2016.2617341
3. Tsiao, J.C.-S., Tong, P.P., Chao, D.Y.: Natural-Language Voice-Activated Personal Assistant, United States Patent (10), Patent No.: US 7,216,080 B2 (45), 8 May 2007
4. Sirbi, K., Patankar, A.J.: Personal assistant with voice recognition intelligence. Int. J. Eng. Res. Technol. **10**(1), 416–419 (2017). ISSN 0974-3154
5. Kawamura, T., Ohsuga, A.: Flower voice: virtual assistant for open data
6. Elshafei, M.: Virtual personal assistant (VPA) for mobile users. Mitel Networks (2000–2002)
7. Chung, H., Iorga, M., Voas, J., Lee, S.: Alexa, can I trust you? In: 2017 IEEE Computer Security (2017)
8. Cowan, B.R.: What can i help you with?: infrequent users' experiences of intelligent personal assistants. In: 2015 IEEE 10th International Conference on Industrial and Information Systems, ICIIS 2015, Sri Lanka (2015)
9. Weeratunga, A.M., Jayawardana, S.A.U., Hasindu, P.M.A.K, Prashan, W.P.M., Thelijjagoda, S.: Project Nethra - an intelligent assistant for the visually disabled to interact with internet services. In: 2015 IEEE 10th International Conference on Industrial and Information System (2015)

10. López, G., Quesada, L., Guerrero, L.A.: Alexa vs. siri vs. cortana vs. Google assistant: a comparison of speech-based natural user interfaces. In: Nunes, I. (ed.) AHFE 2017. AISC, vol. 592, pp. 241–250. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-60366-7_23
11. Zhao, Y., Li, J., Zhang, S., Chen, L., Gong, Y.: Domain and speaker adaptation for Cortana speech recognition. In: ICASSP
12. Bellegarda, J.R.: Spoken language understanding for natural interaction: the siri experience. In: Mariani, J., Rosset, S., Garnier-Rizet, M., Devillers, L. (eds.) Natural Interaction with Robots, Knowbots and Smartphones, pp. 3–14. Springer, New York (2014). https://doi.org/10.1007/978-1-4614-8280-2_1
13. Google: Google Assistant. https://assisatnt.google.com
14. Purington, A., Taft, J.G., Sannon, S., Bazarova, N.N., Taylor, S.H.: Alexa is my new BFF: social roles, user satisfaction, and personification of the amazon echo. ACM, 6–11 May 2017. ISBN 978-1-4503-4656-6/17/05
15. Lopez, G., Quesada, L., Guerrero, L.A.: Alexa vs Siri vs Cortana vs Google assistant: a comparison of speech-based natural user interfaces. Conference Paper, January 2018
16. Kepuska, V., Bohouta, G.: Next generation of virtual personal assistants (Microsoft Cortana, Apple Siri, Amazon Alexa and Google Home). In: IEEE Conference (2018)

Signal and Image Processing

Pre-processed Hyperspectral Image Analysis Using Tensor Decomposition Techniques

R. K. Renu(✉), V. Sowmya, and K. P. Soman

Center for Computational Engineering and Networking (CEN),
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
mail2renurk@gmail.com

Abstract. Hyperspectral remote sensing image analysis has always been a challenging task and hence there are several techniques employed for exploring the images. Recent approaches include visualizing hyperspectral images as third order tensors and processing using various tensor decomposition methods. This paper focuses on behavioural analysis of hyperspectral images processed with various decompositions. The experiments includes processing raw hyperspectral image and pre-processed hyperspectral image with tensor decomposition methods such as, Multilinear Singular Value Decomposition and Low Multilinear Rank Approximation technique. The results are projected based on relative reconstruction error, classification and pixel reflectance spectrums. The analysis provides correlated experimental results, which emphasizes the need of pre-processing for hyperspectral images and the trend followed by the tensor decomposition methods.

Keywords: Remote sensing image · Tensor decomposition
Multilinear Singular Value Decomposition
Low Multilinear Rank Approximation · Relative reconstruction error
Pixel reflectance spectrums

1 Introduction

Hyperspectral images (HSI) are acquired using air-borne or space-borne hyperspectral sensors [13] such as NASA's AVIRIS (Airborne Visible/Infrared Imaging Spectrometer) and Hyperion Imaging Spectrometer. Hyperspectral sensors are capable of capturing a pixel at hundreds of contiguous wavelengths and forms a large image cube of reflectance values. The study has revealed that the hyperspectral data is spatial and spectral correlated [13] and the need of spatio-spectral processing for HSI [8]. There are various feasible approaches in literature for spatio-spectral HSI processing and one among them is processing HSI as a third order tensor.

Spectral-spatial dimensionality reduction and HSI classification is based on tensor modelling is proposed by Bourennane et al. where he used Adaptive Multidimensional Weiner filtering (AMWF), a denoising multidimensional filter [3].

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 205–216, 2019.
https://doi.org/10.1007/978-981-13-5758-9_18

Zhang et al. proposed Higher Order Singular Value Decomposition (HOSVD) as an effective compression algorithm for hyperspectral images [13]. Zhang used Khatri-Rao product for reconstruction of the decomposed tensor to reconstruct them back to the original size. [12] projected Additive Morphological Decomposition (AMD) algorithm for HSI and the comparison of classical Principal Component Analysis (PCA) with Tensor-Principal Component Analysis (TCA) based on HSI classification.

Lower Rank Tensor Approximation (LRTA) has been used by Renard et al. [9] for efficient denoising of spatial data and simultaneously reducing the spectral dimension with low rank approximations. Rank-1 tensor decomposition technique is employed by Guo et al. as a denoising tool for HSI. Most of the tensor decomposition algorithms, encompasses that these algorithms can be used as denoising and dimensionality reduction/compression of HSI.

In this paper, HSI is modelled as a third order tensor and behaviour of tensor decomposition methods such as, Multilinear Singular Value Decomposition (MLSVD) and Low Multilinear Rank Approximation (LMLRA) on HSI is analyzed. The tensor decompositions is applied for raw hyperspectral data (without any pre-processing), data normalized hyperspectral data and Least Square Fast Fourier Transform (LS-FFT), a spectral denoising algorithm proposed in a prior work by Chippy et al. [4]. The experimental analysis is projected based on relative reconstruction error obtained after decomposition and reconstruction of HSI, Support Vector Machine (SVM) classification results and pixel reflectance spectrums.

The following section details about the pre-processing techniques employed and the tensor decomposition methods used. The experimental flow of techniques is explained in the Sect. 3. Section 3 also exhibits the experimental results and the analysis derived from the experimental results obtained. The section includes dataset briefing, fixing of compression size for tensor decompositions based on relative reconstruction error and the results obtained from SVM classification and the pixel reflectance spectrums for hyperspectral data with and without pre-processing. Section 4 concludes with the summary of analysis derived from the experimental results and discusses about the future work possible in the discussed research area.

2 HSI Pre-processing Techniques and Tensor Decomposition Methods

2.1 HSI Pre-processing Techniques

The pre-processing of HSI is always an active research area. Data normalization and LS-FFT denoising technique are chosen as two pre-processing techniques for the scope of this paper.

Data Normalization. HSI data is normalized with respect to spectral dimension. Data is normalized to the range of minimum reflectance value and maximum

reflectance value captured for that pixel. A vector "x" can be normalized to its min-max range using,

$$x'_i = ((x_i - min(x))/(max(x) - min(x))) * ((max(x) - min(x)) + min(x)) \quad (1)$$

where, x'_i is the normalized vector of x and i ranges to the length of the vector.

LS-FFT Denoising. Least Square based denoising technique for one-dimensional signals [10] is proposed by Ivan W Selesnick as,

$$\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{x}||_2^2 + \lambda ||D\mathbf{x}||_2^2 \quad (2)$$

where '$\mathbf{y}$' is the input noisy signal, '$\mathbf{x}$' is the unknown denoised signal and 'D' is the second derivative coefficient matrix; λ is the control parameter, which balances the importance of removing noise with retaining the information in the Eq. (2). The solution of the optimization problem mentioned in the Eq. (2) is derived as,

$$\mathbf{x} = (I + \lambda D^T D)^{-1}\mathbf{y} \quad (3)$$

Chippy et al. introduced a variant approach to the Ivan W Selesnick's optimization problem (Eq. 2) to avoid the complex inverse calculation as in the Eq. 3. Chippy adopted denoising in the frequency domain [4] by rewriting the Eq. 2 as,

$$\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{x}||_2^2 + \lambda ||\mathbf{s} \otimes \mathbf{x}||_2^2 \quad (4)$$

where, $\mathbf{s} \otimes \mathbf{x}$ represents the convolution of $\mathbf{s}$ and $\mathbf{x}$. Solving the convolution in frequency domain becomes multiplication.
For LS-FFT of matrix, the optimization problem can be formulated as,

$$\min_{\mathbf{X}} ||\mathbf{Y} - \mathbf{X}||_2^2 + \lambda ||\mathbf{SX}||_2^2 \quad (5)$$

where, the discrete Fourier transform of $\mathbf{y}$, $\mathbf{x}$, and $\mathbf{s}$ are represented as '$\mathbf{Y}$', '$\mathbf{X}$' and '$\mathbf{S}$' respectively. The solution is the Inverse Discrete Fourier Transform of the Eq. 5 is,

$$\mathrm{X_k} = \mathrm{Y_k}/(1+\lambda|S_k|^2) \quad (6)$$

where k varies from 1 to length of the signal.

2.2 Tensor Decomposition Methods

Multilinear Singular Value Decomposition (MLSVD). Multilinear Singular Value decomposition is a speculation of Singular Value Decomposition for higher dimensional datasets [5]. A third order tensor of size $(I_1 \times I_2 \times I_3)$ is decomposed into a set of core tensor and three orthogonal factor matrices. A pixel at $i1, i2, i3$ location of the tensor 'a' can be represented as [5],

$$a_{i_1 i_2 i_3} = \sum_{j_1=1}^{I_1} \sum_{j_2=1}^{I_2} \sum_{j_3=1}^{I_3} s_{j_1 j_2 j_3} u^{(1)}_{i_1 j_1} u^{(2)}_{i_2 j_2} u^{(3)}_{i_3 j_3} \quad (7)$$

where, $u^{(1)}, u^{(2)}, u^{(3)}$ are the orthogonal factor matrices along the three dimensions/modes and s is the core tensor. For a tensor 'T', MLSVD can be defined as shown in Fig. 1

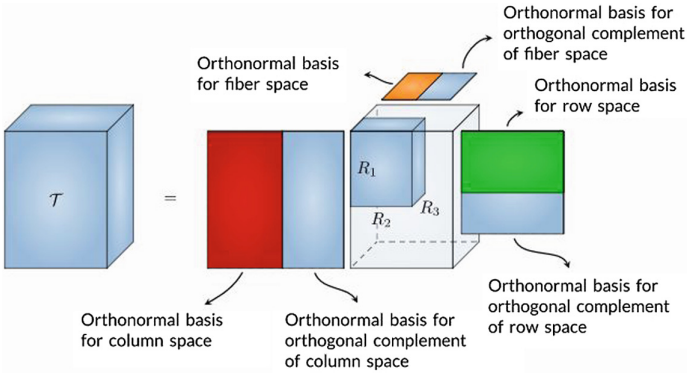


Fig. 1. Multilinear Singular Value Decomposition (Image Courtesy: [2])

and can be written as,

$$T = S._1U^{(1)}._2U^{(2)}._3U^{(3)} \tag{8}$$

where S is the core tensor and $U^{(1)}$, $U^{(2)}$, $U^{(3)}$ are the orthonormal bases for three different subspaces of third order tensor. The $._1, ._2, ._3$ or $\times_1, \times_2, \times_3$ representation mentions the first order, second order and third order (n-order) tensor products respectively [7].

Low Multilinear Rank Approximation (LMLRA). Low Multilinear Rank Approximation is comparable with Multilinear Singular Value Decomposition and is different with the way of computation and the optimality of the techniques [1]. LMLRA is computed in two stages, processing the underlying theory and refining the registered yield from the initial step utilizing the mentioned algorithm. The approximation technique tries to limit the frobenius error in each iteration. [6] The low rank approximation techniques Higher Order Orthogonal I and trust-region-based algorithm is discussed by Ishteva and reasons that the trust-region-based algorithm focalizes to the arrangement speeder.

LMLRA Reconstruction Technique. The decomposed tensor can be reconstructed by using Low Multilinear Rank Approximation (LMLRA) reconstruction/regeneration technique. The technique follows a simple Khatri-Rao product (n-order tensor product) of core tensor and factor matrices with respect to its dimensions. The HSI tensor is reconstructed to its original size for comparative analysis.

2.3 Analysis Methods

- *Relative Reconstruction Error*: Relative reconstruction error is the frobenius norm of the difference in the reconstructed image from the original image to the frobenius norm of the original image [2] and is computed as,

$$RRE = \frac{||Original\ Image\ -\ Reconstructed\ Image||_F}{||Original\ Image||_F} \tag{9}$$

  where RRE denotes Relative Reconstruction Error; $||x||_F$ denotes the frobenius norm of x.
- *SVM Classification*: Hyperspectral data classification follows pixel-wise classification. Each pixel is mapped to a corresponding label. SVM is a linear binary classifier, that classifies two classes with respect to minimize the error function. SVM can be extended for multi-class classification as one-on-one or one-on-all methods.
- *Pixel Reflectance Spectrum*: Pixel reflectance spectrums are graphical plots of reflectance value plotted against the wavelength/band number of a pixel, which is the spectral signature of the pixel from a remote sensing perspective.

3 Experimental Procedure, Results and Observations

The raw hyperspectral image is compressed using MLSVD and reconstructed to its original size, whose relative reconstruction error (computed using Eq. 9) is noted. The reconstructed image is then analyzed based on the variation in SVM classification results, overall classification accuracy, class-wise accuracy and classification maps, and Pixel Reflectance Spectrums for a pixel chosen from the noisy part of the image. This flow of experiments is repeated for LMLRA compression also, and results are tabulated. As shown in the Fig. 2, the whole set of experiments for both MLSVD and LMLRA is repeated for normalized hyperspectral data and LS-FFT denoised hyperspectral data.

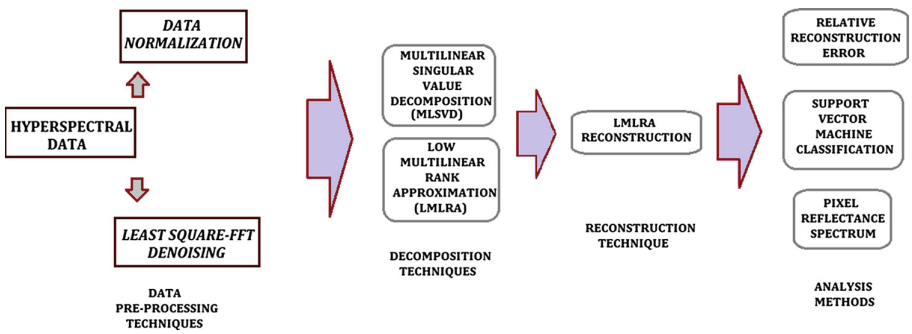


Fig. 2. Experimental flow - block diagram

3.1 Dataset/Hyperspectral Data

Experiments are carried out on the standard dataset of hyperspectral images: Indian Pines, Salinas Scene and Pavia University [4,11]. Indian Pines is captured by NASA AVIRIS sensor. Indian Pines was captured over the Indian Pines in North-Western Indiana with 145×145 pixels and 224 bands in the range of 400–2500 nm. Salinas Scene is also acquired by NASA AVIRIS sensor. Salinas Scene was captured over Salinas valley, California with a spatial resolution of 3.7 m pixels. The area covered consists of 512 scan-lines by 217 samples over 224 spectral bands. Pavia University was acquired by the ROSIS sensor over the region Pavia in Northern Italy consisting of 610×340 pixels (excluding the blacked out pixels) and 103 bands.

3.2 Tuning of Compression Size Based on Relative Reconstruction Error

The compression size tabulated in the Table 1, is fixed based on the graphical plot plotted for relative reconstruction error against the compression size ranging till the minimum size of the image in all dimensions. A graphical plot result is depicted in the Fig. 3 for the dataset Indian Pines, where the relative reconstruction error tends to become constant at the dimension 45 and hence the compression size is chosen as $45 \times 45 \times 45$ for the experiments performed.

The corresponding relative reconstruction errors for the dataset and tuned compression size and the rate of compression are also recorded in the Table 1.

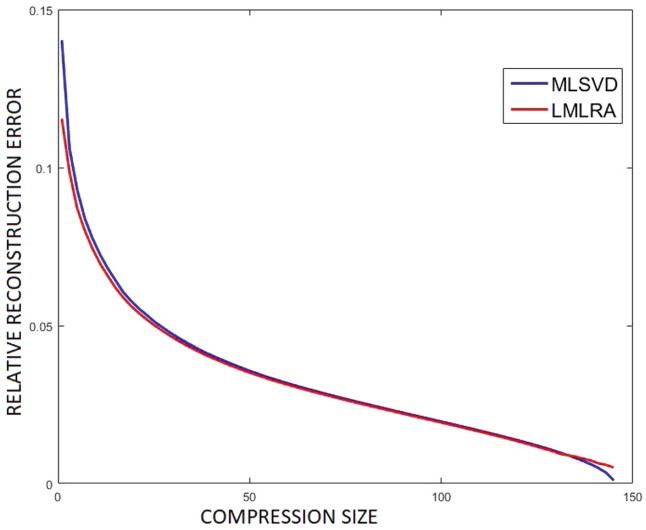


Fig. 3. Graphical plot for tuning of compression size for Indian Pines dataset

Table 1. Tuning results and corresponding compression rate

Dataset & size	Compression size & Rate	Decomposition methods	Relative Reconstruction Error		
			Without pre-processing	With data normalization	With LS-FFT denoising
Indian Pines ($145 \times 145 \times 220$)	$45 \times 45 \times 45$ {97.5%}	MLSVD	0.0380	**0.0380**	**0.0166**
		LMLRA	0.0378	**0.0378**	**0.0165**
Salinas Scene ($512 \times 217 \times 224$)	$75 \times 75 \times 75$ {98%}	MLSVD	0.0482	**0.0482**	**0.0497**
		LMLRA	0.0478	**0.0478**	**0.0493**
Pavia University ($610 \times 340 \times 103$)	$70 \times 70 \times 70$ {97.8%)	MLSVD	0.1792	**0.1792**	**0.1187**
		LMLRA	0.1771	**0.1771**	**0.1173**

The compression rate is computed as,

$$Compression\ Rate = (1 - \frac{No.\ of\ pixels\ of\ Compressed\ Image}{No.\ of\ pixels\ of\ Original\ Image}) \times 100 \quad (10)$$

From the Table 1, the relative reconstruction error for data without pre-processing and normalized data are more or less same while there is a huge variation in reconstruction error in case of LS-FFT denoised data.

3.3 SVM Classification Results

Classification of raw data gives dissatisfying results (refer Table 2), hence pre-processing the data becomes vital. As per the overall classification accuracy in

Table 2. SVM classification results

Dataset & size	Decomposition methods	Overall classification accuracy (%)			
		Original image	Reconstructed image		
			Without pre-processing	With data normalization	With LS-FFT denoising
Indian Pines ($145 \times 145 \times 220$)	MLSVD	77.33	80.96	**82.50**	**88.60**
	LMLRA		81.72	**82.46**	**88.14**
Salinas Scene ($512 \times 217 \times 224$)	MLSVD	86.74	93.86	**93.89**	**94.10**
	LMLRA		93.60	**93.61**	**94.65**
Pavia University ($610 \times 340 \times 103$)	MLSVD	91.15	94.45	**93.92**	**96.50**
	LMLRA		94.36	**94.41**	**96.32**

the Table 2, it can be noted that relative reconstruction error and Overall Classification Accuracy are inversely proportional which is the expected behaviour. The results also confirms with LS-FFT denoising as an efficient denoising technique and the compression algorithms as lossless. Since the decomposition and reconstruction has only aided in the removal of noise by adding smoothness and has not deprecated the information contained within the image.

Sample classification map shown in the Fig. 4 confirms the above derived observations. It can be noted that MLSVD and LMLRA approaches the image in two different manner, so that predictions in one class for MLSVD is better than LMLRA and vice-versa. The same can be noted in all the three datasets. The choice of decomposition algorithm wholly depends on the data and should be experimentally chosen. From the overall classification accuracy and classification maps, MLSVD provides better results for both Indian Pines and Pavia University, while LMLRA provides better results for Salinas Scene.

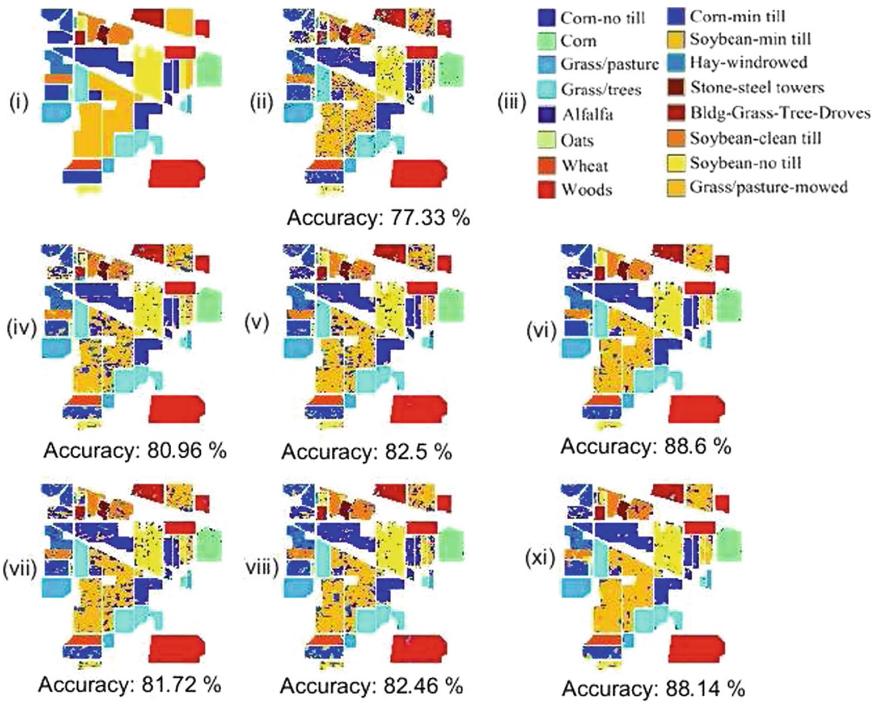


Fig. 4. Sample classification map (Dataset: Indian Pines): (i) Ground truth; (ii) Original image; (iii) Class labels; (iv) MLSVD without pre-processing; (v) MLSVD after data normalization; (vi) MLSVD after LS-FFT denoising; (vii) LMLRA without pre-processing; (viii) LMLRA after data normalization; (ix) LMLRA after LS-FFT denoising

Table 3. Class-wise accuracy for various processing - Indian Pines

Original image		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Class 12	Class 13	Class 14	Class 15	Class 16
		51.22	75.88	68.27	48.36	94.25	93.61	84	98.37	27.78	70.74	72.48	62.55	92.39	93.59	54.76	83.33
Without pre-processing	MLSVD	73.17	80.93	71.49	69.95	94.71	99.09	80	99.53	88.89	77.94	67.90	73.03	100	92.78	71.76	89.29
	LMLRA	75.61	77.04	74.56	74.65	89.66	99.85	80	97.21	50	70.51	75.92	72.66	100	97.45	73.78	90.48
With data normalization	MLSVD	**58.54**	**82.49**	**73.49**	**81.22**	**91.49**	**97.87**	**100**	**98.84**	**72.22**	**73.26**	**73.43**	**76.59**	**98.37**	**95.61**	**81.84**	**88.10**
	LMLRA	**85.37**	**78.99**	**75.77**	**70.89**	**92.64**	**99.54**	**80**	**99.30**	**83.33**	**74.40**	**75.46**	**75.84**	**94.02**	**94.64**	**76.37**	**96.43**
With LS-FFT denoising	MLSVD	**95.12**	**85.53**	**88.62**	**80.28**	**98.16**	**99.39**	**84**	**98.60**	**33.33**	**80.34**	**83.34**	**82.40**	**97.83**	**96.60**	**93.08**	**97.62**
	LMLRA	**85.37**	**88.48**	**80.05**	**86.38**	**92.41**	**99.09**	**76**	**99.07**	**61.11**	**76.23**	**86.60**	**79.21**	**100**	**96.49**	**85.88**	**98.81**

Table 4. Class-wise accuracy for various processing - Salinas Scene

Original image		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9	Class 10	Class 11	Class 12	Class 13	Class 14	Class 15	Class 16
		99.50	99.94	99.44	98.96	99.34	99.94	99.88	66.48	99.80	97.93	98.75	99.65	99.27	98.23	56.08	99.20
Without pre-processing	MLSVD	99.45	99.79	98.76	99.28	98.34	99.94	99.53	87.84	99.66	97.66	97.92	99.71	97.82	97.61	77.45	98.77
	LMLRA	99.67	99.73	99.55	97.77	99.38	99.94	99.81	90.83	99.64	97.86	96.77	99.31	99.27	94.39	70.94	98.46
With data normalization	MLSVD	**99.78**	**99.76**	**99.44**	**97.93**	**98.38**	**99.92**	**99.72**	**91.06**	**99.98**	**98.92**	**97.29**	**99.48**	**99.03**	**97.61**	**68.66**	**99.26**
	LMLRA	**99.67**	**99.91**	**99.21**	**98.09**	**97.22**	**98.96**	**99.75**	**90.87**	**99.73**	**98.17**	**96.98**	**99.65**	**99.27**	**99.07**	**70.46**	**98.40**
With LS-FFT denoising	MLSVD	**100**	**99.94**	**100**	**98.17**	**99.25**	**99.94**	**99.32**	**91.62**	**99.93**	**99.42**	**98.34**	**100**	**99.15**	**97.09**	**70.02**	**98.83**
	LMLRA	**99.94**	**99.88**	**99.89**	**99.76**	**99.34**	**99.94**	**99.44**	**91.01**	**99.93**	**98.03**	**98.02**	**99.88**	**99.15**	**97.72**	**76.78**	**99.02**

Table 5. Class-wise accuracy for various processing - Pavia University

Original image		Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Class 7	Class 8	Class 9
		90.90	95.50	75.91	91.77	99.26	87.58	80.87	81.17	99.88
Without pre-processing	MLSVD	93.72	98.59	83.91	94.49	99.34	95.82	91.65	77.13	98.83
	LMLRA	93.57	98.77	80.78	93.65	99.5	96.71	86.63	78.21	99.41
With data normalization	MLSVD	**93.21**	**98.34**	**81.63**	**93.33**	**99.92**	**97.24**	**86.72**	**75.14**	**98.36**
	LMLRA	**93.21**	**98.47**	**84.22**	**94.56**	**99.50**	**96.40**	**94.24**	**76.01**	**99.06**
With LS-FFT denoising	**MLSVD**	**95.73**	**99.21**	**90.21**	**95.40**	**99.50**	**98.67**	**95.57**	**84.46**	**98.83**
	LMLRA	**96.97**	**98.96**	**87.45**	**95.72**	**99.92**	**97.70**	**93.90**	**84.25**	**99.65**

For deep analysis, class-wise accuracy is also considered and tabulated for all the three datasets Indian Pines, Salinas Scene and Pavia University in the Tables 3, 4 and 5. Except for very few classes, the accuracy has increased prominently for most of the classes, assisting the results derived from the classification maps and overall classification accuracy.

3.4 Pixel Reflectance Spectrum

Pixel Reflectance Spectrum is the spectral signature of a pixel. Reflectance spectrums helps us to understand the pixel level information and variation along the wavelength spectrum. Figures 5, 6 and 7 represents the reflectance spectrum of a pixel from the dataset Indian Pines, Salinas Scene and Pavia University. It can be observed that the data without pre-processing and data normalized MLSVD

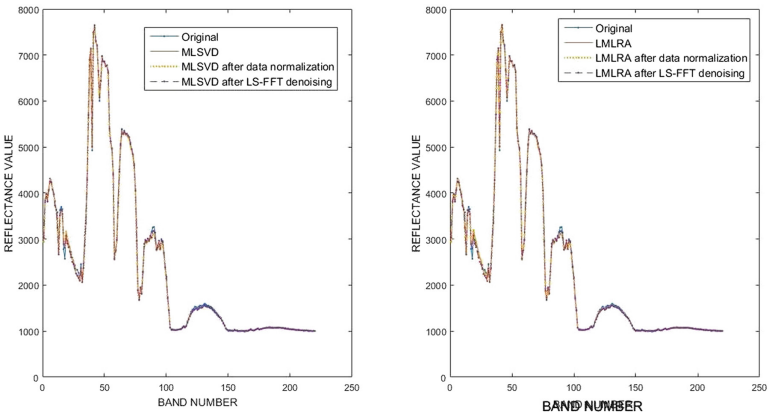


Fig. 5. Pixel reflectance spectrum - Indian Pines (Left: MLSVD, Right: LMLRA)

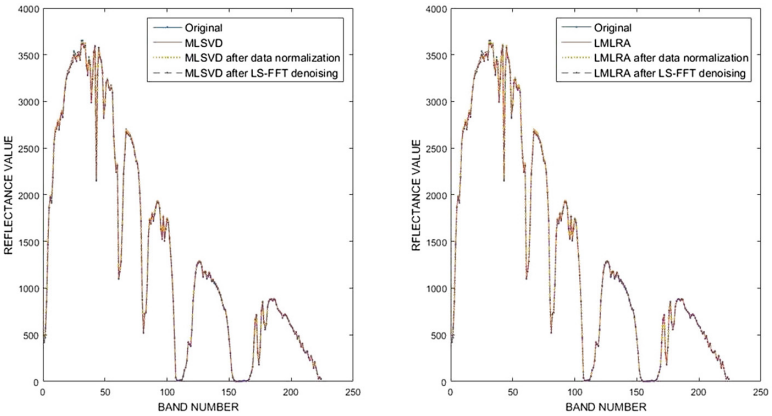


Fig. 6. Pixel reflectance spectrum - Salinas Scene (Left: MLSVD, Right: LMLRA)

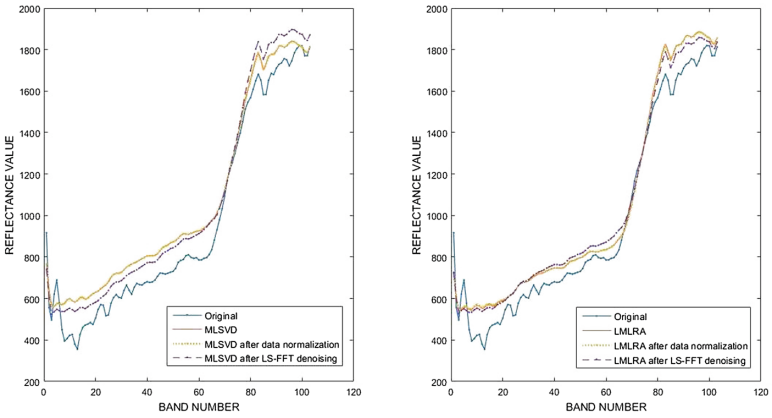


Fig. 7. Pixel reflectance spectrum - Pavia University (Left: MLSVD, Right: LMLRA)

and LMLRA are similar and not differential but varies from the original pixel information denoting the smoothening effect of decomposition algorithm. Spectrum with LS-FFT denoised is prominent and vary from the other reflectance spectrums by the smoothening it provides at highly variation reflection values. This extra smoothening is the denoising effect and reasons the improvement in the classification results and the relative reconstruction error.

4 Conclusion

The paper aims on the comparative study of behaviour analysis of tensor decomposition methods MLSVD and LMLRA with the same reconstruction technique on the raw, data normalized and LS-FFT denoised hyperspectral data. From the experimental results and observations, it can be concluded that the choice of tensor decomposition based on efficiency depends on the data. In terms of computational time, MLSVD is 98% faster than LMLRA. The analysis of other tensor decomposition methods and including other analyzing methods or extending the techniques to other three dimensional data can be considered as the future scope of this paper.

References

1. Multilinear singular value decomposition and low multilinear rank approximation. https://www.tensorlab.net/doc/lmlra.html
2. Multilinear singular value decomposition and low multilinear rank approximation tensorlab demos 3.0 documentation. https://www.tensorlab.net/demos/mlsvd.html
3. Bourennane, S., Fossati, C., Cailly, A.: Improvement of classification for hyperspectral images based on tensor modeling. IEEE Geosci. Remote Sens. Lett. **7**(4), 801–805 (2010)

4. Chippy, J., Jacob, N.V., Renu, R., Sowmya, V., Soman, K.: Least square denoising in spectral domain for hyperspectral images. Procedia Comput. Sci. **115**, 399–406 (2017)
5. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. SIAM J. Matrix Anal. Appl. **21**(4), 1253–1278 (2000)
6. Ishteva, M., Absil, P.A., Van Huffel, S., De Lathauwer, L.: Best low multilinear rank approximation of higher-order tensors, based on the Riemannian trust-region scheme. SIAM J. Matrix Anal. Appl. **32**(1), 115–135 (2011)
7. Itskov, M.: Tensor Algebra and Tensor Analysis for Engineers. Springer, Heidelberg (2007)
8. Patra, S., Bhardwaj, K., Bruzzone, L.: A spectral-spatial multicriteria active learning technique for hyperspectral image classification. IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens. **10**(12), 5213–5227 (2017)
9. Renard, N., Bourennane, S., Blanc-Talon, J.: Denoising and dimensionality reduction using multilinear tools for hyperspectral images. IEEE Geosci. Remote Sens. Lett. **5**(2), 138–142 (2008)
10. Selesnick, I.: Least squares with examples in signal processing. In: Connexions (2013)
11. Srivatsa, S., Sowmya, V., Soman, K.P.: Least square based fast denoising approach to hyperspectral imagery. In: Proceedings of International Conference on Advanced Computing, Networking and Informatics (ICACNI), pp. 22–24. Springer, Heidelberg (2016)
12. Velasco-Forero, S., Angulo, J.: Classification of hyperspectral images by tensor modeling and additive morphological decomposition. Pattern Recogn. **46**(2), 566–577 (2013)
13. Zhang, L., Zhang, L., Tao, D., Huang, X., Du, B.: Compression of hyperspectral remote sensing images by tensor approach. Neurocomputing **147**, 358–363 (2015)

Recovery of a Failed Antenna Element Using Genetic Algorithm and Particle Swarm Optimization for MELISSA

Shweta Vincent[1(✉)], Sharmila Anand John Francis[2], Om Prakash Kumar[3], and Kumudha Raimond[4]

[1] Department of Mechatronics Engineering, Manipal Institute of Technology, MAHE, Karunya, Coimbatore, India
shwetavincent@gmail.com
[2] Department of Computer Science Engineering, King Khalid University, Abha, Saudi Arabia
[3] Department of Electronics and Communication Engineering, Manipal Institute of Technology, MAHE, Coimbatore, India
[4] Department of Computer Science Engineering, Karunya, Coimbatore, India

Abstract. A 2 × 6 planar coaxial cavity horn antenna array has been proposed for the transmitter module of the MELISSA GB-SAR system [7]. This system is installed in Italy for monitoring of land deformations leading to landslides and is required to work round-the-clock for continuous monitoring. Failure of even a single antenna element in the transmitting or receiving module of this system could alter the radiation pattern of the system and could prove to be hazardous. This article performs a comparative analysis of the Genetic algorithm and Particle Swarm Optimization algorithm to recover the failed element in the 2 × 6 antenna array. The results of MatLab simulation prove that both the GA and PSO algorithms converge well to auto-recover the failed element.

Keywords: MELISSA · GB-SAR · Genetic algorithm
Particle swarm optimization · Dolph Chebyshev beamforming

1 Introduction

A GB-SAR (Ground based Synthetic Aperture Radar) system named MELISSA (MIMO Enhanced Linear Short Synthetic Aperture Radar) [1] is being used in Italy for monitoring of landslides and the sinking of the Costa Concordia cruise ship off the coast of France [2]. 12 pyramidal horn antennas placed in a linear geometry are used as the transmitter module of this system. At the receiver's end, 12 Vivaldi antennas placed in a linear geometry operate at a frequency of 12 GHz to image opposite hills in a continuous mode of operation. Figure 1 illustrates the MELISSA GB-SAR system.

This architecture of the MELISSA GB-SAR system has been altered in [3] to yield a 2 × 6 planar antenna array geometry giving a better directivity and lower Peak Side Lobe Ratio (PSLR). This leads to a better quality of the captured image. The altered geometry of the MELISSA's transmitter section is illustrate in Fig. 2.

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 217–228, 2019.
https://doi.org/10.1007/978-981-13-5758-9_19

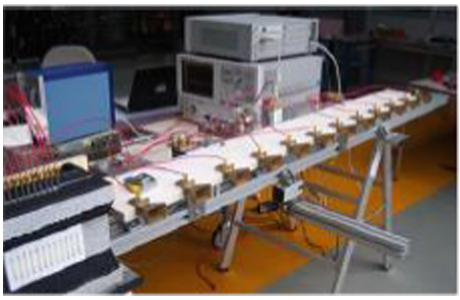


Fig. 1. MELISSA GB-SAR system (Adapted from [1])

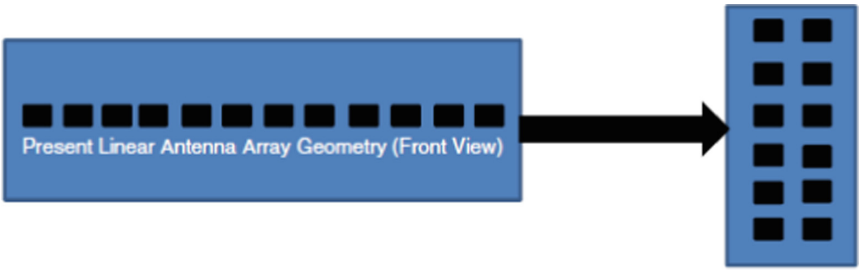


Fig. 2. 2 × 6 planar antenna array geometry for MELISSA's transmitter section (Adapted from [3])

The pyramidal horn of the transmitter has been replaced by the coaxial cavity horn antenna and a 2 × 6 planar geometry of coaxial cavity horns yields the best value of half power beam width (HPBW) without compromising on the directivity [4, 5]. The coaxial cavity horn antenna element is illustrated in Fig. 3.

Further Dolph Chebyshev beamforming has been used to further suppress the side lobes of the transmitter's radiation to improve the HPBW and in turn a lower PSLR to get a better image quality [6, 7].

Failure of one or many of the elements in an antenna array can lead to catastrophes especially in fields such as landslide detection or which require a continuous monitoring.

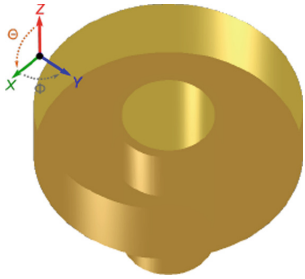


Fig. 3. Single coaxial cavity horn antenna element.

Machine Learning is an extremely useful technology now even being applied to the problem of correcting faulty elements in antenna arrays of such systems.

The advent of the technology of Machine Learning, has revolutionized the way in which systems can respond to certain inputs or even predict future outputs. Machine Learning is a branch of computer science using which computers can 'learn' without being specifically programmed. Machine Learning is diversely being applied in many areas such as medical diagnostics, data analytics, economics, pattern recognition etc. to name a few. The proposal for the use of Machine Learning algorithms for antenna design dates back to the 90 s.

The idea of the usage of genetic algorithms for antenna array design was first proposed by Haupt in [8]. Genetic algorithm has been proposed for array failure correction in digital beamforming arrays [9]. Three mating schemes (adjacent-fitness pairing, best-mate worst and emperor selective) have been proposed. The genetic algorithm designed in [9] enables the non-faulty elements of the antenna array to reconfigure their weights to compensate the loss of a faulty element or a number of faulty elements. A reduction in the side lobes of an antenna array's radiation pattern has been achieved using genetic algorithm in [10]. Decimal linear crossover is employed in order to generate the weights for the antenna elements so as to achieve a desired side lobe level.

Optimum side lobe reduction is also achieved in [11] by using yet another machine learning technique of Particle Swarm Optimization. The authors of this article have demonstrated the achievement of reduced side lobe levels with increased directivity useful for highly directional applications. An algorithm involving PSO with a feedback has been designed in [12]. This new algorithm has been made use of by the authors to synthesize an array pattern using dynamically varying inertia weights.

A comprehensive survey has also been performed in [13] which encompasses all the algorithms of evolutionary learning which have found applications in antenna design. A summary of case-by-case examples is also presented in this article which focuses more on genetic algorithm and particle swarm optimization.

Our article aims at performing a comparative study of both genetic algorithm and particle swarm optimization in the proposed 2×6 MELISSA transmitter module's architecture. A MatLab simulation study has been performed to recover the weight of a single faulty antenna element in the proposed 2×6 array such that the array factor and radiation pattern remain the same as the original array without fault.

Section 2 of this article describes the background theory of the already accomplished 2×6 coaxial cavity horn antenna design using Dolph Chebyshev beamforming. Section 3 of this article focuses on the results obtained by application of the genetic algorithm on the 2×6 array considering one element to be faulty. Particle swarm optimization is applied in Sect. 4 to the same problem discussed in the previous section. Section 5 concludes the article and describes the future work to be undertaken in this regard.

2 Background

The Array Factor (AF) is a constant which determines the radiation pattern characteristics of an antenna array. It is produced by calculating the field characteristics generated by a single element and multiplying it by the number of elements with

weights specified according to a weighting method. Equation (1) describes the AF for a planar $m \times n$ antenna array.

The array factor for a planar $m \times n$ antenna array is mathematically defined as;

$$AF = \sum_{b=0}^{m-1} \sum_{a=0}^{n-1} e^{-j\pi \sin\theta(a\cos\phi + b\sin\phi)} \tag{1}$$

Here, the angles θ and ϕ are the vertical and horizontal scan angles used to describe a point in 3D space in a radiation pattern.

The Dolph Chebyshev beamforming technique is used in [7] for the proposed 2×6 coaxial cavity horn array in MELISSA's transmitter, for suppressing the side lobes completely, thereby increasing the HPBW but decreasing the directivity. The tradeoff between the HPBW and directivity is termed as a Compensation Loss or Distribution Loss (Recommendation ITU-R BT. 1195-1). The amount of Compensation Loss which is permissible varies from one system to another and should be carefully planned while using the Dolph Chebyshev beamforming technique.

Equation (2) shows the AF generated for the proposed 2×6 coaxial cavity horn antenna array. The variable a ranges from 0 to 5 corresponding to the 6 elements in the column of the 2×6 planar array and b ranges from 0 to 1 corresponding to the 2 elements in the row of the 2×6 planar array.

$$AF = \sum_{b=0}^{1} \sum_{a=0}^{5} e^{-j\pi \sin\theta(a\cos\phi + b\sin\phi)} \tag{2}$$

The corresponding weights for this 2×6 coaxial cavity horn antenna array are calculated as shown in Eq. (3) using the Dolph Chebyschev polynomials. The detailed steps of the calculation of the respective weights are presented in [7].

$$T_{m(z)} = \begin{cases} 1 \; for \; m = 0 \\ 1 \; for \; m = 1 \\ 2zT_{m-1}(z) - T_{m-2}(z) \; for \; m = 2, 3, \ldots \end{cases} \tag{3}$$

The final AF term for the 2×6 coaxial cavity horn array is expressed as shown in Eq. (4).

$$\begin{aligned} AF = {} & w_1 \cos g + w_2(4\cos^3 g - 3\cos g) + w_3(16\cos^5 g - 20\cos^3 g + 5\cos g) + w_4(64\cos^7 g - \\ & 112\cos^5 g + 56\cos^3 g - 7\cos g) + w_5(256\cos^9 g - 576\cos^7 g + 432\cos^5 g - 120\cos^3 g + 9\cos g) \\ & + w_6(1024\cos^{11} g - 2816\cos^9 g + 2816\cos^7 g - 1232\cos^5 g + 220\cos^3 g - 11\cos g) \end{aligned} \tag{4}$$

Here $w_1 \ldots w_6$ are the respective Dolph-Chebyshev weights of the antenna elements in the 2×6 array. The variable g is specified by the value (0.33 * 0.5) i.e. ($\lambda/2$ * cos ($\theta/2$)).

The normalized weights obtained for Eq. 4 are tabulated in Table 1. Figure 4 illustrates front view of the 2×6 coaxial cavity horn antenna and the corresponding positions of the weights allotted according to Table 1.

Table 1. Normalized Dolph Chebyschev weights of 2 × 6 planar coaxial cavity horn array

Element position	Weight number	Weight value
(0, 0), (0, 1)	w_1	0.00035
(1, 0), (1, 1)	w_2	0.0073
(2, 0), (2. 1)	w_3	0.249
(3, 0), (3, 1)	w_4	1
(4, 0), (4, 1)	w_5	0.579
(5, 0), (5, 1)	w_6	0.062

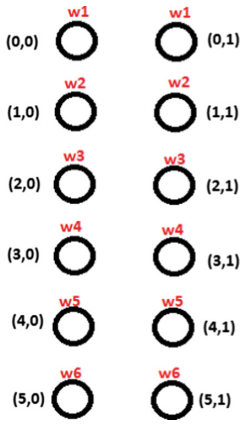


Fig. 4. Weight allocation for the 2 × 6 coaxial cavity horn antenna proposed for MELISSA [7]

In this paper, our aim is to test the application of the Genetic Algorithm and Particle Swarm Optimization algorithm to recover w_4, if there is an element failure and w_4 becomes 0. The next two sections describe the usage of the Genetic Algorithm and Particle Swarm Optimization algorithm to solve this problem.

3 Application of Genetic Algorithm to 2 × 6 Coaxial Cavity Horn Antenna Array for Fault Recovery of the Main Element

3.1 Preliminaries of Genetic Algorithm

Genetic Algorithm draws its inspiration from Charles Darwin's theory of evolution. According to this algorithm, a process of reproduction is carried out among the fittest of individuals in a population in order to bring forth the better (more fit) offspring. The section below describes in detail the terminologies used for the application of the Genetic Algorithm.

Individual: An individual in the Genetic Algorithm is analogous to one out of the many solutions to an optimization problem. In the case of antenna array design, an

individual signifies the weight allotted to the antenna array element which determines its amplitude and phase properties to achieve certain characteristic beam width and directivity.

Gene: A gene is similar to an individual in the form that it describes one of the possible solutions to a problem; but is not the actual solution. A sequence of genes forms a Chromosome, which is the genetic material that the algorithm uses to make predictions.

Population: A set of all possible individuals is termed as a Population. They form a set of all possible solutions to the problem. The 'initial population' and the 'population size' are two facets of the Population. The initial population determines the starting values of weights that are applied to the antenna array elements. The size of the population describes the size of the weight matrix applied on the antenna array. Ideally the initial population should have a large number of individuals in the gene pool to choose from for further calculation and prediction. Also, greater the population size, better is the chance of getting a good solution to the problem as there is a varied number of solutions to choose from.

Halt condition: The condition to halt the algorithm can be determined and set by various factors such as, the number of iterations being lapsed, the most feasible solution being found, all individuals of the population converging to the same solution etc. This is set by the designer before running the Genetic algorithm. Once the halting criterion is met, a new generation evolves and the old generation/population is discarded.

Fitness function/value: The fitness function/value is the most suitable solution to the algorithm to which the individuals in the population have to attain. The fitness value for the process is assigned by the algorithm designer.

Selection: This is the process of selecting the fittest two individuals to perform a crossover to bring forth children. Selection could be any one of the various types: Roulett wheel selection, random selection, rank selection or crossover selection.

Breeding/Crossover: Crossover is the process involving two fit individuals in the population, multiplying to give one or many children who are as fit or even more fit than their parents. This is the process in which the main goal of the Genetic algorithm is achieved. The process of breeding or crossover could be any one of the various types: single point crossover, two-point crossover, uniform crossover etc.

Mutation: After the process of crossover, a process of mutation is performed on the children in order to prevent the solution to be locked in a local minimum. The aim of the algorithm is to attain the global minimum and this is driven by the process of mutation.

3.2 Results for Single Element Failure in 2 × 6 Coaxial Cavity Horn Antenna Array

The central element of the 2 × 6 coaxial cavity is considered faulty and its corresponding weight is reduced to zero. The aim of the Genetic algorithm is to recover the weight of the faulty element based upon the weights of the other elements in the array. The Optimization tool was used in MatLab software to design the algorithm for the above specified task. The forthcoming section describes the values of the parameters chosen to apply the Genetic algorithm for this problem.

- The 'Number of variables' which have to be computed using the algorithm are '1' i.e. the weight of the faulty element at position (3, 0).
- According to the GA problem solver tool, the 'Population type' is taken as 'Double vector' for the integer constraints.
- The parameter of 'Creation function' is chosen as 'Constraint dependent' which causes the initial population to be created. It takes the population to be either uniform, feasible or non-uniform depending upon the constraints of the variables which have to be computed. In the case of the missing weight calculation, the function is assumed to be uniform.
- An initial population of size 20 is chosen with an initial range of the population from 0 to 1.
- The Rank scaling method is chosen for Fitness scaling. This scaling method, scales the raw scores based on the rank of each individual rather than its score. In the Rank scaling method, the fittest individual is given a rank of 1, the next fittest individual is given a rank of 2 and so on.
- The Stochastic uniform method is used as a Selection method. In this method, each parent is allotted a section of length on a line proportional to the expectation of its occurrence. The selection method moves along the line with equal steps with one step for each parent. At each step, the algorithm allocates a parent from the section that it lands on.
- The 'Reproduction' option is chosen to decide upon how two parents create the child. The parameter of 'Elite count' specifies the number of individuals that are guaranteed to survive to the next generation. This value is set to (0.9 * population_size).
- The parameter of 'Crossover fraction' specifies the fraction of the next generation that the crossover produces. This fraction is set to a maximum value of 1.
- The parameter of 'Mutation function' which is responsible for making random changes to the individuals in the population in order to provide genetic diversity, is chosen as 'Adaptive feasible' owing to the linear constraints.
- The parameter of 'Cross over function' is also chosen as 'Constraint dependent' and the parameter of 'Migration' is chosen to be 'Forward'.
- The 'Nonlinear Constraint Algorithm' is set as 'Augmented Lagrangian' with an initial penalty of 10 and penalty factor of 100.
- As a part of the 'Stopping Criterion', the number of generations chosen for stopping are set to 100.
- The number of 'Stall generations' is set to 200. This value specifies that if the average change in the fitness value over 200 generations is lesser than the Fault tolerance specified, then the algorithm stops.

Figure 5 illustrates the flowchart depicting the operation of the Genetic algorithm and Fig. 6 depicts the results obtained by running the algorithm to recover the value of w_4 element in the 2×6 coaxial cavity horn antenna array. It is clear from Fig. 3, that the value of w_4, is computed as 1.17699 which is in close proximity to the desired value of w_4, which is 1.

Therefore it is concluded that the Genetic algorithm is a feasible option for recovery of a single antenna element failure in an antenna array. The next section of this article

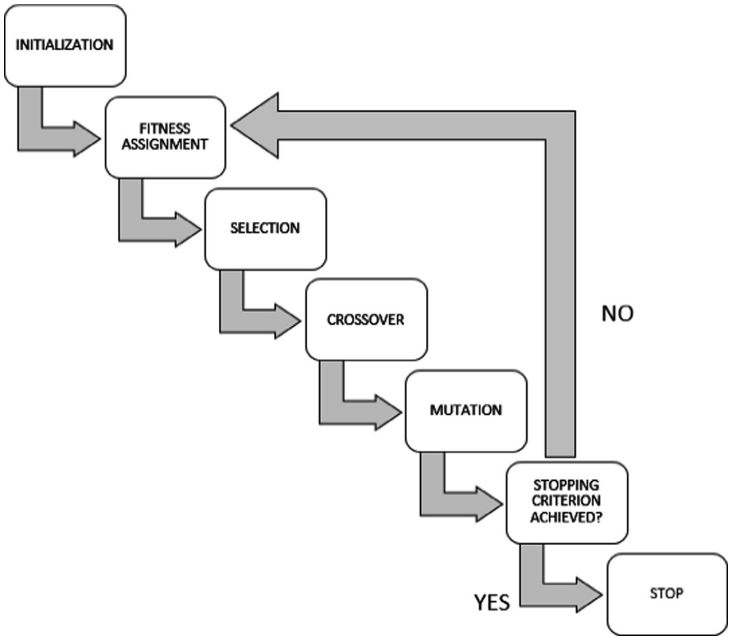


Fig. 5. Flowchart of Genetic algorithm

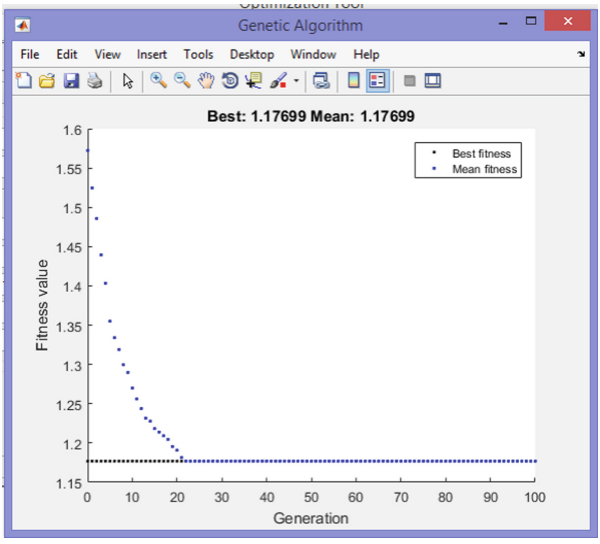


Fig. 6. Recovery of faulty element weight using Genetic algorithm

describes the use of yet another evolutionary computing algorithm which is Particle

Swarm Optimization. The Particle Swarm Optimization algorithm is used for the same problem of the recovery of the weight of a faulty central element in the 2 × 6 coaxial cavity horn antenna array.

4 Application of Particle Swarm Optimization to 2 × 6 Coaxial Cavity Horn Antenna Array for Fault Recovery of the Main Element

4.1 Preliminaries of Particle Swarm Optimization (PSO) Algorithm

PSO is an algorithm designed to mimic the movement of birds or a school of fish [14]. In the PSO algorithm every candidate solution is represented as a particle. A collection of moving/flying particles symbolize changing solutions in a designated area, which move to reach a position of global optimum. In comparison to the Genetic Algorithm, the Particle Swarm Optimization algorithm runs faster to find the solution closest to the optimum [15]. The following terminologies are used to describe the PSO algorithm.

Particle: One of the possible candidate solutions to a problem is termed as a particle in the PSO algorithm. It moves around the space allotted with a designated velocity in order to arrive at a global optimum.

pBest: This is the best known position of a particle at a given point in time.

gBest: This is the globally best known position of the system comprising if a number of particles.

pVelocity: This the particle velocity which gets updated in every iteration depending on the current position of the particle, its current velocity, its inertia weight and its distance from the global optimum. The following section describes the values assigned to the parameters for the working of the PSO algorithm to recover the value of w_4.

- The algorithm is set to run for 30 iterations.
- The parameters of 'Inertia', 'Correction factor' and 'Number of Swarms' are set to the values, 1, 2, and 50 respectively.
- Initially a 50 × 7 matrix termed as Swarm, consisting of all zeros is initialized.
- Columns 1 and 2 of this matrix correspond to the 'Objective function' value or in the case of our simulation, they correspond to the value of the Array Factor.
- Columns 3 and 4 are the updated values of the best position of particle X, which is the weight of the faulty element according to our problem.
- Columns 5 and 6 are the updated values of the velocities of the particle X, which in our case remains zero throughout the simulation.
- Finally Column 7 corresponds to the 'Best' value or the expected value of X, which is set to 1.
- Equation (5) describes the changes in the velocity of particle X.

$$X\_velocity = (rand \times inertia \times X\_velocity) + (correction\_factor \times rand \times (X\_position - Array\_factor)) + (correction\_factor \times rand \times (Global\_best - Array\_factor)) \quad (5)$$

The variables *rand*, *inertia*, *Array_factor* and *Global_best* have been defined earlier in the algorithm. These are used to compute the velocity of the particle X. The PSO algorithm runs to recover this weight to a value of 1.2064. The velocities of the particles in this case remain zero until convergence. The algorithm runs for a set number of 50 epochs. Figure 7 illustrates the PSO algorithm working using a flowchart and Fig. 8 illustrates the recovered weight using the PSO algorithm.

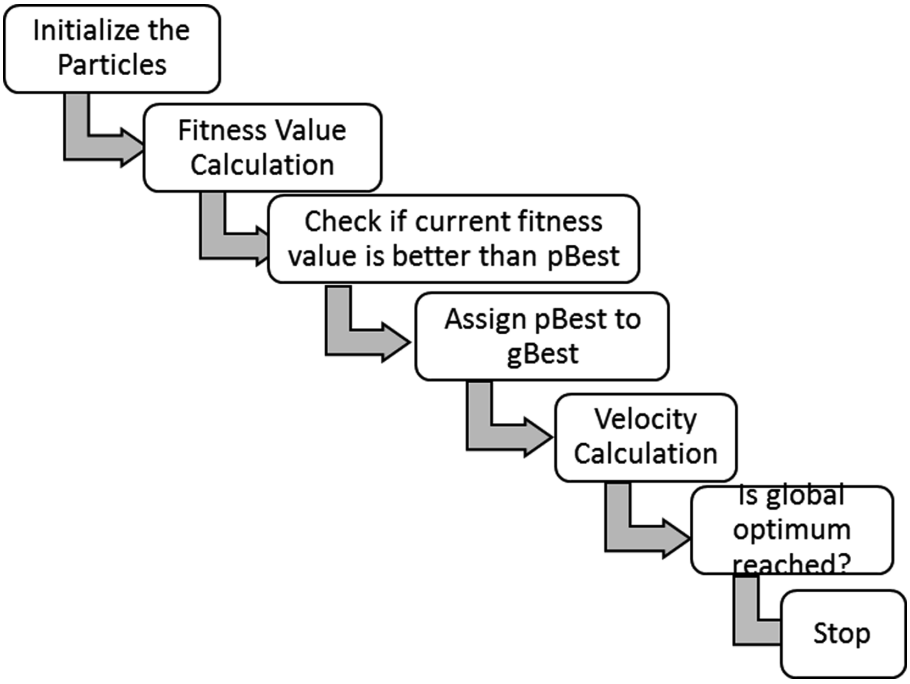


Fig. 7. Flowchart of Particle Swarm Optimization

The recovered weight of the faulty element using Particle Swarm Optimization is 1.2064 which not as precise as that recovered using the Genetic algorithm. This occurs due to the coarse tuning of the PSO algorithm with limited number of weights to choose from.

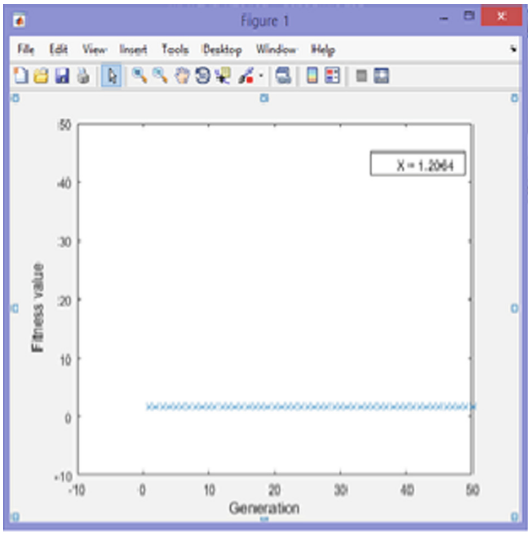


Fig. 8. Recovery of faulty element weight using Particle Swarm Optimization algorithm

5 Conclusion

This article presented a case study of the usage of the Genetic algorithm and Particle Swarm Optimization algorithm to compute the weight of a faulty central element in the proposed 2 × 6 planar coaxial cavity horn antenna array for the MELISSA GB-SAR system. It is proved using simulations in MatLab, that both the algorithms converge well to recover the weight of the faulty element. Genetic algorithm converges better in comparison to Particle Swarm Optimization, and PSO runs faster than GA. Future work in this regard would be to fine tune the algorithms to achieve a near 100% accuracy for the weight calculation of the faulty central element. It is also proposed to design the algorithms for a multi-element failure case of the 2 × 6 array and study the convergence rates as well as the accuracies of these algorithms.

References

1. Tarchi, D., Oliveri, F., Sammartino, P.F.: MIMO radar and ground-based SAR imaging systems: equivalent approaches for remote sensing. IEEE Trans. Geosci. Remote Sens. **51**(1), 425–435 (2013)
2. Broussolle, J., et al.: MELISSA, a new class of ground based InSAR system. An example of application in support to the Costa Concordia emergency. ISPRS J. Photogrammetry Remote Sens. **91**, 50–58 (2014)
3. Vincent, S., Francis, S.A.J., Rajsingh, E.B.: An alternate antenna array geometry for a GB-SAR system used in landslide monitoring. J. Indian Soc. Remote Sens. **43**(3), 761–768 (2015)

4. Vincent, S., Francis, S.A.J., Kumar, O.P., Rajsingh, E.B.: A comparative study of horn antennas suitable for the transmitting antenna array module of melissa architecture. In: Proceedings of the International Conference on Distributed Computing, VLSI, Electrical Circuits and Robotics (DISCOVER - 2016), 13–14 August 2016. IEEE-Xplore Digital Library, Surathkal (2016)
5. Vincent, S., Francis, S.A.J., Kumar, O.P., Rajsingh, E.B.: Design of a planar antenna array for the transmitting module of MELISSA. Int. J. Appl. Eng. Res. **12**(1), 179–184 (2017)
6. Vincent, S., Francis, S.A.J., Kumar, O.P., Rajsingh, E.B.: Optimization of gain and return loss of a 2 × 6 planar coaxial cavity horn antenna array for MELISSA. In: Proceedings of 2017 IEEE International Conference on Antenna Innovations and Modern Technologies for Ground, Aircraft and Satellite Applications (iAIM - 2017) held in Bangalore, India in Novemeber 2017. IEEE-Xplore Digital Library (2017)
7. Vincent, S., Francis, S.A.J., Kumar, O.P., Raimond, K.: A comparative performance evaluation of beamforming techniques for a 2 × 6 coaxial cavity horn antenna array for MELISSA. In: Ray, K., Sharan, S., Rawat, S., Jain, S., Srivastava, S., Bandyopadhyay, A. (eds.) Engineering Vibration, Communication and Information Processing. Lecture Notes in Electrical Engineering, vol. 478, pp. 65–74. Springer, Singapore (2019). https://doi.org/10.1007/978-981-13-1642-5_6
8. Haupt, R.L.: Thin arrays using genetic algorithms. IEEE Trans. Antennas Propag. **42**, 993–999 (1994)
9. Yeo, B., Liu, Y.: Array failure correction with a Genetic Algorithm. IEEE Trans. Antennas Propag. **47**(5), 823–828 (1999)
10. Yan, K., Lu, Y.: Sidelobe reduction in Array-pattern synthesis using Genetic Algorithm. IEEE Trans. Antennas Propag. **45**(7), 1117–1122 (1997)
11. Rahman, S.U., Cao, Q.: Analysis of Linear Antenna Array for minimum side lobe level, half power beamwidth and nulls control using PSO. J. Microwave Optoelectron. Electromagnet. Appl. **16**(2), 577–591 (2017)
12. Han, C., Wang, L.: Array pattern synthesis using Particle Swarm Optimization with dynamic inertia weight. Int. J. Antennas Propag. Hindawi Publishers (2016). Article id 1829458
13. Goudos, S., Kalialakis, C., Mittra, R.: Evolutionary algorithms applied to antennas and propagation: a review of state of the art. Int. J. Antennas Propag. Hindawi Publishers (2016). Article id 1010459
14. Kennedy, J., Eberhart, R.C.: Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, pp. 1942–1948 (1995)
15. Banks, A., Vincent, J., Anyakoha, Ch.: A review of particle swarm optimization. Part I: Background Dev. Natural Comput. **6**, 467–484 (2007)

STCKF Algorithm Based SOC Estimation of Li-Ion Battery by Dynamic Parameter Modeling

R. Ramprasath[1(✉)] and R. Shanmughasundaram[2]

[1] Department of Mechanical Engineering,
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
ramprasathraghavan@gmail.com
[2] Department of Electrical and Electronics Engineering, Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
r_shanmughasundaram@cb.amrita.edu

Abstract. State of Charge (SoC) is the important criterion which reflects the actual battery usage. So, the State of Charge (SoC) has to be precisely estimated for improving the life and the rate of utilization of the battery. During normal operation of the battery, parameters like charge and discharge efficiency, temperature, etc., tend to affect the accurate estimation of SoC. In this paper, for estimating battery SoC with higher accuracy, Strong Tracking Cubature Kalman Filtering (STCKF) is used and the battery model parameters are identified using the method of Recursive Least Square (RLS). Simulation results indicate, STCKF estimates the SoC values as that of Ampere-Hour (AH) method with very minimal error and the dynamically modeled battery parameter values follows the same discharge characteristics as that of real batteries.

Keywords: Estimating SoC · STCKF algorithm · Lithium-ion battery
Dynamic parameter modeling

1 Introduction

For an efficient battery management system, the most important prerequisite would be the precise appraisement of state of charge and estimating the accurate value is an important issue in battery technology. State of Charge equals the fraction of currently available battery capacity to the rated capacity measured in standardized conditions and it is expressed in terms of percentage [1]. Non-linear characteristics of the battery and various other parameters tend to affect the accurate estimation of SoC [2]. Amending the estimation accuracy of SoC has significant impact in perpetuating the battery's life and raising the battery's utilization ratio.

The AH strategy is the mostly utilized method to appraise battery SoC. Calculations used in the AH strategy is basic, steady and simple to actualize, but as the AH strategy does not use the closed loop calculation and as it has the restriction of examining exactness, it will lead to bigger aggregate mistakes. So, it is more often utilized as a validation strategy for other highly precise SoC prediction methods. Neural network is very helpful for perfect modeling of non-linear characteristics of the battery based on the

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 229–239, 2019.
https://doi.org/10.1007/978-981-13-5758-9_20

input training data. But neural network works accurately only when large number of input training data is fed into the system [3, 4]. Kalman filter is a closed loop recursive strategy. It viably illuminates the problem of blunder amassing, but this strategy can be applied only for linear systems. Extended Kalman Filter (EKF) uses Taylor formula and linearizes the non-linear system, but it introduces the linearization errors [5]. Unscented Kalman Filter (UKF) bases its estimation on unscented transformation and is aptly suited for filtering of nonlinear systems, as it doesn't introduces as much of the linearization error (as introduced by EKF). UKF has higher accuracy too [6, 7]. Even though EKF and UKF are commonly utilized for solving nonlinear systems, they have certain disadvantages. EKF algorithm is less accurate and weakly robust when it is utilized for solving strong non-linear systems. UKF, which adopts the sigma points distribution for approximating the nonlinear systems, does not have proper bases to select the sample points and weights. Moreover, if more than 3 dimensions exist for a state variable, covariance matrix will not meet the requirement of being in a semi-definite condition because of the presence of negative weights. So, UKF has reduced stability under such conditions. Strong Tracking Filter (STF) was developed mainly to overcome the disadvantage of EKF. STF introduces suboptimal fading factor to the EKF strategy and this helps the STF to reduce the residual errors and make it more robust and less sensitive to initial state. But STF doesn't increase the accuracy of the filter. To improvise the stability of UKF, Cubature Kalman Filter (CKF) was proposed [8]. CKF has same procedure as of UKF, but it adopts a batch of cubature set points with equivalent weights. This algorithm has reduced computational complexity and has also proven to be highly accurate and a numerically stable one [9]. Combining STF and CKF, Strong Tracking Cubature Kalman Filtering (STCKF) has been proposed, which overcomes the disadvantages of all the currently available SoC estimation techniques [10].

Standard battery modeling, which is done by fitting the data collected through various experiments will results in inaccurate battery SoC estimation. But, limited memory recursive least square algorithm has the ability to attain the exact model parameters, i.e., matching the actual battery characteristics, within a very short convergence time. Main contribution of this paper is that the Li-ion battery's State of Charge is accurately estimated using the combined approach of Strong Tracking Cubature Kalman Filtering (STCKF) algorithm and dynamical modeling of the battery parameters. The latter is obtained with the help of recursive least square algorithm. Finally, the proposed method is verified along with the plots from the available test vehicle and also compared with the State of Charge estimated using the Unscented Kalman Filtering algorithm.

The paper is arranged as mentioned in the upcoming order: In Sect. 2, modeling of Lithium-Ion battery is explained. Section 3 deals with the implementation of Strong Tracking Cubature Kalman Filtering (STCKF) algorithm. Section 4 explains the identification of battery parameters using Recursive Least Square (RLS) algorithm. Section 5 describes the simulation done in MATLAB and its respective results, which proves that the proposed method can accurately estimate the State of Charge of the battery. In Sect. 6, conclusions are provided based on the simulation results.

2 Modeling of Lithium-Ion Battery

Battery models should be in a way that it should be able to exactly replicate the characteristics of the battery. In this paper, second order RC equivalent circuit model, which has high accuracy, straight forward structure, clear physical meaning, and accurate parameter recognizable character is utilized for modeling the Lithium-ion battery. Figure 1 represents the second order RC equivalent circuit model,

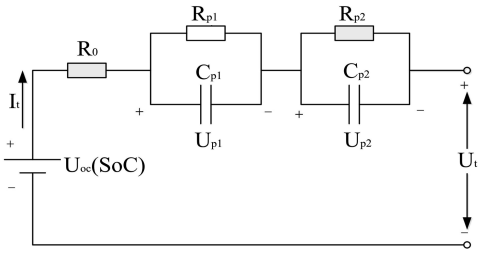


Fig. 1. Second order RC equivalent circuit of Li-ion battery model

It contains battery's Open Circuit Voltage (OCV) U_{OC}, internal resistance R_0, polarization resistances R_{p1} and R_{p2}, polarization capacitances C_{p1} and C_{p2}, where the polarization resistances and polarization capacitances are connected in parallel. Initial values of the battery parameters, R_0, R_{p1}, R_{p2}, C_{p1} and C_{p2} are identified from the linear regression method by application of the single current pulse to a battery and analysis of its voltage response when the current value is zero [11]. From the dynamic characteristics of the circuit, its electrical behavior can be presented as,

$$U'_{p1} = -\frac{U_{p1}}{R_{p1}C_{p1}} + \frac{I_t}{C_{p1}} \tag{1}$$

$$U'_{p2} = -\frac{U_{p2}}{R_{p2}C_{p2}} + \frac{I_t}{C_{p2}} \tag{2}$$

$$U_t = U_{OC}(SoC) - U_{p1} - U_{p2} - R_0 I_t \tag{3}$$

where, U_t indicates the terminal voltage, U_{p1} and U_{p2} denote the voltages across the two RC networks and I_t indicates the load current. U_{OC} denotes the battery Open Circuit Voltage (OCV) and it is expressed in terms of SoC as,

$$U_{OC}(SoC) = K_0 + K_1 SoC + K_2 SoC^2 + K_3/SoC + K_4 \ln(SoC) + K_5 \ln(1 - SoC) \tag{4}$$

where, K_i (i = 0,1,2,…,5) in accordance to the variation of OCV with respect to SoC as presented in Fig. 2 [12].

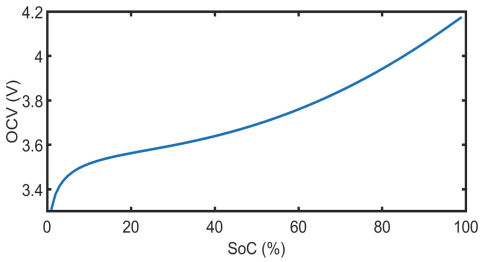


Fig. 2. OCV Vs SoC graph of the battery

From the Coulomb Counting method of SoC measurement, SoC is defined as follows,

$$\mathrm{SoC(t)} = \mathrm{SoC(t_0)} - \frac{\int_{t_0}^{t} \eta i d\tau}{Q_e} \tag{5}$$

where, *SoC(t)* and *SoC(t_0)* indicate the battery's State of Charge measured at the time instances t and t_0 respectively, Q_e indicates the rated battery capacity and *i* denotes the instantaneous current obtained during the time period t_0 to *t*. η represents the charge-discharge efficiency.

On combining (1), (2), (3) and (5), battery's state space equation has been obtained and it is expressed as,

$$\begin{cases} \begin{bmatrix} \mathrm{SoC}(m^+) \\ U_{p1}(m^+) \\ U_{p2}(m^+) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \alpha_{p1} & 0 \\ 0 & 0 & \alpha_{p2} \end{bmatrix} \begin{bmatrix} \mathrm{SoC}(m) \\ U_{p1}(m) \\ U_{p2}(m) \end{bmatrix} + \begin{bmatrix} -\eta \Delta t / Q_e \\ (1-\alpha_{p1}) R_{p1} \\ (1-\alpha_{p2}) R_{p2} \end{bmatrix} I_t(m) + w(m) \\ U_t(m) = U_{OC}(m) - U_{p1}(m) - U_{p2}(m) - I_t(m) R_0 + v(m) \end{cases} \tag{6}$$

where, $\alpha_{p1} = e^{-(\Delta t / R_{p1} C_{p1})}$, $\alpha_{p2} = e^{-(\Delta t / R_{p2} C_{p2})}$, Δt is the sampling time interval, ***w(m)*** represents the noise vector due to processing and *v(m)* indicates the noise due to measurement. State vector, ***x(m)*** is formed with the help of State of Charge (*SoC*), voltage across R_{p1} and C_{p1} (U_{p1}) and voltage across R_{p2} and C_{p2} (U_{p2}). By setting, $u1(m) = I_t(m)$ and $y(m) = U_t(m)$, (6) is abbreviated as,

$$\begin{cases} \mathbf{x}(m^+) = f(\mathbf{x}(m), u1(m)) + w(m) \\ y(m) = h(\mathbf{x}(m)) + v(m) \end{cases} \tag{7}$$

3 STCKF Algorithm for Estimation of Battery SoC

CKF algorithm utilizes the cubature point sets with equivalent weights obtained from the spherical-radical principle. This is mainly done to obtain the non-linear approximation of Bayesian filter integral problem. On adopting the third order cubature rule,

number of cubature points to be selected should be twice that of the number of state variables used. In this paper, number of state variable used is 3. Therefore, number of cubature points will be 6. Calculation of cubature points and their respective weights are presented below,

$$\mathbf{C}^{\mathrm{i}} = \sqrt{\mathrm{n}[\mathbf{1}]_{\mathrm{i}}}, \mu^{\mathrm{i}} = \frac{1}{\mathrm{r}} \quad \mathrm{i} = 1,2,..\mathrm{r} \quad \text{and } \mathrm{r} = 2\mathrm{n} \tag{8}$$

where, $\boldsymbol{C}^{\boldsymbol{i}}$ represents the cubature points, μ^{i} indicates the corresponding weights. $\mathbf{[1]}_i$ denotes the row of the matrix constructed using unit vectors. $\mathbf{[1]}_i$ is presented below,

$$[1]_{\mathrm{i}} = \begin{bmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{bmatrix} \tag{9}$$

Making an assumption that process and measurement noise to be an uncorrelated white Gaussian noise, their statistical properties are presented as $\mathbf{w}(\mathrm{m}) \sim N(0,\mathrm{Q}(\mathrm{m}))$ and $\mathrm{v}(\mathrm{m}) \sim N(0,\mathrm{R}(\mathrm{m}))$

Strong Tracking Filter and Cubature Kalman Filter have been combined together to obtain the STCKF algorithm. The steps of STCKF algorithm are shown below,

Step 1: Initialization of filter

$$\hat{\mathbf{x}}^{+}(1) = \mathrm{Mean}(\mathbf{x}(1)),\ \mathbf{P}^{+}(1) = \mathrm{Mean}[\left(\mathbf{x}(1) - \hat{\mathbf{x}}^{+}(1)\right)\left(\mathbf{x}(1) - \hat{\mathbf{x}}^{+}(1)\right)^{\mathrm{T}}] \tag{10}$$

Step 2: Prediction of State vector

A new time variable, m is introduced, and it will take the value of 2,3,…etc., in order to finish the following steps.

$$\mathbf{P}^{+}(\mathrm{m}^{-}) = \mathbf{S}^{+}(\mathrm{m}^{-})(\mathbf{S}^{+}(\mathrm{m}^{-}))^{\mathrm{T}} \tag{11}$$

$$\mathbf{x}^{\mathrm{i}}(\mathrm{m}^{-}) = \hat{\mathbf{x}}^{+}(\mathrm{m}^{-}) + \mathbf{S}^{+}(\mathrm{m}^{-})\mathbf{C}^{\mathrm{i}} \tag{12}$$

$$\tilde{\mathbf{x}}^{\mathbf{i}}(\mathrm{m}) = \mathrm{f}(\mathbf{x}^{\mathrm{i}}(\mathrm{m}^{-}),\mathrm{u1}(\mathrm{m}^{-})) \tag{13}$$

$$\hat{\mathbf{x}}^{-}(\mathrm{m}) = \frac{1}{2\mathrm{n}}\sum_{\mathrm{i}=1}^{2\mathrm{n}} \tilde{\mathbf{x}}^{\mathrm{i}}(\mathrm{m}) \tag{14}$$

$$\mathbf{P}^{-1}(\mathrm{m}) = \frac{1}{2\mathrm{n}}\sum_{\mathrm{i}=1}^{2\mathrm{n}} \tilde{\mathbf{x}}^{\mathrm{i}}(\mathrm{m})\left(\tilde{x}^{\mathrm{i}}(\mathrm{m})\right)^{\mathrm{T}} - \hat{\mathbf{x}}^{-}(\mathrm{m})(\hat{\mathbf{x}}^{-}(\mathrm{m}))^{\mathrm{T}} + \mathbf{Q}(\mathrm{m}^{-}) \tag{15}$$

where, $\mathbf{P}^{+}(\mathrm{m}^{-})$ is the state error covariance matrix at the previous instance, $\hat{\mathbf{x}}^{-}(\mathrm{m})$ is the further estimated state variable, $\mathbf{P}^{-1}(\mathrm{m})$ is the further predicted state error covariance matrix.

Step 3: Calculation of fading factor, $\lambda(m)$

The fading factor is calculated based on the method proposed in the STF algorithm, and it is computed as follows,

Step 3.1: Estimation of $\boldsymbol{V}(m)$ from the residual error $\boldsymbol{\varepsilon}(m)$

$$\mathbf{V}(\mathrm{m}) = \begin{cases} \boldsymbol{\varepsilon}(1)(\boldsymbol{\varepsilon}(1))^{\mathrm{T}} & \mathrm{m} = 1 \\ \frac{\rho \mathbf{V}(\mathrm{m}^{-}) + \boldsymbol{\varepsilon}(\mathrm{m})(\boldsymbol{\varepsilon}(\mathrm{m}))^{\mathrm{T}}}{1+\rho} & \mathrm{m} \geq 2 \end{cases} \tag{16}$$

where, ρ is the forgetting factor and its value lies between 0 and 1.

Step 3.2: Calculation of $\mathbf{P}^{\mathrm{l}}(\mathrm{y})$ and $\mathbf{P}^{\mathrm{l}}(\mathrm{xy})$

$$\mathbf{P}^{-\mathrm{l}}(\mathrm{m}) = \mathbf{S}^{-\mathrm{l}}(\mathrm{m})\left(\mathbf{S}^{-\mathrm{l}}(\mathrm{m})\right)^{\mathrm{T}} \tag{17}$$

$$\mathbf{x}^{\mathrm{i,l}}(\mathrm{m}) = \hat{\mathbf{x}}^{-}(\mathrm{m}) + \mathbf{S}^{-\mathrm{l}}(\mathrm{m})\mathbf{C}^{\mathrm{i}} \tag{18}$$

$$\tilde{\mathrm{y}}^{\mathrm{i,l}}(\mathrm{m}) = \mathrm{h}(\mathrm{x}^{\mathrm{i,l}}(\mathrm{m}), \mathrm{u}(\mathrm{m})) \tag{19}$$

$$\hat{\mathrm{y}}^{\mathrm{l}}(\mathrm{m}) = \frac{1}{2\mathrm{n}} \sum_{\mathrm{i}=1}^{2\mathrm{n}} \tilde{\mathrm{y}}^{\mathrm{i,l}}(\mathrm{m}) \tag{20}$$

$$\mathbf{P}^{\mathrm{l}}(\mathrm{y}) = \frac{1}{2n} \sum_{i=1}^{2n} \tilde{\mathrm{y}}^{\mathrm{i,l}}(\mathrm{m})\left(\tilde{\mathrm{y}}^{\mathrm{i,l}}(m)\right)^{T} - \hat{\mathrm{y}}^{\mathrm{l}}(\mathrm{m})\left(\hat{\mathrm{y}}^{\mathrm{l}}(\mathrm{m})\right)^{\mathrm{T}} + \mathrm{R}(\mathrm{m}) \tag{21}$$

$$\mathbf{P}^{\mathrm{l}}(\mathrm{xy}) = \frac{1}{2\mathrm{n}} \sum_{\mathrm{i}=1}^{2\mathrm{n}} \mathbf{x}^{\mathrm{i,l}}(\mathrm{m})\left(\tilde{y}^{\mathrm{i,l}}(\mathrm{m})\right)^{\mathrm{T}} - \hat{\mathbf{x}}^{-}(\mathrm{m})\left(\hat{y}^{\mathrm{l}}(\mathrm{m})\right)^{\mathrm{T}} \tag{22}$$

where, $\mathbf{P}^{\mathrm{l}}(\mathrm{xy})$ represents the covariance, $\mathbf{P}^{\mathrm{l}}(\mathrm{y})$ indicates the output variable step prediction variance and $\hat{\mathrm{y}}^{\mathrm{l}}(\mathrm{m})$ denotes the predicted output variable.

Step 3.3: Computation of Matrices $\boldsymbol{M}_m$ and $\boldsymbol{N}_m$

$$\mathbf{N}_{\mathrm{m}} = \mathbf{V}(\mathrm{m}) - \left(\mathbf{P}^{\mathrm{l}}(\mathrm{xy})\right)^{\mathrm{T}}\left(\mathbf{P}^{-\mathrm{l}}(\mathrm{m})\right)^{-1}\mathbf{Q}(\mathrm{m})\left(\mathbf{P}^{-\mathrm{l}}(\mathrm{m})\right)^{-1}\left(\mathbf{P}^{\mathrm{l}}(\mathrm{xy})\right)^{\mathrm{T}} - \beta \mathrm{R}(\mathrm{m}) \tag{23}$$

$$\mathbf{M}_{\mathrm{m}} = \mathbf{P}^{\mathrm{l}}(\mathrm{y}) - \mathbf{V}(\mathrm{m}) + \mathbf{N}_{\mathrm{m}} + (\beta - 1)\mathrm{R}(\mathrm{m}) \tag{24}$$

To obtain smoother state estimation and for avoiding the effect of excessive regulation because of fading factor, the softening factor (β), is introduced.

Step 3.4: Calculation of suboptimal solution for fading factor, $\lambda(m)$

$$\lambda(\mathrm{m}) = \begin{cases} \lambda(0) & \lambda(0) \geq 1 \\ 1 & \lambda(0) < 1 \end{cases}, \quad \lambda(0) = \frac{\mathrm{tr}[\mathrm{N}_{\mathrm{m}}]}{\mathrm{tr}[\mathrm{M}_{\mathrm{m}}]} \tag{25}$$

Step 4: Measurement update
$\boldsymbol{P^{-}(m)}$ is calculated by substituting $\lambda(m)$ into $\boldsymbol{P^{-l}(m)}$.

$$\mathrm{P}^{-}(\mathrm{m}) = \frac{\lambda(\mathrm{m})}{2\mathrm{n}} \sum_{\mathrm{i}=1}^{2\mathrm{n}} (\tilde{\mathbf{x}}^{\mathrm{i}}(\mathrm{m}) * (\tilde{\mathbf{x}}^{\mathrm{i}}(\mathrm{m}))^{\mathrm{T}} - \lambda(\mathrm{m}) * \hat{\mathbf{x}}^{-}(\mathrm{m}) * (\hat{\mathbf{x}}^{-}(\mathrm{m}))^{\mathrm{T}}) + \mathbf{Q}(\mathrm{m}^{-}) \quad (26)$$

Updated values of output variable step prediction variance, $\mathbf{P}(\mathrm{y})$ and covariance, $\mathbf{P}(\mathrm{xy})$ based on the fading factor, $\lambda(m)$ is obtained by the substitution of $\boldsymbol{P^{-}(m)}$ in the places of $\boldsymbol{P^{-l}(m)}$ in the Eqs. (17) to (22), resulting in the below equations.

$$\mathbf{P}^{-}(\mathrm{m}) = \mathbf{S}^{-}(\mathrm{m})(\mathbf{S}^{-}(\mathrm{m}))^{\mathrm{T}} \quad (27)$$

$$\mathbf{x}^{\mathrm{i}}(\mathrm{m}) = \hat{\mathbf{x}}^{-}(\mathrm{m}) + \mathbf{S}^{-}(\mathrm{m})\mathbf{C}^{\mathrm{i}} \quad (28)$$

$$\tilde{\mathrm{y}}^{\mathrm{i}}(\mathrm{m}) = \mathrm{h}(\mathrm{x}^{\mathrm{i}}(\mathrm{m}), \mathrm{u1}(\mathrm{m})) \quad (29)$$

$$\hat{\mathrm{y}}(\mathrm{m}) = \frac{1}{2\mathrm{n}} \sum_{\mathrm{i}=1}^{2\mathrm{n}} \tilde{\mathrm{y}}^{\mathrm{i}}(\mathrm{m}) \quad (30)$$

$$\mathbf{P}(\mathrm{y}) = \frac{1}{2\mathrm{n}} \sum_{\mathrm{i}=1}^{2\mathrm{n}} \tilde{y}^{\mathrm{i}}(\mathrm{m})(\tilde{y}^{\mathrm{i}}(\mathrm{m}))^{\mathrm{T}} - \tilde{y}(\mathrm{m})(\tilde{y}(\mathrm{m}))^{\mathrm{T}} + \mathrm{R}(\mathrm{m}) \quad (31)$$

$$\mathbf{P}(\mathrm{xy}) = \frac{1}{2\mathrm{n}} \sum_{\mathrm{i}=1}^{2\mathrm{n}} \mathbf{x}^{\mathrm{i}}(\mathrm{m})(\tilde{y}^{\mathrm{i}}(\mathrm{m}))^{\mathrm{T}} - \hat{\mathbf{x}}^{-}(\mathrm{m})(\hat{y}(\mathrm{m}))^{\mathrm{T}} \quad (32)$$

Step 5: Calculation of STCKF gain, **K(m)**

$$\mathbf{K}(\mathrm{m}) = \frac{\mathbf{P}(\mathrm{xy})}{\mathbf{P}(\mathrm{y})} \quad (33)$$

Step 6: State and Error Covariance update

$$\hat{\mathbf{x}}^{+}(\mathrm{m}) = \hat{\mathbf{x}}^{-}(\mathrm{m}) + \mathbf{K}(\mathrm{m})(\mathrm{y}(\mathrm{m}) - \hat{y}(\mathrm{m})) \quad (34)$$

$$\mathbf{P}^{+}(\mathrm{m}) = \mathbf{P}^{-}(\mathrm{m}) - \mathbf{K}(\mathrm{m})\mathbf{P}(\mathrm{y})(\mathbf{K}(\mathrm{m}))^{\mathrm{T}} \quad (35)$$

4 Battery Parameters Identification Using RLS Algorithm

Considering $\mathrm{E_t(m)} = \mathrm{U_{OC}(m)} - \mathrm{U_t(m)}$, (6) could be re written as,

$$\mathrm{E_t(m)} = \psi_1 \mathrm{I_t(m)} + \psi_2 \mathrm{U_{p1}(m^{-})} + \psi_3 \mathrm{U_{p2}(m^{-})} + \psi_4 \mathrm{I_t(m^{-})} + \psi_5 \mathrm{I_t(m^{-})} \quad (36)$$

$$\begin{cases} \psi_1 = R_0 \\ \psi_2 = \alpha_{p1} \\ \psi_3 = \alpha_{p2} \\ \psi_4 = (1 - \alpha_{p1})R_{p1} \\ \psi_5 = (1 - \alpha_{p2})R_{p2} \end{cases} \tag{37}$$

Equation (36) could be re-written as (38). Parameters mentioned in (39) could be obtained utilizing RLS algorithm.

$$y(m) = \varphi(m)\boldsymbol{\psi}(m) \tag{38}$$

$$\begin{cases} \varphi(m) = [\, I_t(m) \quad U_{p1}(m^-) \quad U_{p2}(m^-) \quad I_t(m^-) \quad I_t(m^-) \,] \\ \boldsymbol{\psi}(m) = [\, \psi_1 \quad \psi_2 \quad \psi_3 \quad \psi_4 \quad \psi_5 \,]^T \end{cases} \tag{39}$$

Process for implementing the RLS algorithm are mentioned below,

$$y(m) = \varphi(m)\boldsymbol{\psi}(m) \; + \; e(m) \tag{40}$$

$$e(m) = E_t(m) - \varphi(m)\boldsymbol{\psi}(m^-) \tag{41}$$

$$\mathbf{K}(m) = \frac{\mathbf{P}(m^-)(\varphi(m))^T}{\gamma + \varphi(m)\mathbf{P}(m^-)(\varphi(m))^T} \tag{42}$$

$$\mathbf{P}(m) = \gamma^{-1}(\mathbf{P}(m^-) - \mathbf{K}(m)\varphi(m)\mathbf{P}(m^-)) \tag{43}$$

$$\boldsymbol{\psi}(m) = \boldsymbol{\psi}(m^-) + \mathbf{K}(m)e(m) \tag{44}$$

where, $\varphi(m)$ and $\boldsymbol{\psi}(m)$ are the input vector and coefficient vector at the m^{th} instance, $\boldsymbol{P}(m)$ is the state error covariance matrix, γ is the forgetting factor and $y(m)$ is $E_t(m)$. Equation (44) is used to estimate the battery's model parameters.

5 Simulation and Analysis

In this paper, battery current, actual battery voltage, and actual battery SoC as referred in Figs. 3, 4 and 5 are measured from an electric vehicle using a Lithium-ion battery with a rated capacity of 63 Ah. Platform for simulation has been established in MATLAB. The initial values of the battery parameters are set to $R_0 = 0.0013\ \Omega$, $R_{p1} = 0.0267\ \Omega$, $C_{p1} = 0.0077$ F, $R_{p2} = 0.0267\ \Omega$, $C_{p2} = 0.0077$ F, $x(0) = [1\ 0\ 0]^T$, $P(0) = \begin{bmatrix} 10^{-3} & 0 & 0 \\ 0 & 10^{-3} & 0 \\ 0 & 0 & 10^{-3} \end{bmatrix}$, $Q(0) = \begin{bmatrix} 10^{-2} & 0 & 0 \\ 0 & 10^{-2} & 0 \\ 0 & 0 & 10^{-2} \end{bmatrix}$, $R(0) = 10^{-3}$ and $\eta = 1$.

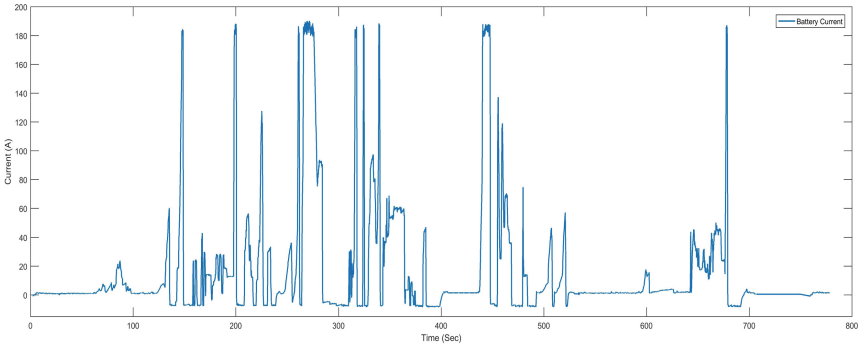


Fig. 3. Current profile of the battery

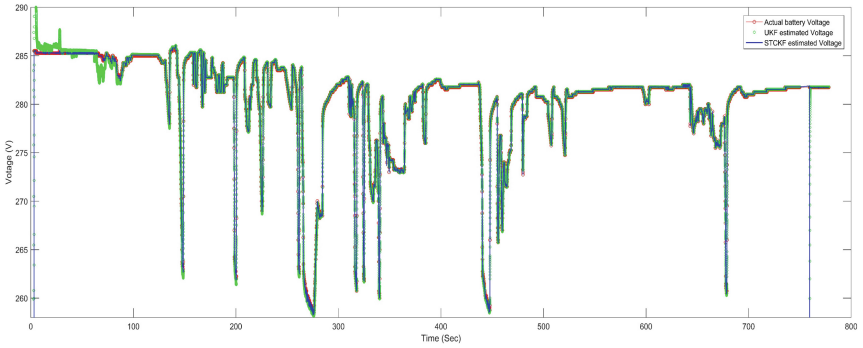


Fig. 4. Actual, UKF estimated and STCKF estimated battery voltage

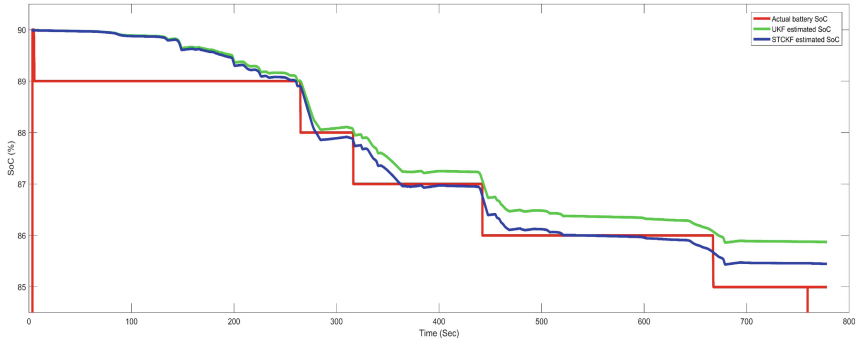


Fig. 5. Actual, UKF estimated and STCKF estimated battery SoC

The battery's current profile is shown in Fig. 3. Battery's actual voltage, UKF estimated voltage and the STCKF estimated voltage are presented in Fig. 4. Actual SoC and the SoC values calculated using UKF and STCKF estimation algorithm are shown in Fig. 5. The error between vehicle measured battery voltage and UKF estimated battery voltage and the error between vehicle measured battery voltage and the STCKF estimated battery voltage are presented in Fig. 6. Similarly the errors in UKF and STCKF estimated battery SoC with respect to the vehicle measured battery SoC are shown in Fig. 7. The above figures clearly depict that the STCKF estimated State of Charge and battery voltage, closely follow the vehicle measured State of Charge and battery voltage respectively. In addition to this, Root Mean Square Error (RMSE) of STCKF and UKF estimation methods have been calculated to verify their accuracy. RMSE values of the battery voltage and SoC based on the STCKF estimation technique are 1.4062 and 1.5031 respectively. RMSE values of the UKF estimated battery voltage and SoC are 2.8212 and 1.5926 respectively. The results specify that the STCKF method predicts the battery SoC and battery voltage with higher accuracy compared to that of UKF method.

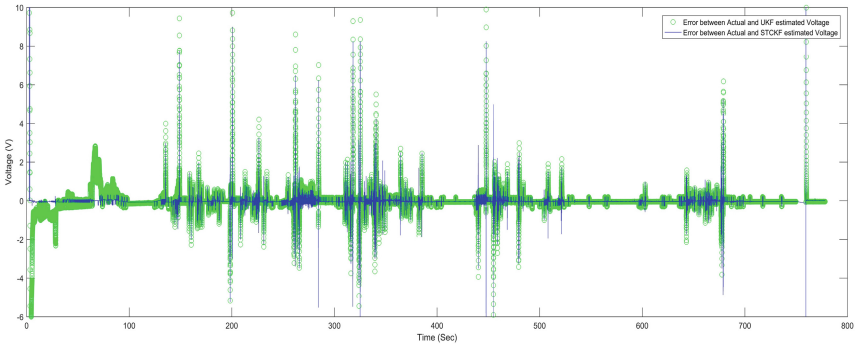


Fig. 6. Errors in voltage estimation using UKF and STCKF algorithms

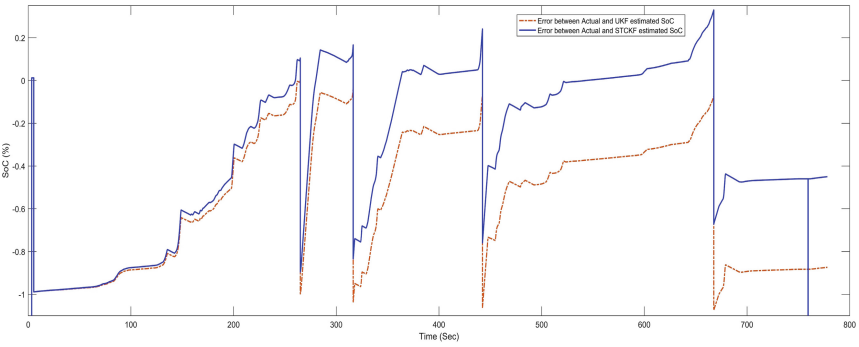


Fig. 7. Errors in SoC estimation using UKF and STCKF algorithms

6 Conclusion

In this paper, second-order RC equivalent circuit model is used to analyze the charging and discharging phenomenon of Lithium-ion batteries. Recursive Least Square (RLS) algorithm is utilized for identification of battery parameters and the battery model parameters tend to get updated continuously in order to attain real time battery characteristics. With the help of dynamical parameter modeling, STCKF algorithm is utilized in this paper for accurate SoC estimation. Simulation results prove that the STCKF algorithm has higher accuracy compared to UKF algorithm. Our future work will focus on the implementation and the execution of the presented method on hardware, used for the Battery Management System (BMS).

References

1. Zhou, X., Zhao, Y.: Study on the SOC estimation of power battery for electric vehicle. In: Mechanical Science and Technology for Aerospace Engineering, vol. 33, no. 2, pp. 263–266 (2014)
2. Li, Z., Lu, L., Ouyang, M.: Comparison of methods for improving SOC estimation accuracy through an ampere-hour integration approach. J. Tsinghua Univ. (Nat. Sci. Ed.) **50**, 1293–1296 (2010)
3. Zhu, H., Liu, Y., Zhao, C.: Parameter identification and SOC estimation of lithium ion battery. J. Hunan Univ. **41**, 37–42 (2014)
4. Muthumanikandan, S., Shanmughasundaram, R.: Estimation of state of charge of lithium ion battery using artificial neural networks. Int. J. Control. Theory Appl. **9**, 4331–4338 (2016)
5. Krishnakumar, A., Shanmughasundaram, R.: Simplified SOC estimation by EKF in Li-ion cell. J. Adv. Res. Dyn. Control. Syst. **3**, 616–622 (2018)
6. Wu, T., Hu, L.: Study of SOC estimation algorithm of power battery based on UKF. Power Electron. **48**(4), 23–26 (2014)
7. Xu, Y., Shen, Y.: Improved battery state-of-charge estimation based on Kalman filter. J. Beijing Univ. Aeronaut. Astronaut. **40**, 855–860 (2014)
8. Arasatranam, I., Haykin, S.: Cubature Kalman filters. IEEE Trans. Autom. Control **54**(6), 1254–1269 (2009)
9. Ning, X., Ye, C., Yang, J.: Cubature Kalman filtering for orbit determination of space targets. Chin. J. Radio Sci. **29**, 27–32 (2014)
10. Li, Z., Yang, W., Ding, D.: Strong tracking cubature Kalman filter for real-time orbit determination for impulse maneuver satellite. In: Proceedings of the 36th Chinese Control Conference, vol. 32, no. 31, pp. 5258–5263 (2017)
11. Shugang, J.: A parameter identification method for a battery equivalent circuit model. In: Proceedings of SAE World Congress, no. 2011–01–1367 (2011)
12. Zhang, X., Wang, Y., Chen, Z.: Model-based remaining discharge energy estimation of lithium-ion batteries. In: 3rd International Conference on Control, Automation and Robotics, pp. 510–513 (2017)

Machine Learning and Data Mining Methods in Testing and Diagnostics of Analog and Mixed-Signal Integrated Circuits: Case Study

Sergey Mosin(✉)

Kazan Federal University, Kazan, Russia
smosin@ieee.org

Abstract. Artificial intelligence methods are widely used in different interdisciplinary areas. The paper is devoted to application the method of machine learning and data mining to construction a neuromorphic fault dictionary (NFD) for testing and fault diagnostics in analog/mixed-signal integrated circuits. The main issues of constructing a NFD from the big data point of view are considered. The method of reducing a set of essential characteristics based on the principal component analysis and approach to a cut down the training set using entropy estimation are proposed. The metrics used for estimating the classification quality are specified based on the confusion matrix. The case study results for analog filters are demonstrated and discussed. Experimental results for both cases demonstrate the essential reduction of initial training set and saving of time on the NFD training with high fault coverage up to 100%. The proposed method and approach can be used according to the design-for-testability flow for analog/mixed-signal integrated circuits.

Keywords: Machine learning · Data mining · Testing · Diagnostics
Analog and mixed-signal IC · Entropy · Principal component analysis
Fault coverage · Neuromorphic fault dictionary

1 Introduction

The concept "Industry 4.0" is based on active use of the digital technologies in all sectors of economy. Developing the state-of-the-art computer systems with parallel and distributed architectures, high-performance processors and large volume of physical memory leads to active use and implementation of the artificial intelligence (AI) methods and algorithms for a wide range of scientific and industrial fields.

The area of AI includes the paradigms of Data Mining, Machine Learning, Big Data and etc., which were considered a long time as a theoretical mathematical apparatus, but nowadays are the typical background for development of the practice oriented applications.

The Data Mining (DM) and Machine Learning (ML) methods are widely used in different areas, providing acquisition of a new knowledge about nature, business, social, physical and other processes and phenomena or ensuring the possibilities to

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 240–255, 2019.
https://doi.org/10.1007/978-981-13-5758-9_21

solve the multi-dimensional and often ill-conditioned problems for the reasonable time with the required quality. The up-to-date tools for a data analysis, prognostics and trends generation mandatory use the DM and ML methods.

The DM and ML methods are also actively used in the complex tasks of *testing and diagnostics* for analog/mixed-signal integrated circuits (AMS IC), which are very important from the quality assurance and financial points of view. An efficient fault sampling and clustering methodology to construct compact fault dictionaries (FD) for complex AMS ICs with significantly reduced fault simulation effort is proposed in [1]. The development of a modular analog circuit fault diagnostic system based on neural networks with ability to identify faulty components or modules in an analog circuit by analyzing its impulse response is presented in [2]. A fault diagnosis scheme for analog integrated circuits based on an assemblage of learning machines that are trained beforehand to support through diagnosis decisions is discussed in [3]. Paper [4] presents a data-driven fault diagnosis method for analog circuits based on the robust competitive agglomeration. A method for diagnosing single-soft fault in analog circuits with low testability is proposed in paper [5]. Authors in [6] propose a new fault diagnostic system for analog circuits based on Local Mean Decomposition (LMD) for output response decomposition and learning vector quantization (LVQ) neural network.

The description of main challenges in testing and diagnostics for the AMS IC, the potential possibility of the DM and ML methods to reduce complexity of the problem of testing and diagnostics as well as essence of practical use some methods are the objects of this paper, which is structured as the following. Section 2 describes main issues of the testing and fault diagnostics for the AMS IC. The use of Principal Component Analysis (PCA) for reduction the set of essential characteristics is presented in Sect. 3. The approach to a cut down the training set based on entropy estimation is discussed in Sect. 4. Metrics for estimation the quality of machine learning are considered in Sect. 5. Experimental results for the circuits of analog filters are demonstrated in Sect. 6.

2 Main Issues of Testing and Fault Diagnostics for Analog and Mixed-Signal Integrated Circuits

Analog/mixed-signal ICs in comparison with digital circuits define more complexity for testing and fault diagnostics because of the following specifics: (1) Continuous processing of analog signals; (2) Nonlinear and functionally complex dependencies between input and output signals; (3) Effect of internal components' tolerances on output signals' value; (4) The sensitivity of the output responses of the circuit to the deviations of the nominal parameters of internal components as well as external environmental parameters; (5) The absence of effective models of defects and faults for analog circuits, etc.

The methods of testing and fault diagnostics for AMS IC are realized using the specification-driven and data-driven concepts. In the first case the tests are generated to detect faulty circuits, which violate one or more parameters in the circuit specifications [7]. In accordance with the second concept, the decision on the status of the circuit under test (CUT) is made on the basis of statistical information on the operation of a fault-free circuit and a circuit with possible existing faults [8].

The increase in the functional and structural complexity of the AMC ICs determines the rising of complexity for corresponding fault diagnostics. More complexity, more potential faults, which require extra time for simulation and lead to extreme increasing the dimension of fault dictionaries. Fault simulation of analog circuits using modern computing systems based on the parallelism paradigm can be performed in a reasonable time by using several computing nodes and/or cores [9]. Nevertheless, such simulation takes essential computational and time costs. The fault-free and faulty behavior of the circuit has to be simulated using Monte Carlo analysis in order to take into account the tolerances of internal components. Not one response, but the family of responses is generated for each state of the circuit. A huge dataset with responses in test nodes is produced in the result of simulating a fault-free circuit and all considered faults.

The transition from the traditional fault dictionary represented in the form of a parametric table [10] to the neuromorphic fault dictionary (NFD) functioning in associative mode [11] was made possible by the use of machine learning methods. The NFD ensures a time reduction for detection and diagnostics of a potential fault. Meanwhile, there are traditional issues related with a neural network (NN) training, such as selecting an architecture of NN including a type of network, number of layers and neurons, type of a transfer function, etc.); selecting the training set; reduction of size for a training set, etc.

Overlapping or intersection of responses for different states brings the indeterminacy into the NN training and the states classification. Extraction of the ambiguity groups that include the states with similar behavior is important task for quality training the NN.

The big number of an output response's characteristics used for training the NN can lead to the problem of convergence or overfitting. Therefore, selection of essential characteristics that increase the distinguishability of the considered circuits' states is required for construction an efficient NFD. Design-for-testability (DFT) paradigm is recommended for an active use in practical development of reliable and easy-testable CUT. DFT provides a reduction of total time for design and test of the developed electronic circuits [12]. The use of DFT in the early stages of design is particularly important for the implementation of highly reliable AMS IC with guaranteed quality, since it allows taking into account the specifics of the project and promptly applying principal corrections to improve a testability of the CUT with minimal time and cost overheads. So, the DM and ML methods may be integrated into DFT-flow to improve the process of NFD construction. Two solutions compatible with DFT-flow of AMS IC: method of reduction the set of essential characteristics based on the PCA and approach to a cut down the training set using entropy estimation are considered below.

3 Data Preprocessing and Principal Component Analysis

Discretization of the transient response at output node and extraction of essential characteristics are required as preliminary pre-processing because of influence of nonlinear distortions, noise, shifting of phase, etc. Wavelet decomposition is offered for representation of a continuous output response as an analog signal in the time-domain by discrete values in the frequency-domain.

Wavelet decomposition (WD) describes of the one-dimensional signal as a generalized series or the Fourier integral (1) over a system of the basis functions (2) generated from the mother wavelet $\psi(t)$ by the time shift operations (b) and the time scale change (a). The factor $1/\sqrt{a}$ provides the independence of the norms of these functions to the scaling number a [13]

$$W(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}^{*}(t)dt, \tag{1}$$

$$\psi_{ab} = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right). \tag{2}$$

$$WT : \mathbf{R} \to \mathbf{X}, \tag{3}$$

$$\mathbf{R} = [\mathbf{R}_i], \mathbf{R}_i = \left[\mathbf{r}_i^m\right], \mathbf{R} \in \Re_{NR\times t}, \tag{4}$$

where $\mathbf{r}_i^m$ is a vector by length t containing the time samples of the output response; t is the number of time points; $i = 1..N + 1$ corresponds to the number of the circuit states; m defines the number of Monte Carlo iterations, $NR = m\ N + m$; $\mathbf{X}$ is the matrix of the WD-coefficients. In result of the wavelet decomposition a continuous output response of the CUT is represented by a tuple of the scaling factor (a) and the shift (b): $\mathbf{x} = [a; b]$. The set of considered responses is a matrix of WD-coefficients:

$$\mathbf{X} = [\mathbf{x}_i], i = 1..N, \mathbf{X} \in \Re_{N\times M}. \tag{5}$$

The distinguishability in the responses corresponding to different states of the CUT can be provided by a subset of the WD-coefficients from the vector $\mathbf{x}$. The use of non-essential or minor coefficients for training the NN leads to an increase in complexity and computational costs. Moreover, the use of non-essential or minor coefficients affects the convergence and overfitting of the NN. Therefore, the dimension reduction for the vector $\mathbf{x}$ due to selecting the essential coefficients for distinguishability of various states of the CUT is an important task. The principal component analysis (PCA) is offered to reduce the dimension of the input vectors:

$$PCA : \mathbf{X} \to \hat{\mathbf{X}},\ \mathbf{X} \in \Re_{N\times M},\ \hat{\mathbf{X}} \in \Re_{N\times \hat{M}},\ M > \hat{M}. \tag{6}$$

The PCA provides the data mapping of one dimension onto a subspace of less order. The resulting residuals $\mathbf{E}$ may be considered as a noise without meaningful information. Data matrix $\mathbf{X}$ consists of I rows corresponding to the samples and J columns with independent variables. The formal variables $\mathbf{t}_a$, $a = 1, .., A$ as a linear combination of the initial variables x_j, $j = 1, .., J$ represent the principal components.

$$\mathbf{t}_a = \mathbf{p}_{a1}\mathbf{x}_1 + \mathbf{p}_{a2}\mathbf{x}_2 + \ldots + \mathbf{p}_{aJ}\mathbf{x}_J. \tag{7}$$

A defines the number of principal components ($A < I,\ A < J$).

Based on (7) the matrix **X** is described by the product (8)

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} = \sum_{a=1}^{A} \mathbf{t}_a \mathbf{p}_a^t + \mathbf{E}, \tag{8}$$

where $\mathbf{T} \in \Re_{I \times A}$ represents the matrix of scores, $\mathbf{P} \in \Re_{J \times A}$ – matrix of loads and $\mathbf{E} \in \Re_{I \times J}$ – matrix of residuals. The matrix of scores **T** represents the mapping of the original samples as *J*-dimensional vectors onto the *A*-dimensional subspace of the principal components. The coordinates of the samples in the new coordinate system are represented in rows of matrix **T**. The mappings of all samples onto new coordinate axis are orthogonal and represented in the columns of matrix **T**. The matrix of loads **P** represents the transition matrix from the original *I*-dimensional space of variables $\mathbf{x}_j$ into the *A*-dimensional subspace of the principal components. The coefficients relating the variables t and x according to (7) are the elements of rows in the matrix **P**.

PCA is realized iteratively by sequential inclusion of new principal components (PC) one after another [14]. Selection of a sufficient number of PCs and determination of a condition for stopping the iterative procedure are important issues. The aggregate sample of data will be incomplete with a small number of PCs, at large a revaluation arises.

The following criterion is suggested to select a sufficient number of PCs:

$$\arg\min_{PC}(ERV_{PC} > \ = ERV_{TH}), \tag{9}$$

where ERV_{PC} – explained residual variance; ERV_{TH} – threshold value.

Matrix $\hat{\mathbf{X}}$ contains the estimates obtained in result of processing the training set **X** during the PCA modelling.

The result of modeling by the PCA is a matrix $\hat{\mathbf{X}}$ containing the estimates found from the model built on the training set **X**. The matrix of residuals includes the deviations of the obtained estimates from the verification values

$$\mathbf{E} = \mathbf{X} - \hat{\mathbf{X}}. \tag{10}$$

The explained residual variance in (9) is calculated using elements of matrix **E**.

$$ERV = 1 - \sum_{i=1}^{I}\sum_{j=1}^{J} e_{ij}^2 / \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}^2, \ e_{ij} \in \mathbf{E}, \ x_{ij} \in \mathbf{X}. \tag{11}$$

So, the PCA reduces the number of coefficients in the input vector used for the NN training ($\hat{M} < M$).

4 An Approach to the Cut Down the Training Set Based on Entropy Calculation

The results of simulation CUT in the time domain considering fault-free and faulty behavior are used as a raw data for construction of the NFD. The Monte Carlo analysis (MCA) is used for taking into account the effect of tolerances on a behavior of the CUT. Every states of the CUT (fault-free and all considering faults) are simulated by the MCA multiple times. The transient responses are collected at test nodes. The direct responses are not appropriate for training the NN. Often, a spectral transformation of continuous signal in time-domain into discrete signal in the frequency-domain is realized. The wavelet decomposition is widely used for such purpose (1)–(3). The vector of coefficients $\mathbf{x}_i = \left[x_{ij}\right]$, $i,j \in \mathbb{Z}_+$, $x_{ij} \in \mathbb{R}$ is generated as a result for each i-th response, where i – number of the circuit responses, j – number of WD-coefficients for each response. Matrix $\mathbf{X} = [\mathbf{x}_i]$ (5) represents the behavior of the CUT including fault-free and faulty states. Each considered k-th state of the circuit is kept in corresponding submatrix $\mathbf{X}^k \in \Re_{m_k \times j}$, where $m_k \in \mathbb{Z}_+$ defines the number of iterations for the MCA.

In general case, the behavior of the CUT for different states may be similar, i.e.

$$\exists(i,j), \quad \mathbf{X}^i \cap \mathbf{X}^j \neq \emptyset, \quad i \neq j, \quad i,j \in \mathbb{Z}_+. \tag{12}$$

Then, final resolvability will be limited by the membership in an ambiguity group (AG). The condition defining an unambiguous diagnostics is the following:

$$\mathbf{X}_i^p \neq \mathbf{X}_j^r, \quad \forall i,j,p,r, \quad i \in \left[1..m_p\right], \quad j \in [1..m_r], \quad p \neq r. \tag{13}$$

Vector $\mathbf{x}_i$ may include non-essential coefficients from the fault diagnostics point of view. Reducing the uncertainty and complexity of the NN training using the vector $\mathbf{x}_i$ can be achieved through the selection essential characteristics that affect the increase in the distinguishability of various states for the CUT. The entropy-based approach is offered for selecting the appropriate coefficients in vector $\mathbf{x}_i$.

Entropy calculation is relied on the following transformations of matrix $\mathbf{X}$:

1. Generation of ambiguity groups for each j-th column by including into subset $\mathbf{AG}_j^k$ the coefficients with close values, where $k = 1..n_j$ and n_j corresponds to the number of ambiguity groups in the column j.
2. Replacement the coefficients at each column by the values of the AG numbers (k) according to a membership.

The matrix $\mathbf{S} = \left[s_{ij}\right], s_{ij} \in \mathbb{Z}_+$ including values of the AGs numbers is constructed in result of the ambiguity grouping $AG : \mathbf{X} \rightarrow \mathbf{S}$.

Let N_{pr}, $p \in [1..k]$ is the number of elements in the ambiguity group $\mathbf{AG}_r^p$.

The probability of appearance any state in a subset $\mathbf{AG}_r^p$ can be calculated as a ratio N_{pr}/m_i, m_i corresponds to the number of rows in $\mathbf{S}$. In this case, expression (14) can be used for calculating the entropy for each j-th column in $\mathbf{S}$

$$E_j = -\left[\frac{N_{1j}}{m_i}\log\left(\frac{N_{1j}}{m_i}\right) + \frac{N_{2j}}{m_i}\log\left(\frac{N_{2j}}{m_i}\right) + \ldots \right.$$
$$\left. \ldots + \frac{N_{kj}}{m_i}\log\left(\frac{N_{kj}}{m_i}\right)\right] = \log(m_i) - \frac{1}{m_i}\sum_{i=1}^{k} N_{ij}\log\left(N_{ij}\right). \tag{14}$$

The quantity of information based on coefficients in j-th column and significant for the fault diagnostics becomes maximum when entropy rate (ER) is minimal because the number of rows m_i is a fixed number.

$$\arg\max_j\left(E_j\right) = \arg\min_j\left(ER_j\right), \tag{15}$$

$$ER_j = \sum_{i=1}^{k} N_{ij}\log\left(N_{ij}\right). \tag{16}$$

The following algorithm describes steps for selection of the significant coefficients using entropy:

1. Generation of matrix $\mathbf{S} = \left[s_{ij}\right]$ by the membership numbers of coefficients from $\mathbf{X}$ at corresponding AGs.
2. Calculating the cardinality of each AG for every j-th column in $\mathbf{S}$.
3. Computation of the entropy rates $ER(j)$ for all columns in $\mathbf{S}$.
4. Include the j-th coefficient which provide minimal value of the entropy rate into sample set and exclude the j-th column from subsequent computations.
5. Rearrange the matrix $\mathbf{S}$ according to order of AGs for the chosen coefficient. Exclude from $\mathbf{S}$ all rows the cardinality of AG in chosen j-th column for which is equal 1. So, k submatrices $\mathbf{S}^i$ ($i = 1..k$) are generated in result for k ambiguity groups in the j-th chosen column of $\mathbf{S}$.
6. Computation of the entropy rates $ER(j)$ for the remained coefficients considering the AGs independently for each of generated submatrices $\mathbf{S}^i$.

$$ER(j) = \sum_{i=1}^{k} ER^i(j), \tag{17}$$

where $ER^i(j)$ is the ER estimated for the submatrix $\mathbf{S}^i$.

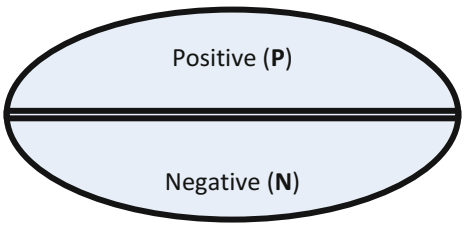
7. Stop the process if $ER(j)$ is equal to 0 or $ER(j)$ has the same value for two consecutive iterations for all j. Else return to step 4.

5 Metrics for the Machine Learning Quality Estimation

Neural network can be considered as a particular case of binary classifier, which generates "Yes/No" status at the output as the reaction on the input pattern. In the ideal case the quality of classification by NN can be estimated by the classical matching the statistical hypotheses (Fig. 1).

The performance of real NN classifier is evaluated by a test dataset, which contains the observed states that are compared with the predicted states. In common case, the predicted states usually match with only part of the observed states. There are four possible outcomes, which are used for construction the confusion matrix (Fig. 2):

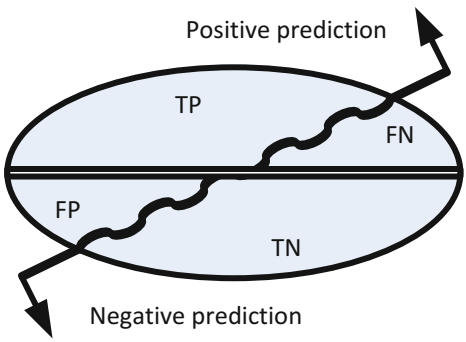
- True Positive (TP) defining the correct positive prediction,
- False Positive (FP) defining the incorrect positive prediction,
- True Negative (TN) defining the correct negative prediction,
- False Negative (FN) defining the incorrect negative prediction.



		Hypothesis	
		P	**N**
Observation	**P**	Fault-Free	β-error
	N	α-error	Faulty

Fig. 1. The chart and rules for the matching the statistical hypotheses.

The confusion matrix contains values of four outcomes obtained for a binary classifier. As rule, they are called simply TP, TN, FP and FN, but not "the number of True Positive, etc."



		Predicted	
		P	**N**
Observed	**P**	TP	FN
	N	FP	TN

Fig. 2. The chart of classification and corresponding confusion matrix.

The confusion matrix is used for estimating different metrics of the classification quality. There are the following metrics:

Error rate (ERR) is a ratio of aggregate incorrect predictions to the total number of applied dataset. The best value of ERR is 0.0, whereas the worst one is 1.0

$$ERR = \frac{FP + FN}{TP + TN + FN + FP} = \frac{FP + FN}{P + N}. \tag{18}$$

Accuracy (ACC) is the ratio of aggregate correct predictions to the total number of the applied dataset. The best value of ACC is 1.0, whereas the worst one is 0.0.

$$ACC = \frac{TP + TN}{TP + TN + FN + FP} = \frac{TP + TN}{P + N}. \tag{19}$$

Sensitivity (SN) is a ratio of correct positive predictions to the total number of positives. SN is called as a true positive rate or recall. The best value of SN is 1.0, whereas the worst one is 0.0.

$$SN = \frac{TP}{TP + FN} = \frac{TP}{P}. \tag{20}$$

Specificity (SP) is the ratio of correct negative predictions to the total number of negatives. SP may be called also as a true negative rate. The best value of SP is 1.0, whereas the worst one is 0.0.

$$SP = \frac{TN}{TN + FP} = \frac{TN}{N}. \tag{21}$$

Precision (PREC) or positive predictive value (PPV) is a ratio of correct positive predictions to the total number of positive predictions. The best value of PREC is 1.0, whereas the worst one is 0.0.

$$PREC = \frac{TP}{TP + FP}. \tag{22}$$

False positive rate (FPR) is a ratio of incorrect positive predictions to the total number of negatives. The best value of FPR is 0.0 whereas the worst one is 1.0.

$$FPR = \frac{FP}{TN + FP} = \frac{FP}{N} = 1 - SP. \tag{23}$$

There are two metrics used for estimating the quality of NN training and the NFD operating for fault diagnostics. The first indicator is *fault coverage*

$$FC = \frac{N_{detected}}{N_{total}}, \tag{24}$$

where $N_{detected}$ – the number of detected faults, N_{total} – the total number of considered faults.

The second indicator is *coverage of a detached faults*

$$FC_i = \frac{N^i_{detected}}{N^i_{total}}, \tag{25}$$

where $N^i_{detected}$ – the number of detected i-th faults, N^i_{total} – the total number of i-th faults.

If fault coverage provides inappropriate value less to the required level, then, firstly, the set of considered WD-coefficients are revised and, secondly, the NN should be retrained.

6 Experimental Case Study

Application of the neural network for constructing a fault dictionary is considered for examples of analog filters. The Sallen-Key filter and the Second-order band-pass filter are used as a case study for efficiency estimation of the PCA-based method on reduction the set of essential characteristics and the approach to a cut down the training set based on entropy. The MATLAB Neural Network Toolbox and computer based on Intel® Core™ i7-4770 CPU @ 3.40 GHz and RAM 8 GB were used at numerical experiments for constructing and training the NFD.

6.1 Reduction the Set of Essential Characteristics Using PCA

The study of the PCA-based method was carried out for an analog Sallen-Key filter as the typical benchmark circuit (Fig. 3). The raw data for experiment are specified in Table 1.

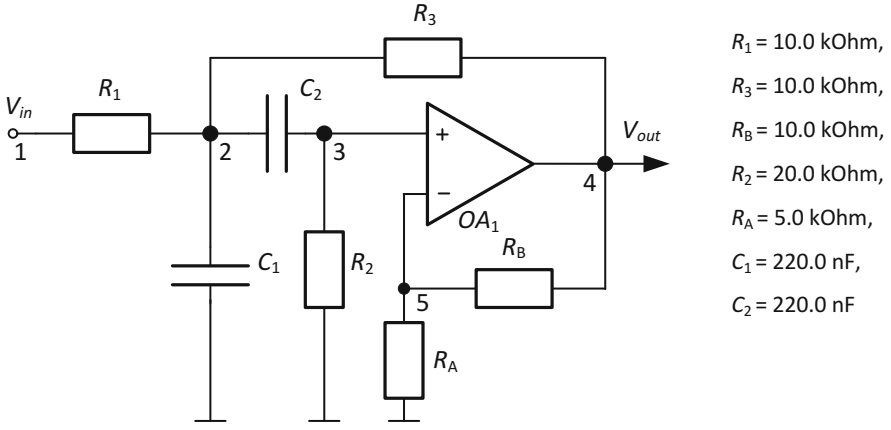


Fig. 3. Sallen-Key analog filter.

Table 1. Initial data for simulation of the Sallen-Key analog filter.

Parameter	Value
Input test signal:	
Waveform	Sine-wave
Amplitude	1 V
Frequency	72 Hz
Components' tolerances	±10%
Number of faults:	
Catastrophic	14
Parametric	14
Transient-process time	80 ms

The open circuit and short circuit effects were considered as the catastrophic faults relative to all components in the circuit. Deviations ±50% of the nominal value for internal components defined the effect of parametric faults. Monte Carlo analysis in the time-domain was used for simulating one fault-free and 28 faulty states of the filter. 20 000 iterations were used for Monte Carlo simulation of fault-free state and 5 000 iterations for every faulty state of the filter. Node 4 was used as a test node for measuring the transient output responses.

The output responses in time-domain were converted into discrete coefficients in frequency-domain by wavelet decomposition on the basis of fourth-order Daubechy wavelet (db4). Decomposition of each response has provided vector with 148 WD-coefficients. The training set was constructed by 600 responses of 20000 for fault-free state and by 200 responses of 5000 for every faulty state. The final matrix of WD-coefficients for training purpose $\mathbf{X}_{train}$ has a dimension 6200 by 148 elements. The remained responses not included into the training set was used as a test set $\mathbf{X}_{test}$.

The principal component analysis for $\mathbf{X}_{train}$ based on Eqs. (7), (8) and (11) has provided constructing the matrices of scores and loadings and estimating the explained residual variance. The threshold $ERV_{TH} \geq 0.95$ was assigned for limitation the minimal number of PCs. In result, the final training set $\hat{\mathbf{X}}_{train}$ was limited by only 62 coefficients for every response. So, the use of 95% of meaningful information has led to reducing the dimension of the actual training set by almost 2.4 times.

Implementation of the NFD is based on three-layer architecture of the NN. 62 neurons were used in the input layer according to an actual number of coefficients in the training set. 29 neurons corresponding to the number of considered states of the filter were used in the output layer of the NFD. Table 2 represents the time characteristics for the NN training.

The adequacy of the transition from the architecture 2 to architecture 1 was confirmed by comparing the values of fault coverage for the testing (FC_{test}) and diagnostics (FC_{diag}) modes estimated for detached faults from the test set $\hat{\mathbf{X}}_{test}$. Table 3 represents values of the overall FC for testing and diagnostics modes evaluated for different architectures of the NFD.

Table 2. Time characteristics on training of the NN for Sallen-Key analog filter.

# of NN	1	2
Architecture of neural network[a]	62:20:29	148:20:29
Used method(s)[b]	WD + PCA	WD
ERV	0.9505	–
Time per training epoch, s.	1.16	609.79

[a]Demonstrates the number of neurons in the first:second: third layers of the NN.
[b]WD: wavelet decomposition of output responses, WD +PCA – principal component analysis of matrix with WD-coefficients of the output responses.

The obtained results demonstrate an effectiveness of the proposed PCA-based method on reduction the set of essential characteristics for the output responses, and consequently effected a decreasing both time and computational costs for the NFD training. The limitation of the number of PCs by threshold $ERV_{TH} \geq 0.95$ provides reducing the dimension of the training set by 2.39 times and decreasing the training time by more 10 times. In result, fault coverage for the detached faults consists of up to 100% and overall fault coverage consists of up to 99.9% for the obtained neuromorphic fault dictionary.

Table 3. Fault coverage for testing and diagnostics modes of Sallen-Key analog filter.

# of NN	Testing			Diagnostics		
	FC_{all}	FC_{cat}	FC_{par}	FC_{all}	FC_{cat}	FC_{par}
1	0.985	0.999	0.972	0.901	0.953	0.848
2	0.986	0.999	0.972	0.905	0.947	0.862

FC_{all} is the estimated fault coverage for all types of considered faults;
FC_{cat} is the estimated fault coverage for all considered catastrophic faults;
FC_{par} is the estimated fault coverage for all considered parametric faults.

6.2 Entropy-Based Approach to a Cut Down the Training Set

The study of the approach to a cut down the training set based on entropy estimation is performed for the benchmark circuit an analog Second-order band-pass filter (Fig. 4). The raw data for experiment are specified in Table 4.

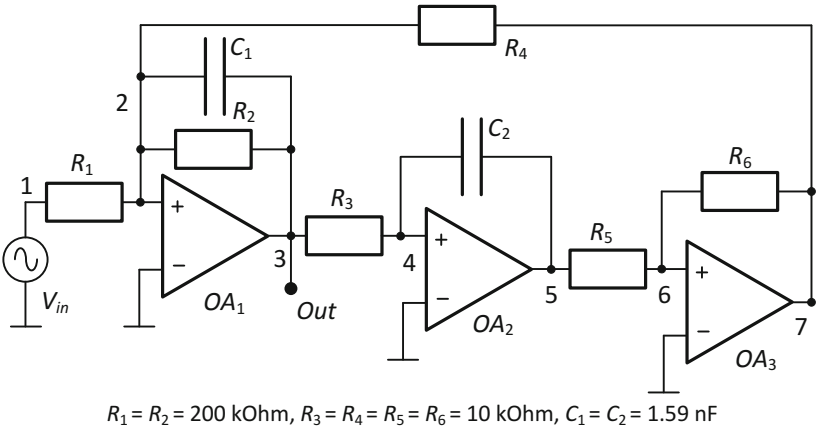


Fig. 4. Second-order band-pass filter.

Table 4. Initial data for simulation of the second-order band-pass analog filter.

Parameter	Value
Input test signal:	
Waveform	Sine-wave
Amplitude	1 V
Frequency	10 kHz
Components' tolerances:	
Resistors	±5%
Capacitors	±10%
Number of faults:	
Catastrophic	16
Parametric	16
Transient-process time	6 ms

The open circuit and short circuit effects were considered as the catastrophic faults relative to all components in the circuit. Deviations ±50% of the nominal value for internal components defined the effect of parametric faults. Monte Carlo analysis in the time-domain was used for simulating one fault-free and 32 faulty states. 20000 iterations were used for Monte Carlo simulation of fault-free state and 5000 iterations for every faulty state of the filter. Node 3 was used as a test node for measuring the transient output responses.

The output responses in time-domain were converted into discrete coefficients in frequency-domain by wavelet decomposition on the basis of fourth-order Daubechy wavelet (db4). Decomposition of each response has provided vector with 108 WD-coefficients. The training set was constructed by 600 responses of 20000 for fault-free state and by 200 responses of 5000 for every faulty state. The final matrix of

WD-coefficients for training purpose $\mathbf{X}_{train}$ has a dimension 7000 rows by 108 elements. The remained responses not included into the training set was used as a test set $\mathbf{X}_{test}$.

Implementation of the NFD is based on three-layer architecture of the NN. The number of neurons in the input layer corresponds to the quantity of coefficients in the row of the matrix with a training sets. The original matrix $\mathbf{X}_{train}$ provides 108 neurons in the input layer of the NN. Only 11 essential coefficients were selected in result of applying the proposed approach based on entropy calculation. So, a size of the training set was reduced by 9.82 times.

The following 20 ambiguity groups were extracted from the training set during entropy computation:

AG_1 = {FF}, AG_2 = {R_1_short},
AG_3 = {R_2_short, C_1_short},
AG_4 = {R_3_short, R_4_short, R_5_short},
AG_5 = {R_6_short, C_2_short, R_3_open, R_4_open, R_5_open, C_2 + 50%},
AG_6 = {R_1_open}, AG_7 = {R_2_open}, AG_8 = {R_6_open},
AG_9 = {C_1_open}, AG_{10} = {C_2_open}, AG_{11} = {R_1 + 50%, R_2 – 50%},
AG_{12} = {R_2 + 50%}, AG_{13} = {R_3 + 50%, R_4 + 50%, R_5 + 50%},
AG_{14} = {R_6 + 50%}, AG_{15} = {C_1 + 50%}, AG_{16} = {R_1 – 50%},
AG_{17} = {R_3 – 50%, R_4 – 50%, R_5 – 50%}, AG_{18} = {R_6–50%},
AG_{19} = {C_1 – 50%}, AG_{20} = {C_2 – 50%}.

It means, the WD-coefficients for the states (faults) included in the same ambiguity group are similar or close each other for all samples and distinguish of such states is impossible. Therefore, the NN is trained for fault classification accurate within the ambiguity groups. So, the number of output neurons in the NN is equal 20. The training time for original training set and after reduction is presented in Table 5.

Table 5. Time characteristics on training of the NN for the second-order band-pass analog filter.

# of NN	1	2
Architecture of neural network[a]	11:20:20	108:20:20
Used method(s)[b]	WD + En	WD
Time per training epoch, s.	1.16	609.79

[a]Demonstrates the number of neurons in the first: second: third layers of NN.
[b]WD: wavelet decomposition of output responses, WD +En – entropy estimation of the matrix with WD-coefficients of the output responses.

Table 6 represents values of the overall FC for the testing and diagnostics modes evaluated for different architectures of the NFD.

Table 6. Fault coverage for testing and diagnostics modes of the second-order band-pass analog filter.

# of NN	Testing			Diagnostics		
	FC_{all}	FC_{cat}	FC_{par}	FC_{all}	FC_{cat}	FC_{par}
1	0.9530	0.9927	0.9203	0.8557	0.9610	0.7691
2	0.9571	0.9931	0.9211	0.8509	0.9344	0.7674

FC_{all} is the estimated fault coverage for all types of considered faults;
FC_{cat} is the estimated fault coverage for all considered catastrophic faults;
FC_{par} is the estimated fault coverage for all considered parametric faults.

The proposed approach has demonstrated a high efficiency in a cutting down the dimension of the training set in 9.82 times due to selection of essential coefficients from the fault diagnostics point of view, and decreasing the time on training the NN in 525.68 times. In result, fault coverage for the detached faults consists of up to 100% and overall fault coverage consists of up to 99.27% for the obtained neuromorphic fault dictionary.

7 Conclusion

The methods of machine learning are widely used for constructing a NFD, which ensure fault diagnostics of the AMS IC in the associative mode. Many problems of neural network training on a large volume datasets may be solved by reduction of the training sets dimension and consideration only essential coefficients. The solutions proposed in the paper ensure an efficient selection of the essential coefficients and a training set reduction that provide construction of the effective NFD for testing and fault diagnostics in the AMS IC.

Experimental results for both cases demonstrate the essential reduction of initial training set and saving time on the NFD training with high fault coverage. The case study for Sallen-Key filter and Second-order band-pass filter provides coverage up to 100% of detached faults and at least up to 99% of the overall fault coverage. The obtained results are promising for a wide active application of the offered solutions.

The proposed method and approach may be used according to the design-for-testability flow for analog/mixed-signal integrated circuits.

Acknowledgement. The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

References

1. Chakrabarti, S., Chatterjee, A.: Compact fault dictionary construction for efficient isolation of faults in analog and mixed-signal circuits. In: Proceedings of 20th Anniversary Conference on Advanced Research in VLSI (ARVLSI 1999), Atlanta, GA, pp. 327–341 (1999). https://doi.org/10.1109/ARVLSI.1999.756057
2. Aminian, M., Aminian, F.: A modular fault-diagnostic system for analog electronic circuits using neural networks with wavelet transform as a preprocessor. IEEE Trans. Instrum. Meas. **56**(5), 1546–1554 (2007). https://doi.org/10.1109/TIM.2007.904549
3. Huang, K., Stratigopoulos,H.G., Mir, S.: Fault diagnosis of analog circuits based on machine learning. In: Proceedings of Design, Automation and Test in Europe Conference and Exhibition (DATE 2010), Dresden, pp. 1761–1766 (2010). https://doi.org/10.1109/DATE.2010.5457099
4. Lang, R., Xu, Z., Gao, F.: Data-driven fault diagnosis method for analog circuits based on robust competitive agglomeration. J. Syst. Eng. Electron. **24**(4), 706–712 (2013). https://doi.org/10.1109/JSEE.2013.00082
5. Sun, Y., Chen, G., Li, H.: Fuzzy clustering and SVM method for diagnosing analog circuits single soft fault. J. Comput.-Aided Des. Comput. Graph. **20**(5), 612–617 (2008)
6. Xiong, J., Tian, S., Yang, C., Yu, B., Chen, S.: Analog fault feature extraction and classification based on LMD and LVQ neural network. In: Proceedings of International Conference on Algorithms, Methodology, Models and Applications in Emerging Technologies (ICAMMAET), Chennai, pp. 1–6 (2017). https://doi.org/10.1109/ICAMMAET.2017.8186706
7. Variyam, P.N., Chatterjee, A.: Specification-driven test design for analog circuits. In: Proceedings of IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems, Austin, TX, pp. 335–340 (1998). https://doi.org/10.1109/DFTVS.1998.732183
8. Vasan, A.S.S., Long, B., Pecht, M.: Diagnostics and prognostics method for analog electronic circuits. IEEE Trans. Industr. Electron. **60**(11), 5277–5291 (2013). https://doi.org/10.1109/TIE.2012.2224074
9. Mosin, S.: Automated simulation of faults in analog circuits based on parallel paradigm. In: Proceedings of IEEE East-West Design and Test Symposium (EWDTS), Novi Sad, pp. 1–6 (2017). https://doi.org/10.1109/EWDTS.2017.8110133
10. Slamani, M., Kaminska, B.: Analog circuit fault diagnosis based on sensitivity computation and functional testing. IEEE Des. Test Comput. **9**(1), 30–39 (1992). https://doi.org/10.1109/54.124515
11. Mosin, S.: A technique of analog circuits testing and diagnosis based on neuromorphic classifier. Adv. Intell. Syst. Comput. **425**, 381–393 (2016). https://doi.org/10.1007/978-3-319-28658-7_33
12. Mosin, S.G.: Design-for-testability automation of mixed-signal integrated circuits. In: Proceedings of IEEE 26th International SOC Conference (SOCC 2013). Erlangen, Germany, pp. 244–249 (2013). https://doi.org/10.1109/SOCC.2013.6749695
13. Debnath, L., Shah, F.: Wavelet Transforms and Their Applications. Birkhäuser, Boston (2015). https://doi.org/10.1007/978-0-8176-8418-1
14. Jolliffe, I.T.: Principal Component Analysis. Springer, New York (2002). https://doi.org/10.1007/b98835

Dimensionality Reduction by Dynamic Mode Decomposition for Hyperspectral Image Classification Using Deep Learning and Kernel Methods

K. S. Charmisha(✉), V. Sowmya, and K. P. Soman

Center for Computational Engineering and Networking (CEN),
Amrita School of Engineering, Amrita Vishwa Vidyapeetham,
Coimbatore 641112, India
charmisha99@gmail.com

Abstract. Hyperspectral images are remotely sensed high dimension images, which capture a scene at different spectral wavelengths. There is a high correlation between the bands of these images. For an efficient classification and processing, the high data volume of the images need to be reduced. This paper analyzes the effect of dimensionality reduction on hyperspectral image classification using vectorized convolution neural network (VCNN), Grand Unified Recursive Least Squares (GURLS) and Support Vector Machines (SVM). Inorder to analyze the effect of dimensionality reduction, the network is trained with dimensionally reduced hyperspectral data for VCNN, GURLS and SVM. The experimental results shows that, one-sixth of the total number of available bands are the maximum possible reduction in feature dimension for Salinas-A and one-third of the total available bands are for Indianpines dataset that results in comparable classification accuracy.

Keywords: GURLS · SVM · Hyperspectral image classification
Dimensionality reduction · Dynamic mode decomposition · Libsvm
Gurls · Classification accuracies

1 Introduction

Hyperspectral images (HSI) are remotely sensed images, which consists of large number of pixels, where each pixel have narrow spectral bands with lots of spectral information. Human eye can detect a small portion of the bands in electromagnetic spectrum. The electromagnetic spectrum contains the bands of x-rays, ultra-violet, radio waves, infra-red radiations and micro waves [1]. The hyperspectral images (HSI) contain enormous amount of information of all these bands.

In the field of remote sensing, HSI classification is one of the major areas of research [5]. Deep learning based methods have achieved great benchmark results

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 256–267, 2019.
https://doi.org/10.1007/978-981-13-5758-9_22

in many fields. Convolution neural networks (CNN) gives better classification rate on vision related tasks [2]. The advantages of using CNN compared to conventional techniques are of its self organisation, real time operations and adaptive learning.

The main challenges of hyperspectral images are, its huge dimension. The bands of HSI are highly correlated. One way to reduce the dimension of these correlated bands is by projecting the data into a new lower dimension without any feature loss. The most commonly used dimension reduction techniques are: Principal Component Analysis (PCA), Singular Value Decomposition (SVD), which are standard reduction techniques used to reduce the unwanted information.

A new hyperspectral dimension reduction technique called Dynamic Mode Decomposition (DMD) is recently incorporated for HSI classification [6]. This method is used to find spectral variations in the HSI bands. Here, kernel based classification methods are also used for the classification of HSI images. Dimensionally reduced hyperspectral images are classified using different kernels available in standard software packages such as LIBSVM and Grand Unified Regularized Least Squares (GURLS). The main objective of this work is to analyze the effect of dimensionality reduction on hyperspectral image classification using VCNN, SVM and GURLS. By computing the total number of parameters required for training the network before and after the reduction, VCNN is analyzed. GURLS and SVM techniques are used to provide a precise classification accuracy.

The organization of the paper is as follows: Sect. 2, provides a background study of VCNN, DMD, GURLS and SVM. Section 3, gives the methodology or the overview of overall flow of algorithm. Section 4, the experimental results are analyzed. Section 5, provides the conclusion.

2 Background Study

2.1 Vectorized Convolution Neural Network (VCNN)

The 3D hyperspectral data m x n x b, where m denotes scan lines, n denotes samples and b bands, which contains hundreds of spectral bands can be illustrated as 2D curves [4]. For the purpose of classification, CNN is used. CNN is a feed forward neural network, which is very much similar to ordinary neural network. CNN contains many number of layers depending on the number of layers and the filters the performance of the network can be estimated. Computation time and complexity increases as the number of layers and filters in an network increases. The input to the network is the pixel vectors, hence called the VCNN. The architecture of VCNN is shown in [4]. The network consists of five layers, an input layer (L1), a convolutional layer (C1), a maxpooling layer (P1), and a fully connected layer (F1) and an output layer (L2). The detailed description of the network is shown in [3].

Total number of trainable parameters for VCNN are calculated as:

$$20 \times (k_1 + 1) + (20 \times n_3 + 1) \times n_4 + (n_4 + 1) \times n_5 \tag{1}$$

where k_1 denotes the size of filters and is given as $k_1 = n_1/9$ and n_1 is the input (number of bands). $n_3 = n_2/k_2$, here, k_2 denotes the kernel size and $n_2 = (n_1 - k_1/s) + 1$, s denotes the stride, equal to one. $n_4 = 100$ and n_5 denotes the number of classes.

2.2 Dynamic Mode Decomposition (DMD)

DMD is used to analyze and understand the high complex non linear systems. For a given non-linear system, DMD is used to find the dynamicity. Spectral decomposition of this map results in an eigen value and eigen vector representation (referred to as dynamic modes) [7]. 2-D matrix is created by vectorizing and appending each band as columns. DMD process is well explained in paper [6]. The DMD process is explained through the Eq. 2.

Computation of $\tilde{S}$ is shown in [4].

$$\tilde{S} = (V\Sigma^{-1})^{-1} S (V\Sigma^{-1}) = U^H X_2^m V\Sigma^{-1} \tag{2}$$

From Eq. 2, the final output matrix contains the information of the most informative bands at the top and the least informative at the bottom (descending order).

2.3 LIBSVM

LIBSVM is used for the classification purpose. This method works on the concept of the hyperplane, which separates the data belonging to different classes. The data points that are nearer to the hyperplane are called support vectors. Support Vector Machine (SVM) is suitable for only linearly separable data. The main aim of SVM is to maximize the margin between the hyperplanes, due to which there is less misclassification of the data. The various kernels and kernel functions are used for mapping to the higher dimensional space. The type of kernel selection is data dependent. The most commonly used kernels are Linear, Polynomial, Radial Basis Function (RBF). The SVM finds its best use in the following points shown below.

- It is used in the formulations of different SVM'S.
- Cross validation for any model selection is done using SVM.
- SVM is much efficient in multi-class classification.
- Estimating the probability is one of the key factors of SVM classification.
- Weighted SVM techniques are also used for the unbalanced data classification.

2.4 GURLS

GURLS - Grand Unified Regularized Least Squares is a software library, which is used for the supervised learning techniques. To solve the machine learning problems such as regression and classification, the properties of regularized least square algorithm is used. For multi category problems, this library is used. The advantages of the GURLS are:

- Distribution of the task execution.
- Mapping of the storage.
- Automatic parameter selection for model selection.
- Quick training and testing procedures with large number of data-points.
- Handling of large dataset with memory mapped storage.
- Ease of use of the data.
- Good estimated results in high dimensional multi output problems.

In GURLS library, Xtr, Ytr, Xte and Ypr represents training data, training labels, testing data and predicted labels respectively.

In GURLS library, Xtr, Ytr, Xte and Ypr represents training data, training labels, testing data and predicted labels respectively. During the training time the weight matrix W is calculated from X_{tr} and Y_{tr}. Weight matrix is used to predict the output labels during the testing time.

$$Y = WX_{te} \tag{3}$$

3 Methodology

The main objective of this work is to analyze the effect of dimensionality reduction on hyperspectral image classification using VCNN, SVM, GURLS. At the initial stage the network is trained with the raw data, and is classified using VCNN, SVM and GURLS. For VCNN each band of HSI is vectorized to form a 2D image of size(b x mn), here b is the bands, and m, n are the rows and columns in a band. Before feeding to VCNN, normalization of the data is done in the range of $[-1, +1]$. The entire data is split into aspects: 80% of the data is for used for training and the remaining 20% is for testing (Fig. 1).

The VCNN network is tuned by varying the hyper-parameters like number of iterations and learning rate. Dynamic mode decomposition (DMD) technique is used to reduce the bands. Initially 50% of the bands are reduced, later this reduction process is extended till maximum possible reduction of the bands is obtained without any information loss for the feature extraction. By computing the total number of trainable parameters of the network before and after dimensionality reduction, VCNN can be analyzed. Classification accuracy is computed for the raw data by VCNN. GURLS and SVM classification techniques are used for both raw and maximum possible reduced dimension data to evaluate the classification results on comparison with VCNN.

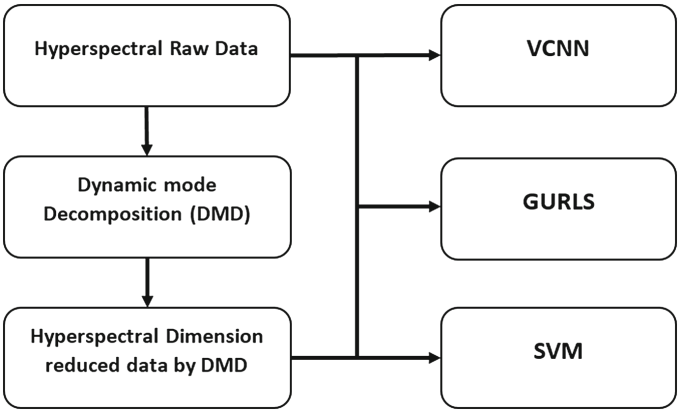


Fig. 1. Block diagram of the proposed hyperspectral image classification

4 Classification accuracies

Classwise accuracy is the ratio of correctly classified data in each class to the total number of data in each class. Overall accuracy gives the total number of correctly classified pixels.

$$Classwiseaccuracy\,(CA) = \frac{No:of\ pixels\ \ correctly\ \ classified\ \ in\ \ each\ \ class}{Total\ no:\ \ of\ pixels\ in\ \ each\ \ class} \times 100$$

$$Overallaccuracy\,(OA) = \frac{Total\ \ No:of\ pixels\ \ correctly\ \ classified}{Total\ no:of\ pixels} \times 100$$

5 Experimental Results

5.1 Dataset

Indianpines dataset is used for the analysis. The details of the dataset are shown as follows:

- Spatial configuration is 143×143 pixels.
- Spectral configuration is 220 bands.
- Wavelength range is of 0.4–2.2 μm.
- The groundtruth has 16 classes.
- Eight classes are considered for the analysis.
- The classes are choosed based on maximum pixels ratio.

Salinas-A is the second data set used for analysis, the details of the data set are shown as follows:

- Spatial configuration is 86×83 pixels.
- Spectral configuration is 224 bands.

- Wavelength range is of 0.4–2.5 μm.
- Six classes are considered for the analysis.
- The classes are chose based on maximum pixels ratio.

5.2 Results and Analysis

The network is tuned by setting two core parameters-learning rate and number of iterations. The parameter is varied one at a time say: when learning rate is varied, the number of iterations are fixed, similarly vice versa is applicable. We vary the parameters till a high classification accuracy is obtained. First, the learning rate is fixed to a small number, as the lower learning rate always avoids over-fitting. Then, we vary the number of iterations until a high classification accuracy is obtained. This process is repeated for both the datasets.

Table 1. Comparison of classification accuracies and total number of trainable parameters for Indianpines dataset

Indianpines dimension reduction	Classification accuracy (%)	Trainable parameters
220-(raw data)	84.63%	79,408
110-(1/2 of total bands)	83.62%	67,168
73-(1/3 of total bands)	83.67%	45,088
55-(1/4 of the total bands)	**80.32%**	**33,048**

From the above Table 1 it is seen that, when this dimension reduction is further carried on for one-fourth (55) of the total bands, we observe that, same (CA) is retained until one-third of the total bands reduction (73), whereas for one-fourth of the total band reduction (55), we are able to reduce the trainable parameters to a great extent of 33,048, but the CA is around 80.32%, which shows that further reduction in the feature dimension led to misclassification. Similarly from the Table 2, when this dimension reduction is further carried on for one-seventh (32) of the total bands, we observe that, same classification

Table 2. Comparison of classification accuracies and total number of trainable parameters for Slainas-A dataset

Salinas-A dimension reduction	Classification accuracy (%)	Trainable parameters
224-(raw data)	91.36%	81,206
112-(1/2 of total bands)	91.33%	68,866
74-(1/3 of total bands)	91.2%	46,786
56-(1/4 of the total bands)	91.02%	36,746
44-(1/5 of total bands)	90.7%	28,706
37-(1/6 of total bands)	90.56%	24,706
32-(1/7 of total bands)	**87.2%**	**22,686**

accuracy is retained until one-sixth of the total bands reduction (37), whereas for one-seventh of the total band reduction (32), we are able to reduce the trainable parameters to a great extent of 22,686 but the CA is around 87.2%, which shows that further reduction in the feature dimension led to misclassification. Hence, we consider the maximum possible extent of dimension reduction of the bands for the both datasets is **73** for Indianpines and 37 for Salinas-A.

1. From Fig. 2, it is observed that the maximum possible classification accuracy **91.3%** is obtained for the learning rate of **0.02** amongst all other learning rates for the raw data consisting of 224 bands. Hence, the learning rate is fixed to **0.02**. Similarly, for the reduced dimension of 37 bands (one sixth of the total bands) the learning rate of **0.06** attains the high classification accuracy of **90.56%**. So, it is fixed to **0.06**.
2. From Fig. 5, it is observed that the maximum possible classification accuracy of **84.63%** is obtained for the learning rate **0.07** amongst all other learning rates for the raw data consisting of 220 bands. Hence, the learning rate is fixed to **0.07**. Similarly for the reduced dimension of the 73 bands (one third of the total bands) the learning rate of **0.08** attains the high classification accuracy of **83.67%.** So, it is fixed to **0.08**.
3. Now, the learning rate is fixed and the number of iterations are varied. From Fig. 3 at the iteration of **13,000**, we get a high classification accuracy of **91.3%** for the raw data of 224 bands as compared to all other iterations tuned. Hence, the iteration value of **13,000** is fixed. Similarly, for the iteration of **14,000**, we achieve a high classification accuracy of **90.56%** for the reduced bands of 37 (one-sixth of the total bands).
4. From Fig. 4 at the iteration of **12,000**, we get a high classification accuracy of **84.63%** for the raw data of 220 bands as compared to all other iterations tuned. Hence, the iteration value of **13,000** is fixed. Similarly, for the iteration of **15,000**, we achieve a high classification accuracy of **83.67%** for the reduced bands of 73 (one-sixth of the total bands).
5. The layer parameters of Salinas-A raw dataset (224 bands) for VCNN are: $n_1 = 224$, $k_1 = 24$, $n_2 = 201$, $k_2 = 5$, $n_3 = 40$, $n_4 = 100$, $n_5 = 6$. The total number of trainable parameters are **81,206** computed using Eq. 1.
6. The layer parameters of Salinas-A dimension reduced dataset (37 bands-one sixth of total bands) for VCNN are: $n_1 = 37$, $k_1 = 4$, $n_2 = 34$, $k_2 = 3$, $n_3 = 11$, $n_4 = 100$, $n_5 = 6$. The total number of trainable parameters are **24,706** computed using Eq. 1.
7. The layer parameters of Indianpines raw dataset (220 bands) for VCNN are: $n_1 = 220$, $k_1 = 24$, $n_2 = 197$, $k_2 = 5$, $n_3 = 39$, $n_4 = 100$, $n_5 = 8$. The total number of trainable parameters are **79,408** computed using Eq. 1.
8. The layer parameters of Indianpines dimension reduced dataset (73 bands-one third of the total bands) for VCNN are: $n_1 = 73$, $k_1 = 8$, $n_2 = 66$, $k_2 = 3$, $n_3 = 22$, $n_4 = 100$, $n_5 = 8$. The total number of trainable parameters are **45,088** computed using Eq. 1.

Below shows the outputs of each layer for raw data as well as maximum possible extent of dimension reduced data for Indianpines and Salinas-A (Table 3).

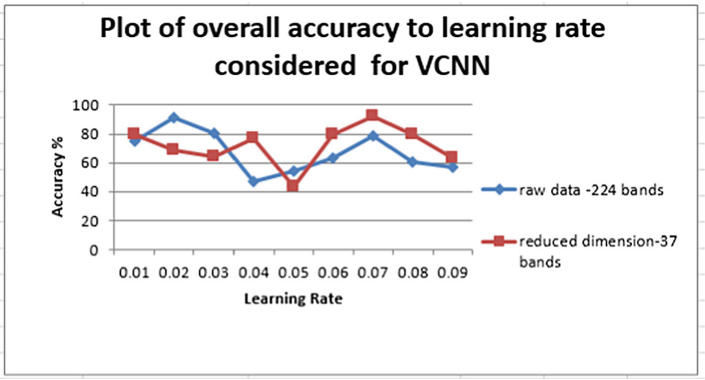


Fig. 2. Plot of overall accuracy vs. learning rate considered for VCNN with and without dimensionality reduction for Salinas-A

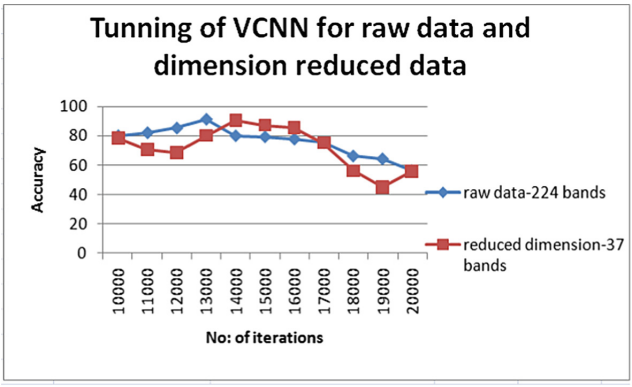


Fig. 3. Plot of overall accuracy vs. number of iterations considered for VCNN with and without dimensionality reduction for Salinas-A

5.3 Classification Accuracy

Through DMD, we are able to reduce the bands to the maximum possible extent, maintaining the same classification accuracy as that of the raw data for both Salinas-A and Indianpines. The maximum reduction for Salinas-A is 37 bands (one-sixth of the total bands) and for the Indianpines is 7 bands (one-third of the total bands).

Now, GURLS and SVM classification techniques are used to compare with the VCNN classification accuracy. Classification accuracy and overall accuracy are calculated as shown below: For GURLS linear kernel is used and for SVM polynomial kernel is used, as they are effective in achieving the high classification accuracy.

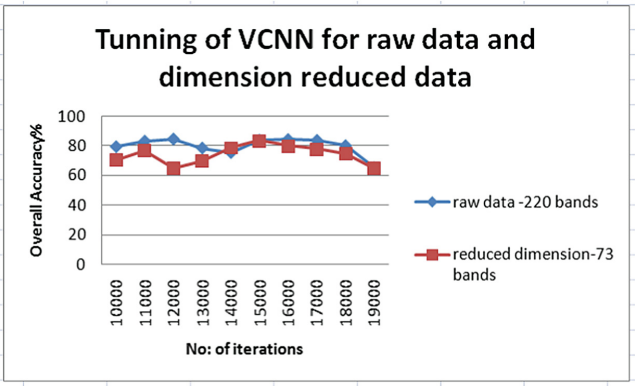


Fig. 4. Plot of overall accuracy vs. number of iterations considered for VCNN with and without dimensionality reduction for Indianpines

Table 3. VCNN outputs for each layer for both the datasets

I/P	Train-para	Convl layer	Max pool layer	Fully conct layer	Train-para	O/P	Train-para
1×220	500	$20 \times (1 \times 197)$	$20 \times (1 \times 39)$	1×100	78100	1×8	808
1×73	180	$20 \times (1 \times 66)$	$20 \times (1 \times 22)$	1×100	44100	1×8	808
1×224	500	$20 \times (1 \times 199)$	$20 \times (1 \times 39)$	1×100	80000	1×6	606
1×37	100	$20 \times (1 \times 34)$	$20 \times (1 \times 11)$	1×100	24100	1×6	606

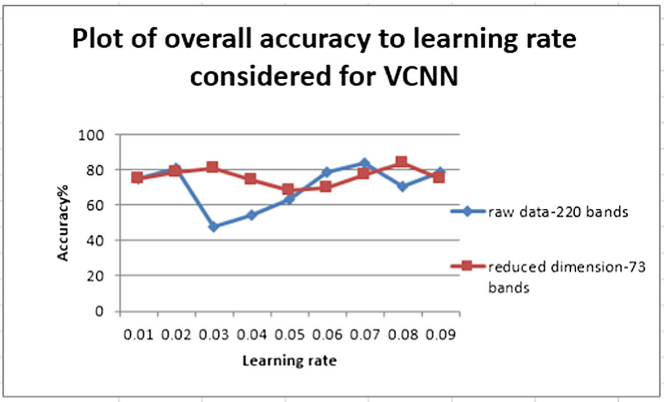


Fig. 5. Plot of overall accuracy vs. learning rate considered for VCNN with and without dimen-sionality reduction for Indianpines

Figures 6 and 7 shows the classwise acccuracies for both raw data and dimension reduced data for Salinas-A and Indianpines dataset. The overall classification accuracies for both the datasets are shown below in Table 4.

Table 4. Comparision of overall classification accuracy with different classification methods with and without dimension reduction of the datasets.

Overall classification accuracy (%)			
DATASET	VCNN	SVM	GURLS
Salinas raw data (224)	91.3 %	99.71%	97.94%
Salinas DMD ($37 = (1/6) \times 224$)	**90.56 %**	**99.3%**	**91.4%**
IndianPines raw data (220)	84.63%	88.49%	78.3%
IndianPines DMD ($73 = (1/3) \times 220$)	**83.67%**	**88.3%**	**75.6%**

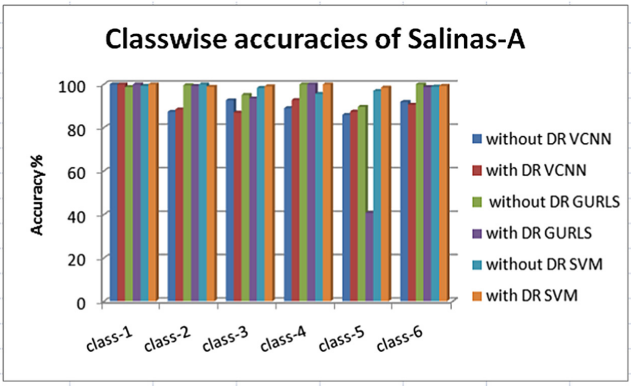


Fig. 6. Classwise accuracies of the Salinas-A for both raw data (224 bands) and DMD reduced data (37 bands)

From the Table 4, it is inferred that SVM and GURLS classifiers were producing high classification accuracy as compared to VCNN for Salinas-A dataset and for Indianpines dataset GURLS shows lower classification accuracy because, of its less classwise accuracies.

From Fig. 8, it is seen that the total number of trainable parameters before and after dimensionality reduction are reduced to a great extent. The classification accuracy obtained before and after dimensionality reduction are comparable to each other with the reduction in the number of trainable parameters. On further reduction to 55 bands (one-fourth of the total bands) for Indianpines, we observe that, the trainable parameters are reduced to a great extent of 33,048 but, the CA is around 80.32%. Similarly for Salinas-A dataset, further reduction to 32 bands (one-seventh of the total bands) we observe that, the trainable parameters are reduced to a great extent of 22,686 but the CA is around 79..32%.

Hence, the maximum possible extent of reduction is of **73** bands (one-third of the total bands) for Indianpines with CA of **83.67%** and for the Slainas-A, it is **37** bands (one-sixth of the total bands) with a CA of **90.56%**.

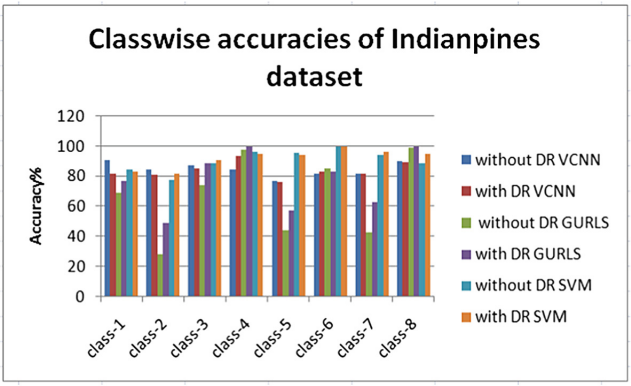


Fig. 7. Classwise accuracies of the Indianpines for both raw data (220 bands) and DMD reduced data (73 bands)

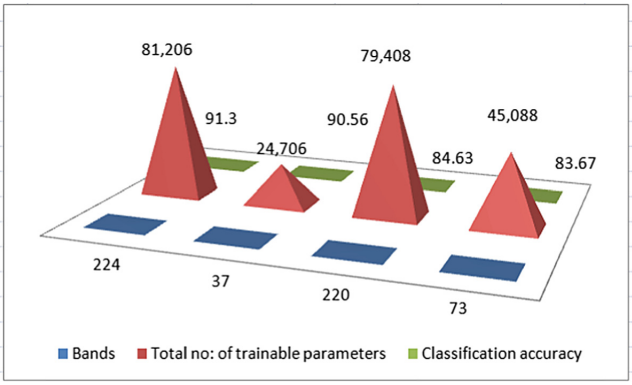


Fig. 8. Comparison of trainable parameters and accuracies with and without dimension reduction for both the datasets

6 Conclusion

The effect of dimensionality reduction on VCNN is analyzed in this work. On obtaining the maximum possible bands reduction, GURLS and SVM classification techniques are used for those reduction bands to compare the class accuracy with that of VCNN. The performance of classifier is evaluated based on accuracy parameters measurement. From the experimental results, for the Salinas-A dataset, it is evident that, even though there is dimension reduction, VCNN and SVM is able to achieve almost the same classification accuracy as that of raw data of HSI. For the Indianpines dataset, we can achieve the same classification accuracy for VCNN and SVM for both raw and dimension reduced data. Here, the total number of trainable parameters are reduced, which had led to the ease of handling data for VCNN. Also, the experimental results shows that, one-sixth

of the total number of available bands are the maximum possible reduction in feature dimension for Salinas-A and one-third of the total available bands are for Indianpines dataset, that results in comparable classification accuracy.

References

1. Aswathy, C., Sowmya, V., Soman, K.P.: ADMM based hyperspectral image classification improved by denoising using Legendre Fenchel transformation. Indian J. Sci. Technol. **8**(24), 1–13 (2015)
2. Chen, Y., Jiang, H., Li, C., Jia, X., Ghamisi, P.: Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans. Geosci. Remote. Sens. **54**(10), 6232–6251 (2016)
3. Dixon, K.D.M., Sowmya, V., Soman, K.P.: Effect of denoising on vectorized convolutional neural network for hyperspectral image classification. In: Nandi, A.K., Sujatha, N., Menaka, R., Alex, J.S.R. (eds.) Computational Signal Processing and Analysis. LNEE, vol. 490, pp. 305–313. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-8354-9_28
4. Hu, W., Huang, Y., Wei, L., Zhang, F., Li, H.: Deep convolutional neural networks for hyperspectral image classification. J. Sens. **2015**, 1–12 (2015). Article ID 258619
5. Koonsanit, K., Jaruskulchai, C., Eiumnoh, A.: Band selection for dimension reduction in hyper spectral image using integrated information gain and principal components analysis technique. Int. J. Mach. Learn. Comput. **2**(3), 248–251 (2012)
6. Megha.P, Sowmya V., Soman, K.P.: Effect of dynamic mode decomposition based dimension reduction technique on hyperspectral image classification. In: International Conference on Nextgen Electronic Technologies: Silicon to Software (ICNETS2). LNEE Springer Proceedings, VIT University, Chennai Campus, India, pp. 23–25 (2017)
7. Schmid, P.J.: Dynamic mode decomposition of numerical and experimental data. J. Fluid Mech. **656**, 5–28 (2010)

Probability-Based Approach for Epileptic Seizure Detection Using Hidden Markov Model

Deba Prasad Dash(✉) and Maheshkumar H. Kolekar

Indian Institute of Technology Patna, Patna, India
{dpdash.srf14,mahesh}@iitp.ac.in

Abstract. Seizure is defined as a sudden synchronous activity of a group of neurons resulting in an electric surge in the brain. Epilepsy is a brain disorder indicated by repeated seizures. Around 10 million people in India are suffering from epilepsy. Electroencephalogram (EEG) signal being low cost and non-invasive in nature can be used effectively for seizure detection. The present work focuses on developing an efficient epileptic seizure detection system using intracranial EEG signals. Dual tree complex wavelet transform is used to decompose the signal into various sub-frequency bands. Probability features are used to extract efficient indicators for seizure and healthy classes. Discriminant correlation analysis is used to increase the difference between different classes as well as reduce the difference between same classes. The fused feature set is clustered using fuzzy c means clustering algorithm. Hidden Markov model discriminates the seizure class with healthy class with good efficiency. Maximum accuracy of 98.57% is achieved for seizure detection with very low execution time.

Keywords: EEG · Epilepsy · Seizure
Dual tree complex wavelet transform
Discriminant correlation analysis · Hidden Markov Model

1 Introduction

EEG and Electrocardiogram (ECG) are used for non-invasive diagnosis of neurological and heart related diseases [14,23,24]. Epilepsy is the most common neurological disorder affecting 70 million people worldwide. Every year around 150,000 people in the United States are diagnosed with epilepsy. Mostly children and old age people get affected by the epilepsy. Seizure is the biomarker for epilepsy. Any person in his entire lifetime may experience a seizure. Seizure repeated minimum twice in period of 24 h is called epileptic seizure. Seizure can be of various types. Around 30% seizures are drug resistive. It is important to detect seizure efficiently. At present doctors manually analyze EEG signal of patient to detect and count seizures. But the process is time consuming and

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 268–281, 2019.
https://doi.org/10.1007/978-981-13-5758-9_23

depends largely on doctor expertise for proper diagnosis. There is a need to develop accurate, reliable and low computation algorithm for auto diagnosis of seizure.

Different algorithms have been proposed for epileptic seizure detection. Time domain features such as different moments, energy, entropy, complexity and fractal dimensions were used for seizure detection [3,16]. Frequency domain features such as spectral moments, power spectral density were evaluated for efficiency in seizure detection [9]. Time-frequency analysis such as discrete wavelet transform and features extracted from wavelet coefficients were widely used for seizure detection [1]. Empirical mode decomposition (EMD) being data independent found good application in developing seizure detection system [7]. But time domain features are not reliable as with little interference of noise the outcome may vary. Different features such as lempel-ziv complexity and mutual information were applied for seizure classification [16,19]. Different similarity index such as fuzzy similarity index, correlation dimension were proposed [20]. EMD has no fixed number of IMFs for different EEG signals even for same length of signal with same sampling frequency. Singular value decomposition (SVD) was also used for epileptic seizure detection. The SVD was applied sequentially on a sliding window of one second width of EEG data and the singular values were obtained and used to indicate sudden changes in the signals. The singular values deviated notably upward from the baseline at the start of epileptic seizure [22]. Support vector machine (SVM) classifier was widely used by many researcher for seizure classification [17]. Apart from SVM, artificial neural network is used for training model for seizure, inter-seizure and healthy EEG signals [10]. But neural networks are computationally expensive and have a high chance of system over-fitting. Recently some of the paper proposes multimodal seizure detection. Elger *et al.* proposed heart rate and accelerometer based epileptic seizure detection [11].

The proposed approach aims at designing an efficient algorithm for seizure detection with less computation time. Feature extraction and classification is based on probabilistic approach both for feature extraction and classification which utilizes the synchronous neural firing during seizure. The paper is organized as follows. Section 2 describes the proposed approach followed by detail explanation of the techniques. Section 4 explains the classification approach and Sect. 5 represents the results and final section concludes the paper with indication of future scope.

2 Proposed Approach

EEG signal is more prone to noises. The first step is to remove higher frequency noises. Infinite impulse response notch filter is used to remove 50 Hz power line interference. Butterworth bandpass filter is used to filter signal above 60 Hz keeping signal up to beta frequency band. Filtered EEG signal is decomposed into 5 sub-bands using dual tree complex wavelet transform (DTCWT). Probability based features and hjorth parameters are extracted from both quantized real

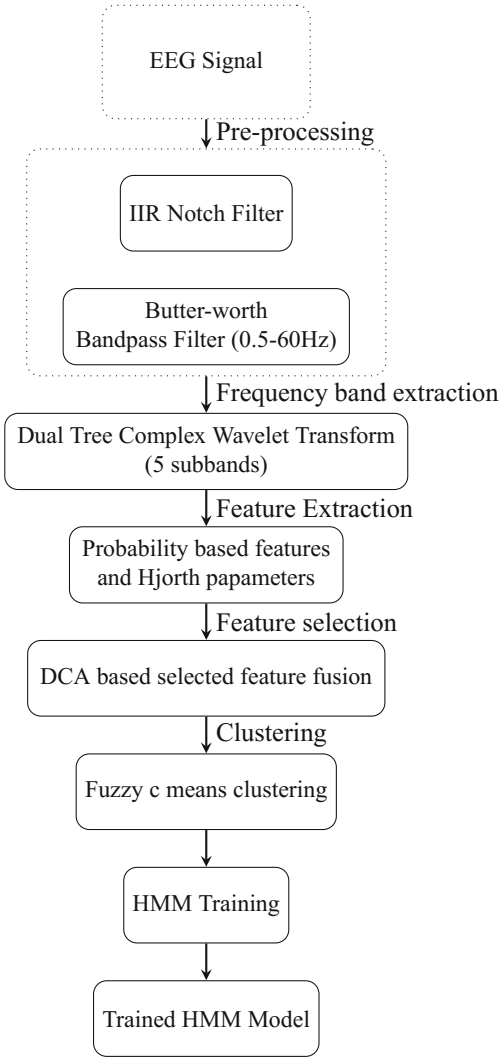


Fig. 1. Proposed HMM training approach

and imaginary wavelet coefficients. Analysis of variance (ANOVA) test is used to select efficient features in each wavelet bands. Discriminant correlation analysis (DCA) is used to fuse features resulting in lowering the dimension of the feature set and increasing the differentiability between two class features. Fuzzy c means clustering algorithm is used to convert features into symbol sequence which is further used by Baum-welch algorithm to train two state ergodic hidden Markov model (HMM) for healthy, seizure and interseizure classes. Figure 1 shows the proposed training approach. During testing phase, test features are combined using DCA coefficients. Manual clustering is performed to get the

symbol sequence of test feature vector by finding distance from test feature vector to the train feature set and assigning cluster number of the train feature vector having minimum distance from the test set. Viterbi algorithm is used to find the state sequence for generated symbol sequence of test feature vector. Figure 2 represents the proposed testing approach. The major contribution of the paper is that the probabilistic algorithm achieved good accuracy with only 30% of training sample evaluated with different combination of EEG classes with less complexity and manageable computation time. System is less complex and converses good in finding HMM model. Probability of different signal elements and features extracted out of it along with other statistical parameter such as Hjorth's parameter are used for classification of seizure, Interseizure and healthy EEG signal.

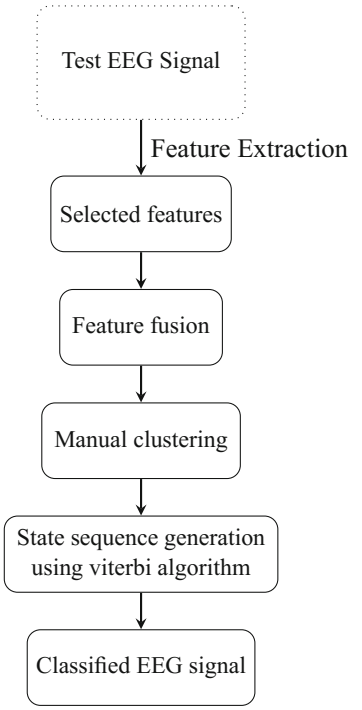


Fig. 2. Proposed HMM testing approach

3 Techniques

3.1 Dual Tree Complex Wavelet Transform

Discrete wavelet transform (DWT) is a wavelet based data transformation technique. It differs from the Fourier transform as it gives both time and frequency domain information at a time. Scale and shifted version of wavelet convolved

with original data gives wavelet coefficients down sampled by factor of 2 for each level of decomposition. Mathematically DWT represents a signal as low pass scaling function ϕ and bandpass wavelet function ψ.

$$\begin{aligned} \psi_{j,k}(t) &= 2^{\frac{-j}{2}} \times \psi(2^{-j}t - k), \\ \phi_{j,k}(t) &= 2^{\frac{-j}{2}} \times \phi(2^{-j}t - k) \end{aligned} \tag{1}$$

where k represents the shifting factor and 2^j represents the scaling factor.

DWT has aliasing, shift variant problems and is less directional sensitive. These problems are solved to some extent by DTCWT [21]. Complex wavelet is used for decomposition of signal. The DTCWT employs two DWTs. One DWT gives the real coefficient and other one gives the imaginary coefficient. The original signal can be reconstructed averaging two branch output. The DTCWT can be explained in matrix form.

$$F = \begin{bmatrix} f_h \\ f_g \end{bmatrix} \tag{2}$$

If vector x is the real signal, then $w_h = f_h x$ represents the real part and $w_g = f_g x$ represents the imaginary part.

In the proposed approach DTCWT is used to extract significant band for seizure classification.

3.2 Probability Features

Probability is an indication of synchronous activity in any system. In the proposed approach difference in synchronous activity between seizures, Interseizure and healthy signals are explored as biomarker for efficient classification. Detailed steps for probability based feature extraction are given in Algorithm 1.

Quantization: Quantization is an essential step to get the appropriate repetition value of each element in signal segment. Data within the range of $x_i \pm 2$ are quantized to a single value. Mathematically it can be presented as:

$$f(x_i) = \begin{cases} k+2 & \text{if } (0+\text{k}) < x_i \leq (2+\text{k}) \\ -k-2 & \text{if } (-2-\text{k}) \leq x_i < (0-\text{k}) \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$\text{k} = 0, 2, \ldots \text{N} - 2$, Where N is the maximum signal value after quantization. In this research N is considered as 1100.

Algorithm 1. Probability based feature extraction

1.Input ← EEG signal X
2.Quantization (Q) ← Threshold based quantization
3.Segmentation ← Segmentation of signal with 50% overlapping
3.Probability ← Probability evaluation of each segment
$P(x(t)), p(x(t+\tau)), p(y(t))$, where t is the array element
4.Conditional probability ←
$p(x(t+\tau)|x(t)), p(x(t+\tau)|x(t), y(t)), p(x(t)|y(t)), p(x(t), x(t+\tau), y(t))$
5.Features ← Shannon, collison and transfer entropy calculations
6.Output ← Probability features

Features: Probability of occurrence of each data point is calculated. Shannon, collision and transfer entropy are evaluated. Mathematically features are represented as:

$$-\sum p(x(t+\tau), x(t), y(t)) \times \log_2(p(x(t+\tau)|x(t), y(t))) \tag{4a}$$

$$-\sum p(x(t+\tau), x(t), y(t)) \times \log_2(p(x(t+\tau)|x(t))) \tag{4b}$$

$$-\sum p(x(t+\tau), x(t), y(t)) \times \log_2 \frac{p(x(t+\tau)|x(t), y(t))}{p(x(t+\tau)|x(t))} \tag{4c}$$

$$-\sum p(x(t)) \times \log_2(p(x(t))) \tag{4d}$$

$$-\sum p(y(t)) \times \log_2(p(y(t))) \tag{4e}$$

$$-\sum \log_2(p(x(t)^2)) \tag{4f}$$

$$-\sum \log_2(p(y(t)^2)) \tag{4g}$$

$$-\sum p(X(t)^2) \times \log_2(p(X(t)^2)) \tag{4h}$$

where x(t), x(t + τ), y(t) are the three segments of EEG signals, τ is the time shifting factor. P(x) is the probability of the signal. In the proposed approach τ is considered to be the 50% of segment length. X(t) is the total EEG signal.

The conditional probability indicates given the particular event the probability of occurrence of other event. Conditional probability in the present context indicates the synchronous activity of neurons in a period of time. $P(x(t+\tau), x(t), y(t))$ is calculated to measure the similarity between three different segments of an ongoing event. Formulae 4(d), 4(e) indicates Shannon entropy for different EEG segments. Formulae 4(f) and 4(g) evaluates the collision entropy for different EEG segments. Formulae 4(a) and 4(b) represents the Shannon entropy for different conditional probabilities. Formulae 4(h) indicates the randomness in the complete EEG signal.

Transfer Entropy: Transfer entropy is a measure of information transfer from one segment to other within the event [26]. It has found a lot of application in neuroscience research. The brain dynamics is non-linear, therefore entire event

is segmented so that each segment can be assumed to be as stationary process. Conditional probability represents the underlying brain dynamics. Let the three segments of EEG signals be x(t), $x(t+\tau)$ and y(t). Mathematical expression is given in the equation 4(c). Conditional probabilities are calculated by the formulae given below.

$$\begin{aligned} p(x(t+\tau)|x(t),y(t)) &= \frac{p(x(t+\tau),x(t),y(t))}{p(x(t),y(t))} \\ p(x(t+\tau)|x(t)) &= \frac{p(x(t),x(t+\tau))}{p(x(t))} \end{aligned} \tag{5}$$

3.3 Hjorth Parameter

Hjorth parameter are indicators of statistical properties of the signal [5]. It has three different parameter namely activity, mobility and complexity. Activity gives the variance of the signal which is a significant parameter to measure the spread of data about their mean. Mobility is the ratio of variance of first derivative of data to the original data. Complexity is the measure of shape of the signal similar to the sine wave. Hjorth parameters are extracted from the entire signal. Mathematically these factors are represented as:

$$\begin{aligned} Activity &= var(X(t)) \\ Mobility &= \sqrt{\frac{var(X'(t))}{var(X(t))}} \\ Complexity &= \frac{mobility(X'(t))}{mobility(X(t))} \end{aligned} \tag{6}$$

where X represents the EEG signal and $X'(t)$ represents the first derivative of signal X(t).

3.4 Fuzzy C Means Clustering

Clustering involves grouping data points based on similarity. Similarity of points is measured by minimum distance from the cluster centre. Clustering of data points is achieved by an iterative process. Fuzzy C means has unsupervised and soft clustering approach [4]. K means clustering fails to perform in overlapping database. So fuzzy c means clustering is preferred in this research for generating symbol sequence. Each feature vector belongs to one cluster based on highest fuzzy membership value. Number of clusters for total feature set is selected empirically. Total feature vectors are divided into 2 clusters. Cluster numbers for feature vectors are considered as the symbol sequence. Symbol sequence for each class of EEG signal is used to train HMM model.

3.5 Feature Level Fusion

DCA [12] is used to fuse the efficient features selected by ANOVA test. DCA aims to maximize the difference between different classes and increase similarity between same classes. Algorithm 2 explains the details of the DCA fusion algorithm. Real and imaginary coefficients of selected bands are fused and later combined to form single column feature vector. The features are simply added with equal weight for each feature.

Algorithm 2. Feature fusion based on DCA

1.Input ← Optimal feature vectors f, f1
2.Find total mean $\bar{f}$ and individual class mean $\acute{f}$
3.Between class scatter matrix ← $\phi_{bf} = \sqrt{n_i}(\acute{f}_i - \bar{f})$, n=number of samples
4.Calculate the co-variance matrix $s_{bf} = \phi_{bf}^T \times \phi_{bf}$
5.Maximum diagonal eigen vector of the s_{bf} arranged in decreasing order ← E_{vector}
6. Most significant eigen vectors are obtained by $S_{eigen} = \phi_{bf} \times E_{vector}$
7.Normalize the matrix to create orthonormal eigen vectors E_n
8. Find diagonal: $s_{diag} = diag(E_n)$
9.Calculate the transformation weight matrix $w_{bf} = s_{eigen} * s_{diag}^{-1/2}$
10. Repeat the step from 1-9 for calculation of weight matrix for feature set f1:w_{bf1}
11 Transform the feature set $f^{new} = w_{bf} \times f$, $f1^{new} = w_{bf1} \times f1$
12.Singular value decomposition of covariance matrix of transformed feature vector is calculated: S
13.Final weights of f and f1 transformation are calculated
$w_{cf} = w_{bf} * S^{-1/2}, w_{cf1} = w_{bf1} * S^{-1/2}$
14.Final f and f1 transformation is achieved by:
$F_s = \acute{w_{cf}} \times f^{new}, F1_s = \acute{w_{cf1}} \times f1^{new}$
15.Final transformation matrix is defined as:
$A_f = \acute{w_{cf}} \times \acute{w_{bf}}, A_{f1} = \acute{w_{cf1}} \times \acute{w_{bf1}}$
16.Output ← Fused features: F1+F2+...Fn

4 Classification

HMM is based on Markov chain rule. Markov chain rule states that probability of moving to next state depends on only present state and not on other previous states. Two state ergodic HMM is used for modeling the seizure, inter-seizure and healthy EEG events. Baumwelch algorithm is used for training HMM. Figure 3 represents the initial state transition matrix of two state hidden Markov model. Initial transition probability for same state to state is assumed to be 0.7 which is higher compared to between state transitions. Fuzzy c means clustering algorithm [13] is used to obtain symbol sequence from the fused feature set. Total feature set is divided into two clusters. In testing phase distance of test feature vector with train feature vector is evaluated. Based on minimum distance symbol

sequence is obtained. Viterbi algorithm is used to evaluate the state sequence for the test symbol sequence. Detail Baumwelch and Viterbi algorithms are mentioned in the paper [6]. In this classification approach only 30% of data is being considered for training purpose. Rest signals are used for testing the accuracy of system. Efficiency of the proposed approach is evaluated for various class combinations.

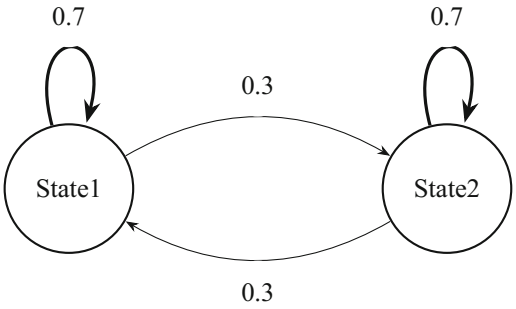


Fig. 3. Initial transition matrix for 2 state HMM model (State 1, State 2): healthy, seizure, interseizure EEG signal

5 Results

5.1 Database

In the present work, online benchmark seizure EEG database [2] is used containing five EEG datasets (Z-S), each dataset containing 100 EEG segments. The signal duration of each study is 23.6 s with sampling frequency of 173.61 Hz. The detail of database used is shown in Table 1. Total length of data is 4097 samples. For dual tree complex wavelet transform application 4096 (2^{64}) samples are considered for processing.

Table 1. Detail of experimental database

Database	Class	Description
Z	Healthy	Surface EEG recording of healthy subjects (eye open)
O	Healthy	Surface EEG recording of healthy subjects (eye closed)
N	Inter-seizure	Seizure free segment recorded from hippocampal formation of the opposite hemisphere of the brain
F	Inter-seizure	Seizure free segment recorded in the epileptogenic zone
S	Seizure	Seizure segment

5.2 Feature Selection and Fusion

Table 2 represents the features selected for various class combinations at different levels. ANOVA test is used to select the efficient features. Entropy features such as Shannon, collision and transfer entropy are efficient in classifying healthy-seizure EEG signals. It is observed that EEG signal features extracted from lower frequency wavelet coefficients that is 4 and 5 level decompositions are efficient for healthy-seizure classification. The frequency range at level 4 wavelet decomposition is approximately 0.5–4 Hz. Similarly in seizure verses non-seizure classification Shannon entropy and transfer entropies are efficient at level 4. It is important to note that healthy-interseizure classification the Shannon entropy is efficient. The maximum difference between features of healthy-interseizure EEG signal is achieved for activity hjorth parameter. Selected features are fused to generate an efficient feature set which can differentiate different classes with higher accuracy. All the efficient features are combined using DCA feature fusion approach. After DCA application features are linearly added to each other. Table 3 shows the ANOVA test P value of fused feature set and classification accuracy for different class combinations. The features are significantly separated from each other. Maximum ANOVA P value is obtained for Z-S features. Table 3 shows classification accuracy obtained for different class combinations. Mathematically accuracy, specificity and sensitivity are defined as:

$$\begin{aligned} Accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\ Specificity &= \frac{TN}{TN + FP}, \; Sensitivity = \frac{TP}{TP + FN} \end{aligned} \tag{7}$$

Where TP = true positive, TN = true negative, FP = false positive, FN = false negative.

Table 2. Selected features at different wavelet decomposition level with equation number

Class	Level of wavelet decomposition	Features (with equation number)
Z-S	4	Shannon entropy 4(a), 4(b), Transfer entropy 4(c), collision entropy 4(g)
	5	Shannon Entropy: 4(a)
ZO-S	4	Shannon entropy 4(a), 4(b), Transfer entropy 4(c)
FN-S	3	Collision entropy: 4(f), 4(g)
	4	Shannon Entropy: 4(a)
ZO-FN	2	Shannon entropy, 4(d), 4(e)
	4	Hjorth Parameter: Activity (6)
ZOFN-S	4	Shannon entropy 4(a), 4(b) Transfer entropy 4(c)

Maximum accuracy is achieved for healthy-seizure classification. In medical application both higher true positive and higher true negatives are desired. Sensitivity of the system is indicated by higher true positive and specificity of the system is indicated by higher true negative values. As can be interpreted from the Table 3 healthy-seizure (ZO-S) is classified with 94.28% sensitivity and 95.71% specificity. Seizure-interseizure (FN-S) is classified with 90% sensitivity and 99.28% specificity. Seizure-non-seizure is classified with 91.42% sensitivity and 95.71% specificity. Overall it can be concluded that the system is efficient in healthy-seizure-interseizure classification with equal or above 90% accuracy, specificity and sensitivity for various class combinations. Figure 4 represents the fuzzy c means clustering output between seizure verses non-seizure class after feature fusion approach. There is a definite difference between the seizure and non-seizure signal features. The cluster index number generated for each class is unique resulting in higher classification accuracy. Generated index numbers are used to train HMM. Table 4 gives the comparison table of different state of art methods with accuracy. Proposed approach achieved higher accuracy compared to other methods. In state of art methods [6,8] shown above have used 70%–30% or 60%–40% for training and testing feature sets. In our proposed approach

Table 3. ANOVA P value and classification accuracy

Class	ANOVA P value	Accuracy (%)	TP/FN	TN/FP
Z-S	2.43×10^{-186}	**98.57**	68/2	70/0
ZO-S	3.59×10^{-163}	95.23	66/4	134/6
ZO-FN	1.00×10^{-129}	90.00	130/10	112/28
FN-S	1.14×10^{-149}	**95.71**	63/7	139/1
ZOFN-S	7.13×10^{-160}	**94.28**	64/6	268/12

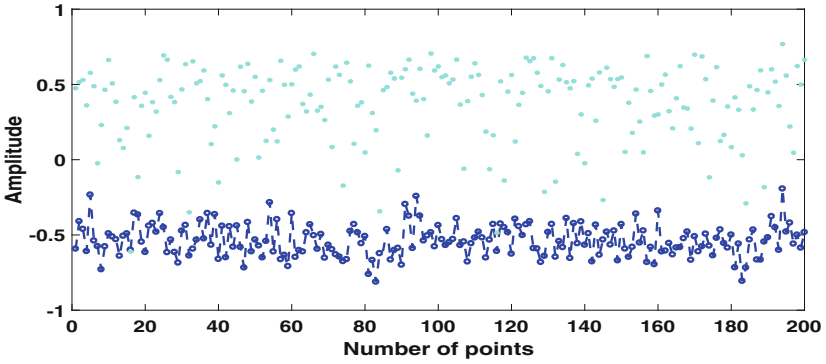


Fig. 4. Fuzzy c means clustering of ZOFN-S classes (Dark blue line indicates the seizure class, light blue dots represents the non-seizure class) (Color figure online)

only 30% data is used for training purpose. Overall system complexity is less and execution time is calculated to be 0.102 s. Analysis is implemented using licensed Matlab 2015a software with Intel core i5 system.

Table 4. Accuracy comparison of state-of-the-art and proposed method for Z-S classification

Author	Techniques	Accuracy (%)
Liu et al. [18]	Increment entropy SVM	97.6
Swami et al. [25]	Empirical mode decomposition 3D Phase trajectory Probabilistic neural network	98.00
Deshmukh et al. [8]	DWT Statistical features SVM	95.00
Kolekar et al. [16]	Entropy, Complexity SVM	91.25
Dash et al. [6]	Average Entropy, Hjorth parameters HMM	96.67
Proposed approach	DTCWT probability features, hjorth parameter HMM	**98.57**

6 Conclusion and Future Scope

Efficient epileptic seizure detection approach is proposed using DTCWT and probabilistic features, hjorth parameters. HMM is used for classification purpose. Lower frequency band features are observed to be efficient in seizure classification. DCA based fusion approach found to be effective in reducing feature dimension and maintaining the differences between classes. The system achieved 98.57% accuracy for seizure-healthy classification. System has less execution time which makes it suitable for real time application. In future different features will be evaluated to further increase the accuracy of the system. Probabilistic classifier such as Bayesian classifier [15] will be evaluated for its efficiency in seizure detection.

References

1. Adeli, H., Ghosh-Dastidar, S., Dadmehr, N.: A wavelet-chaos methodology for analysis of EEGs and EEG subbands to detect seizure and epilepsy. IEEE Trans. Biomed. Eng. **54**, 205–211 (2007)
2. Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C.E.: Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Phys. Rev. E **64**, 061907 (2001)
3. Baldominos, A., Ramón-Lozano, C.: Optimizing EEG energy-based seizure detection using genetic algorithms. In: IEEE Congress on Evolutionary Computation, pp. 2338–2345 (2017)
4. Bezdek, J.C., Ehrlich, R., Full, W.: FCM: the fuzzy c-means clustering algorithm. Comput. Geosci. **10**, 191–203 (1984)
5. Damaševičius, R., Maskeliūnas, R., Woźniak, M., Połap, D.: Visualization of physiologic signals based on Hjorth parameters and Gramian angular fields. In: IEEE World Symposium on Applied Machine Intelligence and Informatics, pp. 000091–000096 (2018)
6. Dash, D.P., Kolekar, M.H.: Epileptic seizure detection based on EEG signal analysis using hierarchy based hidden Markov model. In: International Conference on Advances in Computing, Communications and Informatics, pp. 1114–1120 (2017)
7. Dash, D.P., Kolekar, M.H.: EEG based epileptic seizure detection using empirical mode decomposition and hidden Markov model. Indian J. Public Health Res. Dev. **8**(4), 897 (2017)
8. Deshmukh, P., Ingle, R., Kehri, V., Awale, R.: Epileptic seizure detection using discrete wavelet transform based support vector machine. In: International Conference on communication and Signal Processing, pp. 1933–1937 (2017)
9. Dheepa, N.: Automatic seizure detection using higher order moments & ANN. In: IEEE International Conference on Advances in Engineering, Science and Management, pp. 601–605 (2012)
10. Dhif, I., Hachicha, K., Pinna, A., Hochberg, S., Mhedhbi, I., Garda, P.: Epileptic seizure detection based on expected activity measurement and neural network classification. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 2814–2817 (2017)
11. Elger, C.E., Hoppe, C.: Diagnostic challenges in epilepsy: seizure under-reporting and seizure detection. Lancet Neurol. **17**, 279–288 (2018)
12. Haghighat, M., Abdel-Mottaleb, M., Alhalabi, W.: Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition. IEEE Trans. Inf. Forensics Secur. **11**(9), 1984–1996 (2016)
13. Inan, Z., Kuntalp, M.: A study on fuzzy c-means clustering-based systems in automatic spike detection. Comput. Biol. Med. **37**(8), 1160 (2007)
14. Jha, C.K., Kolekar, M.H.: ECG data compression algorithm for tele-monitoring of cardiac patients. Int. J. Telemed. Clin. Pract. **2**, 31–41 (2017)
15. Kolekar, M.H., Sengupta, S.: Bayesian network-based customized highlight generation for broadcast soccer videos. IEEE Trans. Broadcasting **61**(2), 195–209 (2015)
16. Kolekar, M.H., Dash, D.P.: A nonlinear feature based epileptic seizure detection using least square support vector machine classifier. In: IEEE TENCON Region 10 Conference, pp. 1–6 (2015)
17. Kumar, A., Kolekar, M.H.: Machine learning approach for epileptic seizure detection using wavelet analysis of EEG signals. In: International Conference on Medical Imaging, m-Health and Emerging Communication Systems, pp. 412–416 (2014)

18. Liu, X., Jiang, A., Xu, N.: Automated epileptic seizure detection in EEGs using increment entropy. In: Canadian Conference on Electrical and Computer Engineering, pp. 1–4 (2017)
19. Malladi, R., Johnson, D.H., Kalamangalam, G.P., Tandon, N., Aazhang, B.: Data-driven estimation of mutual information using frequency domain and its application to epilepsy. In: Asilomar Conference on Signals, Systems, and Computers, pp. 2015–2019 (2017)
20. Ouyang, G., Li, X., Guan, X.P.: Use of fuzzy similarity index for epileptic seizure prediction. IEEE World Congress Intell. Control Autom. **6**, 5351–5355 (2004)
21. Selesnick, I.W., Baraniuk, R.G., Kingsbury, N.C.: The dual-tree complex wavelet transform. IEEE Signal Process. Mag. **22**, 123–151 (2005)
22. Shahid, A., Kamel, N., Malik, A.S., Jatoi, M.A.: Epileptic seizure detection using the singular values of EEG signals. In: International Conference on Complex Medical Engineering, pp. 652–655 (2013)
23. Sharma, N., Kolekar, M.H.: Diagnosis of vascular cognitive impairment using EEG. Indian J. Public Health Res. Dev. **8**(4) (2017)
24. Sharma, N., Kolekar, M.H., Chandra, S.: The role of EEG signal processing in detection of neurocognitive disorders. Int. J. Behav. Healthcare Res. **6**, 15–27 (2016)
25. Swami, P., Gandhi, T.K., Panigrahi, B.K., Bhatia, M., Anand, S.: Detection of ictal patterns in electroencephalogram signals using 3D phase trajectories. In: IEEE India Conference, pp. 1–6 (2015)
26. Ye, X., Tian, T., Xu, T., Wang, J.: Analysis of beta wave epileptic EEG signals based on symbolic transfer entropy. In: International Conference on Artificial Intelligence and Industrial Engineering (2015)

Bit-Plane Specific Measures and Its Applications in Analysis of Image Ciphers

Ram Ratan[1(✉)] and Arvind[2]

[1] Scientific Analysis Group, Defence Research and Development Organization, Delhi, India
ramratan_sag@hotmail.com
[2] Department of Mathematics, Hansraj College, University of Delhi, Delhi, India
arvind_ashu12@rediffmail.com

Abstract. The paper presents bit-plane specific new measures to visualize the extensive statistical detail of an image. We compute the frequency of ones, maximum run length and correlation among rows (columns) in each bit-plane of an image. The computed measures give row-wise and column-wise structural detail at bit-plane level and help an interpreter to analyze given image deeply for its effective interpretation and understanding. In this paper, the application of these measures is shown in cryptography to statistically analyze the image ciphers. The simulation study shows that the proposed measures are very useful and can be applied in various image processing applications for pattern recognition and understanding of visual objects.

Keywords: Bit-plane measures · Image analysis · Image cipher
Image quality measures · Qualitative measures · Quantitative measures

1 Introduction

Good quality of an image is the basic need of the customers for better understanding and interpretation of image contents. In current era of information technology, visual data is being used widely in various image processing applications such as remote sensing, medical imaging, entertainment, military communications, information storage and retrieval where good quality images are required. For assessing the quality of an image, one needs some measures which give qualitative (subjective) and quantitative (objective) detail. In qualitative measures, one perceives the visual appearance and in quantitative measures, one observes the value of some measures computed for given image. The perception of an human is highly adaptive which extracts structural details and measures perceived image distortion easily. For example the presence of distortion such as blur in an image can be identified by the human easily but it is very difficult for

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 282–297, 2019.
https://doi.org/10.1007/978-981-13-5758-9_24

a computer. Qualitative measures are expensive and need human intervention to estimate image quality. Quantitative measures are statistical in nature and can be obtained automatically and efficiently. We need such quantitative measures which give structural image detail and correlate with human perception to assess image quality with good approximation of perceived image.

Various qualitative and quantitative measures are reported in the literature [1–9]. The basic qualitative measure is the visual perception of an image in which humans assess the quality by visualizing its appearance and assign the mean opinion score (MOS) [10]. Some quantitative measures such as histogram, correlation are reported for their use in analysis of image quality [11–16]. Another measure is the gray level co-occurrence matrix (GLCM) which contains pair-wise co-occurred pixel values occurred in an image [17,18]. The plot of co-occurrence exhibits pixel values occurred pairwise in an image. Some other measures on an image and reference image are the error measures such as histogram dissimilarity (HD), mean absolute error (MAE), mean square error (MSE), peak signal to noise ratio (PSNR) and pixel change rate (PCR) which give information about how a given image differs from reference image [19–24]. These measures are computed on pixel values and do not provide any detail at bit-plane level of an image to analyze it deeply.

In this paper, we propose bit-plane specific new image measures in which we compute row-wise and column-wise frequency of ones, maximum run length and correlation between adjacent rows (columns) in each bit-plane of an image. Plots of these measures exhibit overall detail which enables to analyze and interpret an image at bit-plane levels. The range of each measure is also computed to see the variations in every bit-plane of an image. The usefulness of these measures is demonstrated in this paper in the area of cryptology to analyze image ciphers statistically. An image cipher transforms a plain image into an encrypted image by applying mathematical operations on image pixels. A plain image looks intelligible but an encrypted image looks random and unintelligible. For the detail on cryptography and image ciphers, one may refer any book on cryptography [25–27]. For a strong image cipher, the proposed measures should exhibit uniform distributions of bits, less values of maximum run length of consecutive bits and less values of correlation in the rows (columns) of bit-planes of encrypted images. The values of these measures should vary in narrow range in bit-planes of encrypted images obtained by strong image cipher.

The paper is organized as follows: Sect. 2 gives a brief on some image quality measures. New image quality measures specific to bit-planes are presented in Sect. 3. The plots of bit-plane measures and image characteristics are presented in Sect. 4. The usefulness of proposed measures in the area of cryptography to analyze image ciphers and identify weak/strong image ciphers is shown in Sect. 5. The paper is concluded in Sect. 5 followed by the references.

2 Some Common Image Measures

An image is the two-dimensional (2D) projection of real world objects. Mathematically, it is represented in the form of numbers and known as digital image. It is defined as a 2D bounded matrix or a function say $f(x, y)$ which represents the brightness or gray level at location (x, y), x and y are the indexes indicating pixel location in the image plane of size $M \times N$. The value of $f(x, y) \in (0, ..., 2^l - 1)$, l is the number of bits which represents gray level, $M \geq x \geq 1$ and $N \geq y \geq 1$. For detail on image and its processing, one may refer any book on image processing such as [24–26]. Let $f(x, y)$ and $f'(x, y)$ be the input and output images. The measures commonly used in various applications of image processing to assess the quality of an image are briefly given in following paragraphs:

(i) Visual perception: of an image is assessed by the human by observing its visual appearance. The quality of an image is scored as mean opinion score (MOS). The MOS can be scored as excellent, very good, good, fair, poor, very poor depending on visual appearance of an image observed. The overall MOS can be judged or measured by considering MOS scored by the observers.

(ii) Histogram: A histogram plot shows pixel values distribution of an image. The frequency of pixel values are obtained by counting their occurrences in an image.

(iii) Co-occurrence: The co-occurrence shows pairwise frequency of co-occurred pixel values in an image. The co-occurrence plot is obtained by plotting the co-occurred pixel values. The co-occurrence matrix for an image is measured as
$C(f) = \sum(c(i, j) + 1)$ if $((f(x, y) = i)$ and $(f(x, y + 1) = j))$, $1 \leq x \leq M$, $1 \leq y \leq N$, $0 \leq i \leq 255$, $0 \leq j \leq 255$.

(iv) Pixel Change Rate (PCR): The pixel change rate is the percentage number of pixels differs between output and input images. It is measured as
$PCR(f, f') = 100 \times (\sum((c+1))/(M \times N)$ if $(f(i, j) \neq f'(i, j)))$, $1 \leq i \leq M$, $1 \leq j \leq N)$, initially $c = 0$.

(v) Mean Square Error (MSE): The mean square error is the error between input and output images. It is measured as
$MSE = (\sum(f(i, j) - f'(i, j))^2)/(M \times N)$, $1 \leq i \leq M$, $1 \leq j \leq N$.

(vi) Histogram Dissimilarity (HD): The histogram dissimilarity is the average difference between occurrences of pixel values between input and output images. It is measured as
$HD = (\sum |(fq(i) - fq'(i))|)/256$, $0 \leq i \leq 255$, $fq(i)$ is the occurrence of ith gray level.

(vii) Mean Absolute Error (MAE): The mean absolute error is the average absolute difference between input and output images. It is measured as
$MAE = (\sum |(f(i, j) - f'(i, j))|)/(M \times N)$, $1 \leq i \leq M, 1 \leq j \leq N$,

(viii) Peak Signal to Noise Ratio (PSNR): It is measured as
$PSNR = 10 \log_{10}(K^2/MSE)$, $K = 2^{l-1}$ is the maximum pixel value.

(ix) Correlation (C): The correlation coefficient is measured as $C = (\sum(f(i,j)-A)\times(f'(i,j)-A')/(\sum(f(i,j)-A)^2\times\sum(f'(i,j)-A')^2))/(M\times N), 1\le i\le M, 1\le j\le N$, A and A' are the mean of pixel values for f and f', $A=(\sum f(i,j))/(M\times N)$ and $A'=(\sum f'(i,j))/(M\times N)$, $1\le i\le M$, $1\le j\le N$.

3 New Measures Proposed

A gray level image can be represented in a number of bit-planes. For example, a 8-bit gray level image $f(i,j)$ is represented in 8 bit-planes, say $BP_1(i,j)$, $BP_2(i,j)$, $BP_3(i,j)$, ..., $BP_7(i,j)$, $BP_8(i,j)$ where $BP_1(i,j)$ is the MSB bit-plane and $BP_8(i,j)$ is the LSB bit-plane. As an illustration, Monalisa and Baboon images, each of size 256×256, and its bit-planes are shown in Fig. 1(a) and (b) respectively. According to visual perception, the bit-planes shows that the MSB bit-planes possess more intelligible and structured information whereas the LSB bit-planes possess more unintelligible and unstructured information of images.

We propose bit-plane specific measures as row-wise and column-wise frequency of ones; row-wise and column-wise maximum run length; and correlation between adjacent rows and adjacent columns in each bit-plane. For computing these measures, we first find bit-planes $b(i,j,k)$, $1\le k\le 8$ of an image $f(i,j)$ of size $N\times N$ by converting each pixel value $x(i,j)$ into 8 binary digits as following:

```
Algorithm for converting gray level image into bit-planes
    for i = 1 : N
    for j = 1 : N
    {  x = dec2bin(f(i,j),8);
       for k = 1 : 8
       b(i,j,k) = x(k);     }
(Here, k represents bit-plane number and dec2bin() is the
 MATLAB function which converts decimal number into binary digits.)
```

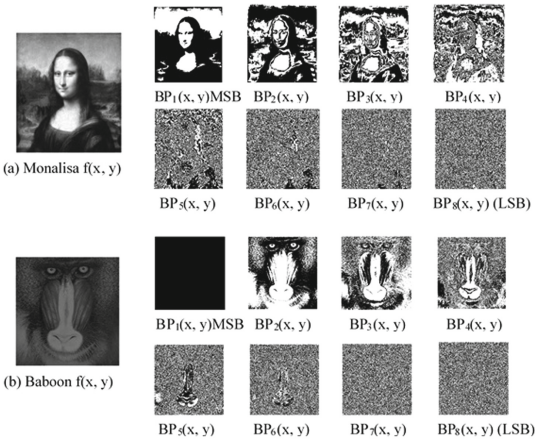


Fig. 1. Images and its bit-planes: (a) Monalisa (b) Baboon

The measures are now presented in the following paragraph:

3.1 Row-Wise (Column-wise) Frequency of Ones in Bit-Planes

For each bit-plane, the frequency of ones is computed in every raw (column) of an image. The frequency of ones in a row (column) is the number of occurrences of ones in a row (column). The frequencies of ones in row i and column j in bit-plane k are denoted by $rf(k,i)$ and $cf(k,j)$ respectively and computed for all the bit-planes as per Algorithm-1 given in Table 1.

Similarly, the frequency of zeros can also be measured.

3.2 Row-Wise (Column-wise) Maximum Run Length in Bit-Planes

For each bit-plane, the maximum run length is computed in every row (column) of an image. A run is the occurrence of consecutive ones in a sequence of binary digits. The run length is the number of consecutive ones occurred in a run and maximum run length is the largest length of a run. The maximum run lengths in row i and column j in bit-plane k are denoted by $rr(k,i)$ and $cr(k,j)$ respectively and computed for all the bit-planes as per Algorithm-2 given in Table 1.

Similarly, the maximum gap length can also be measured. Gap means the occurrence of consecutive zeros and maximum gap length is the largest length of a gap.

Table 1. Algorithms for computing row-wise (column-wise) bit-plane measures in an image

Algorithm-1: Compute frequency of ones in rows (columns)	Algorithm-2: Compute maximum run length in rows (columns)	Algorithm-3: Compute correlation between adjacent rows (columns)

for k = 1 : 8
for i = 1 : N
{rf(k,i) = 0;
cf(k,i) = 0;}
for i = 1 : N
for j = 1 : N
for k = 1 : 8
if (b(i,j,k) == 1)
{rf(k,i) = rf(k,i)+1;
cf(k,j) = cf(k,j)+1;}
	for k = 1 : 8
	for i = 1 : N
	{rr(k,i) = 0;
	c = 0;
	for j = 2 : N
	{if (b(i,j,k) == b(i,j-1,k))
	& (b(i,j-1,k) == 1)
	c = c+1;
	else if (rr(k,i) < c)
	rr(k,i) = c+1;
	c = 0;}}
	for k = 1 : 8
	for j = 1 : N
	{cr(k,j) = 0;
	c = 0;
	for i = 2 : N
	{if (b(i,j,k) == b(i-1,j,k))
	& (b(i-1,j,k) == '1')
	c = c+1;
	else if (cr(k,j) < c)
	cr(k,j) = c+1;
	c = 0;}}
		for k = 1 : 8
		for i = 1 : N-1
		{rc(k,i) = 0;
		for j = 1 : N
		if b(i,j,k) == b(i+1,j,k)
		rc(k,i) = rc(k,i)+1;}
		}
		for k = 1 : 8
		for j = 1 : N-1
		{cc(k,j) = 0;
		for i = 1 : N
		if b(i,j,k) == b(i,j+1,k)
		cc(k,j) = cc(k,j)+1; }
		}
		for k = 1 : 8
		{for i = 1 : N
		rc(k,i) = (rc(k,i)-(N-rc(k,i)))/N;
		for i = 1 : N
		cc(k,i) = (cc(k,i)-(N-cc(k,i)))/N;}

3.3 Correlation Between Adjacent Rows (Columns) in Bit-Planes

For each bit-plane, the correlation is computed in every adjacent rows (columns) of an image. Correlation measure gives an indication about similarity or dissimilarity between two objects.

The correlation between two sequences $A1$ and $A2$ of binary digits each of length L is computed as $C(A1, A2) = (numberofbitsmatch - numberofbitsmismatch)/L$. The value of $C(A1, A2)$ varies in the range from -1 to $+1$. If $C(A1, A2) = 0$ then $A1$ and $A2$ are not correlated, if $C(A1, A2) = -1$ then $A1$ and $A2$ are correlated inversely and if $C(A1, A2) = 1$ then $A1$ and $A2$ both are same.

Adjacent correlations between two rows, i and $i+1$, and two columns, j and $j+1$, in bit-plane k are denoted by $rc(k, i)$ and $cc(k, j)$ and computed for all the bit-planes as per Algorithm-3 given in Table 1.

All the measures discussed above are computed using MATLAB Programming. The visual representation of these measures exhibits detailed information for each bit-plane of an image. Dynamic range of these measures for every bit-plane for rows and columns is also computed to see the variations in the measures for given images.

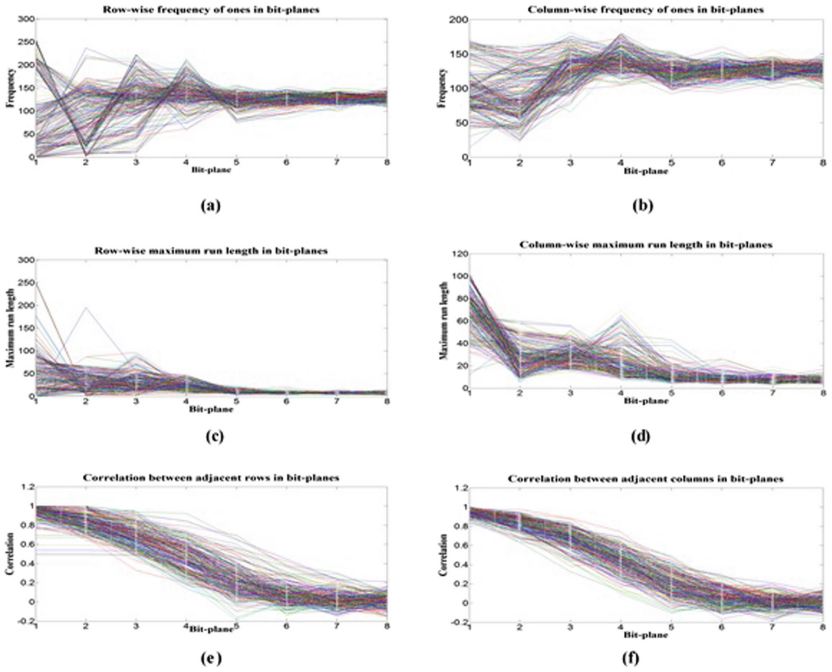


Fig. 2. Bit-plane measures for Monalisa image: (a) Row-wise frequency of ones (b) Column-wise frequency of ones (c) Row-wise maximum run length (d) Column-wise maximum run length (e) Correlation between adjacent rows (f) Correlation between adjacent columns

4 Bit-Plane Specific Image Characteristics

The bit-plane specific measures as discussed above are computed for a number of images. For illustration, the results obtained are demonstrated for Monalisa and Baboon images each of size 256×256 as shown in Figs. 2 and 3 respectively. The dynamic ranges of these measures computed for plain Monalisa and Baboon images for each bit-plane are given in Table 2. Based on Figs. 2, 3 and Table 2, we find some observations as mentioned in the following paragraphs:

(i) The adjacent bits, in rows (columns) in MSB bit-planes vary smoothly but vary frequently in LSB bit-planes.
(ii) The row-wise (column-wise) frequency of ones (zeros) in LSB bit-planes is near to half the number of bits in row (column) but in MSB bit-planes is not and it may go upto the size of row (column) of an image.
(iii) The row-wise (column-wise) maximum run length in MSB bit-planes is larger and lesser length in LSB bit-planes. Similarly, for the gap length in MSB and LSB bit-planes.
(iv) The row-wise (column-wise) correlation between adjacent rows (columns) in MSB bit-planes is higher and lesser in LSB bit-planes.

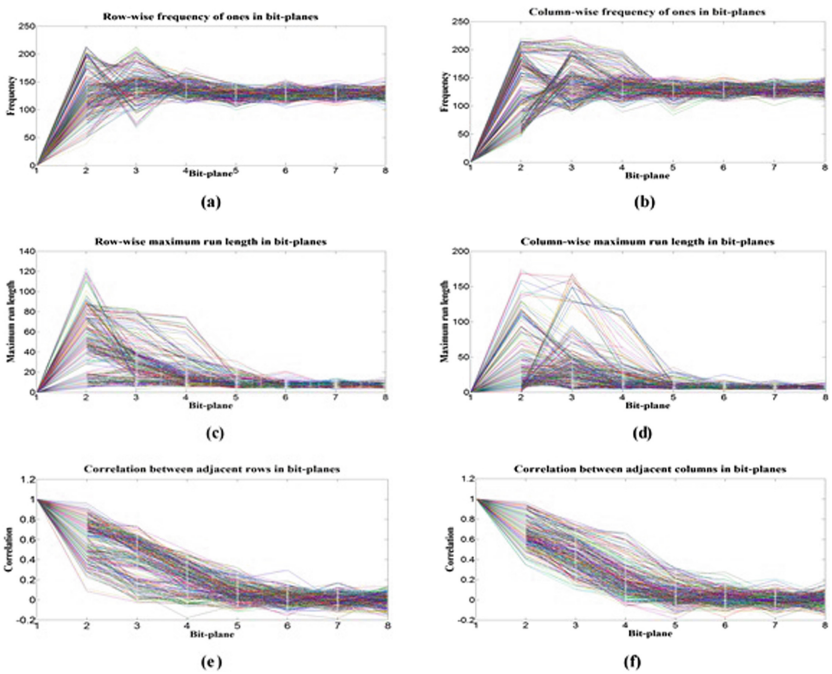


Fig. 3. Bit-plane measures for Baboon image: (a) Row-wise frequency of ones (b) Column-wise frequency of ones (c) Row-wise maximum run length (d) Column-wise maximum run length (e) Correlation between adjacent rows (f) Correlation between adjacent columns

Table 2. Dynamic range of bit-plane measures for Monalisa and Baboon images

Image	Bit-plane measure	Range of Measures in Bit-planes (BP_1-BP_8)							
		BP_1	BP_2	BP_3	BP_4	BP_5	BP_6	BP_7	BP_8
Monalisa	Row-wise frequency of ones	255 000	237 001	223 010	212 059	156 075	153 084	150 100	151 108
	Column-wise frequency of ones	170 014	163 021	181 065	180 104	154 092	160 094	156 100	151 100
	Row-wise maximum run-length	252 001	195 001	096 003	047 006	023 005	020 005	014 004	016 004
	Column-wise maximum run-length	102 012	060 006	056 015	070 008	043 005	026 004	019 004	017 004
	Row-wise correlation	1.000 0.492	1.000 0.492	0.945 0.328	0.923 0.141	0.695 −0.188	0.445 −0.141	0.304 −0.148	0.211 −0.156
	Column-wise correlation	1.000 0.836	0.977 0.633	0.883 0.328	0.742 0.102	0.492 −0.102	0.297 0.164	0.211 −0.188	0.172 −0.156
Baboon	Row-wise frequency of ones	000 000	214 039	214 067	175 110	151 100	154 104	153 108	158 103
	Column-wise frequency of ones	000 000	220 045	224 086	200 089	153 084	157 108	150 101	151 105
	Row-wise maximum run-length	000 000	123 004	082 003	075 004	031 004	021 003	014 003	015 003
	Column-wise maximum run-length	000 000	174 000	168 004	117 004	035 004	025 003	017 003	017 003
	Row-wise correlation	1.000 1.000	0.961 0.086	0.734 −0.040	0.492 −0.180	0.320 −0.118	0.297 −0.156	0.172 −0.188	0.133 −0.172
	Column-wise correlation	1.000 1.000	0.969 0.344	0.773 0.102	0.664 −0.063	0.320 −0.188	0.273 0.172	0.211 −0.180	0.188 −0.156

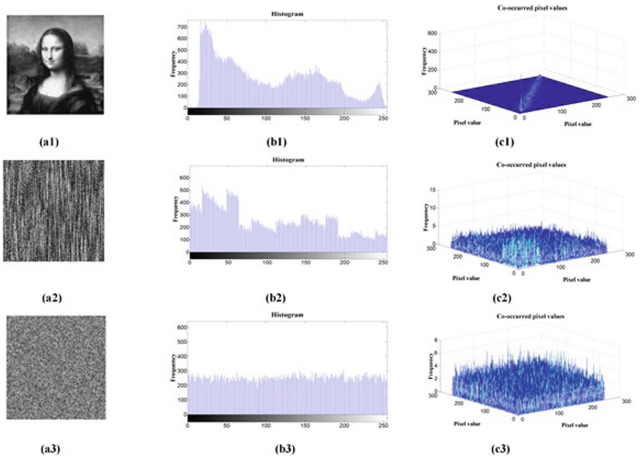


Fig. 4. Histogram and co-occurrence measures: (a1) Monalisa image (a2) encrypted image with weak cipher (a3) encrypted image with strong cipher; (b1)–(b3) histogram of (a1)–(a3); (c1)–(c3) co-occurrence plot of (a1)–(a3)

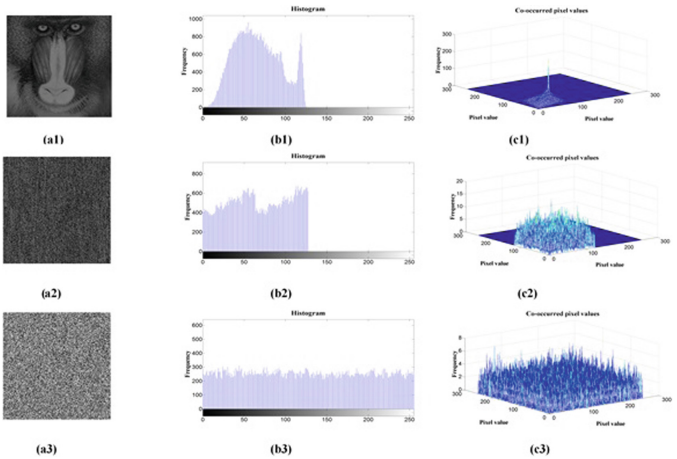


Fig. 5. Histogram and co-occurrence measures: (a1) Baboon image (a2) encrypted image with weak cipher (a3) encrypted image with strong cipher; (b1)–(b3) histogram of (a1)–(a3); (c1)–(c3) co-occurrence plot of (a1)–(a3)

Table 3. Dynamic range of bit-plane measures for encrypted Monalisa image

Type of encryption	Bit-plane measure	Range of measures in Bit-planes (BP_1-BP_8)							
		BP_1	BP_2	BP_3	BP_4	BP_5	BP_6	BP_7	BP_8
Weak encryption	Row-wise frequency of ones	114 084	106 071	149 107	167 120	142 097	149 105	150 105	148 109
	Column-wise frequency of ones	111 085	110 069	152 107	169 122	145 094	145 104	147 107	155 106
	Row-wise maximum run-length	011 003	010 003	015 004	017 005	015 004	014 004	014 004	014 004
	Column-wise maximum run-length	037 007	012 004	026 005	024 004	015 004	014 004	014 004	016 004
	Row-wise correlation	0.617 0.352	0.430 0.133	0.273 −0.031	0.297 −0.117	0.164 −0.195	0.141 −0.180	0.148 −0.211	0.188 −0.156
	Column-wise correlation	0.555 −0.211	0.344 −0.109	0.211 −0.180	0.195 −0.156	0.156 −0.235	0.156 −0.211	0.148 −0.180	0.250 −0.195
Strong encryption	Row-wise frequency of ones	147 097	149 104	151 102	146 110	145 101	147 101	145 105	151 108
	Column-wise frequency of ones	154 108	153 107	148 108	149 103	145 105	148 106	150 104	151 100
	Row-wise maximum run-length	016 004	014 004	014 004	015 004	013 004	015 004	013 004	016 004
	Column-wise maximum run-length	016 004	014 004	014 004	015 004	013 004	015 004	013 004	016 004
	Row-wise correlation	0.180 −0.164	0.141 −0.164	0.156 −0.172	0.227 −0.156	0.195 −0.172	0.195 −0.180	0.148 −0.172	0.211 −0.156
	Column-wise correlation	0.133 −0.172	0.172 −0.133	0.125 −0.164	0.195 −0.156	0.180 −0.156	0.172 −0.148	0.141 −0.125	0.172 −0.156

Above observations reflect that the row-wise and column-wise measures exhibit a lot of details, i.e., the structural detail of an image at bit-plane level. The values of these measures are very distinct and significant in MSB bit-planes but are not so distinct and significant in LSB bit-planes. The values of these measures vary for image-to-image depending on its contents.

5 Use of Measures in Analysis of Image Ciphers

For quality assessment, analysis and identification of image ciphers, we consider those measures which exhibit visual detail of image characteristics as histogram, co-occurrence of gray levels and the new bit-plane specific measures presented in this paper. An image cipher is a cryptographic system which uses for achieving image security to safeguard visual information from adversaries. An image cipher encrypts plain image by the process of encryption using a key and gives encrypted image. The process of encryption is performed by a crypto algorithm which is designed based on strong mathematical functions and primitives. For a good image cipher, the encrypted image should possess very good cryptographic characteristics, *i.e.*, an encrypted image should look like random mess of pixels

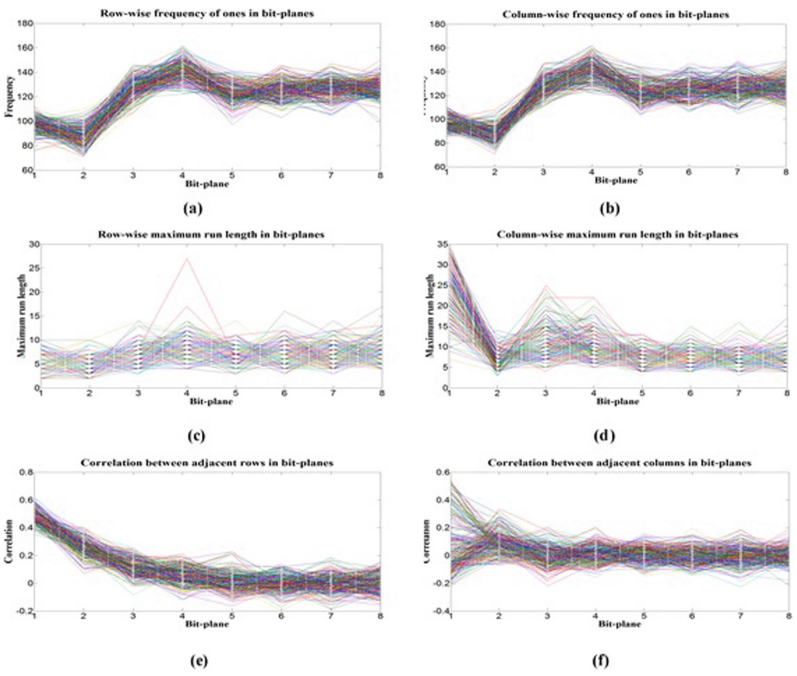


Fig. 6. Bit-plane measures for Monalisa encrypted image by weak cipher: (a) Row-wise frequency of ones (b) Column-wise frequency of ones (c) Row-wise maximum run length (d) Column-wise maximum run length (e) Correlation between adjacent rows (f) Correlation between adjacent columns

and unintelligible to the observer. Moreover, the values of the measures should also be flattened and one cannot draw any inference from these measures about plain image or encryption/decryption key information.

We consider two type of image ciphers: weak cipher and strong cipher. Mathematically, strong and weak ciphers can be defined based on crypto primitives, crypto parameters, key length, key variability, characteristics of cipher output, etc. In this paper we consider the characteristics of cipher output to analyze strong and weak ciphers statistically. A strong cipher should possess very good cryptographic characteristics and a weak cipher should possess poor cryptographic characteristics. For a strong cipher, pixel values in the image plane and bits in the bit-planes of encrypted images should be uniformly and randomly distributed. Moreover, the encrypted images obtained by strong cipher should look as random mess and do not exhibit any prominent measure which could be exploited by an adversary to extract any meaningful information about images, keys and crypto algorithm.

To present the appearance of measures computed in encrypted images, we consider rotation based image ciphers [28,29] in which we rotate bit-planes in weak cipher and transform MSB planes using inversion and xoring with Boolean

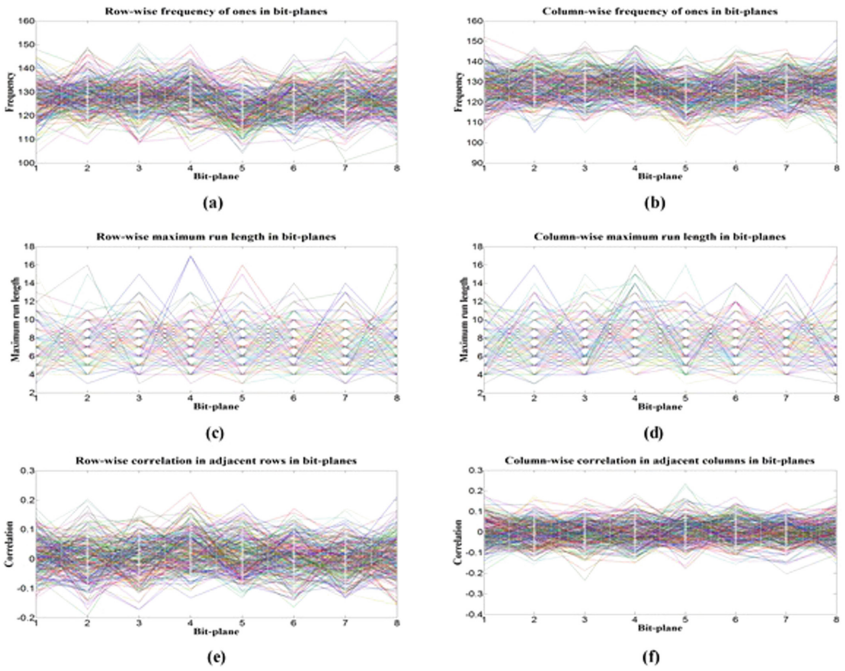


Fig. 7. Bit-plane measures for Monalisa image encrypted by strong cipher: (a) Row-wise frequency of ones (b) Column-wise frequency of ones (c) Row-wise maximum run length (d) Column-wise maximum run length (e) Correlation between adjacent rows (f) Correlation between adjacent columns

output from LSB planes prior to bit-planes rotations in strong cipher. We take the encrypted images of Monalisa and Baboon and computed all the measures. The plots of histogram and co-occurrence values for Monalisa and Baboon images are shown in Figs. 4 and 5. It is seen from Figs. 4 and 5 that the appearance of encrypted images for a weak cipher is not completely random and unintelligible as compared to a strong cipher. The histogram plots show that the gray levels in plain Monalisa image cover full range whereas in plain Baboon image cover nearly the half range but their encrypted images cover full range of pixel values. The frequency and co-occurrence distributions for the encrypted images from a strong cipher exhibit more flattened distributions compared to a weak cipher.

The frequency of ones, maximum run length and adjacent correlation for rows (columns) in all the bit-planes measured for encrypted images of Monalisa and Baboon obtained by weak and strong ciphers are shown in Figs. 6, 7, 8 and 9. The dynamic range of measures for encrypted Monalisha and Baboon images are shown in Tables 3 and 4 respectively. It is seen from Figs. 6, 7, 8, 9 and Tables 3, 4 that the values of measures for a strong cipher exhibits random and flat distributions, *i.e.*, row-wise (column-wise) frequency of ones is almost half the number of pixels in columns (rows), maximum run length is very less and

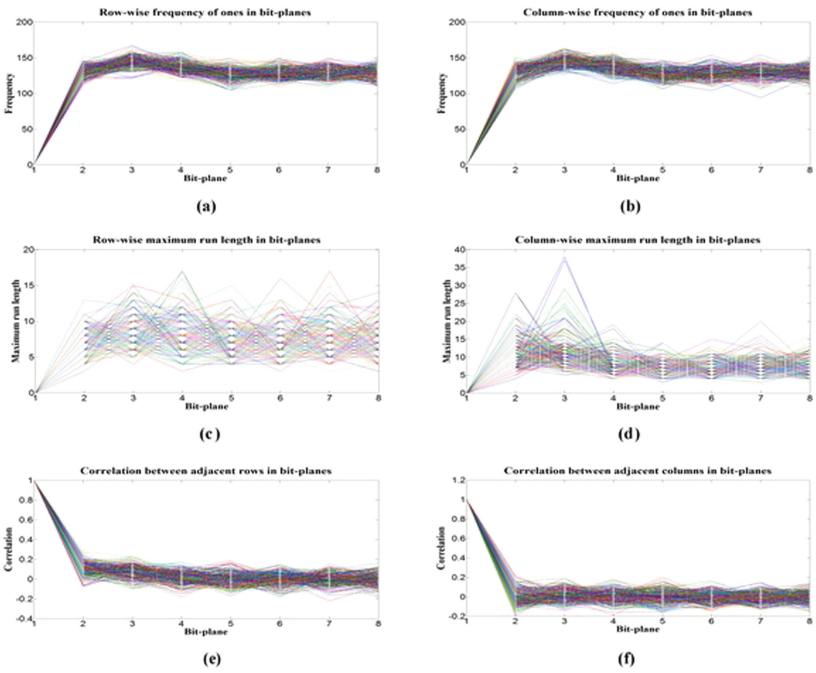


Fig. 8. Bit-plane measures for Baboon image encrypted by weak cipher: (a) Row-wise frequency of ones (b) Column-wise frequency of ones (c) Row-wise maximum run length (d) Column-wise maximum run length (e) Correlation between adjacent rows (f) Correlation between adjacent columns

Table 4. Dynamic range of bit-plane measures for encrypted Baboon image

Type of encryption	Bit-plane measure	Range of measures in Bit-planes (BP_1- BP_8)							
		BP_1	BP_2	BP_3	BP_4	BP_5	BP_6	BP_7	BP_8
Weak encryption	Row-wise frequency of ones	001 001	150 109	167 126	160 116	145 104	150 105	151 105	152 102
	Column-wise frequency of ones	000 000	145 114	161 128	158 115	150 099	148 111	149 104	149 107
	Row-wise maximum run-length	001 001	015 004	019 005	017 004	013 004	015 004	013 004	015 004
	Column-wise maximum run-length	001 001	031 005	032 005	016 004	014 004	018 004	015 004	016 004
	Row-wise correlation	1.000 1.000	0.242 −0.063	0.234 −0.148	0.211 −0.172	0.148 −0.195	0.164 −0.164	0.141 −0.133	0.195 −0.172
	Column-wise correlation	1.000 1.000	0.203 −0.258	0.211 −0.133	0.180 −0.156	0.148 −0.195	0.203 −0.180	0.188 −0.164	0.141 −0.188
Strong encryption	Row-wise frequency of ones	146 106	149 104	148 102	149 106	151 103	149 110	153 106	158 103
	Column-wise frequency of ones	158 108	151 106	148 102	151 101	148 105	155 110	152 104	151 105
	Row-wise maximum run-length	015 004	014 004	014 004	019 004	015 004	017 004	015 004	015 004
	Column-wise maximum run-length	016 004	014 004	017 004	017 004	016 004	015 004	018 004	017 004
	Row-wise correlation	0.148 −0.180	0.172 −0.133	0.188 −0.172	0.172 −0.156	0.156 −0.172	0.203 −0.141	0.172 −0.172	0.133 −0.172
	Column-wise correlation	0.149 −0.188	0.203 −0.172	0.172 −0.211	0.156 −0.156	0.180 −0.234	0.188 −0.234	0.223 −0.195	0.188 −0.156

Table 5. Characteristics of measures for weak and strong image ciphers

Quality measures	Weak cipher	Strong cipher
Visual perception	Encrypted images appear unintelligible	Encrypted images appear totally unintelligible
Histogram	Appear not flattened	Appear totally flattened
Co-occurrence plot	Appear not flattened	Appear totally flattened
Row (column) wise frequency of ones in bit-planes	Vary in MSB bit-planes	Nearly half the number of bits of column (row) in all the bit-planes
Row (column) wise maximum run length in bit-planes	Vary in MSB bit-planes	Very less variations in all the bit-planes
Row (column) wise correlation between adjacent rows (columns) in bit-planes	Vary in MSB bit-planes	Very less variations in all the bit-planes

correlation between adjacent rows (columns) is very less in all the bit-planes as compared to a weak cipher. For a weak cipher, the frequency, maximum run length and adjacent correlation among rows (columns) in MSB planes have larger values. The characteristics of these measures observed are summarized in Table 5.

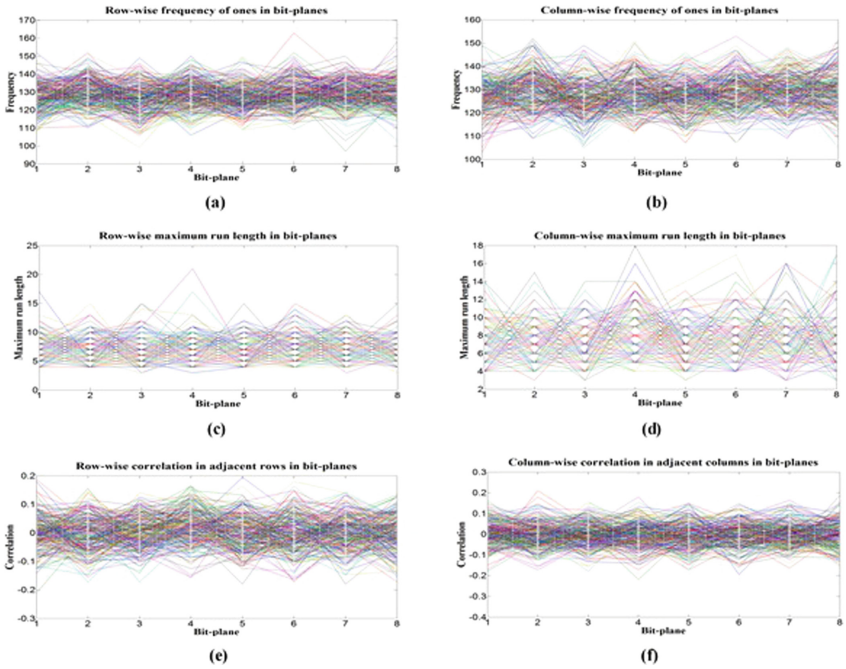


Fig. 9. Bit-plane measures for Baboon image encrypted by strong cipher: (a) Row-wise frequency of ones (b) Column-wise frequency of ones (c) Row-wise maximum run length (d) Column-wise maximum run length (e) Correlation between adjacent rows (f) Correlation between adjacent columns

6 Conclusion

Bit-plane specific measures have been presented in this paper to analyse the characteristics of the images. These measures compute the frequency of ones, maximum run length and correlation among rows (columns) for each bit-plane of an image. It has been shown that these measures exhibit extensive structural detail of an image. It has also been shown that the characteristics of image ciphers can be examined in-depth at bit-plane level with proposed measures and can be used to identify strong and weak ciphers. Apart from the applications in cryptography, the bit-plane specific measures can be used in various other applications of image processing for pattern recognition and understanding of visual data.

References

1. Eskicioglu, A.M., Fisher, P.S.: Image quality measures and their performance. IEEE Trans. Commun. **43**(12), 2959–2965 (1995)
2. Sheikh, H.R., Bovik, A.C.: Image information and visual quality. IEEE Trans. Image Process. **15**(2), 430–444 (2006)
3. Janssen, T.J.W.M., Blommaert, F.J.: Computational approach to image quality. Displays **21**(4), 129–142 (2000)
4. Avcibas, I., Sankur, B., Sayood, K.: Statistical evaluation of image quality measures. J. Electron. Imag. **11**(2), 206–223 (2002)
5. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
6. Lin, W., Kuo, C.C.: Perceptual visual quality metrics: A survey. J. Vis. Commun. and Image Representation **22**(4), 297–312 (2011)
7. Bouyer, P., et al.: Quantitative analysis of real-time systems using priced timed automata. J. Commun. ACM **54**(9), 78–87 (2011)
8. George, A.G., Prabavathy, A.K.: A survey on different approaches used in image quality assessment. J. Emerging Technol. and Advanced Eng. **3**(2), 197–203 (2013)
9. Keelan, B.W.: Handbook of Image Quality: Characterization and Prediction. Marcel Dekker Inc., New York (2002)
10. Streijl, R.C., Winkler, S., Hands, D.S.: Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives. Multimedia Syst. **22**(2), 213–227 (2016)
11. Lahouhou, A., Viennet, E., Beghdadi, A.: Selecting low-level features for image quality assessment by statistical methods. J. Comput. Inf. Technol. CIT **18**(2), 183–189 (2010)
12. Huber, P., Ronchetti, E.: Robust Statistics. Wiley, New York (2009)
13. Doane, D.P., Seward, L.E.: Measuring skewness: a forgotten statistics. J. Stat. Educ. **19**(2), 1–18 (2011)
14. Neto, A.M., Victorino, A.C., Fantoni, I., Zampieri, D.E.: Automatic regions-of-interest selection based on Pearsons correlation coefficient. In: IEEE International Conference on Intelligent Robots and Systems, California, U.S., pp. 33–38 (2011)
15. Usama, M., Khan, M.K., Alghathbar, K., Lee, C.: Chaotic-based secure satellite imagery cryptosystem. Comput. Math. Appl. **60**(2), 326–337 (2010)
16. Flusser, J., Suk, T.: Pattern recognition by affine moment invariants. Pattern Recogn. **26**(1), 167–174 (1993)
17. Alaa, E., Hasan, D.: Co-occurrence matrix and its statistical features as a new approach for face recognition. Turk J. Elec. Eng. Comp. Sci. **19**(1), 97–107 (2011)
18. Haddon, J.F., Boyce, J.F.: Co-occurrence matrices for image analysis. IEEE Electron. Commun. Eng. J. **5**(2), 71–83 (1993)
19. Zhou, W., Bovik, A.C.: Mean Squared Error: love it or leave it? A new look at Signal Fidelity Measures. IEEE Sig. Processing Mag. **26**(1), 98–117 (2009)
20. Chai, T., Draxler, R.R.: Root mean square error (RMSE) or mean absolute error (MAE)? Arguments against avoiding RMSE in the literature. Geoscientif Model Dev. **7**(1), 1247–1250 (2014)
21. Zhang, N., Vladar, A.E., Postek, M.T., Larrabee, B.: A kurtosis-based statistical measure for two-dimensional processes and its application to image sharpness. Proc. Sect. Phys. Eng. Sci. Am. Stat. Soc., 4730–4736 (2003)

22. Ratan, R.: Securing images using inversion and shifting. In: Deep, K., Nagar, A., Pant, M., Bansal, J. (eds.) SocProS 2011. AISC, vol. 131, pp. 401–412. Springer, New Delhi (2012). https://doi.org/10.1007/978-81-322-0491-6_38
23. Katzenbeisser, S., Petitcolas, F.A.P.: Information Hiding Techniques for Steganography and Digital Watermarking. Artech House, Norwood (2000)
24. Jain, A.K.: Fundamentals of Digital Image Processing. Prentice Hall, Englewood Cliffs (1995)
25. Menezes, A.J., Vanstone, S.A., Van Oorschot, P.C.: Handbook of Applied Cryptography. CRC Press, Boca Raton (1996)
26. El-Samie Fathi, E.A., et al.: Image Encryption: A Communication Perspective. CRC Press, Boca Raton (2017)
27. Stinson, D.R.: Cryptography: Theory and Practice. Discrete Mathematics and Its Applications. Chapman & Hall/CRC Press, Ontario (2005)
28. Zhang, W., Wong, K., Hai, Y., Zhu, Z.: An image encryption scheme using lightweight bit-level confusion and cascade cross circular diffusion. Optics Commun. **285**(9), 2343–2354 (2012)
29. Fu, C. and Zhu, Z.: A chaotic encryption scheme based on circular bit shift method. In: International Conference for Young Computer Scientists, pp. 3057–3061. IEEE Computer Society, Los Alamitos (2008)

Classification of Colposcopic Cervigrams Using EMD in R

Kumar Dron Shrivastav[1], Ankan Mukherjee Das[1], Harpreet Singh[2], Priya Ranjan[3], and Rajiv Janardhanan[1(✉)]

[1] Amity Institute of Public Health, Noida, India
{kdshrivastav,amdas,rjanardhanan}@amity.edu
[2] Indian Council of Medical Research, New Delhi, India
hsingh@bmi.icmr.org.in
[3] Amity School of Engineering & Technology,
Amity University, Noida, Uttar Pradesh, India
pranjan@amity.edu

Abstract. Cervical cancer is one of the most common cancer among women world-wide which can be cured if detected early. The current gold standard for cervical cancer diagnosis is clumsy and time consuming because it relies heavily on the subjective knowledge of the medical professionals which often results in false negatives and false positives. To reduce time and operational complexities associated with early diagnosis, we require a portable interactive diagnostic tool for early detection, particularly in developing countries where cervical cancer incidence and related mortality is high. Incorporation of digital colposcopy in place of manual diagnosis for cervical cancer screening can increase the precision and strongly reduce the chances of error in a time-specific manner. Thus, we propose a robust and interactive colposcopic image analysis and diagnostic tool, which can categorically process colposcopic images into Type I, Type II and Type III cervigrams and identify lesions in least amount of time. Furthermore, successful binning of diagnosed cervigrams into digital colposcopic library and incorporation of a set of specific parameters that are typically referred to for identification of transformation zone and SCJ (squamo columnar junction) with the help of open source Programming language - "R" is one of the major highlights of the application. The software has the ability to automatically identify and quantify the morphological features, color intensity, sensitivity and other parameters digitally which may improve and accelerate screening and early diagnosis, ultimately leading to timely treatment of cervical cancer.

Keywords: Cervical cancer · Cervigrams · Early detection
Digital pathology · Screening

This work is supported by a fellowship grant by Indian Council of Medical Research, New Delhi, India. The initial set of colposcopy Images were acquired from Batra Hospital & Medical Research Centre.

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 298–308, 2019.
https://doi.org/10.1007/978-981-13-5758-9_25

1 Introduction

Cervical cancer has increasingly become the fourth most common cancer among women world over, affecting over five lakh women annually [1]. The incidence is higher in countries like India and Africa where more than three fourth of invasive cervical cancers occur with a recorded estimated 67,477 disease-related deaths. Whilst cervical cancer incidence is on the rise, predominantly in low and middle income countries, however, during the last decades, successful implementation of screening and treatment programmes has quite drastically reduced the disease burden in the developed nations. Cervical cancer originates as a few abnormal cells in the initial stage and spreads rapidly if left untreated. The World Health Organization (WHO) recommends that countries ensure that women from 30 to 49 years are screened at least once in their lifetime [2]. In countries like India most cervical cancer cases are diagnosed at a later stage when curative treatment becomes ineffective. Thus, early detection of cervical cancer has pertinent implications for successful treatment in reducing disease related morbidity and mortality. In recent past, several tools and techniques have been developed for early detection of cervical cancer to minimize the burden of medical professionals but none have been incorporated in clinical practice nor have these tools been evaluated or utilized for implementation of cervical cancer screening programmes. This paper for the first time proposes to employ computer aided detection (CAD) techniques for the purpose of assisting pathologists in the early detection of cancer by processing and analyzing Colposcopic cervigrams. Colposcopy is a diagnostic tool wherein visual examination of the cervix of a woman is performed mainly by a gynecologist which aims at the detection of abnormalities present in forms of color and morphology features. This requires precision, experience and rigorous amount of knowledge about: Normal Transformation zone and Abnormal Transformation Zone, Normal SCJ (Squamo Columnar Junction) and abnormal SCJ, CIN1 (Cervical Intra epithelial Neoplasia), CIN2, CIN3 which is also known as per-cancerous. Acetic acid is used on cervix for changing the color of affected area into white, which separates the abnormal areas or lesion from normal region. The trivial task which consumes the maximum time is identification of aceto-white region through visual inspection and categorization of malignancies on the basis of their stage and grade. In practice it sometimes leads to misdiagnoses and missed diagnosis because it relies heavily on the subjective knowledge of medical professionals without any finite quantitative measurement available for colposcopic image analysis. This uneven diagnosis often leads to poor clinical prognosis. Thus, the prime goal of the study is to develop an application which is free of cost, easily available and accessible in remote areas with minimum resources for early and quick detection of cervical cancer lesions through the use of a digital colposcopy tool clearly separating the abnormal from the normal zones quantifiably.

2 Methodology

We used 'R' Version 3.4.4(2018-03-15) as programming language which is an open source software equipped with highly advanced image analysis tools. MySQL Version 14.14 Distrib 5.5.60 for creation of image repository for further processing and LINUX version 14.04 as base platform was also utilized. Initially representative Cervigrams were acquired using Digital Colposcope (Mobile ODT, Israel) from anonymized subjects routinely evaluated for cervical lesions at the Out Patient Department of Batra Hospital and Medical Research Centre (BHMRC), New Delhi. The images retrieved from the cervigrams were preprocessed before being analyzed using segmentation algorithms. One of the biggest challenge marginalizing the accuracy of the digitally acquired images pertains to the lack of pre-processing algorithms for segregating right set of data features, so as to ensure data reduction and redundancy in the image structures associated with the cervical cancer lesions. This is where major research effort is required to clearly identify reliable and robust visual markers which remain invariant under different optical transformations and even in the presence of unwanted objects in images like hair and other body parts. Such a process will eliminate artifacts and other irrelevant structures to seamlessly generate structures for grading measurements. The automation of this feature ensures that the algorithm will detect structural aberrations in cervigrams in a precise and timely manner. This paper is focused on binning of different colposcopic images into its categories (endocervix, ectocervix, and third type having endo and ectocervix) using image repositories provided by Mobile ODT, Intel & Kaggle [3]. EMD (Earth Movers Distance) algorithm is being used for comparison of unknown images with three different sets of cervigrams acquired from https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/data on 02-01-2018.

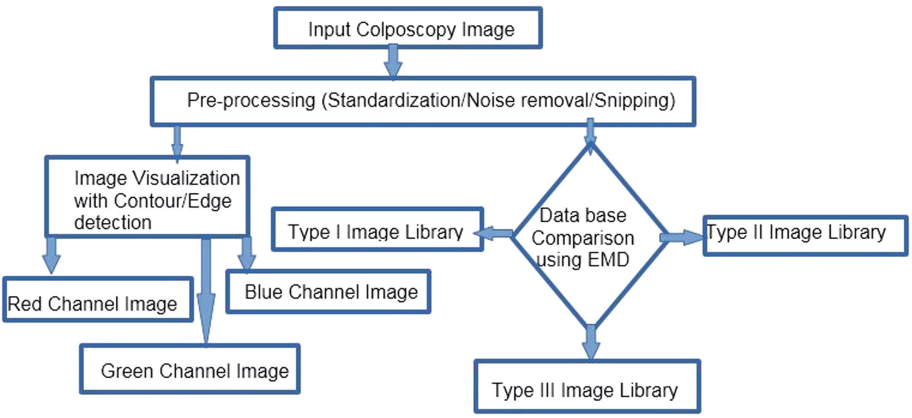


Fig. 1. Data flow diagram (DFD) of proposed system

2.1 Why Use Earth Mover's Distance

The Earth Mover's Distance (EMD) is used for content-based image retrieval. The EMD is based on the minimal cost that must be paid to transform one distribution into the other, in a precise sense, and was first proposed by Peleg, Werman, and Rom. For image retrieval, we combined this idea with a representation scheme for distributions that is based on vector quantization. This combination leads to an image comparison framework that often accounts for perceptual similarity better than other previously proposed methods [4]. It is more robust than histogram matching techniques operating on variable length representations of the distributions that ameliorates quantization and other binning problems typically associated with histograms. Briefly, we describe the steps for the per-processing algorithm used to screen and process the images for Earth Movers Distance (EMD) based segregation.

The anonymized cervigram images acquired from BHMRC was re-sized to 30×30 pixels each. In order to have a precision oriented segregation of cervix, green channel of the image was taken due to the pink coloration of the cervix. The green channel was used to extract the features of cervix by computational method inbuilt into the per-processing algorithm. The image features were subsequently clustered to get the best output, which was converted into mask and subsequently processed with "Canny" edge detection algorithm to extract and segregate images based upon the color coding of the abnormal lesions. The green channel of the image of the lesions having highest color intensity was taken as the abnormal lesion with contour. The pre-processed images were subjected to a classifier algorithm called earth movers distance (EMD). The classifier algorithm functioned as a metric to gauge the progression based upon the similarity/dissimilarity index between the structural features of cervix. Whilst, our preliminary data based upon a small sample set of images is showing congruence to the proposed hypothesis, this association would be non-linear when being sampled in a larger population. The automated preprocessing and segmentation of the cervigrams acquired from a digital colposcope would help in the establishment of a novel triage based method to not only prioritize the treatment options but also give a grounded and realistic picture about the prevalence of this disease. A grand view of algorithmic processing can be roughly summarized in Fig. 1. Figures 2, 3, 4 and finally Fig. 5 show different stages of processing in computational environment of R. Figure 6 shows different orientation for cervix which needs to be detected and categorized for a given colposcopic image.

3 Computational Results

Data represented in Table 1 indicates the EMD comparison matrix between random unknown colposcopic images and classified images of category 1, 2, 3, 4 and 5 (assumed category) acquired for primary testing. After using EMD algorithm comparison of every unknown colposcopic images with already classified colposopic images were made and numerical values were tabulated indicating exact

match between Testimg1 and Cat1 with the numerical value zero, signifying Testimg1 belongs to category1 images. This was a pioneer research and test for further EMD based classification as shown in Table 1. Later, we acquired 1481 colposcopic images from https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/data [3] comprising three groups (Group 1, Group2, Group3) and created MySQL libraries of already classified images based on Fig. 7 into three groups and then compared unknown colposcopic images group with the libraries as shown in Fig. 6, one after another for identification and classification into Ectovervix, Endocervix and the third type. By using EMD for quantifiable comparison, we were successfully able to classify unknown images into Ectocervix, Endocervix and third type of cervix.

Results obtained from our software forms the rationale for recognizing the statutory importance of artificial intelligence aided integretomics based strategies for developing niche specific triage methods for prioritizing and tailoring treatment options. It will also provide affordable and accessible options for patients inhabiting remote areas to interact with the physicians through low end mobile healthcare technologies. Use of low end mobile health based strategies will help in spreading awareness at a community level on the niche specific risk factors contributing to the demographic spread and poor clinical prognosis of cervical cancer. This will not only contribute in overcoming region specific social stigmas and taboos associated with prevalence of cervical cancer but also facilitate remote connect of cancer afflicted patients with the treating physicians.

The final output as shown in Fig. 5 which has a special zoom in feature for clear visualization and understanding of more deeper insights of ROI (Region of Interest) and/or aceto-white area with the fast and robust feature of R that makes it feasible to match unknown images with the already classified and trained image repository for quick and easy classification into Ectocervix, Endocervix and the third type. The automated preprocessing and segmentation of the cervigrams acquired from a digital colposcope would help in the establishment of a novel triage (priority identifier for patients) based method which is in process to not only prioritize the treatment options but also give a grounded and realistic picture about the prevalence of the disease.

Detection Results

Table 1. EMD value after comparison between cervical Images.

	cat1	cat2	cat3	cat4	cat5
Test Image1	0	33.74772	27.92506	36.17544	18.46595
Test Image2	33.60843	0	33.60843	6.737279	21.31866
Test Image3	27.99849	12.20216	0	16.44538	13.54746
Test Image4	36.1909	6.740369	16.46318	0	24.19017
Test Image5	18.4986	21.5642	13.54746	24.18698	0

Fig. 2. Original colposcopy image

4 Discussion and Putting It in Perspective!

Majority of the cervical cancer cases and deaths occur in low-income and middle income countries due to non-addressal of the existent knowledge gaps among the community and healthcare providers with under planned and ineffective cancer screening strategies. Our study highlights the importance of a digitally aided diagnostic tool for pre-processing cervigrams for quantifiable detection of cervical cancer lesions using Earth Mover Distance (EMD). The data from the study shows high validity while matching and binning anonymous colposcopy images to different categories based on the quantitative values generated through the algorithm. The test images are said to match the specific category if the EMD value advances to zero. The algorithm has been successful in segregating the acquired images to the three different cervix groups and is sufficiently quick.

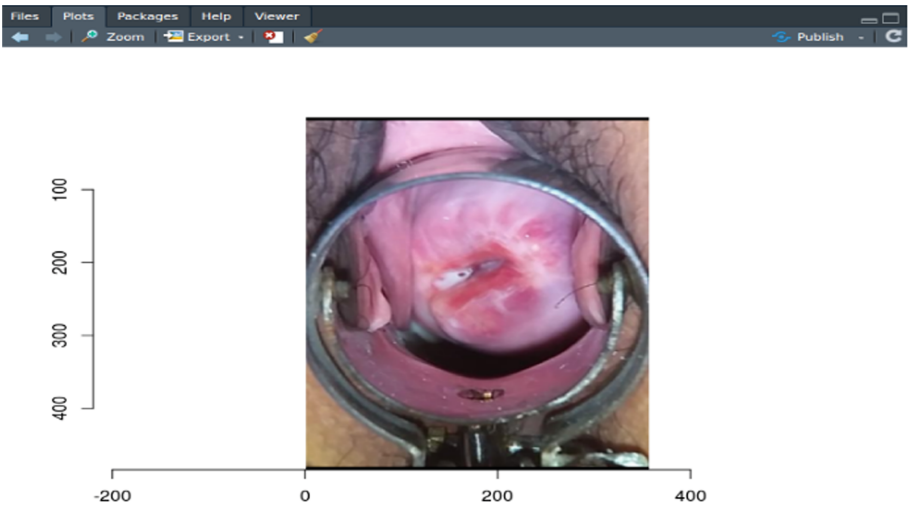


Fig. 3. Preprocessed colposcopy image in R

According to recent guidelines, the World Health Organization (WHO) recommends a "screen and treat" approach using visual inspection with acetic acid (VIA) in low resource settings [5]. Comparatively, colposcopy practices have not yet been strongly looked into considering the advances made in developing algorithms for cancer screening [6,7]. The high burden of cervical cancer across the Indian sub-continent, the social challenges afflicting women along with

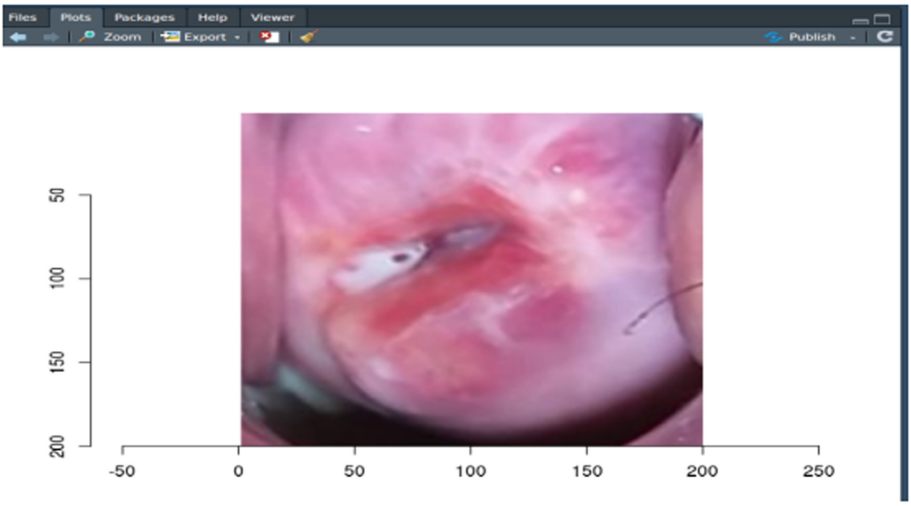


Fig. 4. Cropped image

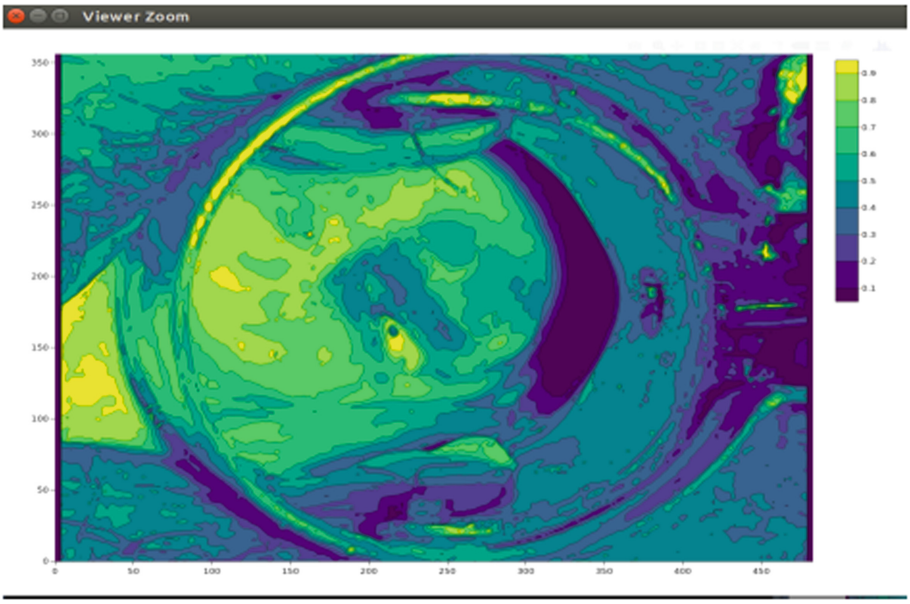


Fig. 5. Green channel with edge detection of cervix and its surroundings (Color figure online)

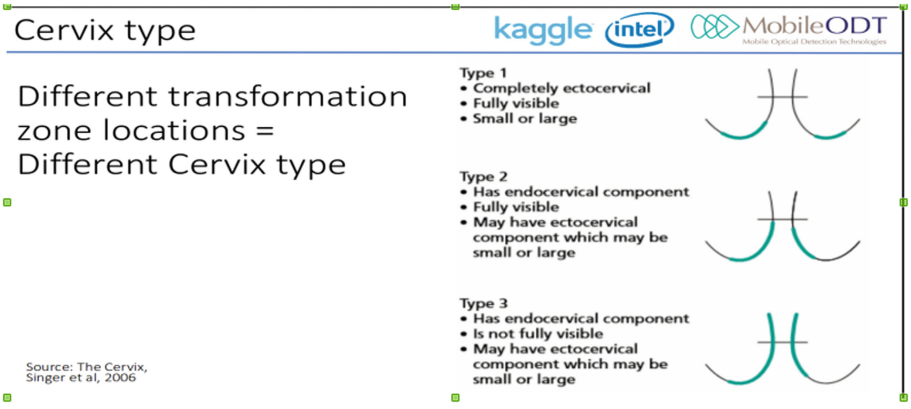


Fig. 6. Classification of different types of cervix

our limited understanding for highlighting triage methods in prioritizing specific interventions dramatically effects our plan for evolving concepts and sustainable strategies towards management of cervical cancer in LMICs [8]. An existing lacuna with respect to the majority of published studies, pertains to their restrictive use for either computer aided diagnosis systems based on either cytology or colposcopy image analysis. However recently, few papers have been published on

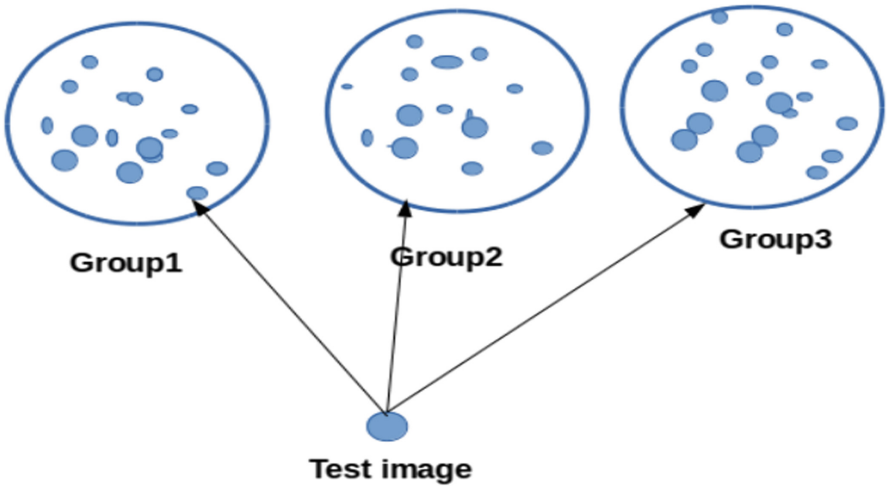


Fig. 7. Comparison of random images with already categorized image repository

artificial neural networks (ANNs) for improved detection, treatment, and follow-up support of cancer [8–10]. Portable digital colposcopes which are commercially available: the Gynius AB's Gynocular [11–13] and Mobile ODT's EVA [14–16] have encountered a number of problems with image quality as they use onboard smartphone camera. When it comes to visual inspection by a camera with direct human eye inspection then the clarity of image needs to be very clear and compact. Specular reflection and blurred images with color mismatch from human eye view of cervix are other challenges associated with these portable colposcope [17,18]. This is one of the areas where the results of this paper focuses and sufficiently fulfills the required need for improving image quality before processing for identification of cancer lesions.

Diagnosis of cancer is a very sensitive and crucial task. We require an accurate and specialized expertise to confirm cancer report which is not available everywhere, especially in remote areas and the tumor progresses to advanced stage when curative treatment is no longer an option. The prevalence of misdiagnosis and missed diagnosis is also high in absence of portable, accurate and robust diagnostic tool. The R software that we are using has advanced image analysis libraries available for medical image analysis [19], which deals with the quantification of the Colposcopic cervigrams, thus can reduce the burden and time of medical professionals by providing accurate and clear view of colposcopy images with numerical data for comparison and classification.

In summary, our findings after colposcopy image analysis using EMD in R and color coding of images with contour enables the thought of more precise, clear, accurate and fast representation and classification of Colposcopic cervigrams with minimum resources utilized. The software is better than existing colposcopy software in terms of robustness, cost, accuracy and amount of resource

required. The study focuses further to investigate the potential role of a novel intelligent bioinformatics CDSS (Clinical Decision Support System) which intelligently combines the results of various diagnostic techniques used in the modern laboratory in order to provide clinicians with patient-specific predictions of diagnostic or prognostic outcomes and thus can identify women who are at true risk of developing cervical cancer.

Acknowledgement. The authors duly acknowledge the support of mobileODT for the provision of cervigrams obtained through mobileODT EVA system at Batra Hospital & Medical Research Centre, New Delhi, India under the supervision of Dr. Shelly Batra.

References

1. Human Papillomavirus and Related Diseases Report WORLD version posted at www.hpvcentre.net on 27 July 2017 (2017). http://www.hpvcentre.net/statistics/reports/XWX.pdf
2. Screening as well as vaccination is essential in the fight against cervical cancer. http://www.who.int/reproductivehealth/topics/cancers/fight-cervical-cancer/en/. Accessed 09 July 2018
3. Intel & Mobile ODT Cervical cancer Screening. https://www.kaggle.com/c/intel-mobileodt-cervical-cancer-screening/data. Accessed 10 July 2018
4. Yossi, R., Carlo, T., Leonidas, J.G.: The Earth Mover's Distance as a metric for image retrieval. Int. J. Comput. Vis. **40**(2), 99–121 (2000)
5. WHO guidelines for screening and treatment of precancerous lesions for cervical cancer prevention. http://apps.who.int/iris/bitstream/10665/94830/1/9789241548694_eng.pdf?ua=1. Accessed 21 May 2017
6. Saslow, D., Solomon, D., Lawson, H.W., et al.: American Cancer Society, American Society for Colposcopy and Cervical Pathology, and American Society for Clinical Pathology screening guidelines for the prevention and early detection of cervical cancer. CA Cancer J. Clin. **62**(3), 147–172 (2012)
7. Massad, L.S., et al.: 2012 updated consensus guidelines for the management of abnormal cervical cancer screening tests and cancer precursors. Obstet. Gynecol. **121**(4), 829–846 (2013)
8. Ginsburg, O., et al.: The global burden of women's cancers: a grand challenge in global health. Lancet **389**(10071), 847–860 (2017)
9. Lin, H.-C., Wu, H.-C., Chang, C.-H., Li, T.-C., Liang, W.-M., Wang, J.-Y.W.: Development of a real-time clinical decision support system upon the web MVC-based architecture for prostate cancer treatment. BMC Med. Inform. Decis. Mak. **11**(16), 1–11 (2011)
10. Karakitsos, P., et al.: Identification of women for referral to colposcopy by neural networks: a preliminary study based on LBC and molecular biomarkers. J. Biomed. Biotechnol. **2012**, 1–8 (2012). https://doi.org/10.1155/2012/303192
11. Nessa, A., et al.: Evaluation of stationary colposcope and the gynocular, by the Swede score systematic colposcopic system in VIA positive women: a crossover randomized trial. Int. J. Gynecol. Cancer **24**(2), 339–345 (2014)
12. Ngonzi, J., et al.: Agreement of colposcope and Gynocular in assessment of cervical lesions by Swede score: a randomized, crossover pilot trial. J. Lower Genital Tract Dis. **17**(4), 372–377 (2013)

13. Peddie, D.: Cervical cancer screening in Kiribati. Pac. J. Reprod. Health **1**(3), 132–134 (2016)
14. Kass, A., Slyper, R., Levitz, D.: Optical design of low cost imaging systems for mobile medical applications. In: Optics and Biophotonics in Low-Resource Settings, vol. 9314, pp. 93140B-1–93140B-6 (2015)
15. Mink, J.W., Wexler, S., Bolton, F.J., Hummel, C., Kahn, B.S., Levitz, D.: Initial clinical testing of a multi-spectral imaging system built on a smartphone platform. In: Optics and Biophotonics in Low-Resource Settings II, vol. 9699, p. 96990R. International Society for Optics and Photonics (2016)
16. Bolton, F.J., Weiser, R., Kass, A.J., Rose, D., Safir, A., Levitz, D.: Development and bench testing of a multi-spectral imaging technology built on a smartphone platform. In: Optics and Biophotonics in Low-Resource Settings II, vol. 9699, p. 969907. International Society for Optics and Photonics (2016)
17. Rashmi, B., et al.: Feasibility of using mobile smartphone camera as an imaging device for screening of cervical cancer in a low-resource setting. J. Postgrad. Med. Educ. Res. **50**, 69–74 (2016)
18. Catarino, R., et al.: Smartphone use for cervical cancer screening in low-resource countries: a pilot study conducted in Madagascar. PloS one **10**(7), 1–10 (2015)
19. Muschelli, J.: Neuroconductor: an R platform for medical imaging analysis. Biostatistics (2018)

Automated Analytical Model for Content Based Selection of Web Services

S. Subbulakshmi[1], K. Ramar[2], Aparna Omanakuttan[1(✉)], and Arya Sasidharan[1]

[1] Department of Computer Science and Application, Amrita Vishwa Vidyapeetham, Amritapuri, India
subbulakshmis@am.amrita.edu, aparnaomanakuttan95@gmail.com, aryaponni95@gmail.com
[2] Department of Computer Science and Engineering, Einstein College of Engineering, Tirunelveli, India
kramar.einstein@gmail.com

Abstract. There are various inbound web services which prescribe services to clients. Specialists are more engaged in making framework for proposal of web service (WS) which limit the intricacy of selection process and improve the quality of service (QOS) suggestion. Our work implements a framework which recommends web services using an analytical model based on the contextual information provided by the service providers. This system helps users obtain high quality service automatically. Adaptive work performs feature reduction, similarity and ranking of WS. The important feature reduction process helps identify attribute values with maximum accuracy which results in proper evaluation of data. Efficient selection of WS for service composition requires better methods which properly calculate the similar values. A similarity helps to identify the closest services as per the requirement in the process of service composition. Ultimately, the system automatically selects the set of web services with highest similarity scores from the optimized set of web service description.

Keywords: SVD · Analytical model · Content based · Recommendation
Feature reduction

1 Introduction

Web services are functional modules widely used in the process of service composition for the implementation of complex business logic. Today, huge numbers of internet service providers are attempting to make their services accessible through the Internet, providing a wide range of functionalities. This has lead to a steep increase in the number of web services. In the contemporary webservice era, the selection of the most effective service is becoming an intricate exercise. The subtle elements of the existing webservices determined by the service providers are accessible in the registry of webservice Universal Description, Discovery, and Integration (UDDI). The service requester utilizes the Web Service Description Language (WSDL) to get into the WS as per their necessities. Users/Service Composition Agents on all occasions opt for

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 309–321, 2019.
https://doi.org/10.1007/978-981-13-5758-9_26

services that make up their demand. The recommendation is methodical only when the opted web services are efficient. Efficient selection requires the reduction of services that are less relevant or ineffective.

Most of the recommendation systems focused on recommending services based on content or collaborative filtering techniques or both [1]. Content based technique; makes available services options to a target user that are similar to the requirements or preferences of the user. Collaborative filtering technique on the other hand identifies users having similar behaviour and a set of similar web services and, recommends what similar users have liked or opted for. Recommendation based on quality factors is also considered. Quality factors are nothing but the non-functional characteristics of the service experienced by typical users [2] such as query response time, (download) data throughput, etc. The proposal of the quality WS incorporates determination, disclosure furthermore, sifting of the required services.

In this paper, we present a new concept of handling the scalability snag using an analytical method called singular value decomposition (SVD). SVD is one of the best matrix factorization technique which used to reduce the dimensionality of the input dataset. Feature reduction is the core characteristic of SVD, which selects only the significant properties of webservices while keeping out unwanted features. It is important to know the actual attribute values, which are essential to select the quality services.SVD produces more accurate results when compared with other factorization techniques like PCA principal component analysis and NMF non-negative matrix factorization technique.

The following imperative advance in the wake of utilizing the feature reduction is to pick the most efficient services. This is done using two linkage algorithms under agglomerative approach in hierarchical clustering. The first single linkage algorithm calculates the minimum distance using the Euclidean distance formula and forms clusters. The second is ward linkage which merges web services build on the peak value of an objective function. The results are depicted in dendrogram based on distance. Finally, ranking of services is done based on the order from the former result.

The rest of paper includes the study of related works in Sect. 2 which describes the popular methods in web service recommendation, proposed work in Sect. 3 using SVD feature reduction technique and its illustration, followed by results in Sect. 4 and closing remarks in Sect. 5.

2 Literature Review

In the course of recent years, numerous looks have been done to suggest web services. Prominent approaches associated with this are content based and collaborative filtering. Combined forms of these two methods are also popular. The present arrangement of substance and community-oriented based approach recognizes the best web services in view of the client choice and other client's inclinations for related WS.

One popular hybrid approach which unifies both the filtering techniques [1]. This used content and collaborative filtering techniques, considering rating information and semantic substance information of a web, utilizing a probabilistic generative model.

All the more, particularly, it is based on a three-way angle which demonstrate the straightforwardly speaks to inconspicuous client preference as an arrangement of inactive factors.

Related to the above work in, a framework has been introduced which actualizes a structure for proposal of customized WS [3] combined with the quality improvement, utilizing the quality features accessible in WS Ontology. It encourages clients to obtain the best suggestion by expending the relevant data and the nature of WS. Diverse systems for figuring thing similitudes were also proposed [4]. They analysed that thing based calculations gave significantly better execution than client based calculations.

Another approach considering users' individual preferences, [5] a CF-based web service proposal technique utilizing clients' singular inclination on non-functional properties is proposed. Along with the QOS attribute values of called services, to promptly recommend services to an active user is proposed. The proposed technique precisely processes the similarity between clients and services by joining clients' individual preference on non-utilitarian characteristics in the proposed similarity work. Similar to this [2] propose a novel suggesting system in view of Important Client Group (ICG) fusing customary synergistic sifting calculation to accomplish a strong web benefit proposal.

Amazon proposed thing-based, community-oriented shifting and they have customized the suggestion [6] for every one of its clients. In another framework, they have consolidated thing-based shared sifting and socioeconomics [7] for suggesting the WS. This half and half approach helps in staying away from issues like adaptability and sparsity and it gives better performance than the past proposal frameworks. But this work does not generally respond to each client in a similar way, as the non-functional elements influence every client's condition.

Another area related to this take the substance of the URLs perused by the client [8] and run a bunching calculation to create the bunches of points perused for the current timeframe. Past usage of clients was also taken in one approach to recommend using collaborative filtering [9]. Recommendation systems are extensively [10] used in internet business applications. The proposed framework manages the online recovery of books by extricating data from the client audits and gives scores to each book.

A survey of recommended methods is proposed [11] which are useful in understanding content based filtering, Collaborative undertakings and their principal challenges, for example, information sparsity, and grey sheep. Subsequently, a hybrid approach was introduced. A survey is performed related to methods and algorithms which were used for web service recommendation by different authors.

In all the existing work, the web services are selected based on the web services description which represents both relevant and irrelevant features. In this proposed system we designed a methodology which uses singular value decomposition to select only the relevant service description by eliminating the irrelevant features. This optimization process of the service description helps to select the best web services.

3 Proposed Methodology

Determination of the improved web benefit is the primary goal of proposed work as the current web period is overflowed with various web services given by different service providers. Choosing the exceptionally reliable web service is a vital need for the individuals. For our strategy (SVD_SL) using SVD and Single linkage, we have taken a domain with tremendous services. These web services are arranged with attributes. Fig. 1 depicts architecture of our work. Each domain is identified from WSDL using queries. After this filtration, WS details are given to SVD for feature reduction. The current result is then passed to linkage algorithms for the final ranking.

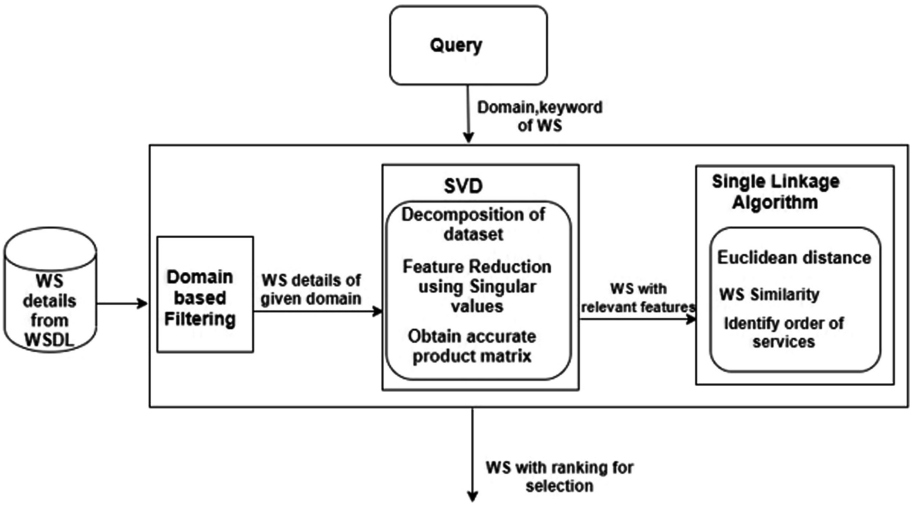


Fig. 1. Methodology for SVD_SL Web service recommendation

Here travellers are taken as our domain and each traveller is flooded with different facilities. Each attribute is weighted based on the quality factors, functions and other add on properties. Functions were given more weightages comparatively. Picking significant service is a troublesome assignment.

Proposed Algorithm SVD_SL

Input: Given WS details W and domain name D.

1. Domain identification.
   - Filter W with D to get web service details W'.
2. SVD feature reduction.
   - Decomposition of matrix W'.
   - Eliminating irrelevant features using singular value.
   - Obtaining accurate set of services W''.
3. Hierarchical clustering based similarity between services in reduced dataset.
   - Distance measured using Euclidian formula.
   - ClusterWS using similarity matrix.
   - Order identification from dendrogram.

Output : Ranking of WS based on the result of single linkage hierarchical clustering.

Our work uses a feature reduction process using sing. This finds most accurate values corresponding to web services and properties. After feature reduction, similarity algorithm is used to obtain services with maximum features.

3.1 Domain Identification

The introductory task of our work starts with the recognition of domain from the web service details. Details of WS from WSDL are populated into the dataset which includes operations, attributes and quality factors of each service. This task is mandatory since WSDL file is surged with different domains. All service providers publish its service description in the WSDL directory. The WS details dataset is initially filtered based on the domain name given in the query.

3.2 SVD Feature Reduction

The analytical model used for our work is singular value decomposition. This is a proved concept [12] where the result shows that the groupings of the information were five times quicker utilizing SVD based calculation. It is perhaps the most widely used unsupervised machine learning algorithm and has great applications in recommended

systems. SVD is known for dimensionality reduction [13] but it is not same as that of reducing the dimensionality of data. SVD algorithm is as follows:

Algorithm for SVD

Input: Web service description W' matrix of $m * n$,where m represent service and n represents feature of webservices.

1. Decompose the above matrix into three matrices, U, V and $\sum$ using equations:
$$W'^{T} W = V \sum^{T} \sum V^{T} . \quad (1)$$
$$W' V = U \sum . \quad (2)$$
where U is left singular vector, V is right singular vectors, $\sum$ is singular value

3. Eliminating rows from V and column from U based on $\sum$.

4. Obtain the product W'' of U, V, $\sum$ which includes accurate featuresof WS.

Output: Reduced set of web service description with accurate features.

SVD is a method of fragmenting a matrix into three matrices. This calculates most accurate values from our sample dataset. In this method, dimensionality reduction is done by eliminating least singular value from the diagonal matrix. Dimensionality reduction in this work refers not to reduce or eliminate values from dataset. But focus on the generation of accurate values corresponding to each value in the sample dataset. It is directly applied to the dataset which is considered as an input matrix. Using the equations elaborated in the below algorithm, three matrices are obtained. Despite of how many singular values approximately set to zero, the resulting matrix always retains its original dimension. The component of dimensionality reduction is just abused in the decomposed version.

3.3 Similarity Algorithm

After the feature reduction process, there is a requirement for recognizing comparable services from framework. The former result obtained from SVD is now given to find similar scores. This is necessary to get the ranking order of services. Similarity measures provide a way for calculating most similar ones. Here, single linkage and ward linkage clustering are used since it contains merge in each step and two closest members have the smallest distance [14].

Algorithm for Hierarchical clustering based similarity – Single and Ward Linkage

Input : Dataset W'' with accurate set of WS details.

1. Calculate Euclidean distance and create distance matrix.
$$d(x,y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots (x_n - y_n)^2} \quad . \tag{3}$$
x and y represents two web services of the W'' where x_1, x_2 etc. are the features of WS.

2. Identify values similar with two linkage methods.
   - Using single linkage
   $$d(x,y) = \min\big(dist(x[i], y[j])\big) \quad . \tag{4}$$
   for all points i in cluster x and j in cluster y.
   - Using ward linkage
   $$d(x,y) = \sqrt{\frac{|y|+|s|}{T} d(y,s)^2 + \frac{|y|+|t|}{T} d(y,t)^2 - \frac{|y|}{T} d(s,t)^2} \quad . \tag{5}$$
   x is newly joined cluster consisting of clusters s and t. y is the unused cluster.

3. Create two dendrograms $d1$ and $d2$ from two linkages.

4. Identify services from the dendrogram to obtain the ranking from two results.

Output: Final ranking of web services.

This bunching is one of a few strategies for hierarchical clustering [15]. There are many similarity algorithms available at present. But linkage algorithms are desirable for our work so as find the ordering of web services based on the distance measure. In single linkage, distance calculation is based on Euclidean distance formula. Ward linkage is measured using ward variance minimization formula. Dendrograms are the most relevant part from these algorithms. They are in effect, visualization of the cluster arrangements obtained after the single linkage clustering and ward linkage clustering. These algorithms helps to identify the closest similar services gradually based on the distance values calculated. Web services are added to the cluster based on their similarity. Clustered WS depicted in the dendrogram helps to identify the order in which WS are recommended.

3.4 Ranking of Services

The final process is suggestion of high quality services based on the efficiency. This process is done considering contextual information of WS. WS are ordered based on domain names, SVD with single linkage and SVD with ward linkage.The recommendation of WS is done based on the output of the above three results.

The ranking of WS is proved to be accurate with the following root mean square error (RMSE) formula:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(A_i - F_i)^2}{n}}. \quad (6)$$

where A_i is the actual ordering of web services observed in the real world scenario with the services in travellers domain provided by different providers, F_i is the SVD_SL ranking of WS and n is the number of ranked web services. This method compares the actual with the predicted results. The prediction error in SVD_SL is proved to be minimum within the acceptable range. Therefore, ranking of SVD with single linkage algorithm which is closest to the importance of WS is considered for final recommendation.

4 Experimental Results

In this work, we have introduced a recommendation system based on feature reduction. The experiment is done with a dataset which is actually an extracted set of data, obtained after filtering of domain. These results are the part of service composition. This section is abstract for a user requesting the system.

We have taken travellers as our domain to find the suitable ranking of services. Each row includes the details of different web service and each column includes the attributes/features of web services.

Table 1 shows the sample of the dataset of web services description. In the table P1, P2,…PN represents various attributes of web services which includes domain name, properties of WS, operations and quality factors of WS. Weightages are assigned for attributes. Attribute with operation name are given more weightage, attributes with quality factors are given second highest weightage and others are assigned minimum weightage.

Generally, prediction of high quality services purely based on the contents of web services may not provide accurate results. This ranking cannot be taken for suggesting

Table 1. Sample dataset of WS details

WSDL/Features	P1	P2	P3	P4	P5	P6	P7	P8
W1	4	3	1	1	1	1	1	1
W2	4	3	1	1	1	1	0	0
W3	4	3	1	1	1	1	0	1
W4	4	3	1	1	1	1	1	1
W5	4	3	1	1	1	1	0	0
W6	4	3	1	1	1	0	0	1
W7	4	3	1	1	1	0	0	1
W8	4	3	1	1	1	0	0	1
W9	4	3	1	1	1	1	1	1
W10	4	3	1	1	1	1	0	1

services since the values are not precise. The dataset of WS details are thus improved by using SVD dimensionality reduction process. Table 2 shows the results of the SVD process which represents the most precise values for different attributes.

Table 2. SVD Result

WSDL/Features	P1	P2	P3	P4	P5	P6	P7	P8
W1	4.16	1.18	−0.73	4.06	−0.45	0.6	0.02	0.36
W2	3.91	0.95	−0.58	3.32	−0.25	0.53	−0.13	0.28
W3	4.19	1.18	−0.46	3.31	−0.22	0.13	−0.21	0.91
W4	4.32	1.05	−0.46	3.45	−0.01	0.31	−0.1	1.21
W5	2.87	0.87	−0.42	2.94	−0.21	0.05	0.11	0.42
W6	4.6	1.24	−0.61	3.79	−0.4	0.19	0.18	0.6
W7	4.23	1.37	−0.3	3.47	−0.26	-0.34	−0.11	0.39
W8	4.52	1.23	−0.68	4.06	−0.45	0.41	0	0.72
W9	4.27	1.21	−0.82	3.92	−0.45	0.29	0.28	0.64
W10	2.86	0.76	−0.21	2.34	−0.23	0.3	−0.32	0.88

For collation, output is tested against two linkages. At first, single linkage is applied to the former data which resulted in the clustering of WS based on the minimum distance measure. The clustered order of web services is used for ranking of similar web services. Figure 2 depicts the dendrogram obtained after the single linkage clustering. Web services are added at every phase to similar clusters based on their minimum distance values which is shown in the y axis of the dendrogram. From the output, web service w1 is ranked first as it has the minimum distance and w8 is ranked second. All others are likewise ranked in ascending order of distance.

Figure 3 shows the clustered ordering of web services with ward linkage. Ward linkage merges the web services at every phase into similar clusters based on the optimum degrees of the objective function. When mapping two results with feature based ranking, single linkage dendrogram is closer to the importance of WS description.

All these results are ranked separately for suitable interpretation. The first ranking is based on content based information. The web service with maximum value is ranked first. All other WS are rank ordered in a relative fashion.

The ranking purely on the content details does not provide quality services precisely, which endorses the choice of methods like SVD, PCA etc. Nevertheless, these values can very well serve as approximate predictions. The graph is constructed with three rankings (i) based on content information (ii) SVD with single linkage (iii) SVD with ward linkage.

Figure 4 shows the representation of those results. The ranking based on SVD Single Linkage result is considered for recommendation. It is realised that SVD-SL ranking is closer to the real world predictions done with travel domain data set, as the error rate of prediction using RMSE is only 0.23, which is considered as the minimum error. Knowing the accuracy of values, services are suggested from the former SVD-SL output.

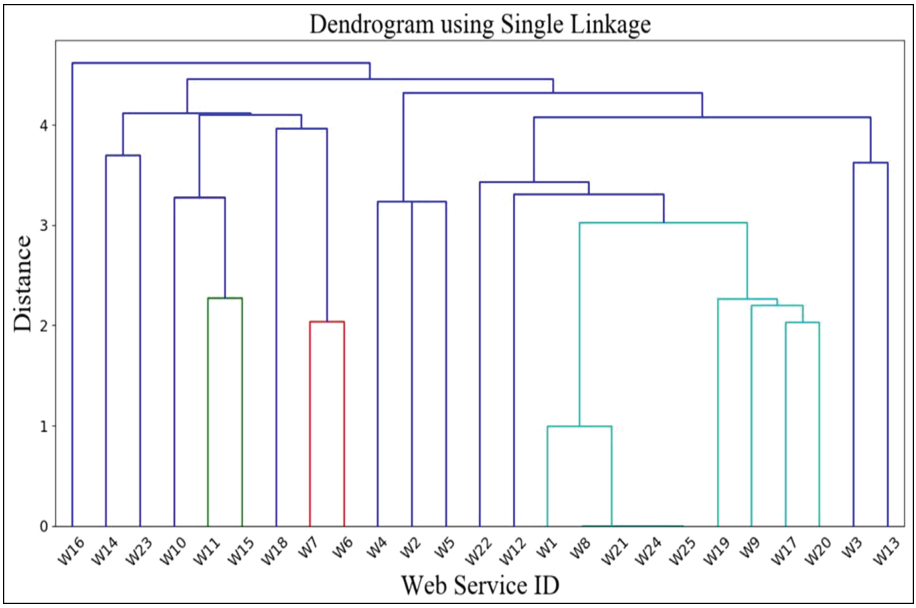


Fig. 2. SVD with Single linkage

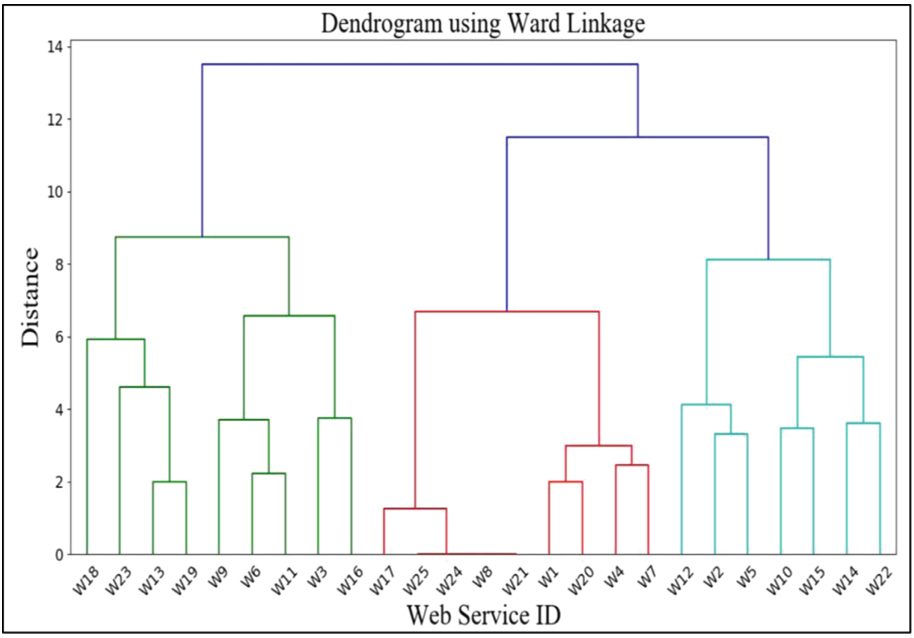


Fig. 3. SVD with Ward linkage

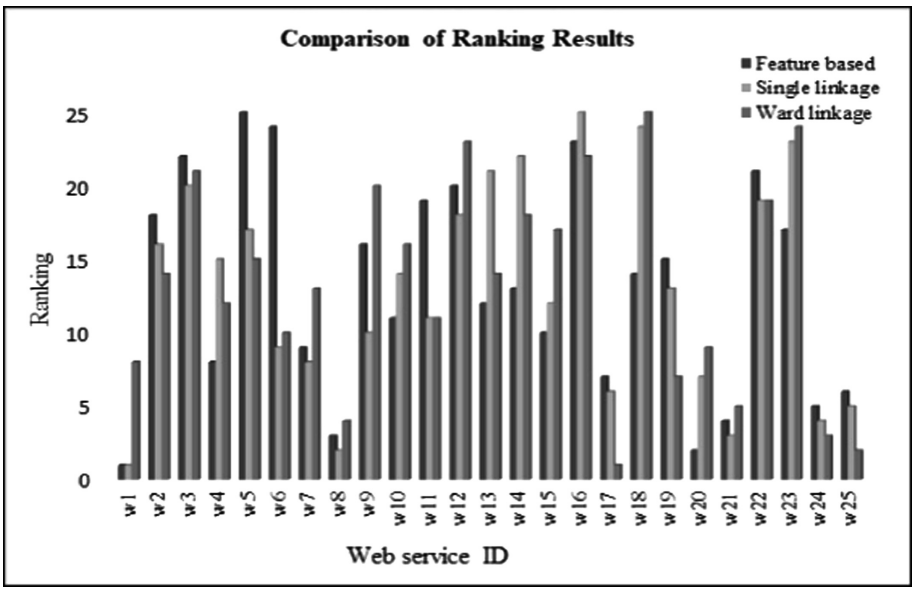


Fig. 4. Graph with comparison of results

Figure 5 shows the final ranked result obtained from SVD-SL algorithm and this could be used for recommending quality services. Each positioning of WS can be clearly observed from the above outcome.

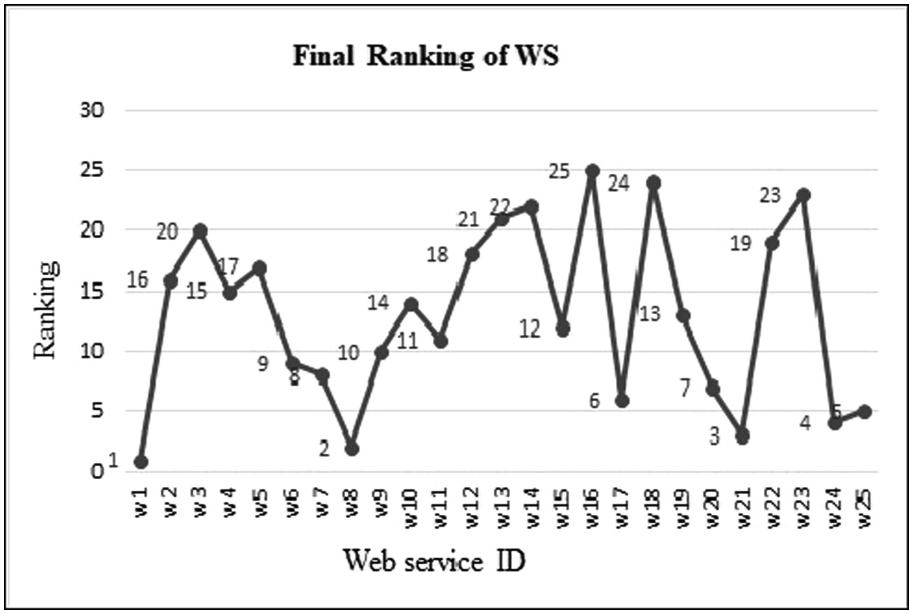


Fig. 5. Ranking order of services based on SVD-SL.

5 Conclusion

All the existing service-oriented web solutions urgently request the development of new procedures for the discovery of context-sensitive web services, with high caliber in standard by way of their functional properties. The essential thought in our approach is to develop a framework which predicts the arrangement of WS that fulfils the client-preferred options, using a feature reduction and similarity approach. Subtle elements of the WS from the WSDL are utilized for the formation of accurate data values. These include functional as well as the non-functional characteristics of webservices provided by WS providers. SVD efficiently calculates the most accurate values for each WS with its features. Then a similarity approach is cranked through the output results in the order of ranking WS. This method have producedthe most accurate selection when it is compared with the similarity measure without feature reduction. The system is validated for its efficiency with the RMSE methods and is proved to be closer with the real world observations.

In future, the service selection process can be enhanced by using trust-based selection process or deep learning methodologies for access to the best services instead of the clustering methodologies.

References

1. Yao, L., Sheng, Q.Z., Ngu, A.H., Yu, J., Segev, A.: Unified collaborative and content based webservice recommendation. IEEE Trans. Serv. Comput. **8**, 453–466 (2014)
2. Yu, L., Gao, M., Xiao, X., Li, X., Xiong, Q.: Important user group based web service recommendation. In: 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), pp. 413–418 (2017)
3. Subbulakshmi, S., Ramar, K., Renjitha, R., Sreedevi, T.U.: Implementation of adaptive framework and WS ontology for improving QoS in recommendation of WS. Intelligent Systems Technologies and Applications 2016. AISC, vol. 530, pp. 383–396. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-47952-1_30
4. Sarwar, B., Karypis, G., Konstan, J., Riedl, J.: Item-based collaborative filtering recommendation algorithms. In: 10th International Conference on World Wide Web, pp. 285–295 (2001)
5. Vadivelou, G.: Collaborative filtering based web service recommender system using users. Satisfaction on QoS attributes. In: International Conference, Inventive Computation Technologies (ICICT), vol. 3, pp. 1–5 (2016)
6. Linden, G., Smith, B., York, J.: Amazon.com Recommendations: item-to-item collaborative filtering. IEEE Internet Comput. **7**, 76–80 (2003)
7. Gupta, J., Gadge, J.: Performance analysis of recommendation system based on collaborative filtering and demorgraphics . In: International Conference on Communication, Information & Computing Technology (ICCICT), pp. 1–6 (2015)
8. Gudla, S.K., Bose, J., Gajam, V., Srinivasa, S.: Relevancy ranking of user recommendations of services based on browsing patterns. In: 16th IEEE International Conference on Machine Learning and Applications (ICMLA), pp. 765–768 (2017)
9. Zheng, Z., Ma, H., Lyu, M.R., King, I.: QoS-aware web service recommendation by collaborative filtering. IEEE Trans. Serv. Comput. **4**, 140–152 (2010)

10. Devika, P., Jisha, R.C., Sajeev, G.P.: A novel approach for book recommendation systems In: 2016 IEEE International Conference on Computational Intelligence and Computing Research, (ICCIC), pp. 1–6 (2016)
11. Arunachalam, N., Amuthan, A., Sharmilla, M., Ushanandhini, K.: Survey on web service recommendation based on user history. In: International Conference on Computation of Power, Energy, Information and Communication (ICCPEIC), pp. 305–309 (2017)
12. Phillips, R.D., Watson, L.T., Wynne, R.H., Blinn, C.E.: Feature reduction using a singular value decomposition for the iterative guided spectral class rejection hybrid classifier. ISPRS J. Photogrammetry Remote Sens. **64**, 107–116 (2009)
13. Reshma, R., Sowmya, V., Soman, K.P.: Effect of Legendre-Fencheldenoising and SVD-based dimensionality reduction algorithm on hyperspectral image classification. Neural Comput. Appl. **29**, 301–310 (2018)
14. Hendrix, W., Palsetia, D., Patwary, M.M.A., Agrawal, A., Liao, W.K., Choudhary, A.: A scalable algorithm for single-linkage hierarchical clustering on distributed-memory architectures. In: IEEE Symposium on Large-Scale Data Analysis and Visualization (LDAV), pp. 7–13 (2013)
15. Sabarish, B.A., Karthi, R., Gireeshkumar, T.: Clustering of trajectory data using hierarchical approaches. In: Computational Vision and Bio Inspired Computing, pp. 215–226 (2018)

HSV Based Histogram Thresholding Technique for MRI Brain Tissue Segmentation

T. Priya and P. Kalavathi(✉)

Department of Computer Science and Applications, The Gandhigram Rural Institute - (Deemed to be University), Gandhigram, Tamil Nadu, India
suga.priya04@gmail.com, pkalavathi.gri@gmail.com

Abstract. Background: To bring it as a human interactive perceive color process, an automatic color model based segmentation of White Matter (WM), Gray Matter (GM), and Cerebrospinal Fluid (CSF) in Magnetic Resonance Brain images is proposed in this paper.

Methods: Preprocessing process is done for the MRI brain images using wavelet based bivariate shrinkage method and Contour based Brain Segmentation method (CBSM). Then segmentation of brain tissues using Hue Saturation Value (HSV) color model Based Histogram Thresholding Technique (HSVBHTT) was applied. Normal and Alzheimer's disease (AD) brain images obtained from Internet Brain Segmentation Repository (IBSR) and Minimal Interval Resonance Imaging in Alzheimer's Disease (MIRIAD) datasets.

Results and Conclusions: The results of proposed method was analyzed with similarity measures and quantitative measures like Jaccard (J), Dice (D), Sensitivity (S) and Specificity (SP) and compared with the manual segmented images which produced better results on segmenting WM, GM and CSF compared to other existing methods.

Keywords: Alzheimer's Disease · Brain tissue segmentation
Histogram Thresholding · HSV color model

1 Introduction

Magnetic Resonance imaging is a standout among the most generally used imaging method to envision the interior and exterior part of the brain. It clearly demonstrates the functions of brain tissue so as to help the doctors to diagnose different types of diseases like Alzheimer's disease, Multiple Sclerosis, Parkinson's diseases and Brain tumor. It supports treatment planning and also for enumeration of tissue volume estimation.

Brain tissue consists of three major classes such as Gray Matter (GM), White Matter (WM) and Cerebrospinal Fluid (CSF). GM is made up of interconnecting unmyelinated neurons which are the areas of regions for nerve connections and processing. Whereas, WM are built with myelinated neurons and it is the information pathway of the brain which sends the signal faster between the distant parts of the brain and the body than unmyelinated neurons. Brain and the spinal cord are encompassed by clear fluid called CSF. It gives many imperative capacities to the central nervous system. Segmentation of brain tissue is an important part in medical imaging field.

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 322–333, 2019.
https://doi.org/10.1007/978-981-13-5758-9_27

It also helps in earlier diagnosis of diseases. In order to classify the brain tissues quantitatively, some preprocessing process need to be done to remove unwanted noise, skull, eye, skin part, marrow from the brain images. In recent years, there are many brain image segmentation approaches were developed [1] in the form of cluster based methods [2, 3], region growing methods [4–6], threshold based methods [7–9], edge based methods [10], hybrid methods [11], classification based methods [12] and color model based methods [13, 14].

Segmentation of brain tissue became a challenging task, now a days, many methodologies and algorithms for brain tissue segmentation has been developed by researchers. Lalaoui et al. [15] have proposed a method which reduces number of iterations and it converge a center quickly with less time. This method was implemented on both real and synthetic images and found to produce better results. Moreover, it is robust to initialization process which allows for fully automated applications. Qin et al. [16] proposed a method with combination of Gaussian Mixture Model and K-Means (KM) to segments brain tissue part, it works well on the T1 weighted images and there will be some difficulties in T2 weighted images. A method in [17] uses level set with sharp peaks for three patch segmentation and it is robustness to noise and gives related accurate result compared to groundtruth.

A new method called Adaptive Fuzzy K-Means in [18] was developed with a combination of Modified KM and Fuzzy C-Means (FCM) applied to MRI brain images and has produced sharper and clearer segmentation results compared to other methods. To avoid time complexity of clustering process, an automatic Kernelized FCM with weighted bias field is used in [19]. In our proposed method, HSV color model with histogram based thresholding technique is used; because this model is sensitive to human perceive color and also it is suitable for manipulation of interactive color images. A method by [20] imported HSI color space based FCM clustering techniques which states that HSI is best suited for detection of simple images with uniform background and the improved FCM gives better segmentation result compared to other techniques. In order to segment normal and pathological tissues in MRI brain images, a method in [21] with KM clustering in HSV transform is used. It gives better segmentation result and also it extracts the correct position of brain tumor. Different color model are used for segmentation process, a model YC_bC_r with local thresholding method [22] is applied to various color images and it produced good results.

Bora et al. [23] have compared the performance of L*A*B and HSV color space, HSV color space will be suitable for segmentation of noisy images. Duan et al., proposed a method in [24] based on HSI color space and region growing techniques for segmenting color white blood cells. Sharma et al., introduced a method [25] called adaptive thresholding to detect color and segmentation of color images and this method works robust on a variety of image samples. A new method in [26] contains two phases in which statistical features are applied to color images to produce feature vectors, and again each of these feature vectors are processed by FCM algorithm. Experimental results show that Modified FCM works more effective in finding homogenous regions and the research shows that color feature hue is more adequate than RGB color features. Wavelet based Image Fusion with Clustering Techniques (WBIFCT) in [27] gives better segmentation results of brain tissues compared to other

techniques. Next sections of the paper are systematized as follows: Detailed explanation of methodology, in Sect. 2. Section 3 signifies exploratory results and discussion and conclusions in Sect. 4.

2 Methodology

Thresholding is one of the efficient methods for automatic segmentation and classification of images based on gray scale values. Furthermore, HSV color model is used, which realizes human visual color and also it will be best suitable for segmentation process. In our proposed method, we combined both the thresholding and HSV color model to yield proper result. Our proposed method has two processes, first process consists of denoising and skull removing from the brain images and the second part is HSV based Histogram Threshold Technique (HSVBHTT). The process of proposed method is shown in following Fig. 1. It will be discussed in detail in the following two process stage.

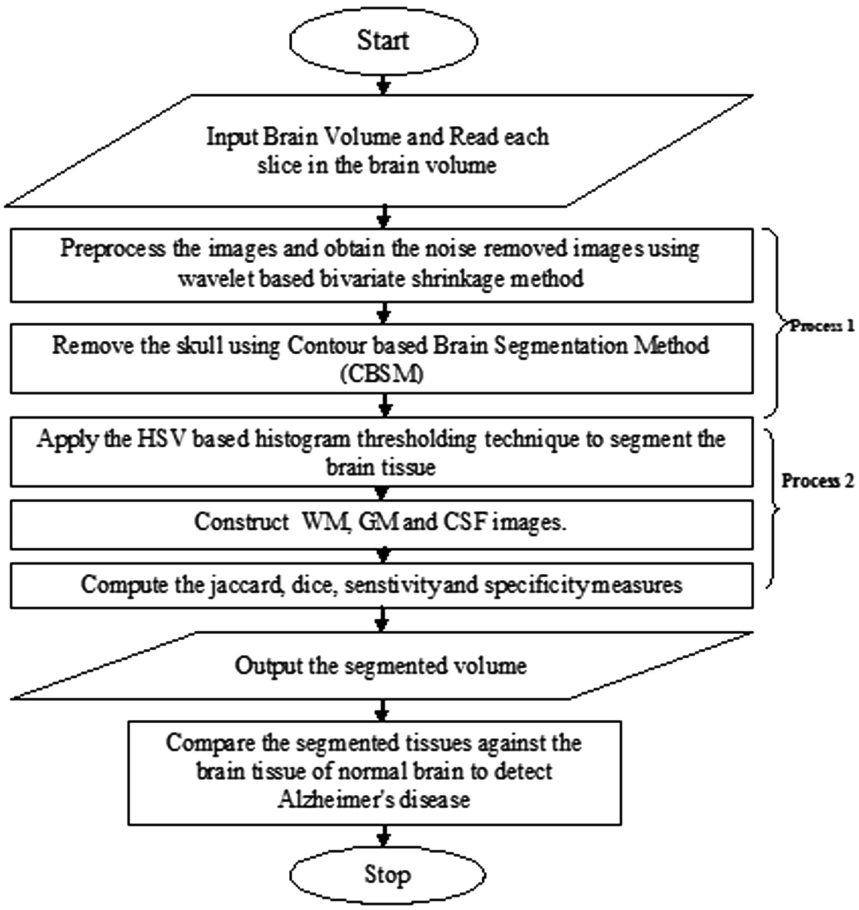


Fig. 1. Flowchart of our proposed method

2.1 Process 1: Eliminating Noise and Separating Brain Portion

In order to get outstanding result of segmentation process, we need to do some pre-processing process such as low frequency noise, and human head part such as skull, eye, marrow and skin part are removed from the brain images using wavelet based bivariate shrinkage method [28] and Contour Based Brain Segmentation Method [29].

2.2 Process 2: HSV Based Histogram Thresholding Technique

To convert gray scale image into color image, a colormap is used. Colormaps can be of any length, but must be three columns wide. In that we have applied HSV color space [30, 31] to the original gray scale images. In HSV, H denotes hue, S denotes saturation and V denotes the value of the image.

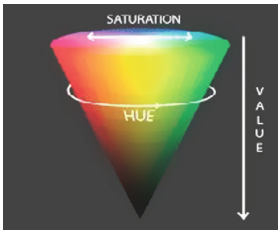


Fig. 2. HSV color model

In the Fig. 2, it shows the single hex-cone model of HSV color model. HSV color wheel is utilized to choose the required shading. In HSV color cone, a circle denotes hue, and triangle for saturation and value. Horizontal axis denotes saturation and vertical axis is the value. If we need to put different color for an image, first of all, we want to select a color from a hue. For saturation and value, we can select the horizontal and vertical angle from the triangle. Hue represents the color and it ranges from 0 to 360°, saturation is the range of gray in color space and it is from 0 to 100%, sometimes it can be calculated between 0 and 1, where the value 0 denotes gray color and 1 denotes the primary color. Value is the brightness of the color space which lies between 0 to 100%. If the value is 0 means it is totally black and if the increasing value indicate the increase in the brightness of the color. In our method, we converted the skull stripped gray image into HSV color image and then we took the circular histogram of Hue component and histogram for the saturation and value of a color image as shown in following Fig. 3. Based on this circular histogram statistics ranging from 0° to 360°, we identified six threshold values as shown in Fig. 3(a).

In the Fig. 3, it shows the histogram of Hue component of HSV image and from that we find the min and max value of Hue channels of an image (H1min, H1max, H2min, H2max, H3min, H3max) and the minimum (Smin, Vmin) and maximum (Smax, Vmax) values for other two channels such as saturation and value were identified from their respective histogram to calculate threshold values for the mask image. The calculated threshold values for the each of the channels for the HSV color model is given below:

$$H = I_{c1}(H1min, H1max, H2min, H2max, H3min, H3max) \quad (1)$$

$$S = I_{c2}(Smin, Smax) \quad (2)$$

$$V = I_{c3}(Vmin, Vmax) \quad (3)$$

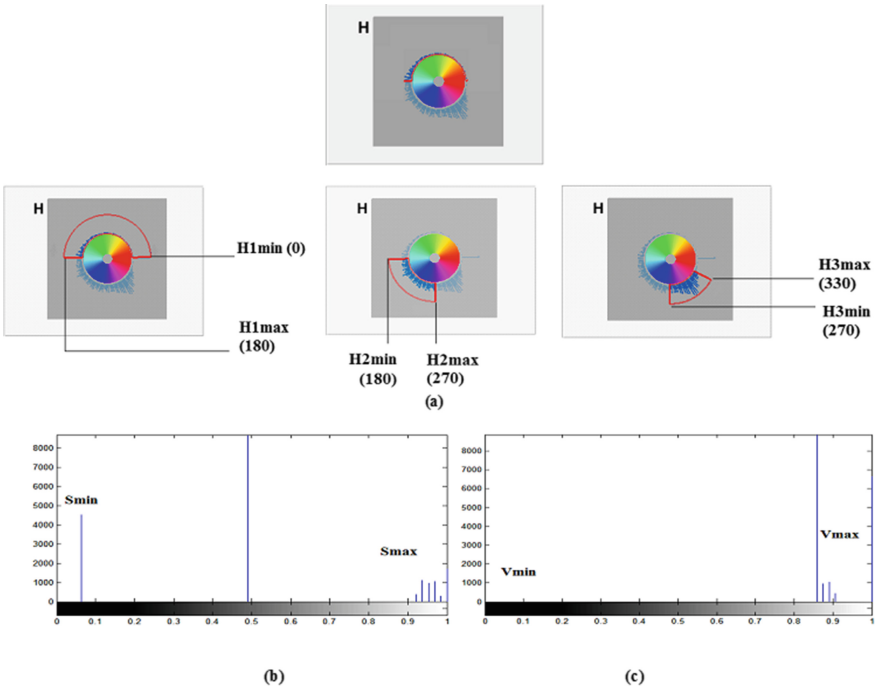


Fig. 3. Histogram of (a) Hue; (b) Saturation; (c) Value

where I_{c1}, I_{c2}, and I_{c3} represents the each color channel of the HSV image, min and max represents the minimum and maximum value of each color channel and we consider these minimum and maximum value as threshold value and apply to the each color channel of color image in order to create mask and this can be calculated by the following equation as.

$$\begin{aligned} \text{Mask} = {} & (I_{c1}(:,:,1) > = \text{H1min}) \,\&\, (I_{c1}(:,:,1) < = \text{H1max}) \,\&\, \ldots \\ & (I_{c2}(:,:,2) > = \text{Smin}) \,\&\, ((I_{c2}(:,:,2) < = \text{Smax}) \,\&\, \ldots \\ & (I_{c3}(:,:,3) > = \text{Vmin}) \,\&\, ((I_{c3}(:,:,3) < = \text{Vmax}) \,\&\, \ldots \end{aligned} \tag{4}$$

It is observed from the above equation, that the CSF tissues of the brain image are present from 0° to 180° as H1min and H1max in the circular histogram of the hue component. Likewise, from 180° to 270° denote GM as H2 min and H2max and WM appears from the degree 270° to 330° as H3min and H3max. From that, we computed a color mask image for segmenting the CSF, GM and WM by overlaying this mask in the original skull stripped image.

3 Results and Discussion

Our proposed technique was executed with two brain datasets, first one is obtained from IBSR [32] which contains twenty volumes of T1 weighted MR brain images in that each volume consists of 2D sequential coronal slices ranges from 60 to 65, thickness of 3.1 mm with dimensions of 256 × 256 pixels of young-middle aged normal individuals in the Center for Morphometric Analysis (CMA) at the Massachusetts General Hospital. Second dataset which contains series of volumetric T1 MRI scans of 46 mild - moderate Alzheimer's subjects and 23 normal controls of different aging group for both male and female is taken from Minimum Interval Resonance Imaging In Alzheimer's Disease (MIRIAD) [33] to segment the brain tissue such as WM, GM and CSF and it is compared with manual segmented images. Manual segmented images [34] are the images which are justified by two principle observers in which observers will manually segmented the brain regions by applying suitable threshold value of WM, GM and CSF and the binary result of this of tissue type was altered by getting relevant data from other scan and it was reviewed and analyzed by other observers to get the gold segmented images. Using these manual segmented and proposed segmented images, performance of proposed strategies was validated by two measures. First one, is the similarity measures like Jaccard (J) and Dice (D) and compared with existing techniques. These evaluation techniques will take two images of gold standard image and proposed image as input and it will produce the value between 0 and 1.

The Jaccard (J) [35] value is calculated by:

$$J(s_1, s_2) = |s_1 \cap s_2| / |s_1 \cup s_2| \tag{5}$$

The Dice (D) [35] value is measured by

$$D(s_1, s_2) = 2|s_1 \cap s_2| / |s_1| + |s_2| \tag{6}$$

where, S_1 represents the aggregate pixels of the image obtained by the proposed segmentation method and S_2 represents the aggregate pixels in the image obtained from ground truth image. Second one is quantitative measures such as sensitivity and specificity. Sensitivity and specificity are mathematical proportions of the execution of a binary classification test, likewise it also called in statistics as classification function.

Sensitivity (S) [36] is the proportion of positive extents that are effectively identified as

$$S = TP/(TP + FP) \tag{7}$$

where TP denotes True Positive and FN denotes False Negative respectively.

Specificity (SP) [36] is the proportion of negative extents that are effectively identified as

$$SP = TN/(TN + FP) \tag{8}$$

where TN represents True Negative and FP represents False Positive respectively.

The images in the dataset are compared with existing methods such as Adaptive Maximum A Posteriori Probability (AMAP), Biased MAP (BMAP), Maximum A Posteriori Probability (MAP), Maximum Likehood (ML), Tree Structure K-Means (TK-Means). The segmentation result of our proposed method for a sample slice is given in the Fig. 4.

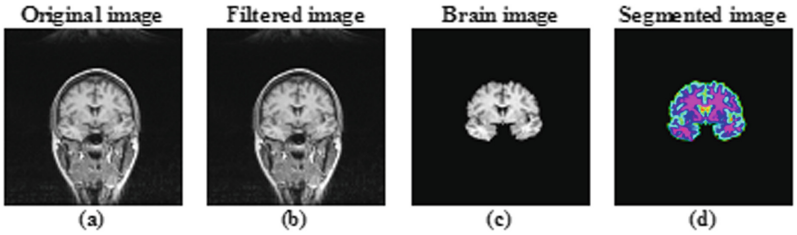


Fig. 4. Brain tissue segmentation of proposed method (a) Original image (b) Filtered image by wavelet based bivariate shrinkage method (c) Skull stripped image by CBSM (d) Segmented result of proposed method (pink denotes WM, blue denotes GM and cyan denotes CSF). (Color figure online)

In Fig. 4(a) denotes original brain image, Fig. 4(b) and (c) represents the filtered image by wavelet based bivariate shrinkage method and skull removed images by CBSM. Figure 4(d) denotes the segmented result of proposed method (here pink denotes WM, blue denotes GM and cyan denotes the CSF respectively). The computed similarity measures and Quantitative measures for twenty volumes of T1-weighted images obtained by our proposed method are given in Table 1.

Alzheimer's disease (AD) [37] is a neuro degenerative disease which was initially reported by Dr. Alois Alzheimer and it is in two forms such as sporadic and familial. Sporadic AD is the most common form of AD which mostly affect the persons at age of 65 and Familial AD is developed by the person at the age of 40's or 50's due to mutated genes. AD cause due to the stacked deposits of cells in the outer layer of brain called plaques and inner surface called tangles which influenced the signal passed between brain cells and transportation of food and energy around the brain cells. Due to destruction of brain cells, there will be shrinkage in the brain tissue and it destroys the short term memory and advanced growth of these disease leads to long term memory loss. To analysis the process of normal and affected regions of brain tissue, we implement our method on brain images of both normal and AD images which are obtained from MIRIAD database and segmented into WM, GM and CSF and the result obtained from this method is shown in Figs. 5 and 6 respectively.

Table 1. Computed Jaccard (J), Dice (D), Senstivity (S) and Specificity (SP) for twenty volumes of T1 weighted MRI Brain images.

BV	Jaccard			Dice			Sensitivity			Specificity		
	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF	WM	GM	CSF
1	0.654	0.587	0.154	0.769	0.683	0.162	0.980	0.975	0.997	0.998	0.998	0.997
2	0.536	0.468	0.139	0.651	0.600	0.141	0.985	0.974	0.995	0.998	0.993	0.995
3	0.552	0.487	0.115	0.668	0.619	0.118	0.983	0.974	0.997	0.997	0.995	0.997
4	0.571	0.474	0.067	0.701	0.637	0.078	0.964	0.971	0.998	0.988	0.996	0.998
5	0.588	0.486	0.048	0.716	0.642	0.056	0.959	0.970	0.998	0.987	0.997	0.998
6	0.718	0.594	0.067	0.818	0.640	0.079	0.972	0.974	0.996	0.996	0.999	0.996
7	0.707	0.544	0.127	0.810	0.647	0.132	0.971	0.974	0.996	0.995	0.998	0.996
8	0.685	0.573	0.033	0.798	0.676	0.065	0.953	0.950	0.997	0.993	0.998	0.997
9	0.657	0.545	0.064	0.754	0.688	0.086	0.964	0.953	0.997	0.997	0.997	0.997
10	0.749	0.687	0.095	0.841	0.733	0.101	0.959	0.951	0.998	0.997	0.998	0.998
11	0.408	0.449	0.083	0.533	0.577	0.098	0.986	0.967	0.995	0.998	0.991	0.995
12	0.538	0.566	0.150	0.650	0.667	0.166	0.982	0.972	0.998	0.997	0.996	0.998
13	0.636	0.589	0.111	0.745	0.641	0.113	0.976	0.975	0.998	0.996	0.998	0.998
14	0.708	0.690	0.159	0.799	0.706	0.170	0.969	0.958	0.997	0.997	0.997	0.997
15	0.715	0.594	0.127	0.716	0.681	0.157	0.964	0.962	0.999	0.994	0.999	0.999
16	0.642	0.579	0.109	0.763	0.689	0.114	0.969	0.959	0.997	0.998	0.996	0.998
17	0.731	0.595	0.127	0.826	0.677	0.134	0.967	0.969	0.998	0.995	0.998	0.999
18	0.751	0.565	0.111	0.842	0.647	0.118	0.966	0.971	0.997	0.995	0.999	0.997
19	0.714	0.628	0.095	0.816	0.695	0.098	0.970	0.961	0.996	0.998	0.998	0.997
20	0.711	0.608	0.079	0810	0.656	0.087	0.966	0.967	0.996	0.996	0.998	0.997
M	**0.649**	**0.565**	**0.103**	**0.751**	**0.660**	**0.114**	**0.970**	**0.966**	**0.997**	**0.996**	**0.997**	**0.998**

*BV-Brain Volume, M-Mean

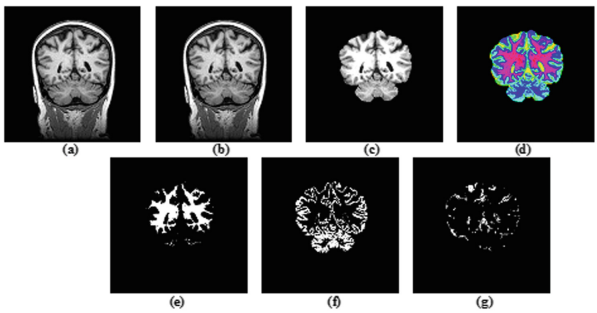


Fig. 5. Brain tissue segmentation of normal image (a) Original normal image (b) Filtered image (c) Skull removed image (d) Mask image: (e)–(g) WM, GM and CSF of normal image.

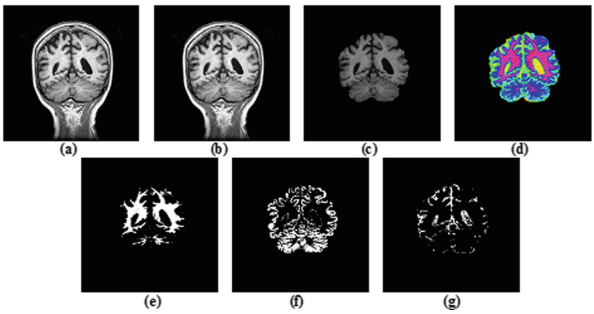


Fig. 6. Brain tissue segmentation of AD affected image (a) Original AD image (b) Filtered image (c) Skull removed image (d) Mask image: (e)–(g) WM, GM and CSF of AD image.

The Figs. (5) and (6) shows the segmented result of WM, GM and CSF of normal and AD brain disease images by our proposed method and it can be observed that there is some shrinkage in the white matter between normal and AD brain images. The computed Jaccard and Dice value are shown in the Table 2.

Table 2. Computed Jaccard (J) and Dice (D) value for normal and Alzheimer disease

Images	Jaccard			Dice			Clinical
	WM	GM	CSF	WM	GM	CSF	
Image 1	0.6762	0.5670	0.0557	0.7467	0.6862	0.1055	AD afflicted
Image 2	0.6558	0.5986	0.0906	0.7248	0.6314	0.1661	AD afflicted
Image 3	0.5760	0.4908	0.0994	0.6465	0.5205	0.1809	AD afflicted
Image 4	0.5977	0.4507	0.1283	0.6690	0.5009	0.2174	AD afflicted
Image 5	0.7477	0.6390	0.0980	0.8185	0.7858	0.1785	AD afflicted
Image 6	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	Normal

From the Table 2, it can be seen that, Jaccard and Dice value of first five images are not having the same value as 1, since it is affected by Alzheimer's disease and the value for remaining one image is 1 so it is a normal image (i.e, the segmented brain tissue of both normal and the test image are same).

The comparative analysis of similarity measures of our proposed method and other techniques like AMAP, BMAP, MAP, ML, TKmean are given in the following graph Fig. 7.

From the graph, it can be seen that our proposed method yields better result compared to other segmentation techniques.

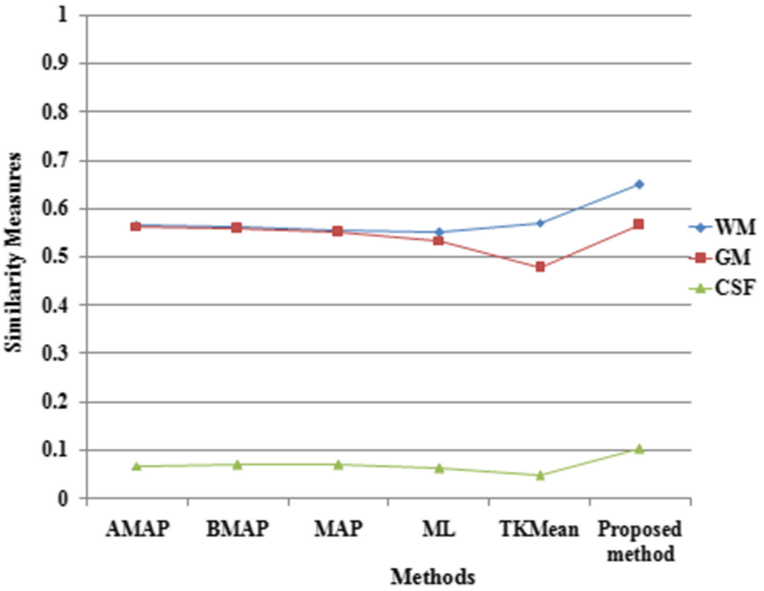


Fig. 7. Comparative analysis of proposed method and other segmentation techniques like AMAP, BMAP, MAP, ML, TKmean.

4 Conclusions

In this paper, to segment the brain tissues WM, GM and CSF automatically a HSV Based Histogram Thresholding (HSVBHTT) Technique is proposed and it is systematically applied on T1 weighted MRI brain images. Comparative analysis of our proposed method and existing methods such as AMAP, BMAP, MAP, ML, TKmean was carried out by using Jaccard (J), Dice (D), sensitivity (S) and Specificity (SP) and gives précised results. For identification of Alzheimer's disease, this method was employed on Alzheimer's disease brain images obtained from MIRIAD database and found acceptable results. In future, this method may be experienced with different color models to get more accurate result.

Acknowledgement. This work was supported by Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India.

References

1. Kalavathi, P., Surya Prasath, V.B.: Methods on skull stripping of MRI head scans images - a review. J. Digit. Imaging **29**, 365–379 (2016)
2. Kalavathi, P., Priya, T.: Performance of clustering techniques on segmentation of brain tissues in MRI human head scans. In: Proceedings of National Conference on New Horizons in Computational Intelligence and Information Systems, pp. 164–170. Excel India Publications, New Delhi (2015)

3. Kalavathi, P., Priya, T.: MRI brain tissue segmentation using AKM and FFCM clustering techniques. In: Proceedings of National Conference on Recent Advances in Computer Science and Application, pp. 113–118. Bonfring Publications, India (2015)
4. Somasundaram, K., Kalavathi, P.: Brain segmentation in magnetic resonance human head scans using multi-seeded region growing. Imaging Sci. J. **62**(5), 273–284 (2014)
5. Somasundaram, K., Kalavathi, P.: A novel skull stripping technique for T1-weighted MRI human head scans, pp. 1–8. ACM Digital Library (2012)
6. Somasundaram, K., Kalavathi, P.: Skull stripping of MRI head scans based on chan-vese active contour model. Int. J. Knowl. Manag. E-Learn. **3**(1), 7–14 (2011)
7. Kalavathi, P.: Brain tissue segmentation in MR brain images using Otsu's multiple thresholding technique, pp. 638–642. IEEE Xplore Digital Library (2013)
8. Otsu, N.: A threshold selection method from gray-level histogram. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
9. Kalavathi, P., Priya, T.: Brain extraction from MRI human head scans using outlier detection based morphological operation. Int. J. Comput. Sci. Eng. **6**(4), 266–273 (2018)
10. Aslam, A., Khan, E., Beg, M.M.S.: Improved edge detection algorithm for brain tumor segmentation. In: Second International Symposium on Computer Vision and the Internet (VisionNet 2015), vol. 58, pp. 430–437. Elsevier (2015)
11. Somasundaram, K., Kalavathi, P.: A hybrid method for automatic skull stripping of magnetic resonance images (MRI) of human head scans, pp. 1–5. IEEE Xplore Digital Library (2010)
12. Renjith, A., Manjula, P., Mohan Kumar, P.: Brain tumor classification and abnormality detection using neuro-fuzzy technique and Otsu thresholding. J. Med. Eng. Technol. **39**(8), 498–507 (2015)
13. Attique, M., et al.: Colorization and automated segmentation of human T2 MR brain images for characterization of soft tissue. PLOS ONE **7**(3), 1–13 (2012)
14. Zhong, S.D., Wei, Y.K., Xie, Z.G.: Method of automatic tongue area extraction in tooth-marked tongue images. Comput. Technol. Dev. **19**(1), 245–247 (2009)
15. Lalaoui, L., Mohamadi, T., Djaalab, A.: New method for image segmentation. Soc. Behav. Sci. **195**, 1971–1980 (2015). World Conference on Technology, Innovation and Entrepreneurship
16. Qin, Z., Wang, F., Xiao, Z., Lan, T., Ding, Y.: Brain tissue segmentation with the GKA method in MRI. In: IEEE International Conference on Signal and Image Processing, pp. 273–276 (2016)
17. Roy, S., Bandyopadhyay, S.K.: A new method of brain tissues segmentation from MRI with accuracy estimation. In: International Conference on Computational Modeling and Security, vol. 85, pp. 362–369. Elseveir (2016)
18. Sulaiman, S.N., Non, N.A., Isa, I.S., Hamzah, N.: Segmentation of brain MRI image based on clustering algorithm. In: Energy, Environment, Biology and Biomedicine, pp. 54–59 (2014)
19. Kannan, S.R., Sathya, A., Ramathilagam, S., Devi, R.: Novel segmentation algorithm in segmenting medical images. J. Syst. Softw. **83**, 2487–2495 (2010)
20. Ganesan, P., Chakravarty, P., Verma, S.: Segmentation of natural color images in HSI color space based on FCM clustering. Int. J. Adv. Res. Comput. Eng. Technol. **3**(3), 618–622 (2014)
21. Verma, R., Singh Rathore, S., Verma, A.: MRI segmentation using K-Means clustering in HSV transform. Int. J. Adv. Res. Comput. Eng. Technol. **4**(10), 3925–3929 (2015)
22. Mandal, A.K., Baruah, D.K.: Image segmentation using local thresholding and Ycbcr color space. Int. J. Eng. Res. Appl. **3**(6), 511–514 (2013)

23. Bora, D.J., Gupta, A.K., Khan, F.A.: Comparing the performance of L*A*B and HSV color spaces with respect to color image segmentation. Int. J. Emerg. Technol. Adv. Eng. **5**(2), 193–203 (2015)
24. Duan, J., Yu, L.: A WBC segmentation method based on HSI color space. In: Fourth IEEE International Conference on Broadband Network and Multimedia Technology, pp. 629–632 (2011)
25. Sharma, P., Abrol, P.: Color based image segmentation using adaptive thresholding. Int. J. Sci. Tech. Adv. **2**(3), 151–156 (2016)
26. Harrabi, R., Ben Braiek, E.: Color image segmentation based on a modified fuzzy C means technique and statistical features. Int. J. Comput. Eng. Res. **2**(1), 120–135 (2012)
27. Kalavathi, P., Priya, T.: Segmentation of brain tissue in MR brain image using wavelet based image fusion with clustering technique. In: Proceedings of National Conference on Computational Methods, Communication Techniques and Informatics, pp. 28–33 (2017)
28. Kalavathi, P., Priya, T.: Noise removal in MR brain images using 2D wavelet based bivariate shrinkage method. Glob. J. Pure Appl. Math. **13**(5), 77–86 (2017)
29. Somasundaram, K., Kalavathi, P.: Contour-based brain segmentation method for magnetic resonance imaging human head scans. J. Comput. Assist. Tomogr. **37**(3), 353–368 (2013)
30. Maiti, I., Chakraborty, M.: A new method for brain tumor segmentation based on watershed and edge detection algorithms in HSV color model. In: Proceedings of National Conference on Computing and Communication Systems (2012)
31. HSV Color Model. http://www.tech-faq.com/hsv.html
32. IBSR data set. http://www.cma.mgh.harvard.edu/ibsr/index.html
33. MIRIAD dataset. http://miriad.drc.ion.ucl.ac.uk/atrophychallenge
34. Anbeek, P., Vincken, K.L., Groenendaal, F., Koeman, A., van Osch, M.J.P., Grond, J.V.D.: Probabilistic brain tissue segmentation in neonatal magnetic resonance imaging. Pediatr. Res. **63**(2), 158–163 (2008)
35. Similarity Measures. http://sve.bmap.ucla.edu/instructions/m
36. Shantha kumar, P., Ganesh kumar, P.: Performance analysis of brain tumor diagnosis based on soft computing technique. Am. J. Appl. Sci. **11**(2), 329–336 (2014)
37. Chui, H.C., Gomez, L.R.: Clinical and imaging features of mixed Alzheimer and vascular pathologies. Alzheimer's Res. Ther. **7**(1), 21 (2015)

Significance of Epoch Identification Accuracy in Prosody Modification for Effective Emotion Conversion

S. Lakshmi Priya(✉) and D. Govind

Center for Computational Engineering and Networking (CEN),
Amrita School of Engineering, Amrita Vishwa Vidyapeetham, Coimbatore, India
lakshmipriyashasi@gmail.com, d_govind@cb.amrita.edu

Abstract. Estimating the accurate pitch marks for prosody modification is an essential step in the epoch based time and pitch scale (prosody) modification of a given speech. In epoch based prosody modification, the perceptual quality of the time and pitch scale modified speech depends on the accuracy with which glottal closure instants (epochs) are estimated. The objective of the present work is to improve the perceptual quality of the prosody modified speech by accurately estimating the epochs location. In the present work the effectiveness of variational mode decomposition (VMD) in spectral smoothing and wavelet synchrosqueezing transform (WSST) in time-frequency sharpening of a given signal is exploited for refining the zero frequency filtering (ZFF) method which is one of the simple and popular epoch extraction method. The proposed refinements to the ZFF method found to provide improved epoch estimation performance on emotive speech utterances where the conventional ZFF method show severe degradation due to rapid pitch variations. Improved mean opinion scores are obtained based on the subjective evaluation tests performed on the prosody modified speech with the epochs estimated using the refined ZFF method. The reason for improved perceptual quality in the prosody modified speech is the better identification accuracy of the estimated epochs using the proposed method as compared to the conventional ZFF method in the case of emotive speech signals.

Keywords: Zero frequency filtering
Variational mode decomposition · Wavelet SynchroSqueezed Transform

1 Introduction

Prosody modification refers to manipulation of pitch, duration and loudness of a given speech signal [1]. The pitch and time-scale (duration) modification together are applied in applications such as emotion conversion [2–4], unit selection speech synthesis [5], voice conversion [6], expressive speech synthesis for story telling applications [7]. Time scale modification is particularly used in the applications such as fast forwarding of voice messages in play back systems [1,8] and voice mail

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 334–346, 2019.
https://doi.org/10.1007/978-981-13-5758-9_28

systems [6], etc. All the above applications require the prosody modification to be performed with reduced perceptual distortions and computational complexity.

From the literature, the prosody modification methods are mainly categorized into pitch-blind methods and pitch synchronous methods [6]. The overlap add (OLA) and synchronous overlap add (SOLA) methods are the popular existing pith-blind techniques for pitch and duration modification of speech [9,10]. The prosody modification achieved in the former case is by adding of overlapping speech segments and later case by synchronously adding the overlapping segments using autocorrelation functions. The pitch consistency and the perceptual distortions are the major problems in both the methods [10]. In pitch synchronous methods, the segments are analyzed around the pitch marks estimated from the speech signals. Among the pitch synchronous overlap add (PSOLA) methods, the time-domain PSOLA (TD-PSOLA), frequency-domain PSOLA (FD-PSOLA) (only for pitch modification) and linear prediction PSOLA (LP-PSOLA) are the widely used prosody modification methods [1,10,11]. The overlapping speech frames around the pitch marks are added in all the PSOLA methods. The perceptual quality of the PSOLA method mostly depend on the accuracy of pitch marks estimation. Considering perceptual relevance of the waveform samples around pitch marks, epoch synchronous methods are introduced. In epoch synchronous prosody modification methods, epochs locations are accurately identified and samples in the epoch interval regions are preserved to ensure the perceptual quality of the synthesized speech [1,6,12]. As the epochs represent the instants at which vocaltract is decoupled with trachea due to the closure of the glottis, the waveform region around these instants are least effected by the high frequency vocaltract interactions. Therefore these regions have high SNRs due to most significant excitations from the glottal source [13]. According to Rao et. al, epoch synchronous or epoch based prosody modification has the following steps:

- Estimation of accurate epoch location from speech
- Method for deriving the modified epoch location according to the desire prosody scale factors
- Waveform synthesis method for generating the prosody modified speech

Since the processing of the original speech is anchored around the epochs, the perceptual quality of the prosody modified speech depends on the accuracy with which epochs are estimated. Prasanna et al. showed the significance of accuracy of the estimated epochs for prosody modification [3]. Prosody modification with improved perceptual quality is obtained in [3] when epochs are estimated from speech signals using zero frequency filtering (ZFF) method as compared to the group delay (GD) based epoch estimation from linear prediction residual used in the epoch based prosody modification proposed by Rao et al. in [1]. The reason for the improvement in the perceptual quality of prosody modified speech is the accuracy of epochs estimated using ZFF method than GD based method. The significance of epoch identification accuracy is further demonstrated in the epoch based prosody modification of telephonic speech signals wherein the epochs

estimation is challenging due to the attenuation of fundamental frequency components [14]. Having said that the accuracy of the estimated epochs is significant for prosody modification, the work presented in this paper is focused on refining the existing best working extraction methods for accurate epoch identification from speech signals.

ZFF of speech, proposed by Murty et al., is one of the existing popular methods for accurate epoch estimation from clean speech signals [15]. Even though other epoch estimation methods like, Speech Event Detection using the Residual Excitation mean-based signal (SEDREAMS) [16], single frequency filtering [17], dynamic plosion index based integrated linear prediction residual (ILPR) [18] etc. also provide accurate estimates of epochs from clean speech signals, ZFF method is selected as the case study for the present work. Apart from the accuracy of the hypothesized epochs, the computational efficiency of the ZFF method is one of the major attractive point behind the choice of the method for the prosody modification application. Like other existing methods, ZFF also show degradations in the epoch estimation performances when operated on emotive speech signals [19,20]. The rapid pitch variation in the emotive speech signals is the reason for the degraded epoch estimation performance. The degradation in the epoch estimation performance indeed affects the quality of prosody modified speech used in emotion conversion applications. Therefore in the present work, the effectiveness of variational mode decomposition (VMD) and wavelet synchrosqueezed transforms (WSST) in improving the time and frequency resolution of rapidly varying source components of emotive speech signals is exploited for refining the epoch estimation using conventional ZFF method.

The rest of the paper is organized as follows: Challenges in the epoch estimation using conventional ZFF method from emotive speech signal are described in Sect. 2. The effectiveness of VMD and WSST is given in Sect. 3. Section 4 describes the experimental results on improved epoch estimation and the effect of accurate epoch estimation in prosody modification. Section 5 summarizes the paper with the scope for future work.

2 Challenges in the Epoch Extraction Using Zero Frequency Filtering from Emotive Speech Signals

2.1 Conventional ZFF Method for Epoch Estimation from Speech

In zero frequency filtering, the speech after differencing operation is passed through two zero frequency resonators (ZFR) (resonators whose center frequency is centered at 0 Hz) which are connected in series. The polynomially growing/decaying output obtained from the cascaded ZFRs are subjected to local mean subtraction using a window length (which is either fixed to be 10 ms or set as the average pitch period of the input speech signal). The resulting signal after the local mean subtraction is termed as zero frequency filtered signal (ZFFS) and epochs are hypothesized as the positive zero crossings of ZFFS.

2.2 Epoch Extraction from Emotive Speech Signals

The epochs estimated using conventional ZFF method show accurate temporal resolution when compared to the ground truth differenced electro-glottogram (EGG) signal peaks for clean speech. However, when the pitch in the signal shows rapid variations as in the case of emotive speech signals, the ZFFS obtained by the local mean subtraction using fixed window length does not capture the source variations accurately and hence spurious epochs get introduced among the actual epochs. Figure 1 compares the waveform, ZFFS and differenced EGG voiced segments of anger and neutral utterances. Spurious zero crossings in the ZFFS segments (in subplot (b)) of anger utterance result in the falsified epoch estimations. The corresponding ground truth differenced EGG segment has no peaks (Fig. 1(c)) corresponding to the spurious zero crossings in Fig. 1(b). Where as neutral ZFFS segment (subplot(e)) has no spurious zero crossings present which result in unambiguous epoch estimation from neutral speech signals.

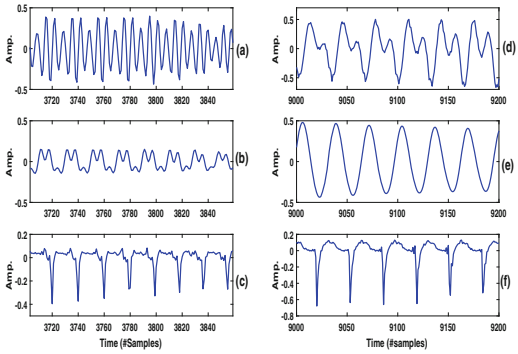


Fig. 1. ZFFS characteristics of emotive speech and neutral speech. (a) A voiced speech segment, (b) corresponding ZFFS, (c) differenced EGG of anger emotive utterance and ((d)−(f)) waveform, ZFFS and differenced EGG voiced segments of neutral speech utterances

Table 1. Epoch estimation performance of conventional ZFF on German emotion speech database [19].

Measures	Emotions				
	Neutral	Anger	Happy	Boredom	Fear
IDR	99.12	87.93	90.66	98.75	94.90
MR	00.08	00.41	00.33	00.04	00.13
FAR	00.79	11.66	09.02	01.20	04.97
IDA	00.32	00.41	00.39	00.35	00.28

Table 1 shows the epoch identification rate (IDR), miss rate (MR), false alarm rate (FAR) and epoch identification accuracy (IDA) obtained for the epochs

estimated from emotive speech utterances of classical German emotion database (EmoDb) having simultaneous ground truth EGG recordings. Table 1 is a re-computed version of the epoch performance results published in [19] for the conventional ZFF method on EmoDb. Based on the Table the average epoch identification rate obtained for emotive utterances is significantly lower than that of the neutral utterances in EmoDb. Similarly, the IDA measures also show larger deviations for emotive speech utterances as compared to neutral utterances. In the context of epoch based prosody modification, IDA measures have to be lower in order to reduce the perceptual distortions in the pitch and time scaling of speech signals.

2.3 Effect of Epoch Identification Accuracy in Epoch Based Prosody Modification

To demonstrate the significance of epoch identification accuracy for prosody modification,the epoch location is randomly varied before deriving the modified epoch locations according to pitch and duration modification factors. By randomly varying epochs, the sequence information present in the original speech will be hampered resulting in the change in the structure of the modified speech and inconsistent prosodic scaling. Figure 2, the pitch and duration modified speech segments by a factor of 2 are plotted. For instance, Fig. 2(b) plots the pitch modified speech segment by a factor of 2 from Fig. 2(a). Figure 2(c) is plotted using the epochs obtained by randomly varying the original epochs. The random variations introduced in the epochs are uniformly distributed within 1 ms. A significant difference in the temporal of the pitch modified waveform can be observed in Fig. 2(c) as compared to Fig. 2(b) (where pitch modification is achieved from original epochs estimated). Similar observation can be seen in the case duration modified speech segments plotted in Fig. 2(d−f) also. The difference in the temporal structure either introduce perceptual distortion or inconsistent prosodic changes in the prosody modified speech.

3 Variation Mode Decomposition and Wavelet Synchrosqueezed Transforms for Improved Time-Frequency Resolution

VMD is an efficient tool like empirical mode decomposition (EMD) which helps to decompose a given time series into a number of component modes having a non overlapping center frequencies [21]. The given signal is divided into component modes in such a way that, the sum of all the constituent modes provide the original time sequence. In one of our earlier works, the epochs are reliably estimated from ZFFS of emotive speech utterances by applying a VMD based smoothing in the spectral domain [20]. The spectral smoothing is achieved by discarding some of the modes (2 high central frequency modes out of 5 modes) corresponding to larger center frequencies. To further ensure the variation around

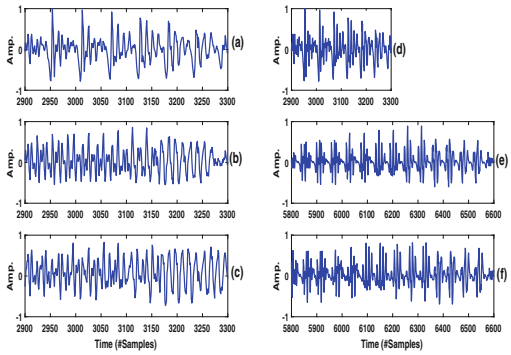


Fig. 2. Plot Showing the effect of epoch deviations in pitch and duration modified speech. (a) Original speech segment, (b) corresponding pitch modified speech by a fixed scaling factor of 2 (doubling pitch), (c) corresponding pitch modified speech obtained by randomly deviating the estimated epochs by 1 ms. (d) The original speech segment (same as in (a)), (e) duration modified speech by a factor of 2 and (f) corresponding duration modified speech by randomly varying estimated epochs by 1 ms

the fundamental frequency, the time domain signal is synthesized from the sinusoidal parameters corresponding to dominant spectral magnitude peak in each frame. The VMD based spectral smoothing of ZFFS is summarized as follows:

$$\underset{\{u_k\},\{\omega_k\}}{min} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \tag{1}$$

The constrained optimization given in the Eq. 1 decomposes the ZFFS spectra into K component modes (value of K is empirically selected as 5) with ω_k as the central frequency of the k^{th} mode. The objective function in the Eq. 1 is optimized subject to the constraint of

$$\sum_{k=1}^{K} u_k = s(t)$$

where u_k is kth component mode and $s(t)$ in the present work is the original spectra of the considered ZFFS frame.

The summation of the smoothed spectra obtained by discarding two higher VMD modes, frame in the time domain is reconstructed using sinusoidal parameters correspond to dominant spectral of the smoothed ZFFS frame.

3.1 Wavelet Synchrosqueezed Transforms for Improved Time-Frequency Resolution

The time-frequency representation of signals using continuous wavelet transform has the drawback of the dependency of center frequency (ω_0) of the mother

wavelets in the frequency scale in the signal analysis [22,23]. For instance if $W_x(a,b)$ is the continuous wavelet transform of the continuous signal $x(t)$, a and b represent the scaling and shifting variables. The $W_x(a,b)$ is computes as given in Eq. 2.

$$W_x(a,b) = \int \frac{1}{\sqrt{a}} x(t)\psi(\frac{t-b}{a}) \tag{2}$$

The frequency evolutions in $W_x(a,b)$ show a spread proportional to $a = \omega_0/\omega$. Accordingly Daubechies proposed a transform method to map the time-frequency representation given by the continuous wavelet transform, $W_x(a,b)$ to (b,$\omega_s(a,b)$), where $\omega_s(a,b)$ is the instantaneous frequency estimated from $W_x(a,b)$ [23]. Where the instantaneous frequency can be computed from the $W_x(a,b)$ as given Eq. 3,

$$\omega_x(a,b) = -i\frac{1}{W_x(a,b)}\frac{\partial}{\partial b}W_x(a,b) \tag{3}$$

While mapping the discrete bins of scale variables a_ks are computed computed for the continuous scale variables. Then the mapping transform is computed at center frequency, ω_l, by centering around the successive frequency bins $[\omega_l - \frac{1}{2}\Delta\omega,\ \omega_l + \frac{1}{2}\Delta\omega]$ and the transform $T(\omega,b)$ is computed as given in the Eq. 4.

$$T_x(\omega_l,,b) = \frac{1}{\Delta\omega} \sum_{a_k:\ |\omega(a_k,b)-\omega_l| \le \frac{\Delta\omega}{2}} W_x(a_k,b)a_k^{-3/2}(\Delta a)_k \tag{4}$$

where $(\Delta a)_k$ is the difference between discrete scales of successive bins and is represented as $(\Delta a)_k = a_k - a_{k-1}$. This mapping of original time-frequency scale to a different time-frequency plane helps to sharpen the time-frequency representation of the given signal.

The synthesis formulae from the $T(\omega,b)$ to original signal $x(t)$ is given as follows:

$$x(b) = \mathrm{Re}\left[C_\psi^{-1}\int_0^\infty W_x(a,b)a^{-3/2}da\right] \tag{5}$$

where $C_\psi = \int_0^\infty \overline{\hat{\psi}(\xi)}\frac{d\xi}{\xi}$ and the chosen mother wavelet ψ has positive frequency axis ($\psi(\xi) \ge 0$).

In the present work while synthesizing the original signal using the synthesis formulae in Eq. 5, the frequencies outside the human pitch range in $T(\omega,b)$ are discarded in order to get the predominant frequency variations due to source characteristics alone.

3.2 Combination of VMD and WSST for Smoothing Zero Frequency Filtered Signal

In the proposed method, the refined ZFFS obtained using VMD spectral smoothing is further processed using WSST for better time-frequency resolution to improve the epoch estimation performance. The proposed method has the following steps:

- The ZFFS is obtained by the conventional ZFF of speech
- The short spectra of short ZFFS segments (25 ms duration) are decomposed into five component modes using VMD
- The spectral smoothing is achieved by adding the first three modes and discarding two modes with higher central frequency
- ZFFS signal is reconstructed by the sinusoidal synthesis using the sinusoidal parameters corresponding to the dominant peaks in each of the smoothed spectral frames
- Reconstructed ZFFS is then subjected to WSST with instantaneous frequency contours obtained for each of the frequency components
- The final refined ZFFS is then obtained from the inverse WSST by removing the instantaneous frequency contours outside the human pitch range (typically within 50 Hz−200 Hz)
- The resulting refined ZFFS provided significantly improved estimation with reduced falsified epochs and better epoch identification accuracy

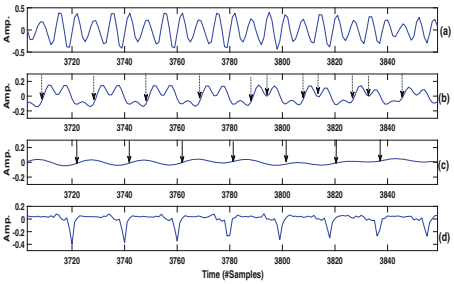


Fig. 3. Plot Showing smooth characteristics of refined ZFFS obtained by the proposed method. (a) Original speech segment from the anger waveform (same as in Fig. 1(a)), (b) corresponding ZFFS segment obtained from conventional ZFF of speech, (c) refined ZFFS segment obtained from the proposed VMD-WSST method and (d) reference difference EGG segment for the ground truth epochs. The arrows in the ZFFS plots indicate the hypothesized locations of epochs at the positive zero crossings.

Figure 3(c) plots the smoothed ZFFS obtained from the proposed VMD-WSST refinement method for an anger waveform segment corresponds to Fig. 1(a) as compared to the conventional ZFFS segment with spurious zero

crossings. The dotted arrows in Fig. 3(b) show the epochs estimated at the positive zero crossings which include falsified estimations also (round 3800 samples in the Figure). The solid arrows in the Fig. 3(c) shows no spurious zero crossings with reduced epoch deviations compared to difference egg peaks in Fig. 3(d). The reduced falsified epochs with better time resolution in the Figure show the effectiveness of VMD-WSST smoothing in improving the epoch estimation performance.

4 Experimental Results

The performance evaluation of the proposed refinements to ZFF method is performed in two stages. In the first stage, the effectiveness of the proposed method is demonstrated from epoch estimation measures such as IDR, MR, FAR and IDA. Secondly, the effect of the epoch identification accuracy is tested in prosody modification by mean opinion based subjective evaluations.

4.1 Epoch Estimation Performance Analysis

CMU-Arctic and German emotion speech (EmoDb) databases having simultaneous speech and EGG recordings are used for epoch estimation performance evaluation. CMU-Arctic database is manly used for the epoch estimation performance for the clean speech signals [24] and EmoDb database is use to test the effectiveness of proposed method in emotion conditions where there are rapid pitch variations [25]. CMU-Arctic database contains 1132 phonetically balanced utterances for each of the three speakers recorded in clean studio quality environments. In EmoDb, there are seven emotions (five distinct emotions are used in present work) of 10 professional speakers in 10 texts. All the utterances of both the databases are downsampled to 8 kHz (CMU-Arctic database is originally recorded at 32 kHz and EmoDb is originally recorded at 16 kHz).

Table 2. Epoch estimation performances of the proposed VMD-WSST refined ZFF method on German emotion speech database and CMU-Arctic database.

Measures	EmoDb-emotions					CMU-arctic database
	Neutral	Anger	Happy	Boredom	Fear	Speech from 3 Spkrs
IDR	99.71	98.35	99.06	99.55	99.20	99.34
MR	00.07	00.25	00.13	00.15	00.01	0.04
FAR	00.22	01.40	00.81	00.30	00.79	0.62
IDA	00.33	00.30	00.30	00.33	00.25	0.31

Table 2 shows the epoch estimation performances of the proposed method on EmoDb and CMU Arctic databases. The EmoDb is used for the performance evaluation to test how effective the proposed VMD and WSST is in improving

the ZFF based epoch estimation for m emotive speech with random pitch variations. From Table 2, apart from the improved performance on emotive utterances, the method also provides equal performance with that of conventional ZFF for clean speech utterances of CMU-Arctic database. Also, apart from improving the identification rate of estimated epochs, the proposed method also improves the temporal resolution (identification accuracy) in emotive utterances.

4.2 Perceptually Improved Prosody Modification

Due to significant improvement in epoch estimation performance of the proposed method in emotive utterances with rapid pitch variations, the corresponding perceptual improvement in prosody modified speech is subjectively compared using epoch estimation method using conventional ZFF and the proposed method. The subjects were asked to give the opinion scores based on the perceptual distortions present. The justification for the ranking are provided in Table 3.

Table 3. Ranking used for judging the processing distortion due to different epoch estimation methods. The ranking is used as used in the paper by Prasanna et al. [3]

Score	Attribute	Level of distortion
1	Unsatisfactory	Very annoying and objectionable
2	Poor	Annoying but not objectionable
3	Fair	Perceptible and slightly annoying
4	Good	Just perceptible but not annoying
5	Excellent	Imperceptible

Three male utterances chosen from anger emotion category present in EmoDb is used for the subjective evaluation. The anger utterances are prosody modified for arbitrary pitch and duration scaling factors. There were a total of 39 pitch and duration modified stimuli generated ($(3X3 + 3X3)$ X 2 Methods +3 Original waveforms) for evaluation. 10 subjects who have prior knowledge of distortions present in speech signal, participated in the subjective evaluation. The subjects were asked specifically to note the perceptual distortions present in each stimulus presented to them during the pilot test. The subjects were asked rank the perceptual distortion in a five point scale as provided in Table 3. The filenames of the prosody modified files are encoded before presenting to the subjects to have an unbiased opinion based on the perceptual distortions irrespective of the methods used for prosody modification. The opinion scores obtained for three files given by 10 subjects for each of the epoch estimation methods are averaged to get the mean opinion scores (MOS).

Table 4 shows a significant improvement in the MOS obtained by using the proposed method of epoch estimation in the prosody modification as compared to the scores obtained for conventional ZFFS. As epoch estimation performances

Table 4. Comparison of significance of differences in MOS scores of pitch and duration modification with epochs obtained from conventional ZFF and proposed method.

Pitch modification		
Scaling factors	Conv. ZFF	Proposed
0.8	1.86	2.52
1.0	3.22	3.26
1.2	2.26	2.61
Duration modification		
0.8	2.96	3.36
1.0	3.23	3.62
1.2	3.04	3.28

of both the methods on anger utterances are almost equal, the perceptual distortions present in prosody modified speech from the two methods sound nearly similar. Hence the subjective evaluation is performed for anger utterances of EmoDb databases.

5 Summary and Conclusions

The work presented in the paper proposed a refinement to the existing ZFF method of epoch estimation for improving the epoch estimation performance from emotive utterances. The VMD based spectral smoothing and the effective time-frequency sharpening using WSST are the reasons for the achieving improved epoch estimation performance from emotive utterances. The improved epoch identification accuracy found to reduce the perceptual distortion which is introduced during the epoch based prosody modification of emotive utterances with rapid pitch variations. The future work should focus in further improving the epoch estimation performance for emotive speech utterances where the perfect epoch extraction like from neutral utterances is still a research issue. Although the MOS for the proposed method is improved, the perceptual quality of the prosody modified speech is only in 'fair' quality level. Hence the future work should also focus on further improving the perceptual quality of the prosody modified speech for effective emotion conversion.

References

1. Rao, K.S., Yegnanarayana, B.: Prosody modification using instants of significant excitation. IEEE Trans. Audio Speech Lang. Process. **14**, 972–980 (2006)
2. Govind, D., Prasanna, S.R.M.: Dynamic prosody modification using zero frequency filtered signal. Int. J. Speech Technol. **16**(1), 41–54 (2013)
3. Prasanna, S.R.M., Govind, D., Rao, K.S., Yenanarayana, B.: Fast prosody modification using instants of significant excitation. In: Proceedings of the Speech Prosody, May 2010

4. Cabral, J.P., Oliveira, L.C.: Emo voice: a system to generate emotions in speech. In: Proceedings of the INTERSPEECH, pp. 1798–1801 (2006)
5. Clark, R.A.J., Richmond, K., King, S.: Multisyn: open-domain unit selection for the festival speech synthesis system. Speech Commun. **49**, 317–330 (2007)
6. Rudresh, S., Vasisht, A., Vijayan, K., Seelamantula, C.S.: Epoch-synchronous overlap-add (ESOLA) for time- and pitch- scale modification of speech signals. arXiv:1801.06492v1, pp. 1–10, January 2018
7. Tao, J., Kang, Y., Li, A.: Prosody conversion from neutral speech to emotional speech. IEEE Trans. Audio Speech Lang. Process. **14**, 1145–1154 (2006)
8. Portnoff, M.R.: Time-scale modification of speech based on short-time fourier analysis. IEEE Trans. Acoust. Speech Signal Process. ASSP **29**, 374–390 (1981)
9. Allen, J.B., Rabiner, L.R.: A unified approach to short-time fourier analysis and synthesis. Proc. IEEE **65**(11), 1558–1564 (1977)
10. Mourlines, E., Laroche, J.: Non-parametric techniques for pitch-scale and time-scale modification of speech. Speech Commun. **16**, 175–205 (1995)
11. Moulines, E., Charpentier, F.: Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. Speech Commun. **9**, 452–467 (1990)
12. Govind, D., Joy, T.T.: Improving the flexibility of dynamic prosody modification using instants of significant excitation. Int. J. Circuits Syst. Signal Process. **35**(7), 2518–2543 (2016)
13. Yegnanarayana, B., Veldhuis, R.N.J.: Extraction of vocal-tract system characteristics from speech signals. IEEE Trans. Speech Audio Process. **6**(4), 313–327 (1998)
14. Adiga, N., Govind, D., Prasanna, S.R.M.: Significance of epoch identification accuracy for prosody modification. In: Proceedings of the International Conference on Signal Processing and Communication (SPCOM) (2014)
15. Murty, K.S.R., Yegnanarayana, B.: Epoch extraction from speech signals. IEEE Trans. Audio Speech Lang. Process. **16**(8), 1602–1614 (2008)
16. Drugman, T., Dutoit, T.: Glottal closure and opening instant detection from speech signals. In: Proceedings of the INTERSPEECH, pp. 2891–2895 (2009)
17. Kadiri, S.R., Yegnanarayana, B.: Epoch extraction from emotional speech using single frequency filtering approach. Speech Commun. **86**, 52–63 (2017)
18. Prathosh, A.P., Ananthapadmanabha, T.V., Ramakrishnan, A.G.: Epoch extraction based on integrated linear prediction residual using plosion index. IEEE Trans. Audio Speech Lang. Process. **21**(12), 2471–2480 (2013)
19. Govind, D., Prasanna, S.R.M.: Epoch extraction from emotional speech. In: Proceedings of the Signal Processing & Communications (SPCOM), pp. 1–5, July 2012
20. Govind, D., Pravena, D., Ajay, G.: Improved epoch extraction using variational mode decomposition based spectral smoothing of zero frequency filtered emotive speech signals. In: Proceedings of National Conference on Communications (NCC). Indian Institute of Technology Hyderabad, February 2018
21. Dragomiretskiy, K., Zosso, D.: Variational mode decomposition. IEEE Trans. Signal Process. **62**(3), 531–544 (2014)
22. Daubechies, I., Maes, S.: A nonlinear squeezing of the continuous wavelet transforms based on auditory nerve model. In: Wavelets in Medicinal Biology. CRC Press (1996)

23. Daubechies, I., Lu, J., Wu, H.T.: Synchrosqueezed wavelet transforms: an empirical mode decomposition-like tool. Appl. Comput. Harmon. Anal. **30**, 243–261 (2011)
24. Kominek, J., Black, A.: CMU-arctic speech databases. In: 5th ISCA Speech Synthesis Workshop, Pittsburgh, PA, pp. 223–224 (2004)
25. Burkhardt, F., Paeschke, A., Rolfes, M., Sendlemeier, W., Weiss, B.: A database of German emotional speech. In: Proceedings of the INTERSPEECH, pp. 1517–1520 (2005)

Design of a Compact High Gain Wide Band Ψ Shaped Patch Antenna on Slotted Circular Ground Plane

Anitha Peram[1(✉)], Agarala Subba Rami Reddy[2], and Mahendra N. Giri Prasad[1]

[1] Department of ECE, JNTUA, Anantapuramu, Andhra Pradesh, India
anitha.peram@gmail.com
[2] G.K. College of Engineering, Sullurpet, Andhra Pradesh, India

Abstract. A compact wideband high gain Ψ shaped patch antenna on circular ground plane is proposed. The patch is designed and fabricated on FR4 substrate with a thickness of 1.6 mm and is compact in size as 122.7 mm^2 including ground plane. The SMA connector with a center radius of 0.36 mm is connected at a coordinate (x = 5 mm, y = 3.3125 mm) as a feeding line to give RF signal as input. A wide impedance bandwidth is obtained due to the slots on the radiating element and by placing shorting post near zero potential fields makes the structure compact. An impedance bandwidth is further enhanced by placing an inverted U-slot on ground plane. The proposed structure is resonating at five different frequencies 1.924 GHz, 2.88 GHz, 5.29 GHz, 8.58 GHz and 12.27 GHz yields an impedance bandwidth of 345 MHz, 92 MHz, 200 MHz and 4.77 GHz respectively. Reflection coefficient (S_{11}) at 8.58 GHz & 12.27 GHz is −36.37 dB and −44.61 dB respectively. The proposed antenna is giving a maximum gain of 6.1 dB and has a stable radiation pattern with in the resonating band. The designed antenna is fabricated and is experimentally validated for the results. It reveals that the proposed antenna is suitable for WLAN and X band applications.

Keywords: Circular ground plane · Shorting post · Inverted U slot
X band · WLAN band

1 Introduction

In the Recent years, broadband micro-strip antennas are attracting more attention in the areas such as wireless local area network (WLAN), Bluetooth, Global Positioning System (GPS), radar, satellite and biomedical telemetry systems. X band (8–12 GHz) technology is widely used in various applications due to high data transmission rates, short range and large bandwidth. Numerous applications of MPA which operates in the X-band region includes in satellite communications, military communications, RADAR, Amateur radio, Precision approach RADAR (PAR) (9–9.2 GHz), traffic light crossing detectors (10.4 GHz), motion detectors (10.57 GHz) etc.

Broadband antennas are of immense demand for these systems. Hence, it is required to develop a miniaturized broadband antenna with high gain, better efficiency and stable

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 347–357, 2019.
https://doi.org/10.1007/978-981-13-5758-9_29

radiation pattern. MPA is one which will fulfill the demands of Wireless communication due to its attractive advantages like less profile, low weight, low cost, easy to fabricate and conformal design. Apart from advantages it has limitations such as lower power handling, lower bandwidth and low efficiency [1]. A high gain, wide band, compact MPA at low cost with simple design is the challenging issues in designing MPA for commercial use. Since improvement in one will degrade the other parameter.

To enhance the bandwidth and gain of patch antenna, scientists and researchers have developed various methods over two decades. The bandwidth of an antenna can be increased by different techniques like, Multilayered configurations [2, 3], however if height is increased beyond 0.1λ0, surface wave excitation results in degradation of the antenna performance. Second, Stacked Multi-resonator and Planar Multi-resonator configurations [4, 5] increases a bandwidth up to 40%, the resulting configuration is larger in size compared to single resonator. Another method of increasing bandwidth is by using foam substrate and thick air [6] the resulting structure is larger in volume. Another miniaturized wide band approach is by reshaping the structure (placing slots) and defected ground plane structure, using shorting pins and shorting walls [8–10]. However, the obtained bandwidths of these antennas are below 30–80%. In [11–15] an X band MPA's were proposed using reshaping and defected ground plane approach, however not met enough bandwidth and gain criteria. Another method for improving the performance is the use of fractal antennas [16]. For further enhancement hybrid fractal slot is proposed [17] with a patch size of 40 × 40 mm i.e. complex in structure, large in size and gain is not stable throughout the band.

For commercial applications of MPA, However enhancing the bandwidth and gain, compactness in structure are the major design considerations as improving of one will degrade the other. This made us to work on broadband Ψ-shaped patch antenna on circular ground plane with an inverted U slot achieves the desired goal.

In this article a design that provides simultaneously a compact, high gain and wide bandwidth MPA for commercial applications. The techniques such as modifying the shape of the patch, placing slots on ground plane and using a shorting pin are merged. Hence, the design employs a modified Ψ-shape patch on an inverted U-slot loaded circular ground plane with a shorting post near to probe feed to meet our goals. Parametric analysis is performed to attain good antenna performance. By optimizing the arm length of Ψ-shape patch, slot and the position of shorting post wider impedance bandwidths of 4.77 GHz in X band and 330 MHz in L band and a peak gain of 6.1 dBi is achieved. The designed antenna is fabricated and the results are validated experimentally which supports majorly X band applications.

2 Antenna Design and Configuration

The geometry of the proposed antenna is illustrated in Fig. 1. An MPA with a modified Ψ-shape patch as radiating element and a circular ground plane with an inverted U-slot is designed. The ground plane and patch are printed on opposite sides of Flame Resistant (FR4) substrate with a thickness of 1.6 mm, relative permittivity of 4.4 and a dielectric loss tangent (tan δ) of 0.02. An inverted U slot is placed below the Ψ patch on a circular ground plane with a diameter of 12.5 mm to get wide bandwidth. Slot and

Patch dimensions are chosen after numerous iterations. Optimized dimensions of the designed antenna are given in Table 1. A 50 Ω co-axial probe feed is used to excite radiation. By placing shorting post a coordinate (x = 5 mm, y = 4.125 mm) with a diameter of 1 mm enhances the bandwidth further. The prototype of the designed patch is illustrated in Fig. 2.

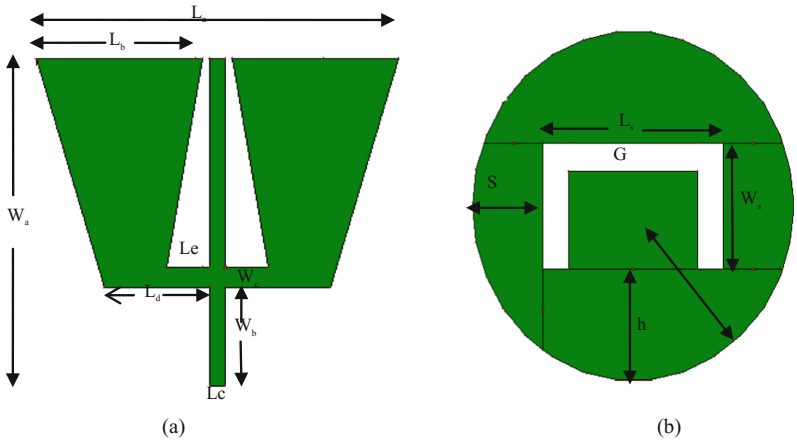


Fig. 1. Geometry of the proposed antenna (a) Radiating patch (b) Ground plane

Table 1. Designed antenna specifications

Parameter	L_a	L_b	L_c	L_d	L_e	W_a	G
Designed antenna	10	4.6	0.4	2.9	1.2	10	1
Parameter	W_b	W_c	W_s	L_s	R	h	S
Designed antenna	3	0.625	4.5	7	6.25	4	2.75

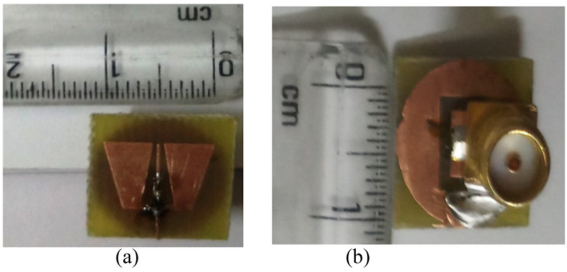


Fig. 2. Proto type of the fabricated structure (a) Top layer (b) Bottom layer

2.1 Parametric Study

To investigate the effect of parameters on antenna performance, some sensitive parameters are identified and studied based on design. Depending on the geometry of

ground plane and patch the frequency response is decided. Hence, to optimize the performance, parametric study have been performed with respect to (a) Effect of shorting post and inverted U slot on ground plane (b) Effect of slot width Ws i.e. width of vertical slots (c) Effect of G i.e. horizontal slot width. All simulation has been carried out by using IE3D simulator.

2.1.1 Effects of Shorting Post and U Slot

To study the effect of shorting post on antenna performance a simulation is carried out without shorting post by keeping all other parameters fixed and is presented in Fig. 3. From the figure it is seen that there is a significant effect on bandwidth. The Ψ patch with inverted U slot on ground plane resonates at 5.2 GHz and 12.71 GHz, which yields negligible bandwidth and reflection co-efficient.

The same antenna without U slot on ground plane and with shorting post at (x = 5 mm, y = 4.125 mm) coordinates, the patch is resonating at four different frequencies as 2.2 GHz, 3.16 GHz, 5.61 GHz and 11.268 GHz yields an impedance bandwidth of 2.56 GHz ($S_{11} < -10$ dB from 10.10 GHz to 12.66 GHz) in X band.

The same antenna with shorting post and an inverted U slot on ground plane resonates at five different frequencies 1.924 GHz, 2.88 GHz, 5.29 GHz, 8.58 GHz and 12.27 GHz. An impedance bandwidth of 345 MHz, 92 MHz, 200 MHz and 4.77 GHz is achieved at their corresponding resonating frequencies. Table 2 shows the effect of shorting post and U slot on reflection coefficient and hence on impedance bandwidth

On comparing, the patch without shorting post and with slot it is resonating at single frequency yields a very narrow bandwidth. With the insertion of shorting post excluding slot resonates at lower frequency and hence results in size reduction. By placing shorting post and slot makes the patch to resonate at two nearby frequencies which enhances the bandwidth. This can be attributed due to strong coupling between the patch and ground plane.

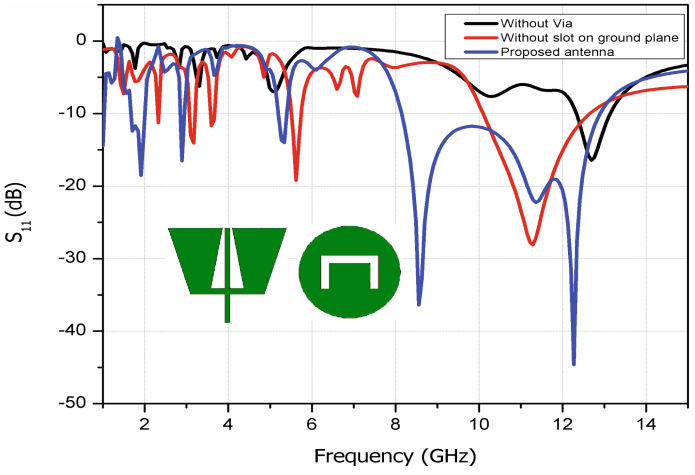


Fig. 3. Simulated S_{11} characteristics by considering the effect of slot and shorting post

Table 2. Comparison results by considering the effects of shorting posts and U slot

Antennas	Return loss (S_{11} < −10 dB)	Bandwidth	Resonating frequency f_0 in GHz
Proposed antenna	8.14 GHz–12.91 GHz	4.77 GHz	8.58 & 12.27
Antenna without slot on ground plane and with via	10.10 GHz–12.66 GHz	2.56 GHz	11.268
Antenna without via and with slot	12.31 GHz–13.14 GHz	0.83 GHz	12.71

2.1.2 Effect of Slot Width Ws; Width of Vertical Slots

With other parameters optimized, for different values of slot width the designed antenna is simulated for reflection coefficient which is shown in Fig. 4. Slot width is varied from 3.5 mm to 5 mm and it can be seen that there is a significant variation in bandwidth and impedance matching. Good impedance matching and bandwidth can be found for a slot width of 4.5 mm further increase in width decreases both bandwidth and matching. A slot width of 4.5 mm produces multiple resonances and thus achieves wide bandwidth with a compact size.

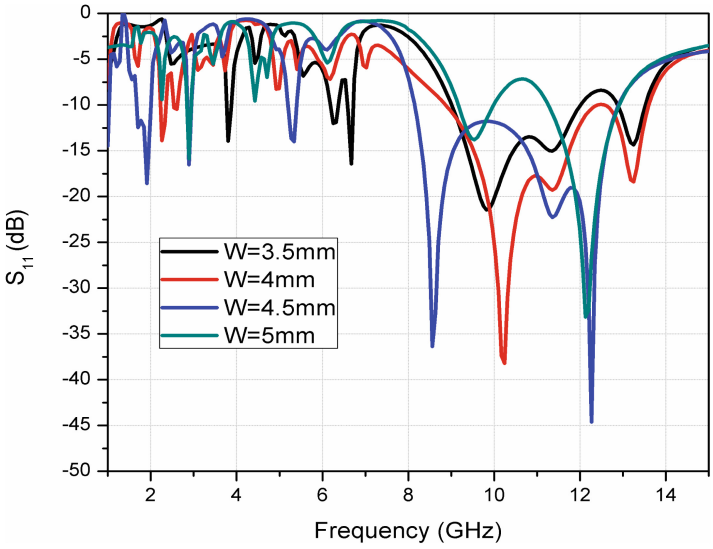


Fig. 4. Simulated S_{11} for different slot widths of the proposed antenna

2.1.3 Effects Gap Between the Slots

From Fig. 5, it can be seen that the variation in slot gap majorly affects impedance matching. Good impedance matching is found for a gap of 1 mm. A coupling between the patch and ground plane increases due to decrease in overlapping area. A slight increase in bandwidth can be seen for a gap of 0.8 mm than the proposed antenna, however reduces impedance matching. Further increase in gap reduces the bandwidth and impedance matching due to increase in overlapping area.

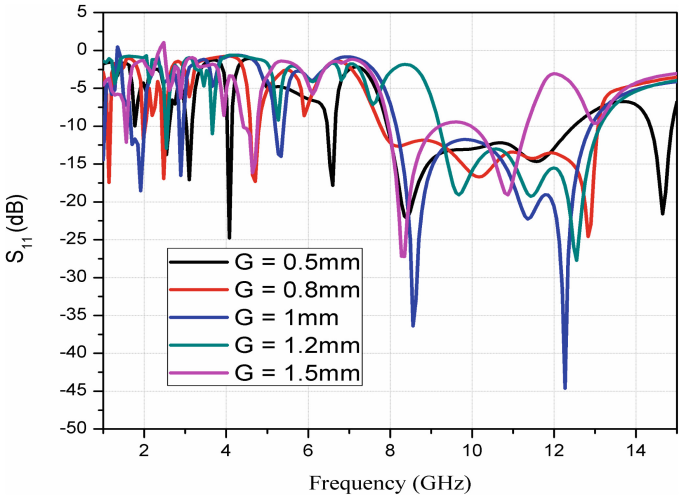
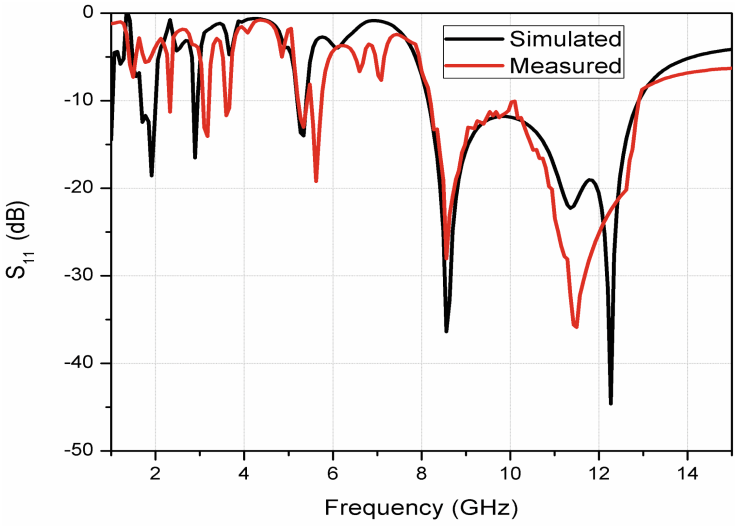


Fig. 5. Simulated S_{11} by considering the effect of slot gap

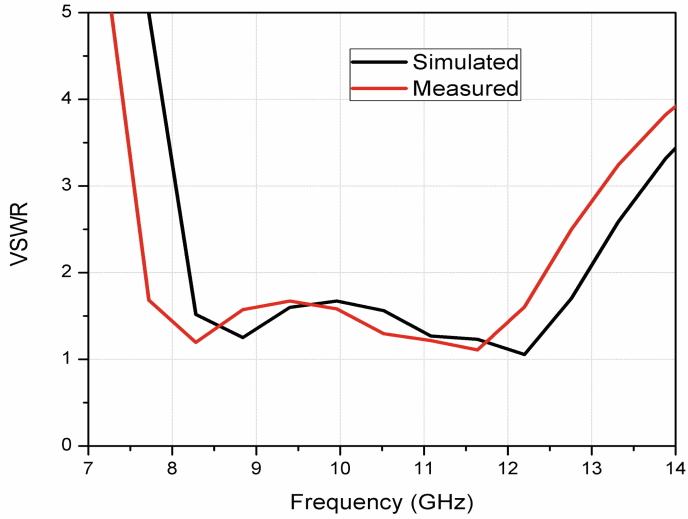
3 Measured Results

The designed antenna is fabricated on FR4 substrate with a thickness of 1.6 mm and is tested for experimental verification through The Agilent N5230A (10 MHz–50 GHz) series network analyzer. The return loss parameters of simulated and measured results comparison is shown in Fig. 6(a). The measured impedance bandwidth is about 4.77 GHz ($S_{11} < -10$ dB from 8.14 GHz to 12.91 GHz). Comparison of simulated and measured results reveals that there is a good correlation throughout the band. Slight variation may be due to fabrication errors. VSWR < 2 for the designed antenna is from 8.14 GHz to 12.91 GHz and are verified through VNA. The designed antenna has a minimum VSWR of 1.055 at 8.58 GHz which is illustrated in Fig. 6(b).

The proposed antenna gain is measured by two antenna method after attaining S_{21} parameters of Horn antenna and AUT using anechoic chamber of 5 m × 3 m × 3 m and is shown in Fig. 7. The simulated and measured gain of the proposed patch antenna



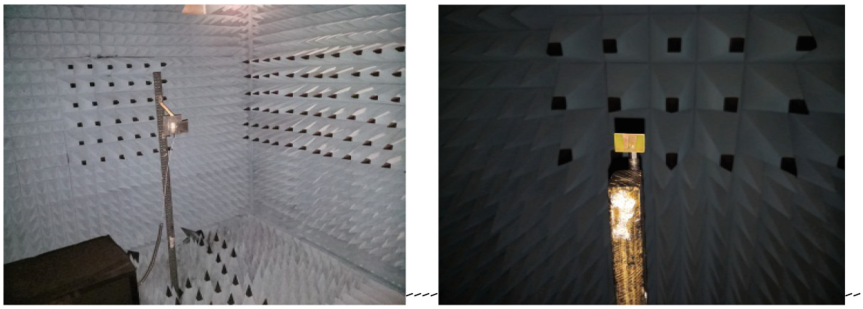
(a)



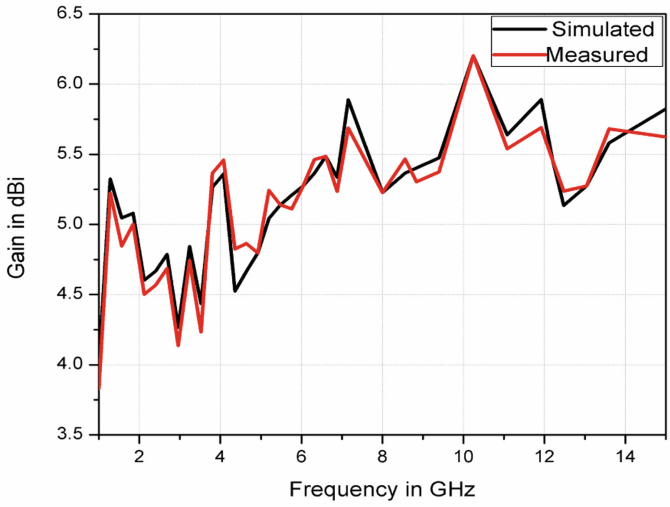
(b)

Fig. 6. Simulated and measured results of the designed antenna (a) S_{11} characteristics (b) VSWR characteristics

ranges from 5.2 dBi to 6.13 dBi in the resonating band (8–13 GHz). For practical applications gain variation should be within 2 dBi in the resonating band which satisfies the requirement. From 2–15 GHz gain variation is between 4–6 dBi. The simulated radiation pattern in E and H planes at resonating frequencies are shown in Fig. 8. The radiation pattern is omnidirectional both in E and H planes.



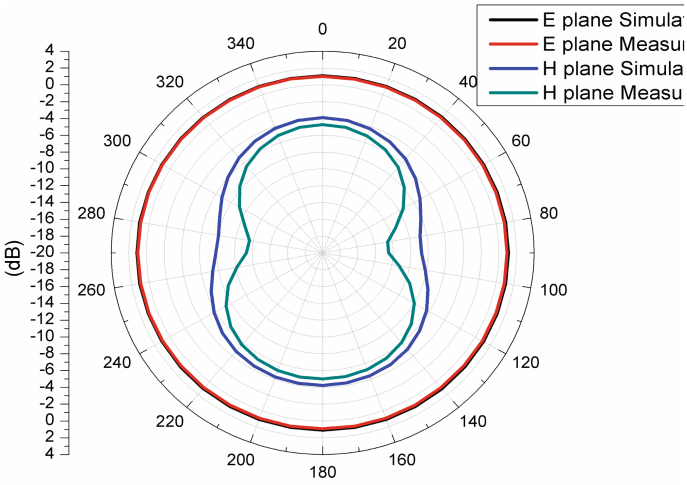
(a) (b)



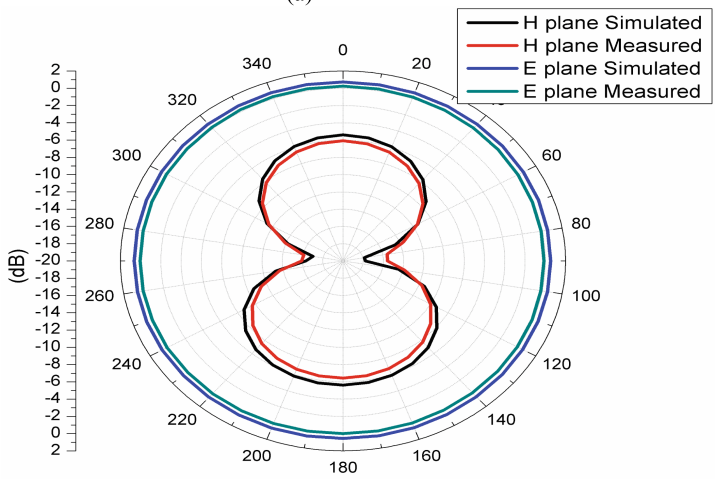
(c)

Fig. 7. Radiation pattern and gain measurement through anechoic chamber (a) Transmitter (b) AUT (c) Simulated and Measured Gain

Current distribution at 8.58 GHz and 12.2 GHz is shown in Fig. 9. It reveals that strong coupling exists between the slot edges and patch shows that performance depends on slot length and width. Current is concentrated at the edges of U-shaped inverted slot and Ψ-shaped patch. Table 3 shows the comparison study of the proposed antenna with respect to literature.



(a)



(b)

Fig. 8. Radiation patterns in E and H planes at (a) 8.58 GHz & (b) 12.27 GHz

Table 3. Comparison study with respect to literature

Parameters	[9]	[6]	[10]	[11]	[12]	[13]	[16]	[17]	Our work
Resonating frequency in GHz	9.56	9.6	9	11.4	9.75	10.53	2.46 & 9.4	5.7, 6.4 & 8.4	**5.3, 8.58 & 12.27**
Bandwidth in GHz	3.04	1.2	0.11	0.0114	1.801	4.06	0.45, 2.44	0.158, 0.1593, 0.2677	**0.2 & 4.77**
Patch size in mm	100 × 9 × 3.2	50 × 50 × 2.4	11.01 × 13.9 × 1.6	4 × 9 × 1.57	17.56 × 18.04 x1.6	8.5 × 7.5 × 4.662	40 × 39 × 1.6	60 × 55 × 1.6	**11.1 × 11.1 × 1.6**
Gain in dBi	6	14	N/A	4.82	3.56	5.23	5.73	3.74	**6.1**

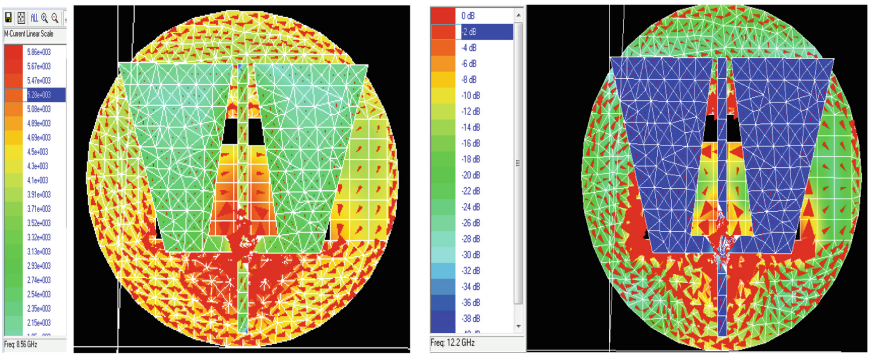


Fig. 9. Current distribution of the proposed patch antenna at 8.56 GHz and 12.2 GHz

4 Conclusion

High gain wide band Ψ shaped patch antenna with an inverted U slot on circular ground plane is presented. The designed antenna is fabricated on FR4 substrate with a thickness of 1.6 mm and is tested for result yields a good impedance bandwidth of 4.77 GHz (8.14 GHz–12.91 GHz) with a maximum gain of 6.1 dB. Compactness is achieved by placing shorting post along the zero potential field makes it easy to integrate. The experimental result shows that there is a good correlation between the simulated and measured ones. The proposed design best suits for WLAN and X band applications.

References

1. Kumar, G., Ray, K.P.: Broadband Microstrip Antennas. Artech House, Boston (2003)
2. Wnuk, M., Przesmycki, R., Nowo-sielski, L., Bugaj, M.: Multilayer microstrip antenna on flat base in the X band (8.5 GHz–12 GHz). In: Progress in Electromagnetics Research Symposium (PIERS), vol. 7, pp. 216–220 (2011)

3. Sarkar, D., Saurav, K., Srivastava, K.V.: Multi-band microstrip-fed slot antenna loaded with split-ring resonator. Electron. Lett. **50**(21), 1498–1500 (2014)
4. Anguera, J., Boada, L., Puente, C., Borja, C., Soler, J.: Stacked H-shaped microstrip patch antenna. IEEE Trans. Antennas Propag. **52**(04), 983–993 (2004)
5. Wong, K.-L., Tung, H.-C., Chiou, T.-W.: Broadband dual-polarized aperture-coupled patch antennas with modified H-shaped coupling slots. IEEE Trans. Antennas Propag. **50**(02), 188–191 (2002)
6. Sharma, A.K., Mittal, A.: Diagonal slotted diamond shaped dual circularly polarized microstrip patch antenna with dumbbell aperture coupling. In: The European Conference on Wireless Technology, IEEE Conference, pp. 463–465 (2005)
7. Mandal, K., Sarkar, P.P.: A compact low profile wideband U-shape antenna with slotted circular ground plane. Int. J. Electron. Commun. (AEÜ) **70**, 36–340 (2016)
8. Verma, S., Ansari, J.A.: Analysis of U-slot loaded truncated corner rectangular microstrip patch antenna for broadband operation. Int. J. Electron. Commun. (AEÜ) **69**, 1483–1488 (2015)
9. Claudius, L., Thomas, F.E.: Unidirectional radiation efficient stacked aperture antenna for X-band application. J. Antennas Wirel. Propag. Lett. **7**, 264–266 (2008)
10. Verma, A., Srivastava, N.: Analysis and design of rectangular microstrip antenna in X band. MIT Int. J. Electron. Commun. Eng. **1**(1), 31–35 (2011)
11. Mazumdar, B.: A compact microstrip antenna for X band application. Int. J. Recent Technol. Eng. **1**, 104–106 (2012)
12. Harrabi, A., Razban, T., Mahe, Y., Osman, L., Gharsallah, A.: Wideband patch antenna for x-band applications. In: Proceedings of the Progress in Electromagnetics Research Symposium, Stockholm, Sweden (2013)
13. Lakrit, S., Ammor, H.: Design on X-band wideband and high-gain multi-layer microstrip antenna. J. Eng. Sci. Technol. Rev. **7**(3), 176–179 (2014)
14. Anitha, P., Reddy, A.S.R., Giri Prasad, M.N.: Design of a compact micro-strip patch antenna with enhanced bandwidth on modified ground plane for x-band applications. J. Eng. Technol. **6**, 112–120 (2018). Special Issue (Emerging Trends in Engineering Technology)
15. Anitha, P., Reddy, A.S.R., Giri Prasad, M.N.: Design of a compact dual band patch antenna with enhanced bandwidth on modified ground plane. Int. J. Appl. Eng. Res. **13**(1), 118–122 (2018). ISSN 0973-4562
16. Sharma, N., Sharma, V.: A design of Microstrip Patch Antenna using hybrid fractal slot for wideband applications. Ain Shams Eng. J. 1–7 (2017). https://doi.org/10.1016/j.asej.2017.05.008
17. Gupta, M., Mathur, V.: Koch boundary on the square patch microstrip antenna for Ultra Wideband applications. Alexandria Eng. J. AEJ **57**(3), 2113–2122 (2017)

Quality and Complexity Measurement of 2D-DCT Archietecture Using Loeffler Algorithm Along with CSD and CSE

Vivek V. Kajagar(✉), Shaik Mohammad Ashraf Ansari, J. N. Swaminathan, and S. Rajasekaran

Department of Electronics and Communication Engineering, Madanapalle Institute of Technology and Sciences, Madanapalle, AP, India
vivekvkajagar5@gmail.com, smd.ashrafansari@gmail.com, sammbuddy@gmail.com, srsmeae@gmail.com

Abstract. To optimize the Discrete Cosine Transform (DCT) in terms of size and to improve the quality of the image in image compression, in this paper, A 2D-DCT using Loeffler algorithm along with Canonical Signed Digit (CSD) and Canonical Sub expression Elimination (CSE) has been proposed. For fast computation 2-DCT/IDCT is executed by using 1D DCT row column method. The performance of the proposed architecture has been evaluated and compared with the other technique.

Keywords: 2D-DCT · Area · PSNR · Loeffler · Image compression

1 Introduction

The modern world is driven by technology and sciences, one such technological aspect is image compression in digital image processing. Image compression helps in compressing the original data without loss of information and fewer data storage. Reduction of data storage means improving the capacity of storage [1]. Through years, in various applications such as space, audio and video applications, image compression has gained a lot of attention in image processing [2–4]. The compressed data can be transferred with less communication cost and time through the internet [5]. In image compression, Discrete Cosine Transform (DCT) is popular, as it results in low power consumption and less computational time in contrast with other conventional transforms [5]. In image standards such as JPEG [6], H.264 [7], MPEG [8], DCT is widely used. To optimize the DCT in terms of size, computational time and power consumption, many algorithms were introduced. To reduce the complexity, a recursive algorithm was investigated. This algorithm results in reducing the number of adder and multipliers [9]. Based on matrix decomposition, a fast DCT algorithm was introduced which results in six times computationally faster than the existing DCT [10]. Distribute arithmetic approach based on bit serial is investigated on DCT. This approach results in complex size [11]. In order to overcome the problem of complexity in terms of area, Multiple Constant Multiplication (MCM) approach based on shift and adds operations

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 358–365, 2019.
https://doi.org/10.1007/978-981-13-5758-9_30

was implemented. This approach results in less power consumptions and image quality [12]. Based on Loeffler algorithm, simulation work using matlab has been investigated and the performance is evaluated in terms of PSNR on JPEG image. By using different algorithms, performance is compared and analyzed to solve 2D-DCT [13]. In the existing approaches, area and PSNR is not fully explored. Performance is analyzed individually in terms of area, power consumption and PSNR. In this paper, a FPGA design based on DCT and IDCT is implemented for image compression. An Loeffler algorithm for 2D-DCT and IDCT is investigated along with CSE and CSD. The performance is analyzed in terms of PSNR and size.

The rest of the paper as follows: Basics of DCT is discussed in Sect. 2. Proposed loeffer algorithm is explained briefly in Sect. 3. The results and observation are discussed in Sect. 4 and finally conclusion and future scope is summarized.

2 Basics of DCT Implementation

Theoretically from Eq. 1, the direct investigation of 2D-DCT performance results in more costs and complexity which is around 1024 multiplication and 896 additions. The other way of implementing 2D-DCT architecture is by cascading of two 1D-DCT and the transposition block based on separability property.

The basic 2D-DCT is formulated in Eq. (1)

$$F(a,b) = D(a)\,D(b)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} F(x,y) \times \cos\left[\frac{(2x+1)a\pi}{2N}\right] \times \cos\left[\frac{(2y+1)b\pi}{2N}\right] \tag{1}$$

where

$$a,b = 0,1,2,3,\ldots\ldots N-1$$

$$D(a) = D(b) = \sqrt{\frac{1}{N}} for\, a,b = 0$$

$$D(a) = D(b) = \sqrt{\frac{2}{N}} for\, a,b = 1,2,3,\ldots\ldots N-1$$

Here $F(x,y)$ size of $N \times N$, represents the input matrix pixels of sub image, size of. $F(a,b)$ are corresponding output coefficients of 2D-DCT. $D(a), D(b)$ represents the normalizing factor. The other way, the 2D-DCT architecture is formed by cascading of two 1D-DCT and the transposition block based on separabilityproperty. The 1D-DCT is formulated in Eq. 2

$$F(a) = D(a)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} F(x) \times \cos\left[\frac{(2x+1)a\pi}{2N}\right] \tag{2}$$

where

$$a, b = 0, 1, 2, 3, \ldots\ldots N - 1$$

$$D(a) = \sqrt{\frac{1}{N}} for\, a = 0$$

$$D(a) = \sqrt{\frac{2}{N}} for\, a = 1, 2, 3, \ldots\ldots N - 1$$

Here $F(x)$ size of $N \times N$, represents the input matrix pixels of sub image, size of. $F(a)$ are corresponding output coefficients of 2D-DCT. $D(a)$ represents the normalizing factor.

3 Proposed 2D-DCT Architecture

The inspiration of proposed architecture of 2D-DCT is to design hardware with low power consumption, computational time and the area. This results in improving the accuracy in image reconstruction. The proposed 2D-DCT architecture is based on separability property, by cascading of two 1D-DCT and the transposition block (Fig. 1).

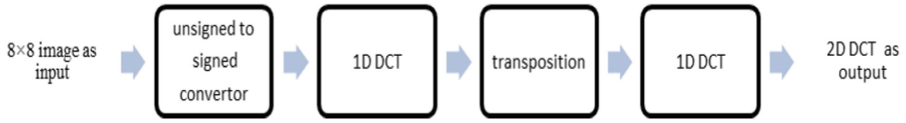


Fig. 1. 2D-DCT architecture

An image of Size 8 × 8 resolution ranging from 0 to 255 is fed as an input. The given input is converted from unsigned to signed values, this conversion is said to be as level shifting by subtracting from 128. The signed converted values are now given as the input to first 1D DCT in either row wise or column wise. The output of the first 1D DCT is given to the transposition block where the transpose of the input takes place. Then the output of this transposition block is given as input to the second 1D DCT to obtain the final 2D DCT. Similar steps are repeated for 2D IDCT. For 2D-DCT/IDCT, it is important to select suitable algorithm for higher order dimensions. To implement the 2D-DCT, a loeffler algorithm has been proposed.

4 Loeffler Algorithm

Loeffler algorithm is divided in four stages starting from 1 to 4. Due to data dependency, these stages are executed in serial mode. The algorithm in stage 2 distributed into two parts, 1. Even 2. odd ones. Even part actually a 4 point DCT and it is again distributed into even and odd slices in stage 3. In Fig. 2, The rotational block in indicates a multiplication by a constant.

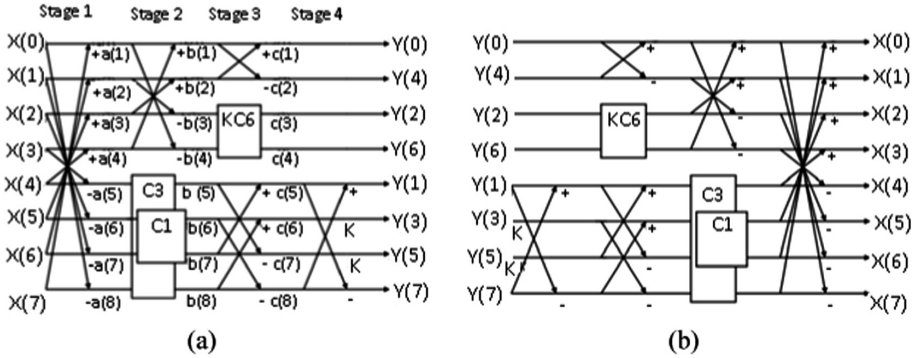
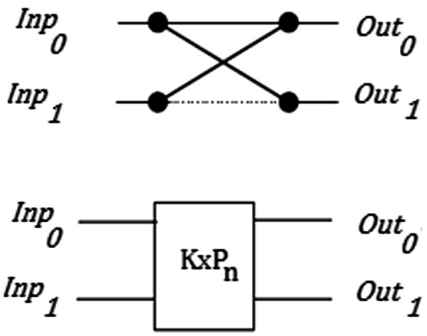


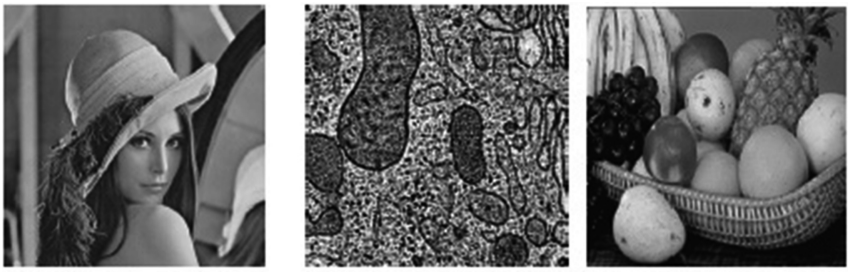
Fig. 2. Flow graph of Loeffler 8 point (a) DCT (b) IDCT

With the help of rotational block, the multiplication and additions can be reduced from 4 multiplications and 2 additions to 3 multiplications and 3 additions. The Fig. 3 describes the operation and equation of butterfly and the rotational block.



$$Out_O = Inp_0 + Inp_1 \tag{3}$$

$$Out_1 = Inp_0 - Inp_1 \tag{4}$$



(a) (b) (c)

Fig. 3. Testing images - (a) Lena (b) cell (c) fruit [15]

$$Out_O = Inp_0\, k\, \cos\left[\frac{n\pi}{16}\right] - Inp_1\, k\, \sin\left[\frac{n\pi}{16}\right] \tag{5}$$

$$Out_O = -Inp_0\, k\, \sin\left[\frac{n\pi}{16}\right] + Inp_1\, k \cos\left[\frac{n\pi}{16}\right] \tag{6}$$

In VLSI design, multiplierless method is popular in low power consumption and area. In order to reduce the complexity and power consumption, the proposed architecture uses loeffler algorithm along with CSD and CSE operation. In CSD, the constant number multiplication is being replaced by adding and shifting operations. Coefficient p is a constant, The CSD is formulated as

$$p = \sum_{i=0}^{N-1} p_i 2^i \, where \, p_i = \{-1, 0, 1\} = \{-, 0, +\} \tag{7}$$

On the other hand, CSE improves the application of adders and shifters by recognizing the common expressions. With the help of CSD, CSE identifies the common sub expression (which means bit pattern occurs more than once in the given expression). By using CSE, common sub expression can be eliminated. This helps in reducing the resource utilization and power consumption in hardware design.

To analyze the execution and advantage of using CSE and CSD. From Fig. 2, let us take the theoretical evaluation of $Y(1)$

$$Y(1) = c(5) + c(8); \tag{8}$$

$$= b(5) + b(7) + b(6) + b(8); \tag{9}$$

$$\begin{aligned} &= a(5) * cos(3\pi/16) + a(8) * sin(3\pi/16) - a(6) * sin(\pi/16) + a(7) * \\ &cos(\pi/16) + a(6) * cos(\pi/16) + a(7) * sin(\pi/16) - a(5) * sin(3\pi/16) + \\ &a(8) * cos(3\pi/16); \end{aligned} \tag{10}$$

$$\begin{aligned} &= 27 * (sb1 + sb4) + 26 * (sb2) - 25 * (sb1 + sb3) + 23 * (sb1 + sb2 + sb3) \\ &+ 21 * (sb1 - sb4) - 20 * (sb2 + sb3); \end{aligned} \tag{11}$$

Where $sb1 = a(5) + a(8); sb2 = a(8) - a(5); sb3 = a(6) - a(7); sb4 = a(6) + a(7)$.

From the equation can be seen $sb1$ occurs 4 times, $sb2$ occurs 3 times, $sb3$ occurs 3 times and $sb4$ occurs 2 times. Thus by implementing CSD and CSE, power consumption and area can be reduced. Now the multipliers above can be replaced by the shift operations yielding

$$Y(1) = (sb1 + sb4) \ll 7 + (sb2) \ll 6 - (sb1 + sb3) \ll 5 + (sb1 + sb2 + sb3) \ll 3 + (sb1 - sb4) \ll 1 - (sb2 + sb3). \quad (12)$$

4.1 Simulation Results

An Loeffler algorithm for 2D-DCT is investigated along with Common Sub expression Elimination (CSE) and Canonical Signed Digit (CSD). The signal flow graph of loeffler algorithm is shown in Fig. 2. In the hardware implementation of the algorithm, it is difficult to design with floating point multiplication. To reduce the complexity of floating point multiplications, constants are used in this algorithm. Constants are multiplied by factor of 2 and the outcomes are divided consequently in order to remove the floating points. Multiplication and division can be easily implemented in hardware by shifting operation. To evaluate the performance in terms of quality, peak to signal noise ratio (PSNR) and mean square error (MSE) are calculated. The PSNR and MSE of a T and S monochrome image of size of $m \times n$ in which one image is considered as noise compare to other and is given as

$$MSE = \frac{1}{64} \sum_{i=0}^{7} \sum_{j=0}^{7} [T(i,j) - K(i,j)]^2 \quad (13)$$

$$PSNR = 10 \log_{10} \left(\frac{MAX_T^2}{MSE} \right) \quad (14)$$

MAX_T of the image represents maximum pixel value, where each pixel has 8 bits/sample i.e. 255 [14]. PSNR of a image or video compression ranges from 30 to 50 dB. To test the loeffler algorithm along with CSD and CSE, in this paper different images with extenstion .bmp format are considered.

In regard with the quality and the size of the image, Minimizing the bit length reduces image quality which inturn reduces the power consumption. There is a tradeoff between power consumption and the image quality, Based on the requirement and application, power consumption and image quality can be considered as either main or secondary aspects. In this paper, the testing images used for simulation and synthesis are same as referred in [15]. An Loeffler algorithm for 2D-DCT is investigated along with CSE and CSD is simulated and synthesized using Xilinx Spartan -3AN device. The synthesis report of FPGA device utilization and simulated PSNR values shown in Figs. 4 and 5.

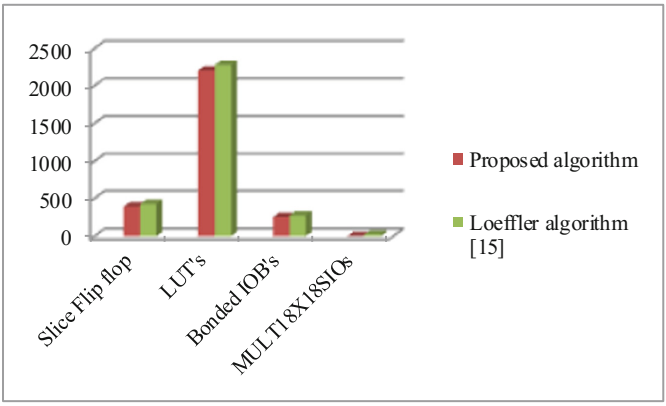


Fig. 4. FPGA device usage of proposed architecture using Loeffler algorithm along with CSE and CSD vs Loeffler algorithm [15].

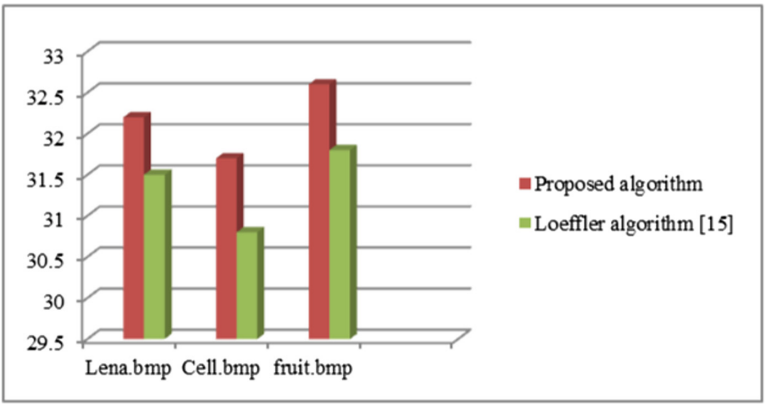


Fig. 5. PSNR plot of proposed algorithm vs Loeffler algorithm [15].

The images, which were considered in [15], the same images have been considered and the PSNR of the images have been evaluated in this paper. From Fig. 5, loeffler algorithm with CSE and CSD the PSNR is higher in comparison with [15]. From Fig. 4, the FPGA utilization is also less when in comparison with the [15].

5 Conclusion

For image compression, A 2- DCT using Loeffler algorithm along with CSD and CSE has been implemented. For fast computation 2-DCT/IDCT is executed by using 1D DCT row column method. The hardware implementation of the proposed architecture is simulated and synthesized using Xilinx Spartan -3AN device. From the results, it is observed that it consumes less area and high PSNR in comparison with other

technique. From the proposed architecture, the size parameter such as IOB's, Multipliers, flipflops, LUT's, occupied slices have been reduced in comparison with the other technique. The PSNR of different images of resolution 8 × 8 has been evaluated. In comparison with other technique, The PSNR values of the proposed techniques are much higher which defines sufficient quality for compression of low-resolution images.

Acknowledgment. Authors would like to acknowledge Dr. N. Vijayabaskara Chowdry, Correspondent, Madanapalle Institute of Technology and Science, Madanapalle.

References

1. Vijay Prakash, A.M., Gurumurthy, K.S.: A novel VLSI architecture for digital image compression using discrete cosine transform and quantization. IJCSNS **10**, 175 (2010)
2. Taher, F., Zaki, A., Elsimary, H.: Design of low power FPGA architecture of image unit for space applications. In: International Midwest Symposium on Circuits and Systems (MWSCAS). IEEE (2016)
3. Singh, K.K., Pandey, D: Implementation of DCT and IDCT based image compression and decompression on FPGA. In: International Conference on Inventive Systems and Control. IEEE (2017)
4. Park, J., Roy, K.: A low complexity reconfigurable DCT architecture to trade off image quality for power. Accessed 3 June 2008
5. Ahmed, N., Natarjan, T., Rao, K.R.: Discrete cosine transform. IEEE Trans. Comput. **23**(2), 90–93 (1974)
6. Wallace, G.K.: The JPEG still picture compression standard. IEEE Trans. Consum. Electron. **38**(1), 18–34 (1992)
7. Wiegand, T., Sullivan, G.J., Bjontegaard, G., Luthra, A.: Overview of the H.264/AVC video coding standard. IEEE Trans. Circuits Syst. Video Technol. **13**(7), 560–576 (2003)
8. Gall, D.L.: MPEG: a video compression standard for multimedia applications. Commun. ACM **34**(4), 46–58 (1991)
9. Hou, H.S.: A fast recursive algorithm for computing the discrete cosine transform. IEEE Trans. Acoust. Speech **35**(10), 1455–1461 (1987)
10. Chen, W.H., Smith, C.H., Fralick, S.C.: A fast computational algorithm for the discrete cosine transform. IEEE Trans. Commun. **25**(9), 1004–1009 (1977)
11. Yu, S., Swartziander, E.E.: DCT implementation with distributed arithmetic. IEEE Trans. Comput. **50**(9), 985–991 (2001)
12. Kim, B., Ziavras, S.G.: Low-power multiplierless DCT for image/video coders. In: Proceedings of IEEE International Symposium on Consumer Electronics, May 2009, pp. 133–136 (2009)
13. Singh, T.V.P.: Matlab implementation of baseline JPEG image compression using hardware optimized discrete cosine transform. Int. J. Eng. Sci. Inven. **3**(8), 47–53 (2014)
14. August, N.J., Ha, D.S.: Low power design of DCT and IDCT for low bit rate video codecs. IEEE Trans. Multimed. **6**(3), 414–422 (2004)
15. Martisius, I., Birvinskas, D., Jusas, V., Tamosevicius, Z.: A 2-D DCT hardware codec based on Loeffler algorithm. Elektronika IR Elektrotechnika (2011). (ISSN 1392)

Investigation of SAR Exposure Assessment in Vital Human Tissues at GSM Frequency

Stephen Jemima Priyadarshini(✉) and D. Jude Hemanth

Department of Electronics and Communication Engineering,
Karunya Institute of Technology and Sciences, Coimbatore, Tamilnadu, India
mailjemi@gmail.com, judehemanth@karunya.edu

Abstract. The human exposure assessment is dependent on the measurement parameter such as Specific Absorption Rate (SAR). SAR investigation highlights the biological impact of human health due to a radiation source. This paper proposes a novel investigatory method for SAR assessment on human tissues such as kidney, heart, lungs, trachea, and nerve. The radiation source is considered is Inverted F Antenna designed to operate at GSM frequency 900 MHz. The tissue structures are designed with equivalent properties. The experimental analysis of exposure on tissues is performed, and SAR is computed. The various response of tissues towards SAR is compared and analyzed with gain and operation characteristics.

Keywords: Specific absorption rate · GSM · IFA · Dielectric properties

1 Introduction

The exposure of Human with a radiating structure is inevitable for everyday life. There is a need for investigation of the response of human tissue with the radiating device. SAR is inevitable for the calculation of the amount of levels the tissue is prone to absorb under exposure conditions. The computation of SAR and estimation is expressed with reference to mass equivalence to 1 g or 10 g. The prominence of Inverted F antenna (IFA) is used in communication applications such as mobile phones and Laptops. The existing SAR measurements are based on monopole and dipole antennas, that includes drawbacks such inability to provide proper shielding and downside of its size dependence to the operating frequency [1]. In this paper, vital tissues such as heart, kidney, Lungs, trachea, nerve and spinal cord are studied. For analysis, a tissue structure is designed as a numerical model with its permittivity, mass density and, conductivity. The IFA is validated for 900 MHz and SAR interactions are analyzed and are computed between the tissue and antenna.

2 The Specific Absorption of Radiation

The rate at which an exposed Human body absorbs the radio frequency (RF) is measured as a dosimetry quantity is denoted by SAR. The absorption of the power by the tissue mass is expressed in units of watts per mass of Kilogram [2]. In the human

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 366–372, 2019.
https://doi.org/10.1007/978-981-13-5758-9_31

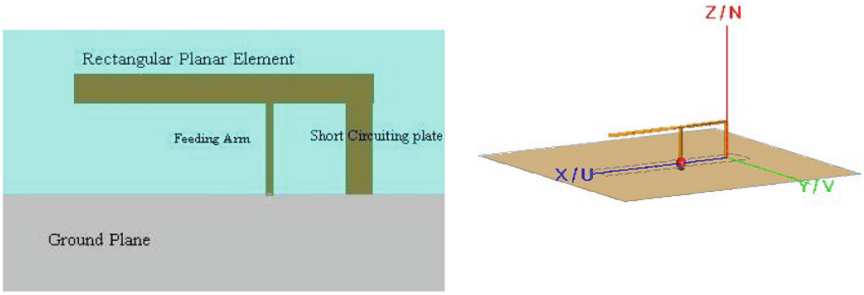
dosimetry studies, SAR is computed by several input quantities such as Power, Current density, change in temperature and Electric Field intensity. In this paper, the computation is based on electric field E(V/m). SAR is expressed by the following Eq. (1).

$$SAR = \frac{\sigma |E|^2}{\rho} (W/Kg) \tag{1}$$

In the above equation, E represents the electric field intensity at a point on the tissue mass (V/m). The σ is the conductivity of tissue mass at 900 MHz with units (S/m). The ρ is the density of the tissue mass equivalence in (Kg/m^3).

3 Inverted F Antenna

Inverted F antenna is a variation of a monopole antenna with a rectangular element which is planar and placed above the ground plane. A plate is purposed for short-circuiting, or pin is present as the intermediate arm and pivotal in the feeding mechanism for the planar element [3]. The wire of the planar element is bent to its length and placed at parallel on the ground substrate. The Antenna's height is reduced and resonant trace length is maintained. The pivotal feature is that its structural capability is to induce capacitance to the input impedance and the stub element do compensation. The inverted F antenna representation is displayed as Fig. 1(a) and validated model designed is shown in Fig. 1(b).



(a) Inverted F Antenna (b) IFA designed using FEKO

Fig. 1. (a) Inverted F antenna (b) IFA designed using FEKO

4 Numerical Modeling

4.1 Modeling Background

There is a strict limitation on experiments performed on the Human body for the exposure studies. The experimentation of SAR cannot be done on pragmatic human conditions; However, the investigation can be evaluated with equivalent numerical modeling in Software tools supporting electromagnetic field calculations. The SAR is

computed by different methods based on mathematical backgrounds, Finite domain time difference and Method of Moments are used widely in SAR computation [4]. In this experimental simulation Method of moments is used due to its simplicity in computation.

4.2 Geometry and Dimensions of IFA

The inverted F antenna is categorized as a linear antenna which can be further modified as plana antennas. The primary advantage of IFA is its capability to reduce backward radiation thus reducing the undesirable electromagnetic absorption on the exposed medium. The IFA also exhibits characteristics of enhancing bandwidth and performance of the radiation structure [5]. The electric field of IFA is dependent on its polarization impact of its combined vertical and horizontal polarization. The design and validation of IFA model is executed using FEKO based on the design values specific to 900 MHz is displayed in Table 1.

Table 1. The design values of IFA for 900 MHz

The design parameters	Design values (mm)
The Length and width of the planar arm (L_1&W_1)	23.8 and 2.5
The Length and width of the folded upper arm (W_2&L_2)	3 and 5.3
The Length and width of the feed arm (W_3&L_3)	0.85 and 5.3
The Feed offset (offset)	6.97

5 The Design of Tissue Model

The tissue structure is designed with a length of 1 mm and width of 0.3 mm. The dielectric parameter when applied to the designed structure, it behaves like an equivalent tissue model. The dielectric values for the particular frequency of 900 MHz are derived from the existing databases [6]. The specific dielectric properties such as conductivity σ of the exposed tissue over 900 MHz and relative permittivity ε_r are used for design and evaluation. For each vital tissue, the dielectric values are different and are shown in Table 2. The setup simulated for the experiment is shown in Fig. 2.

Table 2. Dielectric properties of vital tissues

Tissue	ε_r	σ
Heart	59.892	1.229
Kidney	58.675	1.392
Lung_Inflated	21.999	0.456
Lung_Deflated	51.423	0.857
Trachea	42.007	0.771
Nerve_Spinal_cord	32.530	0.573

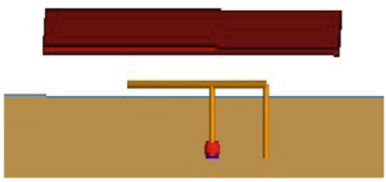


Fig. 2. IFA with tissue equivalent

6 Simulation Setup

The Design tools FEKO is ideal for this experiment and the simulation of computation uses a methodology known as Method of Moments (MOM) The distance between the antenna and the Tissue is kept constant with 0.1 mm for exposure analysis. When the antenna is placed in free space condition, the tissue equivalent is positioned within the exposure area of a field. The SAR interaction occurs between the tissue equivalent and the IFA.

7 Results and Validation

7.1 Operation Characteristics

S11 parameters are a vital parameter for estimation of the operation characteristics of an antenna. It is the total amount of power reflected by the IFA. It is represented in dB. If the reflection coefficient is 0 dB, there is no power radiation occurred. In this experimental setup, the S11 parameter is verified throughout for the operation frequency for each tissue equivalent is considered and also the operation frequency around 900 MHz. Return loss characteristics of all the tissues are tested in the experiment in which a sample operation characteristic of Heart is shown in Fig. 3.

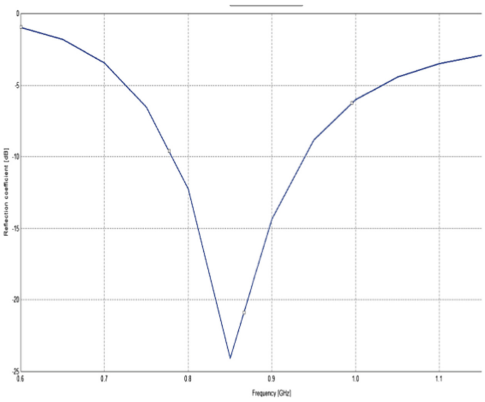


Fig. 3. The Return loss characteristics of Heart (dB)

Table 3. Reflection coefficient

Tissue	S11 (dB)
Heart	−25
Kidney	−25
Lung_Inflated	−19
Lung_Deflated	−28
Trachea	−32
Nerve_Spinal_cord	−23

The reflection coefficient S11 parameters of all the tissues are simulated, and the respective values are obtained and are given by the following Table 3. The above Table 3 indicates the operation in optimal conditions where experiment can be verified. S11 parameters show that the IFA provides sufficient electric field in SAR computation.

7.2 Far-Field in 3D

The setup also investigates the far-field radiation pattern of the tissue equivalent with gain. The far-field is the amount of radiation that can get around the antenna to a distance where the Electric field is exhibited [7]. The far-field in 3D is given in Fig. 4.

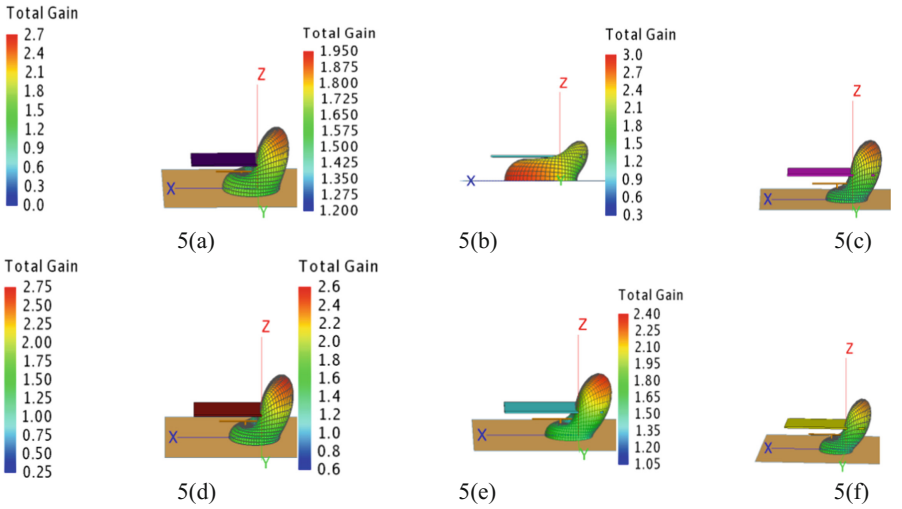


Fig. 4. 3D Farfield characteristics of (a) Heart, (b) Kidney, (c) Lung Inflated, (d) Lung deflated, (e) Trachea and (f) Nerve.

7.3 Gain and Near Field

The gain of IFA determines the extent of the capability of the radiating structure to convert the input power to the radio waves [5]. The gain of the tissue equivalents is compared in Fig. 5. The comparison shows the gain of equivalents is correlative within 3 dB of the IFA gain.

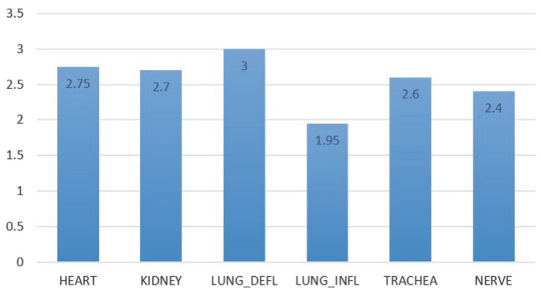


Fig. 5. The gain of the Tissue equivalents (dB)

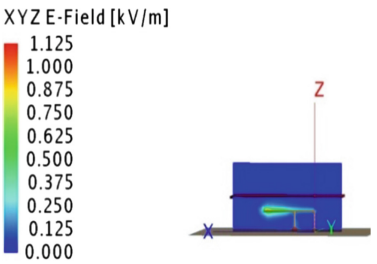


Fig. 6. Near field (KV/m)

Near field is the field around the IFA due to its scattering effect [8]. The Nearfield of the tissue equivalent is the basis for SAR computation is displayed in Fig. 6. The computed near field value is estimated to be 1.125 kV/m.

7.4 SAR Results

The SAR estimation of 1 g equivalence is an average of SAR maximum up to 1 g cube of the tissue mass. For SAR 10 g the total mass is calculated based on point over a tissue mass. In SAR 10 g, an average of SAR maximum is done by summation of each point on the tissue mass. For both SAR 1 g and 10 g the average of the tissue mass is results in a shape of a cubic structure with a dimension that is lesser than the actual cube. As a result, SAR 1 g and 10 g are not possible in computation [9]. Therefore, overall peak SAR is computed and compared for analysis. Figure 7 compares the obtained SAR values for the tissue equivalents. The results indicate each tissue sample varies in response to SAR at a particular frequency. Based on the comparison the Kidney Tissue has maximum SAR exposure of 45 mW, and the Lungs at the deflated position have minimum SAR value of 35 mW at 900 MHz. The Lung inflated has more SAR in comparison with Lung deflated position. Human trachea responds moderately to SAR with 37 mW/Kg. The Nerve which is considered in the set up is a sample of the spinal cord which has a SAR value of 36 mW/Kg. In this comparison Kidney and heart tissues are maximum affected by SAR.

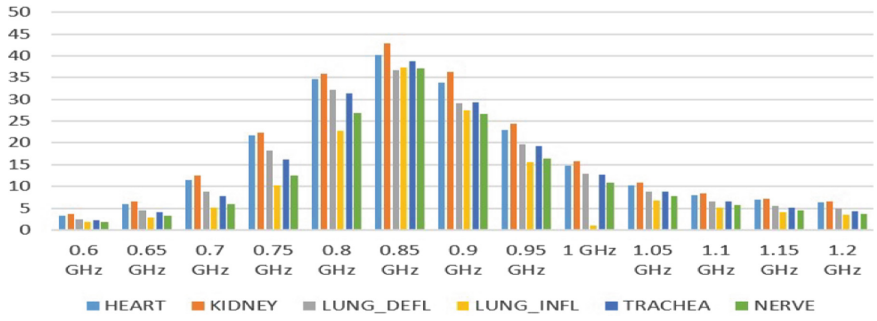


Fig. 7. SAR comparison of tissue equivalents (mW/Kg)

8 Conclusion

The SAR estimation of vital tissues is purposed for the examination of field distribution and effects of SAR impact due to IFA over 900 MHz. The various range between 6 to 12 MHz is further investigated to estimate the relative performance over other operating frequencies is studied. The homogeneous model is designed with the tissue properties of Heart, Kidney, Lungs, Trachea and Nerve. The essential conductive and permittivity properties are applied to the tissue equivalent model. Peak SAR and the Field parameters such as near field and Far-field are calculated for analysis. The evaluation of SAR is done by the MOM methodology that has been simple and more

efficient in computation than the Finite domain Time difference method [10]. The validation of the IFA and the tissue is simulated in CADFEKO [11]. FEKO is proved to be more reliable in computing electromagnetic problems involving scalability factors. It is observed that SAR absorption increases at the operating frequency of 900 MHz. This implied method can be further examined for other tissues for SAR analysis. A spectrum of tissue exposure analysis can be done and compared in detail for future work.

References

1. Simba, A.Y., Hikage, T., Watanabe, S., Nojima, T.: Specific absorption rates of anatomically realistic human models exposed to radiofrequency electromagnetic fields from mobile phones used in elevators. IEEE Trans. Microwave Theory Tech. **57**(5), 1250–1259 (2009)
2. Dautov, O.S., Zein, A.E.M.: Application of FEKO Program to the analysis of SAR on Human Head modeling at 900 and 1800 MHz from a Handset Antenna. In: IEEE International Symposium on Electromagnetic Compatibility and Electromagnetic Ecology, pp. 274–277 (2005)
3. IEEE Standard-1528: IEEE Recommended Practice for Determining the Peak Spatial-Average Specific Absorption Rate (SAR) in the Human Head from Wireless Communications Devices: Measurement Techniques (2003)
4. Abd-Alhameed, R.A., Excell, P.S., Mangoud, M.A.: Computation of specific absorption rate in the human body due to base-station antennas using a hybrid formulation. IEEE Trans. Electromagn. Compat. **47**(2), 3374–3381 (2005)
5. Product standard to demonstrate the compliance of mobile phones with the basic restrictions related to human exposure to electromagnetic fields (300 MHz–3 GHz) European Committee for Electrical Standardization (CENELEC), EN 50360, Brussels (2001)
6. https://www.fcc.gov/general/body-tissue-dielectric-parameters
7. Basic Standard for the Measurement of Specific Absorption Rate Related to Exposure to Electromagnetic Fields from Mobile Phones (300 MHz–3 GHz) European Committee for Electrical Standardization (CENELEC), EN-50361 (2001)
8. Procedure to measure the Specific Absorption Rate (SAR) in the frequency range of 300 MHz to 3 GHz—part1: handheld mobile wireless communication devices, International Electrotechnical Commission, committee draft for the vote, IEC 62209
9. Young, M.: Specific Absorption Rate (SAR) Estimation for Cellular Phone. Association of Radio Industries and businesses, T56,2002 (1989)
10. Jensen, M.A., Rahmat-Samii, Y.: EM interaction of handset antennas and a human in personal communications. Proc. IEEE **83**, 7–17 (1995)
11. http://www.feko.info/contact.html

Investigation of Fractality and Stationarity Behaviour on Earthquake

Bikash Sadhukhan(✉), Somenath Mukherjee, and Sugam Agarwal

Department of Computer Science and Engineering,
Techno India College of Technology, Kolkata 700156, India
bikash.sadhukhan@gmail.com, nitsoml0@gmail.com,
agarwalsugam.99@gmail.com

Abstract. In this paper, an investigation has been made to detect the self-similarity and stationarity nature of magnitude of occurred Earthquake by exploring the fractal pattern and the variation nature of frequency of the essential parameter, Magnitude of occurred earthquake across the different place of the world. The time series of magnitude (19.04.2005 to 07.11.2017), of occurred earthquakes, collected from U.S.G.S. have been analyzed for exposing the nature of scaling (fractality) and stationary behavior using different statistical methodologies. Three conventional methods namely Visibility Graph Analysis (VGA), Wavelet Variance Analysis (WVA) and Higuchi's Fractal Dimension (HFD) are being used to compute the value of Hurst parameter. It has been perceived that the specified dataset reveals the anti-persistency and Short-Range Dependency (SRD) behavior. Binary based KPSS test and Time Frequency Representation based Smoothed Pseudo Wigner-Ville Distribution (SPWVD) test have been incorporated to explore the nature of stationarity/non-stationarity of that specified profile where the magnitude of earthquake displays the indication of non-stationarity character.

Keywords: Earthquake · Hurst Parameter (H)
Visibility Graph Analysis (VGA) · Wavelet Variance Analysis (WVA)
Higuchi's Fractal Dimension (HFD)
Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test
Smoothed Pseudo Wigner-Ville Distribution (SPWVD)

1 Introduction

The shear stress and the shear strain are described in terms of forces produced in Earth's shell. The shear stress is defined as the force per unit area which is functional tangent to a body whether the shear strain is the alteration of a body formed due to shear stress. New advancement in space-based geodesy, GPS and Satellite-Interferometry produces a pure outline of strain buildup and crustal measure (Kanamori and Brodsky 2001). When the stress at point on the Earth's shell surpasses a critical value, an unexpected failure arises. The plane where letdown arises is known as the fault plane and the focus is define by a point where letdown starts. Naturally, an abrupt dislocation of the earth's shell near to the fault plane resulting a disaster that release elastic waves. This natural phenomenon is known as earthquake. In most of the

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 373–385, 2019.
https://doi.org/10.1007/978-981-13-5758-9_32

earthquake, the dislocation happens at a prevailing geological fault, specifically, a plane which is already weak.

Earthquakes are typical instances of composite scenario which reveal scale-invariance and fractality characteristics combinedly. Temporal, spatial and size distribution of earthquake has been taken in consideration during the research experiments of earthquakes, the property of fractality and scale variation nature are being exposed. As the Earthquake predicting is a demanding exploration topic so single prediction technique is not enough being the best. After long observation it is shows that the time series earthquake data, shows a composite pattern which is jointly an association of statistical parameters. The common goal of this kind of analysis is to differentiate the original frame of the earthquake movement in a seismic province, estimate dynamical progression by qualitatively and, quantitatively. Many researchers across the world are continuously being engaged to analyze the probability of occurrences of earthquake using the different methods and to design perfect predictive models which are discussed as follows:

A reasonable level of determination in the occurrences and strength, relapse of earthquakes in Indonesia may rise in the coming years (Priyadarshini 2016). The scaling frequency spectra are being flexible by spectacles of normal and constrained self-organized criticality, spectra of white and color rackets and rare change of Markov and non-Markov effects of long range memory during earthquake (Yulmetyev et al. 2001). The prospects of earthquake predictability research in order to realize practical operational forecasting in the near future has been analyzed (Ogata 2013). The statistics and soft computing techniques to analyze the earthquake have been focused (Preethi and Santhi 2011). Multifractal method has been used to understand the time-based dynamics of the current earthquake in the Corinth- Rift which indicates heterogeneity in clustering and correlations in every time scale that recommend non-Poissonian performance (Michas et al. 2015). Context of automatic predicting of the earthquake, appropriate for instantaneous earthquake monitor and an examination on how various characteristics of the data series are coupled with different prediction algorithms for the best conceivable precision has been proposed (Fong and Nannan 2011). A multi-fractal distribution can also be fitted to the observations using Bayesian techniques in earth quake prediction (Panduyos et al. 2013). Several variables processes having common contact and common attributes the study for earthquake occurrence has been devised (Dutta et al. 2013). Wavelet Transform Modulus Maxima method which designated qualitatively and quantitatively the complex temporal patterns of seismicity has been described (Enescu et al. 2004).

The endeavor has been taken to expose the scaling behavior and the time-based frequency representation of earthquake in different location of the world. Here magnitude of the earthquake dataset (19.04.2005 to 07.11.2017) has been chosen from U.S Geological Survey (https://earthquake.usgs.gov/earthquakes) which can be preserved as the signatory characteristic of any Earthquake events. Magnitudes are generally determined from dimensions of an earthquake's seismic waves as recorded on a seismogram. Based on the measurement and calculation of the type and component of the seismic waves, magnitude scales will vary. Different magnitude scales are essential because of differences in earthquakes, and in the purposes for which magnitudes are used. A concurrent study of Magnitude scale of Earthquake may give apparent scenario

of the Earthquake events, the strength of the earthquake events and the pattern of the Earthquake.

The scaling behavior of the said time series has been exposed by estimating the Hurst exponent of earthquake magnitude time series. Estimation of the Hurst exponent from experimental data leads a very important role in the research of processes which show properties of self-similarity. Several methods like Visibility Graph Analysis (VGA), Wavelet Variance Analysis (WVA) and Higuchi's Fractal Dimension (HFD) used to estimate the Hurst-Exponents to realize the statistical aspects of the signals in diverse scales and to classify whether the signaling system have the characteristics of fractional Brownian motion or fractional Gaussian noise. Hurst Exponent can be calculated using different methods but most of them have certain limitations. To find an indisputable decision about the nature of scaling of the said time series, it will be prudent to use several methods to evaluate the value of Hurst parameter. Three methods, as stated, are being selected to compute the Hurst parameter and thus approve the satisfactory accuracy of the decisions derived from the results.

The study of Earthquake would be incomplete without revealing the frequency fluctuation nature of earthquake magnitude time series data. A stationary signal has persistent frequency components while the non-stationary has ever changing frequency with respect to time.

Due to several the activities and properties of signal are severely distressed by the stationarity or non-stationarity so test of stationarity/non-stationarity is required for the signal. Different approaches are prevailing to test the stationarity/non-stationarity of the signals. The conventional technique and Time-Frequency Representation (TFR) method have been selected in this paper to identify the stationarity and non-stationarity character of that specified profile. Binary based KPSS systems has been used to detect a null hypothesis that an apparent time series is stationary around a deterministic trend (i.e. trend-stationary) against substitute of a unit root. Basically, TFR method is used to traces the period of presence of frequencies and identify the features of the frequency content within the signal. Time Frequency Representation Based Smoothed Pseudo Wigner Ville Distribution (SPWVD) method is being incorporated in this paper to reveal the scenario of frequency requirement of the magnitude of Earthquake. SPWVD methods are being used due to its advantageous characteristics like (1) As the method of the SPWVD is close to the Fast Fourier transform, it can merely be being employed for its intention. (2) Furthermore, the method produces a finite integration when run on a finite time series gesture. If the self-similar time series signal belongs to non-stationary, then it exhibits Fractional Brownian Motion (fBm) and if it explores stationarity then it is called Fractional Gaussian Noise (fGn).

2 Experimental Dataset

Primarily the magnitude of Earthquake (Occurred in different location of the world) datasets has been taken from U.S Geological Survey (https://earthquake.usgs.gov/earthquakes). Magnitude scales will vary based on the various kinds of the seismic waves which were measured and calculated. Several magnitude scales are required

because of differences in earthquakes, and in the purposes for which magnitudes are used. The various magnitude scales correspond to several ways of deriving magnitude. All magnitude scales preserve the logarithmic scale as devised by Charles Richter, are used to mid-range approximately which correlates with the original "Richter" scale. The earthquake of magnitude 5.7–5.9 is moderate which causes slight or no damage, the magnitude of 6–6.9 are strong may cause a lot of damage in very populated areas, magnitude of 7–7.9 are major which causes serious damage and magnitude 8 or more are great which can totally destroy communities near the epicentre.

So, in this work magnitude of Earthquake event whose values are greater than 5.7 and duration of 19.04.2005 to 07.11.2017 have been considered to investigate the scaling pattern as well as nature of frequency fluctuation of the earthquake.

The summary statistics of the experimental data sets are tabulated in Table 1 which basically provide the statistical outline of the signals under examination in this work and Fig. 1 depicts the plotting of dataset respectively.

Table 1. Summary statistics for magnitude of earthquake dataset.

Scores	Magnitude
Mean	6.2374
Median	6.1000
Mode	5.8000
Standard deviation	0.4745
Variance	0.2251
Maximum	9.1000
Minimum	5.7000
Skewness	1.6410
Kurtosis	6.2506

Table 1 indicates that the mean of the magnitudes is 6.2374 which is strong earthquake. The mean may not be a reasonable picture of the data, because the average is easily inclined by outliers. The median is another method to quantity the center of a numerical data set. The median is 6.1 which also indicates a strong earthquake. The mode is 5.8 which indicates that most often the earthquakes are nearly of magnitude 5.8 which is a moderate earthquake. The standard deviation of the dataset is low which specifies that the data points incline to be adjacent to the mean of the set and not banquet out over a broader range of standards. The variance signifies that how far are the magnitudes in the dataset from the mean. Skewness is typically defined as a degree of symmetry or asymmetry in a data set. Symmetric data set will have a skewness of 0. Kurtosis is a measure of the collective sizes of the two tails. It measures the amount of probability in the tails. The value is often related to the kurtosis of the normal distribution, which is equal to 3. Here the kurtosis is greater than 3, which means the dataset includes heavier tails than a normal distribution (more in the tails).

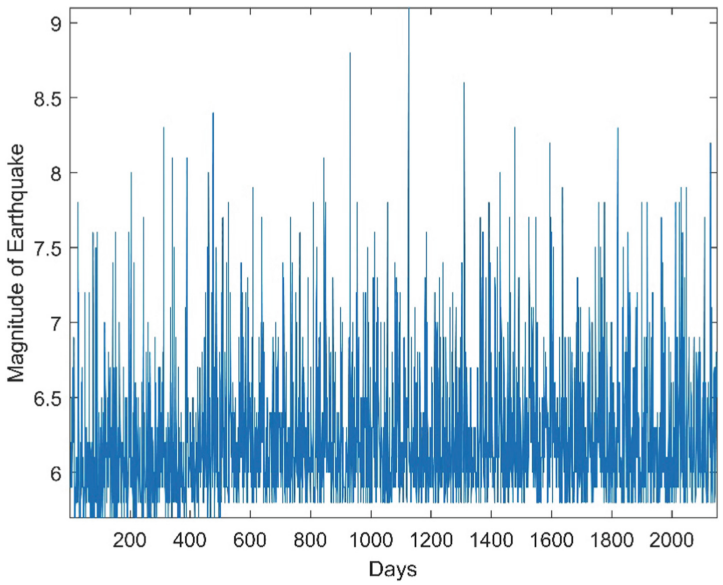


Fig. 1. Plot of magnitude of earthquake dataset.

3 Theory

3.1 Hurst Exponent Estimation

The Hurst exponent applied in numerous fields of applied mathematics, including fractals and chaos theory, long memory processes and spectral analysis. Approximation of the Hurst parameter was formerly developed in hydrology. Nevertheless, the current procedures for approximating the Hurst parameter originated from fractal concepts. It defines the autocorrelations concepts used in time series signal and the proportion at which it is being reduced as the lag between pairs of values rises. It reflects the comparative inclination of a signal either to revert powerfully to the mean or to gathering in a track. Hurst analysis exposes the scaling nature of a time series Y(i) of length (i) and the sampling interval (t). The fractal behavior of the signal is numerically signified through a Hurst parameter (H). The scale of this specified parameter is fluctuating between 0 and 1. Computing the standards of Hurst Parameter, the features within the signal can be predicted. At, $H = 0$ signifies the system having white noise whose autocorrelation function reduces fast with interval, value of H in the series of 0 – ½ specifies a signal with long term basis, converting among the high and low values in nearby sets, sense a high value will possibly be trailed by a low value and the next value at will tend to be high. The range $0 < H < 0.5$ specifies that a time series has negative autocorrelation and anti-persistent in nature i.e. increase in the value is followed by a decrease or vice-versa. The values change with respect to a mean, which indicates that new values possess an inclination to convert to a mean on long-term basis and when the range $0.5 < H < 1$ indicates that a time series has positive autocorrelation

i.e. persistent in nature. In a persistent time series, with an increase in values the short term increases a decrease in values will be followed by a decrease in the short term.

Among different estimators three Hurst estimation methods like Visibility Graph Analysis (VGA), Wavelet Variance Analysis (WVA) and Higuchi's Fractal Dimension (HFD) has been used to detect the persistency or anti persistency nature of the time series more accurately.

As VGA is independent on the fractal scale and the exponent of the power law degree distribution in the visibility graph is linearly related to the H, the VGA is initiated to be vigorous sufficient to invent long range dependency. Wavelet analysis has a benefit to decompose macroeconomic signal, and overall data into their time scale mechanisms. Higuchi's method has been preferred as it is efficient, highly accurate and not very sensitive to artifacts.

4 Methods used

Visibility Graph Analysis (VGA) Method: A visibility graph (Mukherjee et al. 2017; Lacasa et al. 2009; Albert and Barabasi 2002) is attained by mapping time series data into a network corresponding to the subsequent perceptibility standard: Two random data points (ta, ya) and (tb, yb) in the time signal have visibility, and therefore develop two linked nodes in the allied graph, if any other data (tc, yc) such that ta < tc < tb satisfies:

$$\mathrm{y}_{c} < \mathrm{y}_{a} + (y_b - y_a)\frac{t_c - t_a}{t_b - t_a} \tag{1}$$

Now points y_a and y_b is linked as two nodes to create a graph. In this manner the visibility graph is generated by tracing and connecting the perceptibility of similar data points of the supposed signal. the consistency allocation of the visibility graph determines the behavior of the signal. It means a consistent graph designates an interrupted sequence. however, an arbitrary graph suggests for an arbitrary signal. Correspondingly, a scale-independent graph is attained for a scaling sequences with the distribution of power law dependent, $P(x) = y^{-\alpha}$. The α exponent is well-defined as:

$$\alpha = 1 + N\left[\sum_{i=1}^{N} \log\frac{y_i}{y_{min}}\right]^{-1}. \tag{2}$$

Here, N is the total numbers of values,

$y_i\,(i = 1, \ldots\ldots\ldots, N)$ is the signal and y_{min} is the smallest value of y having power law behavior. α is linearly dependent on the Hurst Exponent (H)

$$H = 1 - \frac{\alpha}{2} \tag{3}$$

Wavelet Variance Analysis (WVA) Method: This traditional process (Percival and Guttorp 1994; Percival and Mondal 2012; Ray et al. 2015) can be illustrated as follows:

For time series $f(t)$, Continuous wavelet transforms, or coefficient, described as

$$W(t,a) = \frac{1}{a^{\frac{1}{2}}} \int_{-\infty}^{\infty} g^*_{t,a}(t') f(t') dt' \tag{4}$$

Where g is mother wavelet and g^* represents the complex conjugate of g. A scale or dilatation parameter with t-position or translation parameter; $g^*_{t,a}(t')$ is specified by:

$$g^*_{t,a}(t') = \frac{1}{a^{\frac{1}{2}}} g\left(\frac{t'-t}{a}\right) \tag{5}$$

A Scalo gram is defined by the plot of modulus square off $W(t,a)$ in $t-a$ plane

$$\text{Variance of } W(t,a), v(a) = E(W^2) - E(W)^2 \tag{6}$$

$v(a)$ follow a power law of a and specified by $v(a) \sim a^{\delta}$ if time series is linearly scaled, where δ represents exponent of wavelet variance.

δ is computed by double log plot of $v(a)$ versus a and its range is $-1 \leq \delta \leq 3$. For Fractional Gaussian Noise (FGN) signal, $-1 \leq \delta \leq 1$ and for Fractional Brownian Motion (FBM) $1 \leq \delta \leq 3$.

Wavelet Hurst Exponent (H_w) is represented in terms of δ as:

$$H_{\omega} = \begin{cases} \frac{\delta+1}{2} & -1 \leq \delta \leq 1 (FGN) \\ \frac{\delta-1}{2} & 1 \leq \delta \leq 3 (FBM) \end{cases} \tag{7}$$

Higuchi's Fractal Dimension (HFD) Method: Higuchi's method is a technique use to check the irregularity of time series directly. This procedure has been used to compute the fractal dimension D of a time series. This technique is proposed by Higuchi (Higuchi 1988). To observe the fractality, HDF is calculated (Gomez et al. 2009; Wairimu 2013) for the discrete data point series:

$$S : S(1),\ S(2),\ S(3),\ \ldots, S(n) \tag{8}$$

Where n is the data points number. From the original data series d new data series $S_{\kappa}(d)$ with $\kappa = 1,\ 2,\ 3,\ \ldots d$ are constructed. Where κ is the initial time and d is time interval.

$$S_{\kappa}(d) : S(\kappa),\ S(\kappa+d),\ S(\kappa+2d), \ldots \ldots S\left(\kappa + \left\lfloor \frac{n-\kappa}{d} \right\rfloor d\right) \tag{9}$$

$L_\kappa(d)$ is the length of the time series $S_\kappa(d)$ and define as:

$$L_\kappa(d) = \frac{1}{d}\left\{\left(\sum_{i=1}^{\frac{n-\kappa}{d}} |S(\kappa+id) - S(\kappa+(i-1)d)|\right)\frac{n-1}{\left\lfloor\frac{n-\kappa}{d}\right\rfloor d}\right\} \tag{10}$$

Where the term $\frac{n-1}{\left\lfloor\frac{n-\kappa}{d}\right\rfloor d}$ represents a normalization factor and $L_\kappa(d)$ is the normalized sum of the differences of values and calculated as follows

$$L(d) = \frac{1}{d}\sum_{i=1}^{d} L_\kappa(d) \tag{11}$$

Where $L_\kappa(d)$ is the mean value.

4.1 Detection of Stationarity/Non-stationarity

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test: Kwiatkowski–Phillips–Schmidt–Shin (KPSS) tests are used for testing a null hypothesis to check whether the observable time series is stationary or termed stationary or is non-stationary. This test is used as a complement to the standard tests in analyzing time series properties. The KPSS test is based on linear regression. The time series is broken down into three parts: a deterministic trend (βt), a random walk (r_t), and a stationary error (ε_t), with the regression equation:

$$X_t = r_t + \beta t + \varepsilon_t \tag{12}$$

If the data is stationary, it will have a fixed element for an interceptor the series will be stationary around a fixed level (Wang 2006). The test uses Ordinary Least Square (OLS) to find the equation, which differs slightly depending on whether you want to test for level stationarity or trend stationarity. A simplified version, without the time trend component, is used to test level stationarity.

Smoothed Pseudo Wigner-Ville Distribution (SPWVD) Method: The Wigner Ville Distribution method mainly highlights the energy distribution of the signal. The main goal of distributing energy of the signal over time and frequency is to increase the firmness of the time-frequency spectrum.

Wigner-Ville Distribution is described as:

$$W_x(t,f) = \int_{-\infty}^{\infty} z(t+\gamma/2)z^*(t-\gamma/2)e^{-j2\pi f\gamma}d\gamma. \tag{13}$$

Where z is an analytic signal which is derived from real signaling system.

Now, in term of spectrum WVD is defined as

$$W_X(t,f) = \int_{-\infty}^{\infty} Z(f+\varepsilon/2)Z^*(\omega-\varepsilon/2)e^{-j2\pi\varepsilon\gamma}d\gamma \tag{14}$$

Here * indicates a complex conjugate.

Now to eliminate the interference which represents cross terms or the frequency components which do not exist Cohen's class Distribution Method [32] is to be applied.

Class distribution of The Cohen's method is defined as,

$$C_x(t,v;f) = \iiint f(\epsilon,\gamma)e^{-j2\pi\varepsilon(m-t)}x(m+\gamma/2)x^*(m-\gamma/2)e^{-j2\pi v\gamma}d\varepsilon\ dm\ \ d\gamma. \tag{15}$$

Here $f(\epsilon,\gamma)$ is called 'kernel' because it is a two- variables parameterization function.

The Eq. 15 can be inscribed as

$$C_x(t,v,f) = \iint f(\varepsilon,\gamma)A(\epsilon,\gamma)e^{-j2\pi}(v\gamma+\epsilon t)d\varepsilon\ d\gamma \tag{16}$$

Where the ambiguity function can be defined as

$$A(\epsilon,\gamma) = \int x(m+\gamma/2)x^*(m-\gamma/2)dm. \tag{17}$$

If $f(\varepsilon,\gamma) = 1$, we get the Wigner-Ville distribution.

$$C_z(t,f) = \int z(t+\gamma/2)z^*(t+\gamma/2)e^{-j2\pi f\gamma}d\gamma \tag{18}$$

where z denotes the analytic part of the original signal.

The Fourier transform of the 'kernel' $f(\varepsilon,\gamma) = G(\varepsilon)h(\gamma)$ is given by

$$F(t,v) = FT[f(\varepsilon,\gamma)] = g(t)H(v). \tag{19}$$

The function $h(\gamma)$ denotes the frequency smoothing and the $g(t)$ indicates the temporal smoothing.

SPWVD can be found by replacing the kernel function in the general expression (15)

$$SPWVD = \int h(\gamma)\left[\int g(m-t)x(m+\gamma/2)x^*(m-\gamma/2)dm\right]e^{-j2\pi v\gamma}d\gamma \tag{20}$$

The following kernel function is to be chosen to filter out the interferences independently in the two axis directions and provides more flexibility to the smoothing operation.

$$f(\varepsilon,\gamma) = e^{-\left(\frac{\pi\varepsilon}{\sqrt{2}\sigma_\varepsilon}\right)^2} e^{-\left(\frac{\pi\gamma}{\sqrt{2}\sigma_\gamma}\right)^2} \tag{21}$$

Now substituting this expression of kernel in Eq. (20) we get

$$SPWVD(t,v) = \sqrt{\frac{2}{\pi}} \int e^{-\left(\frac{\pi\varepsilon}{\sqrt{2}\sigma_\varepsilon}\right)^2} Be^{-j2\pi v\gamma} d\gamma \tag{22}$$

Therefore, SPWVD is represented by its the discrete form

$$SPWVD(r,s) = 2\sqrt{\frac{2}{\pi}} \sum_{k=-h}^{h} e^{-\left(\frac{2\pi k}{\sqrt{2}\sigma_\gamma}\right)} \sum_{u=-g}^{g} \sigma_\varepsilon e^{-2(\sigma_\varepsilon u)^2} x(r+u+k)x(r+u-k)e^{-j4\pi sk} \tag{23}$$

5 Results and Discussion

The Hurst exponent values of specified profile, which are computed by the specified methods are being represented below in Table 2.

Table 2. Hurst parameter values for magnitude of earthquake

Methods	Hurst exponent (H)
VGA	0.3631
WVA	0.3776
HFD	0.2990

It represents the value of Hurst exponents is less than 0.5 which indicates the anti-persistency character of the specified time series. It also indicates that future values of that time series have the inclination to return to the value on long-term basis. As there is a probability for the profile to back again to its respective mean, it has been observed that some encouraging forces must be required which evoke the profile close to its mean when the profile depart from the mean. This signifies that some respond system must be functioning negatively which endlessly attempt to stabilize the systems. However, low values of H denote short-range dependency (SRD) which indicates the fractal behavior in short range basis for that specified profile.

A scaling distribution of minor fault-slip regions usually triggers short-period seismic waves which is recognized from the people and seismogram records where each zone radiates short-period waves. Even though plate tectonics demonstrated clear context for explaining the long-range progressions, but the short-range activation of the processes are still ambiguous, which makes earthquakes most unpredictable.

The logical value of binary based KPSS test which applied on the magnitude of earthquake dataset is equal to 1, which denotes the dynamic of earthquake is non-stationary in nature.

KPSS Test returns logical value 1 which exhibits that the system discards the trend of stationary null. So, the magnitude of Earthquake acknowledges the non-stationarity.

Figure 2 shows the result of Time Frequency Representation (TFR) based method SPWVD which has been applied on that specified data set for getting unarguable conclusion regarding the testing of stationarity/non-stationarity behavior of that specified profile. It undoubtedly indicates that the magnitude of Earthquake's frequency is changing with respect to the time. So, the magnitude of Earthquake is non-stationary in nature.

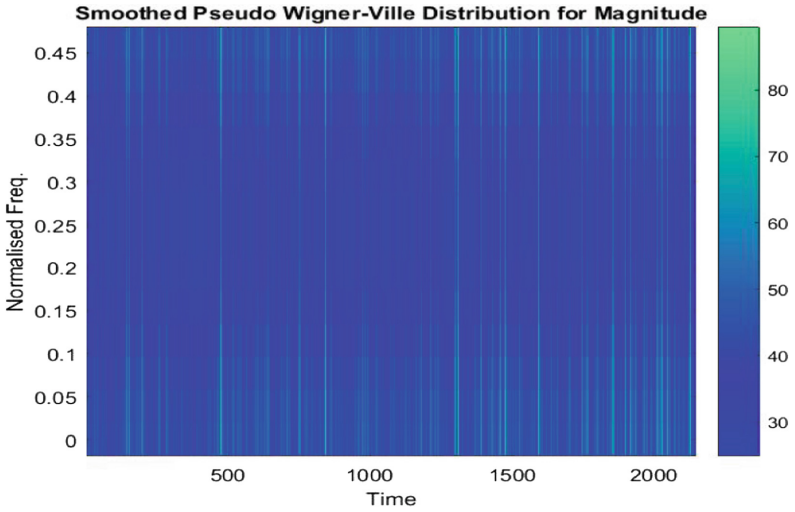


Fig. 2. SPWVD for magnitude of earthquake.

The frequency aspects which are the functions of time and having a dependency on the oscillation of the time is referred as non-stationarity of the signal. So, it can be concluded that the magnitude of occurred earthquake per unit time is dependent on time. It was found that the stress growth rate on earth crust is inconsistent over time. Big earthquake occurred on an area of a fault, alters the pressure in the nearby areas and increases/decreases the seismic activity, based on the fault geometry. The strength of the shell is variable over time. Moreover, earth's shell is significantly weakened by migrating fluids with altering times which accelerates earthquake. The stress fall at the timing of earthquakes also fluctuates from occurrence to occurrence.

6 Conclusion

It is believed that in a nonlinear system, Hurst Parameter is always belonging to greater than ½ or less than ½ but not equal to 1. As the Hurst parameter values of the magnitude of occurred earthquake are below ½, it can be resolved they are nonlinear in nature. Both binary based KPSS test and TFR based SPWVD test reflects non-stationarity nature of the signal. Moreover, the anti-persistent behavior gives outline of some non-positive response system which desires to be discovered further in the consequent works. So, it can be decided that the earthquake is not an arbitrary spectacle rather it is much more composite, non-linear and steady process. As process is long-term and consistent, but substantial progressive disparities of seismicity arise, which makes exact forecast of earthquakes very problematic. A time series is more important in the view of geophysics and economy problems. The noticeable understanding of earthquake and ancestors of earthquakes are used to generate an original method for forecasting the magnitude of forthcoming earthquake.

References

Albert, R., Barabasi, A.: Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 47–97 (2002)

Dutta, P.K., Mishra, O.P., Naskar, M.K.: A review of operational earthquake forecasting methodologies using linguistic fuzzy rule-based models from imprecise data with weighted regression approach. J. Sustain. Sci. Manage. **8**(2), 220–235 (2013)

Enescu, B., Struzik, Z.R., Ito, K.: Wavelet-based multifractal analysis of real and simulated time series of earthquakes. Annuals of Disaster Prevention Research Institute, Kyoto University, pp. 1–14 (2004)

Fong, S., Nannan, Z.: Towards an adaptive forecasting of earthquake time series from decomposable and salient characteristics. In: PATTERNS 2011: The Third International Conferences on Pervasive Patterns and Applications, pp. 53–60. IARIA (2011)

Gomez, C., Mediavilla, A., Hornero, R., Abasolo, D., Fernandez, A.: Use of the Higuchi's fractal dimension for the analysis of MEG recordings from Alzheimer's disease patients. Med. Eng. Phys. **31**(3), 306–313 (2009)

Higuchi, T.: Approach to an irregular time series on the basis of the fractal theory. Phys. D Nonlinear Phenom. **31**, 277–283 (1988). https://doi.org/10.1016/0167-2789(88)90081-4

Kanamori, H., Brodsky, E.E.: The physics of earthquakes. Phys. Today **54**, 34 (2001)

Lacasa, L., Luque, B., Luque, J., Nuno, J.: The visibility graph: a new method for estimating the Hurst exponent of fractional Brownian motion. EPL (Europhys. Lett.) **30001**, 1–5 (2009)

Michas, G., Sammonds, P., Vallianatos, F.: Dynamic multifractality in earthquake time series: insights from the Corinth rift, Greece. Pure. appl. Geophys. **172**(7), 1909–1921 (2015)

Mukherjee, S., Ray, R., Khondekar, M.H., Samanta, R., Sanyal, G.: Characterisation of wireless network traffic: fractality and stationarity. In: ICRCICN 2017, IEEE, pp. 79–83. IEEE, Kolkata (2017)

Ogata, Y.: A prospect of earthquake prediction research. Stat. Sci. **28**(4), 521–541 (2013)

Panduyos, J.B., Villanueva, F.P., Padua, R.N.: Fitting a fractal distribution on Philippine seismic data: 2011. SDSSU Multidiscip. Res. J. **1**(1), 50–58 (2013)

Percival, D.: Estimation of wavelet variance, pp. 619–631 (1995)

Percival, D., Guttorp, P.: Long Memory Process, the Allan Variance and Wavelets, pp. 1–15 (1994)

Percival, D., Mondal, D.: M-estimation of wavelet variance, pp. 623–657. Elsevier (2012)

Preethi, G., Santhi, B.: Study on techniques of earthquake prediction. Int. J. Comput. Appl. **29**(4), 0975–8887 (2011)

Priyadarshini, E.: An analysis of the persistence of earthquakes in Indonesia using rescaled range. Indian J. Sci. Technol. **9**(21), 1 (2016)

Ray, R., Khondekar, M.H., Ghosh, K., Bhattacharjee, A.K.: Memory persistency and nonlinearity in daily mean dew point across India. Theor. Appl. Climatol. **124**, 119–128 (2015)

Wang, W.: Stochasticity, Nonlinearity and Forecasting of Streamflow Processes. IOS Press, Amsterdam (2006)

Wairimu, M.J.: Features Affecting Hurst Exponent estimation on time series. Jomo Kenyatta University of Agriculture and Technology, Juja (2013)

Yulmetyev, R., Gafarov, F., Hanggi, P., Nigmatullin, R., Kayumov, S.: Possibility between earthquake and explosion seismogram differentiation by discrete stochastic non-Markov processes and local Hurst exponent analysis. Phys. Rev. E **64**(066132), 1–14 (2001)

Geometrically Uniform Differential Signalling for MISO/MIMO Channels Using Optimized PM Codebook

K. Susheela(✉) and Prerana G. Poddar

Department of Electronics and Communication Engineering,
BMS College of Engineering, Bangalore, India
susheela109@gmail.com, preranagp.ece@bmsce.ac.in

Abstract. In multiple antenna systems like Multiple Input Single Output (MISO) or Multiple Input Multiple Output (MIMO), the transmitted signal gets distorted due to various factors such as channel fading, multipath propagation, mobility and transmission bandwidth. Using the concepts of differential signalling and permutation modulation which were developed for wireline channels, a new signalling method is proposed in this paper for MIMO channel configurations. The BER performance of proposed system is systematically evaluated under different channel conditions, which includes Additive White Gaussian Channel (AWGN) and multipath Rayleigh fading conditions under MISO and MIMO configurations. The proposed system gives a performance better than conventional Alamouti space time coding scheme, with the same order of computational complexity. It also performs within the theoretical upper bound plotted under AWGN conditions.

1 Introduction

Applications like interconnecting integrated circuits or television to a set top box require wireline links for data communication. This may be done through single ended and differential signalling schemes using Permutation modulation (PM) codebook, which is generated by permuting the initial code vector to generate the entire codebook. The codes generated with this modulation are more efficient than any other known digital modulation scheme. In single ended signalling, each link has one dedicated wire and all links share the ground for return current. But in differential signalling, two wires are dedicated to each transmitter receiver link [1–4].

The wired AWGN channel contributes many other disturbances apart from Gaussian noise like common mode noise, simultaneous switching noise and Inter symbol interference (ISI). All these factors are reduced as follows: Common mode noise is reduced by considering signal components whose sum is equal to zero; simultaneous switching noise is reduced by using constant energy signals; and ISI by using only two amplitude levels at detection. The comparison in [4] shows that differential signalling has more advantages compared to single-ended signalling i.e. it inherently rejects many sources of noise and transmits small signal levels which allows the system to operate at relatively lower power levels. This explanation shows that, vector line coding scheme

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 386–397, 2019.
https://doi.org/10.1007/978-981-13-5758-9_33

is capable of overcoming all the impairments caused by noisy channel. Further the received signal detection is done using Maximum likelihood (ML) detector or hard decision decoding. The advantages of differential signalling scheme and orthogonal property of PM codebook motivates us to investigate the use of this scheme in wireless channel transmission.

Wireless communication systems face a lot of challenge due to large scale propagation effects as well as multipath propagation, mobility, etc. Multiple-Input-Multiple-Output (MIMO) systems have been widely used is recent years to provide a capacity gain and/or diversity gain in wireless fading channels. The fundamental transmit diversity scheme developed by Alamouti and many subsequent designs [5–7] have been used.

In this work, we use the concepts of PM codebook and differential signalling, and develop a signalling scheme for data transmission in wireless Multiple-Input-Single Output (MISO)/MIMO fading channels similar to the Alamouti's method. The proposed transmit/receive design is tested for Bit error rate (BER) performance over AWGN and Rayleigh fading MISO/MIMO environments, and is shown to outperform the existing signalling methods. The computational complexity of the proposed scheme will be same as Alamouti's signalling method.

2 Design of Optimized PM Codebook

A general scheme of vector line coding using PM codebook is shown in Fig. 1. For transmission of 'b' bits over $v \geq b$ wires, a line encoder is used to convert input binary symbols ± 1 to a code vector w with v real components, which is a one to one mapping from input $\{\pm 1\}^b$ to codebook $W \subseteq \mathrm{R}^w$. The generated codebook, which is a permutated form of initial vector w is called PM codebook, and it has an orthogonal property. The transmitted codeword will be added with AWGN and the received signal is represented as vector 'y'. The received signal 'y' is processed by a detector which outputs an estimate $\widehat{b}$ of the information vector [2].

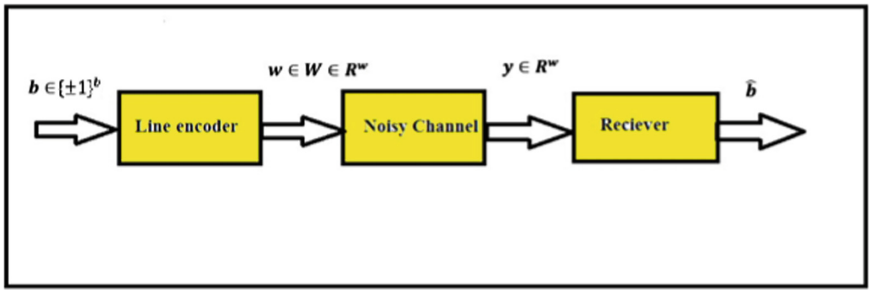


Fig. 1. General scheme for vector line coding

The details of PM codebook design for geometrically uniform differential signalling and its demodulation are discussed in this section.

To design a line encoder a differential signalling scheme is used, which minimizes common mode noise. The transmitted signal components are chosen such that their sum is zero. This means 'b' bit data with signal levels +1 and –1 are to be transmitted over 'b + 1' wires, and a code vector is generated, whose sum is zero.

In PM system, the codebook is chosen in special way by permuting the initial code vector w_1. To do this there are two different ways - Variant I and Variant II [2]. Among these, variant I is computationally simple and hence it is used in our work. Considering the initial vector,

$$w_1 = \{\overbrace{\mu_1, \mu_1, \ldots . \mu_1}^{m_1}, \overbrace{\mu_2, \mu_2 \ldots \ldots \mu_2}^{m_2}, \overbrace{\mu_k, \mu_k \ldots \ldots \mu_k}^{m_k}\} \tag{1}$$

The components in code vector consist of μ_1, μ_2 and μ_k repeated for m_1, m_2 and m_k times respectively. Here it is assumed that the code vector is sorted such that

$$\mu_1 < \mu_2 < \mu_k \tag{2}$$

And, the length of the initial vector is given by

$$n = m_1 + m_2 + \ldots . . m_k \tag{3}$$

where m's and k are all positive integers. The other codewords are obtained by permuting the order of $\mu' s$ in (1) in all possible ways. The possible number of code words is given by

$$\text{No of code words (z)} = n!/(m_1!\, m_2! \cdots m_k!) \tag{4}$$

Where, n = length of the initial vector initial vector w_1.

$m_1, m_2 . . . m_k$, = number of times the distinct component is repeated.

The codebook is represented in 3D view by considering the radius of the sphere which is a modulus of initial vector [3], and then using Peterson projection matrix, it is converted to 2D view for easy analysis [1].

$$A = \begin{matrix} [1+\beta & \beta & \beta. \ldots . & \ldots \beta & \gamma \\ \beta & 1+\beta & \beta. \ldots . & \ldots \beta & \gamma \\ : & : & : & : & \\ \beta & \beta & \beta & 1+\beta & \gamma \\ \gamma & \gamma & \gamma & \gamma & \gamma] \end{matrix} \tag{5}$$

Where, A is the $n \times n$ projection matrix

n = length of code vectors

$\gamma \triangleq 1/\sqrt{n}$

$\beta \triangleq -1/(\mathrm{n} - \sqrt{n})$

The line coding scheme is next reduced to a power of 2. If number of transmitted bits is *'b'*, then the length of codebook using PM should be 2^b. The codebook is reduced using Coxeter matrix groups, which is generated using b additional permutations (or root permutations) from initial vector w_1. This PM codebook is demodulated in receiver section using slicers or detection matrix [3].

We start with one of the design example to get the reduced PM codebook W, and it is used to design signalling scheme for MISO/MIMO channels.

Example 1: Consider b = 2 bits data are to be transmitted with signal levels '+1' or '–1', over b + 1=3 wires. The initial code vector satisfying the condition in (2) may be taken as w_1 = –1 0 1, have total 3 distinct components (–1, 0, 1) which are present one time each. Hence, the number of code words generated is calculated as z = n! / (m_1! m_2! m_3!) = 3! / (1! 1! 1!) = 6 code words. The permutated codebook is represented as,

$$P_m = \begin{matrix} -1 & 0 & 1 \\ -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{matrix}$$

To represent this PM codebook in geometrical approach the radius of the sphere is found to be $r = \sqrt{(1)^2 + 0^2 + (-1)^2} = \sqrt{2}$. By considering this radius the codebook 'P_m' vertices are plotted in 3D and 2D view (using Peterson Projection matrix) which is shown in Figs. 2 and 3 respectively. All the vertices of the codebook 'P_m' lie in a sphere of radius $\sqrt{2}$, and form the end points of a hexagon, which indicates the detection is simpler at receiver section. The geometrical approach helps to visualize the location of various points, and can be used to determine whether detection is simpler at the receiver (simple threshold based) or not.

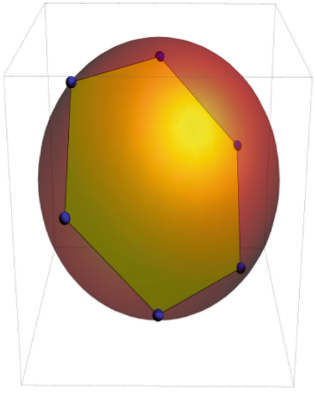


Fig. 2. 3D geometrical representation of PM codebook

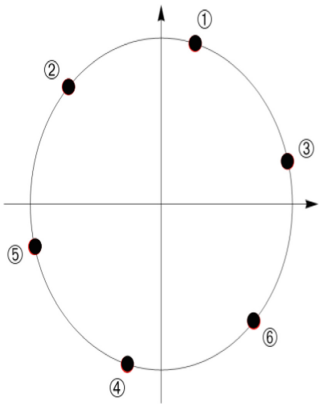


Fig. 3. 2D equivalent representation of PM codebook

For the PM codebook 'P_m', applying the concept of root permutations and reflection matrices, we reduce the PM codebook to W [3]

$$W = \begin{matrix} -1 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & -1 \\ 1 & -1 & 0 \end{matrix}$$

Based on the reduced codebook, a mapping can be created from the two bit input data to the codeword to be transmitted on 3 lines as shown in Table 1.

Table 1. Code vector generated for 2 bit binary data

Input binary data/signal levels	Code vector generated
00 (–1, –1)	−1 0 1
01 (–1, 1)	−1 1 0
10 (1, –1)	1 −1 0
11 (1, 1)	1 0 −1

The received signal can be detected using hard decision schemes and MLE. The hard decision and MLE schemes are explained below.

Hard decision decoding - The noisy signal which is received is given to a three level slicer with thresholds at –0.5 and +0.5 V. That is if the signal received is greater than 0.5 it is considered to be a '1', and if it is between –0.5 to 0.5 it is considered as '0', else it will be considered as '–1'.

Maximum Likelihood detection or Maximum Likelihood estimation (MLE) - MLE attempts to find the parameter values that maximize the likelihood function. Based on this value, received signal is detected which reduces bit error rate compared to Hard decision decoding [9].

For a given statistical model 'x' a maximum likelihood finds the values of the modeled parameter θ that maximizes the likelihood function which is given by,

$$\widehat{\theta} \in \{\arg\max L(\theta; x)\} \tag{6}$$

Theoretical BER in Wired AWGN channel - When the reduced PM codebook is used over an AWGN channel with zero mean and $N_0/2$ variance, showing independent and identical distribution on all wire channels, the probability of error at hard decision receiver is bounded as [3]

$$P_e \leq v\, Q(\alpha_{min}\sqrt{2\eta}) \tag{7}$$

Where,

P_e is bit error probability
$\alpha_{min} = min_j\, \alpha_{1,j}$ and $\alpha_{1,j} = \sqrt{b} \times (\|w_1 - w_j\|/(2 \times \|w_1\|))$, j = 2, ... $b+1$
W_j = neighbor vectors
v is number of $\alpha_{1,j}$ taking value α_{min}
Q is the complementary error function
η is the signal to noise ratio

3 Proposed Signalling Scheme for MISO/MIMO Channels

In this section, we propose two possible methods to transmit in MISO/MIMO wireless fading channel using the PM codebook. Our design uses the reduced PM codebook '*W*' from Sect. 2. The overall block diagram of the proposed system is shown in Fig. 4. The incoming binary data is encoded using permutation modulation and then converted to a form suitable for MISO/MIMO signalling. The receiver reverses the operations performed at the transmitter. The intermediate frequency/radio frequency stages, power amplifier and the low noise amplifier are essential to complete the system, but here only the baseband model is shown for simplicity.

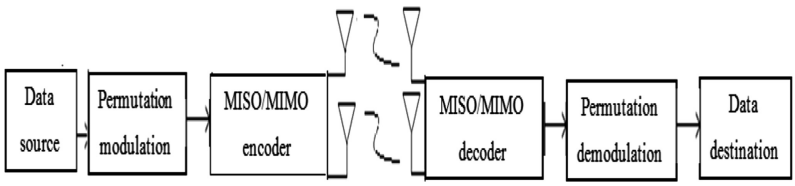


Fig. 4. Complete system block diagram

Method I:

In this method we start with the reduced codebook '*W*' and generate a new PM codebook as follows:

1. Place '*b*' code vectors which are non-orthogonal to each other as first '*b*' rows.
2. Fill the later rows with signals antipodal to the first '*b*' code vectors.
3. The generated new codebook is denoted as W_0. Now the reduced PM codebook '*W*' and the new PM codebook 'W_0' are read out column wise and arranged as a row vector, which is used for transmission using multiple antennas.

Continuing from Example 1 in Sect. 2, *b* is 2, hence, first two rows of codebook should be non-orthogonal vectors, and later rows are vectors antipodal to the first two rows. Based on the reduced PM codebook *W,* the new PM codebook is found as,

$$W_0 = \begin{bmatrix} -1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix}$$

Now the two codebooks are arranged in row vector form, the row vector corresponding to reduced PM codebook *W* is represented as TW_1 and row vector corresponding to new PM codebook W_0 as TW_2.

$$TW_1 = [-1 \quad -1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 0 \quad -1 \quad 1 \quad 0 \quad -1 \quad 0]$$
$$TW_2 = [-1 \quad 1 \quad 1 \quad -1 \quad 0 \quad -1 \quad 0 \quad 1 \quad 1 \quad 0 \quad -1 \quad 0]$$

The signals which are arranged in row vector form are transmitted using two transmitter antennas over fading channel using 2×1 MISO or 2×2 MIMO system, its sequence is shown in Table 2. The sequence of transmitted signals will be as below.

Table 2. Encoding scheme and transmission sequence using Method I

	TX Antenna 1	TX Antenna 2
Time 't'	$TW_1(l)$	$TW_1(l+1)$
Time 't + T'	$TW_2(l+1)$	$TW_2(l)$

In this table, the variable $l = 1, 3, 5, \ldots (2^b(b+1) - 1)$ occupies odd index values, and T denotes symbol duration.

The received signals of method I are used to create the following estimated variables at the receiver

$$\begin{aligned} \widetilde{TW_1}(1) &= h_1 y_1 + h_2 y_2 \\ \widetilde{TW_1}(l+1) &= h_2 y_1 - h_1 y_2 \end{aligned} \tag{8}$$

Where y_1, y_2 are the signals from received antenna at symbol durations 1 and 2 respectively, h_1, h_2 = channel fading co-efficient from Tx antenna 1 and 2 to the receiver respectively.

For a 2 × 2 MIMO system, the transmission scheme remains same. The receiver equations are modified as

$$\begin{aligned}\widetilde{TW_1}(1) &= h_1 y_1 + h_2 y_2 + h_3 y_3 + h_4 y_4 \\ \widetilde{TW_1}(l+1) &= h_2 y_1 - h_1 y_2 + h_4 y_3 - h_3 y_4\end{aligned} \tag{9}$$

Here h_3, h_4 are the channel fading coefficients from Tx antenna 1 and 2 to Rx antenna 2 respectively, y_3, y_4 are the received symbol values on Rx antenna 2 for symbol duration 1 and 2 respectively.

Method II:

The new PM codebook 'W_0' arranged in row vector form 'TW_2' is divided into two row vectors by assigning pairwise alternative symbols to TW_{11} and TW_{12}.

The new PM codebook row vector is repeated here for convenience,

$$TW_2 = [-1 \quad 1 \quad 1 \quad -1 \quad 0 \quad -1 \quad 0 \quad 1 \quad 1 \quad 0 \quad -1 \quad 0]$$

It is split to two sub vectors say TW_{11} and TW_{12} which is given by,

$$\begin{aligned}TW_{11} &= [-1 \quad 1 \quad 0 \quad -1 \quad 1 \quad 0] \\ TW_{12} &= [1 \quad -1 \quad 0 \quad 1 \quad -1 \quad 0]\end{aligned}$$

The sequence of transmitted signals will be as follows (Table 3):

Table 3. Encoding scheme and transmission sequence using Method II

	TX Antenna 1	TX Antenna 2
Time 't'	$TW_{11}(p)$	$TW_{11}(p+1)$
Time 't + T'	$TW_{12}(p+1)$	$-TW_{12}(p)$

Where again, $p = 1, 3, 5, \ldots\ldots((2^b(b+1)/2) - 1)$ and T denotes symbol duration.

The received signals of method II are used to estimate the transmitted signal as follows:

$$\begin{aligned}\widetilde{TW}_{11}(p) &= h_1 y_1 + h_2 y_2 \\ \widetilde{TW}_{11}(p+1) &= -h_2 y_1 + h_1 y_2\end{aligned} \tag{10}$$

Where y and h variables are as defined in Eq. (8). Likewise the equations can be found for 2 × 2 MIMO receiver also.

Equations (8–10) are derived based on the Alamouti's coding scheme in [7]. After estimation, these values can be given to a hard decision detector or further log likelihood detector for soft decoding. The detected bits are compared with transmitted bits to calculate the BER.

4 Results and Discussion

This section gives a detailed discussion of the simulation results and the performance comparison of our proposed methods with existing signalling schemes. The simulations are performed in MATLAB environment with different parameters and their values shown in Table 4. The channel scenarios considered are simple AWGN channel and Rayleigh fading single tap wireless channel (with 2×1 MISO and 2×2 MIMO antenna configurations.

Table 4. Summary of simulation parameters

Parameters	Value/Type
Antenna configuration	2×1 and 2×2
Modulation	Permutation modulation, Alamouti signalling
Detection scheme	Hard decision/Maximum likelihood
Channel	AWGN and Rayleigh Fading, 30 independent realizations

4.1 BER Performance in AWGN Channel

The BER performance of PM with differential signalling scheme is studied in AWGN channel, and the result is plotted in Fig. 5. We can find that the PM based system performs well in AWGN channel, and achieves a BER of 0.001 within 12 dB SNR. The performance achieved by using ML receiver is consistently 3–4 dB better than hard decision coding (in terms of received SNR). The blue curve labelled as AWGN theoretical is obtained using Eq. (7), for the given PM code as per [3]. We can observe that the curve serves as an upper bound to the performance of said receivers.

4.2 BER Performance in Rayleigh Fading MISO/MIMO Channel

The BER performance of PM is studied in wireless 2×1 MISO channel with differential signalling scheme. It was observed that the proposed methods I and II provide identical performance in MISO scenario. As shown in Fig. 6, the proposed PM methods with differential signalling scheme achieve a BER performance (shown as red curve) which is 3 to 4 dB better than the theoretical BER performance using Alamouti's coding (shown as black curve). The theoretical BER performance for Alamouti's coding is plotted using Eq. (11), which corresponds to the error performance of a two branch diversity scheme proposed in [6, 8].

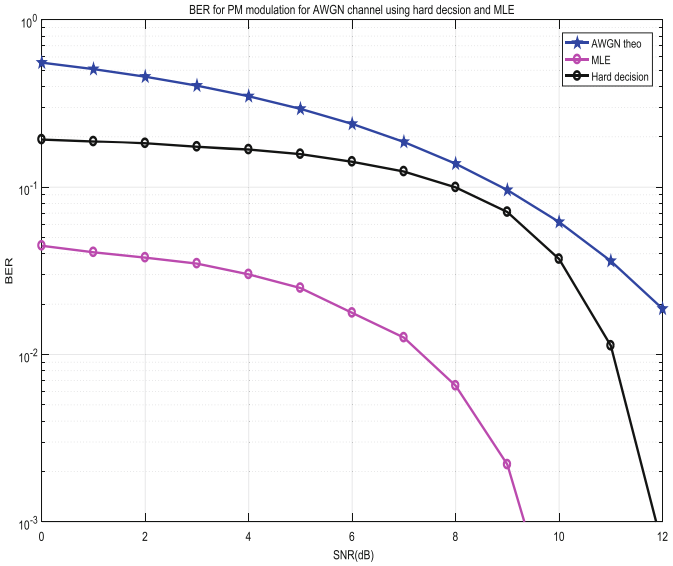


Fig. 5. BER performance using hard decision and MLE scheme

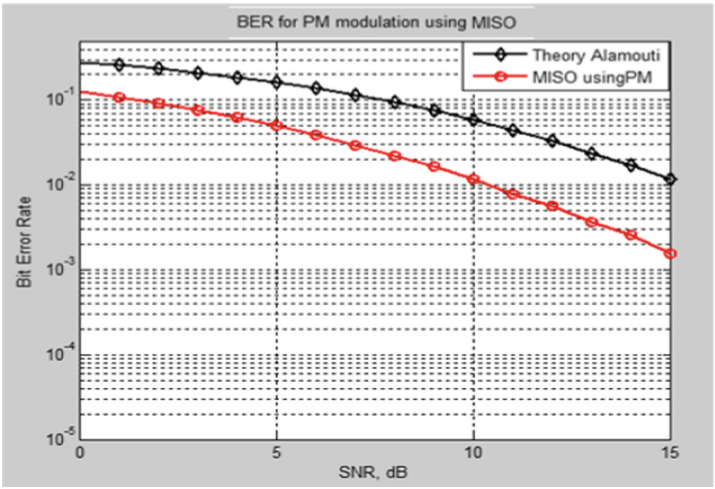


Fig. 6. BER performance on wireless channel using 2 × 1 MISO (Color figure online)

$$P_{e,STBC} = P_{STBC}^2(3 - 2P_{STBC})) \tag{11}$$

Where, $P_{STBC} = \frac{1}{2} - \frac{1}{2}\sqrt{(Eb/N0)/((Eb/N0)+2)}$

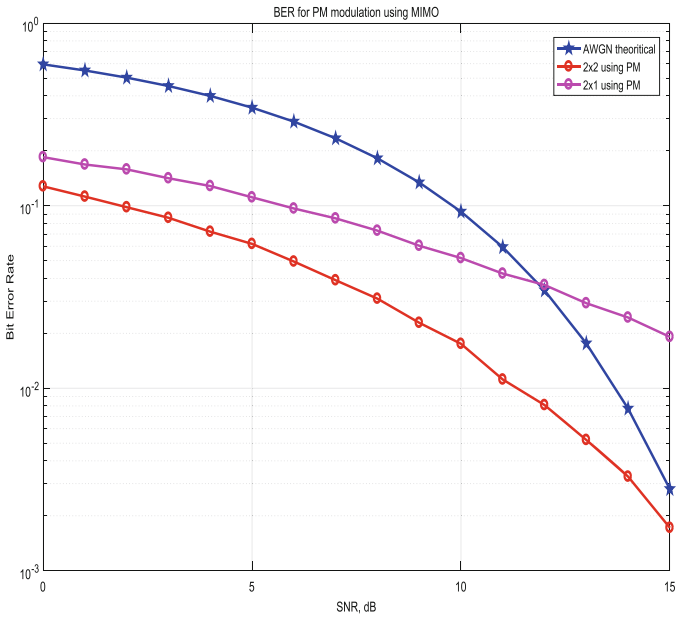


Fig. 7. BER performance on wireless channel using 2 × 2 MIMO

Next, the BER performance of PM with differential signalling scheme is studied in wireless 2 × 2 MIMO channel, and the result is shown in Fig. 7. On comparison, we can say that BER performance of PM in 2 × 2 MIMO is 2 dB better (in terms of SNR) than using in 2 × 1 MISO. In addition to this, we have plotted the theoretical upper bound of Eq. (7), which is derived in [3] for AWGN channel and used in Fig. 5 for comparison. The proposed PM based signalling methods are capable of achieving good error performance even in fading channels, which is within the upper limits set by Eq. (7) for AWGN channel.

5 Conclusion

This work combines the concepts of permutation modulation, differential signalling and space time coding for MIMO systems and proposes new methods for signalling over MISO/ MIMO channels. The proposed method is computationally not complex because the steps of PM codebook design can be performed one time and maintained as a lookup table. The receiver complexity is minimal when using hard decision (simple threshold detection), and as derived from Alamouti's work, ML detection is also achieved through linear processing at receiver. Future work requires generalizing and extending the proposed design for any larger MIMO configuration.

References

1. Biglieri, E., Viterbo, E.: Geometrically uniform differential vector signalling schemes. In: IEEE International Symposium on Information Theory (ISIT), pp. 506–510 (2017). ISSN 2157-8117
2. Biglieri, E., Viterbo, E.: Line coding for differential vector signalling. In: ITA Workshop, San Diego, CA, pp. 1–6 (2017)
3. Biglieri, E., Viterbo, E.: Line codes generated by finite Coxeter groups. IEEE Trans. Inf. Theor. (2017). Early access
4. Abbasfar, A.: Generalized differential vector signalling. In: IEEE International Conference on Communications (ICC 2009), Dresden, Germany, 14–18 June 2009
5. Tarokh, V., Seshadri, N., Calderbank, A.R.: Space–time codes for high data rate wireless communication: performance criterion and code construction. IEEE Trans. Inf. Theor. **44**(2), 744–765 (1998)
6. Alamouti, S.M.: A simple transmit diversity technique for wireless communications. IEEE J. Sel. Areas Commun. **16**(8), 1451–1458 (1998)
7. Tarokh, V., Jafarkhani, H., Calderbank, A.R.: Space–time block codes from orthogonal designs. IEEE Trans. Inf. Theor. **45**(5), 1456–1467, July 1999
8. Barry, J.R., Lee, E.A., Messerschmitt, D.G.: Digital Communication, Third Edn. Springer, Boston. https://doi.org/10.1007/978-1-4615-0227-2, ISBN 978-1-4615-0227-2
9. https://en.wikipedia.org/wiki/Maximum_likelihood_estimation

Validating Few Contemporary Approaches in Image Segmentation – A Quantitative Approach

Syed Fasiuddin(✉)

Blackbuck Engineers Pvt Ltd, Hyderabad, India
syedrvce@gmail.com

Abstract. In this paper, we present an extensive study and quantitative evaluation of six segmentation techniques on images of Berkeley Segmentation Database. Image segmentation plays a vital role in many computer vision applications and benchmarking such algorithms may assist research community in present and future research efforts in the field of image segmentation. Color space models, Hybrid color space and wavelet, Gradient Magnitude Techniques, K – means, C-Means & Fuzzy C-Means (FCM) and Edison's Mean – shift approaches are evaluated using at least six metrics with respect to ground-truth boundaries of entire images in BSD 300/500 dataset images. The results stated here gives useful insights to above mentioned approaches and its significance in aligning upcoming research avenues in image segmentation.

Keywords: Computer vision · Image segmentation · Quantitative analysis
Contemporary approaches

1 Introduction

Image segmentation defined as partitioning an image into several but distinctive regions has become prerequisite in the field of computer vision applications. Most of the segmentation algorithms are designed based on discontinuity and similarity [1]. Discontinuity can be measured in terms of abrupt changes in intensity levels defining edges, whereas similarity measures the degree of coherence level depending upon criteria like thresholding & region merging/splitting [2]. Different regions of color images can be obtained by solving difference equations to obtain normalized cuts [3]. Region variations are computed using graph theoretic algorithm and compared iteratively with its neighboring regions to divide image into its different parts [4]. Mean shift combined with normalized cut simplifies discontinuity property proving reduced complexity during segmentation process [5]. Graph reduction technique developed to reduce cost function which delineates thresholds by marking up intensities belonging to similar objects [6]. Peaks & valleys of histogram intensities can be adjusted to desired number of optimal clusters using Genetic Algorithm and wavelets [7].

Multiphase level set technique is used to generate multi threshold values depicting efficient class/regions in color image [8]. Gabor filters applied to color image on HSV plane preserves pixel level color features such as texture, maximum local energy,

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 398–409, 2019.
https://doi.org/10.1007/978-981-13-5758-9_34

maximum gradient and maximum second moment matrix and further classified using LS-SVM [9]. Local regions obtained after bottom-up region approach can be further represented based on global features in next level and allowed to score predicted values with the probability of region belonging to particular category using region scoring and SVM classifier [10]. Complex hybrid color space was derived after examining the performances of segmentation in different color spaces against k-means & FCM, its impact was quantitatively measured using widely held metrics and state of the art techniques [11].

Color homogeneity of image regions can be exploited by applying simple clustering techniques in different color spaces [16]. Edges are obtained using graph cut technique with min-max flow, classifier combining gradients, color & texture operators predicts the edge strength encoding user perception [17]. HSV color space is predominant in extracting pixel level color homogeneity almost equivalent to human observation of color and texture patterns given to LS-SVM classifier defining Arimoto entropy thresholding to find edges [9].

Local patterns obtained through Hough transform detector are further subjected to max-margin technique to find object center by assigning weights and SVM as classifiers to improve the performance [12]. In [13], a seed was initialized in region and allowed to grow and converge based on maximum density with locality and thereby removing regions which falls below some threshold. Quantitative evaluation of segmentation techniques is carried out in [14], revealing precision and recall as standard metrics to find the likeness between obtained boundaries through segmented images and ground-truth of BSDS 300/500 dataset. Few challenges such as, abrupt changes in image intensity levels and/or with more number of edges. Foreground and background exhibiting uni-modal histogram features sharing similar intensity levels. Difficulty in generating discriminative features. Requires much mathematical effort and computations are mentioned in [15].

However it is evident from the literature that, there is no standard metrics and ground truth datasets for a systematic evaluation and validation of image segmentation techniques. Visual comparisons of results obtained due to current segmentation techniques leads to perceptual conflicts and interest of locating the objects and image boundaries/contours to evaluate the performances and create benchmark. The goal of this paper is to thoroughly evaluate the performances of six different segmentation techniques considering standard metrics satisfying intuitive perceptions of segmentation quality and to provide significant advantage for improving the performances of imminent techniques in image segmentation.

This paper is organized as follows. Section 2 gives an overview of some contemporary approaches used for validation. In Sect. 3, experimental results are discussed. Section 4 provides performance evaluation and conclusion at the end.

2 An Overview of Contemporary Approaches

Before we progress to the analysis and evaluation sections, it is appropriate to concisely know about the segmentation techniques considered for evaluation. In this section, a brief review on contemporary approaches of image segmentation such as Color space

models, Hybrid color space and wavelet, Gradient Magnitude Techniques, K – means, C-Means & Fuzzy C-Means (FCM) and Edison's Mean shift is discussed.

2.1 Color Space Models

Color homogeneity of pixel intensities are studied by transforming color image into different color spaces and obtained hybrid color space model. Color space model is a geometrical representation of color in Cartesian coordinate's space; each coordinates representing colors selected from different color space models. Color spaces such as YCbCr, YUV, HSI & LUV are considered for transformation and wide-ranging observations pertaining to segmentation of homogeneous regions were made in single component analysis and derived hybrid color space models [11].

Impact of color spaces on segmentation can be realized by transforming original image into different color spaces and carefully selecting one or more components to find homogeneity in image regions. Figure 1 shows the original image and resultant segmented images obtained due to single component and multiple component analysis in different color spaces.

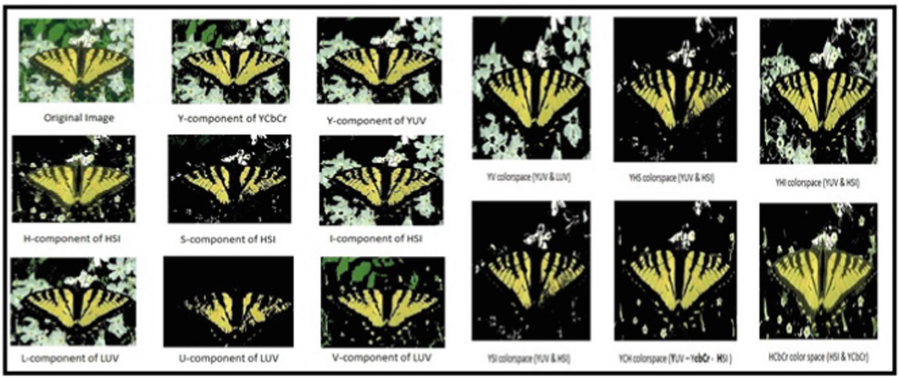


Fig. 1. Original and resultant segmented images obtained due to single and multiple component analysis of different color spaces.

2.2 Complex Hybrid Color Space and Wavelet

After a thorough analysis of color space models (as explained in above section), a complex hybrid color space model is derived. In this model, CbCr - componenets of YCbCr color space is merged with H-component of HSI model to obtain HCbCr and then transformed into LUV color space to construct complex hybrid color coordinates. Transformed image is subjected to multi-resolution analysis to separate foreground from background. This model has become a prerequisite in designing many of image retrieval, face recognition, object detection systems. Figure 2 shows an example image and segmented image using derived color space and wavelet, along with the canny edges [18].

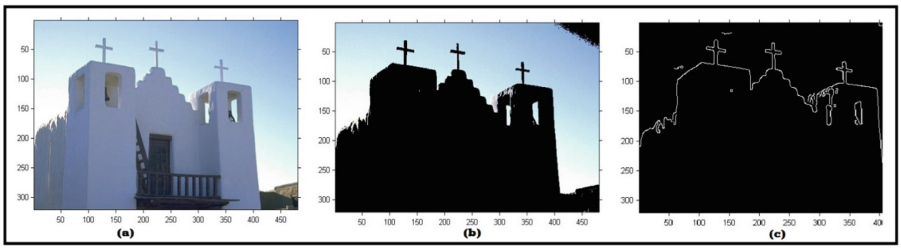


Fig. 2. (a) Original image (b) Segmented image (Complex hybrid color space and wavelet) (c) Canny edge

2.3 Gradient Magnitude Techniques

Gradients exhibit distinctive properties of revealing transitions between pixel intensities and detecting the probability of those pixels as boundaries or contours. We considered four such methods such as Gradient magnitude (pbGM), Multi-scale second moment matrix (pb2MM2), Canny edge detection (pbCanny), Multi-scale gradient magnitude (pbGM2) [19].

pbGM: The probability of the edge is high whenever gradient magnitude is high and gradient direction aligns itself perpendicular to the edge and can be calculated as follows:

$$\text{Gradient Magnitude } (\nabla f) = \sqrt{(\partial \mathrm{f}/\partial \mathrm{x})^2 + (\partial \mathrm{f}/\partial \mathrm{y})^2} = \sqrt{M_x^2 + M_y^2}$$

$$\text{Gradient Direction } ((\nabla f) = \mathrm{a\,tan2}\left(\mathrm{M_y}/\mathrm{M_x}\right)$$

$$\text{Where } \nabla f = \begin{pmatrix} \partial \mathrm{f}/\partial \mathrm{x} \\ \partial \mathrm{f}/\partial \mathrm{y} \end{pmatrix} \text{ is gradient of function '}f\text{'}.$$

pbGM2: Different scales are used to investigate and categorize image features. Boundaries can be obtained by labeling boundaries found due to watershed regions at different levels and identifying well defined edges.

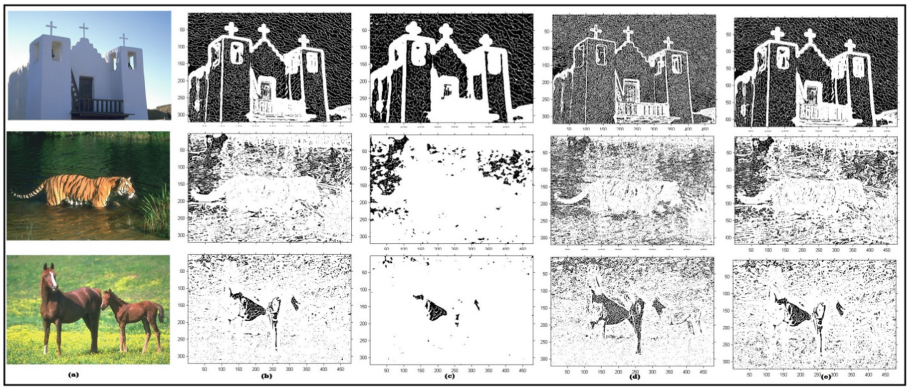


Fig. 3. (a) Original image (b) PbGM (c) Pb2MM2 (d) PbCanny (e) PbGM2

pb2MM2: Probability of detecting edges is more while analyzing gradients at different orientations and scales, hence name multi-scale second moment matrix will detect the boundaries observing the significant angle of directions in gradients and rational regions.

pbCanny: Canny outperforms many recent techniques due to its intrinsic noise removal and edge preserving features. Firstly, image is convolved with Gaussian function to smoothen the image. 1st order gradient is applied to obtain edges, finally non-maximal threshold and thinning is applied to get fine edges.

Figure 3 shows the boundary segmented images obtained due to gradients magnitude methods.

2.4 K – Means

K-means clusters image into 'k' disjoint regions or clusters as specified by the user. Some of the crucial steps are as follows.

1. Randomly select one seed point as cluster center μ_k and initiate threshold variable t_{nk}.
2. Declare objective function **F** as bias verification.
3. Start first phase, with an argument of minimizing the objective function 'F' w.r.t t_{nk} keeping μ_k constant.
4. Start second phase, with an argument of minimizing 'F' w.r.t μ_k and keeping t_{nk} constant.
5. Steps 3 & 4 are repeated for better convergence.

Figure 4 shows original image (1st column) and k-means segmented images (2nd column).

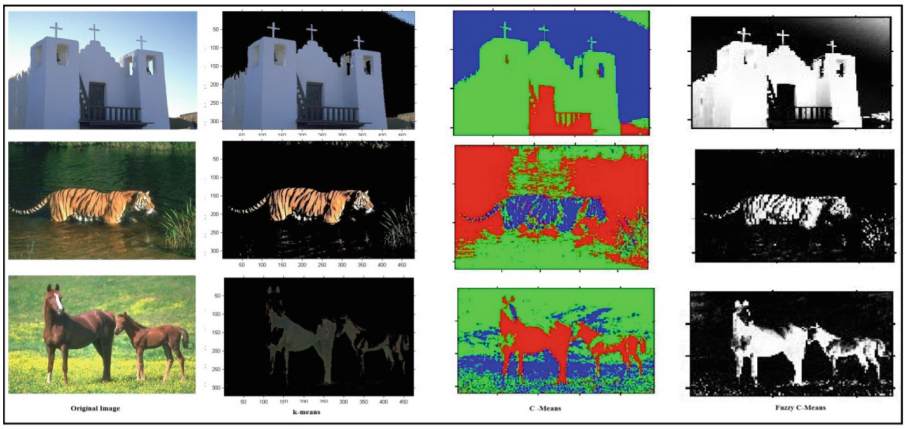


Fig. 4. Column-wise – Original image, K-means, C-means, FCM segmented images respectively

2.5 C-Means and Fuzzy C-Means (FCM)

C-Means: An n-dimensional gray scale image can be divided into "c" classes using histogram of pixel intensities in LUT color space. Important steps in c-means are as follows.

1. Histogram of image intensities is calculated.
2. Initialize centroids for each cluster.
3. Calculate distance between centroids.
4. Update cluster centroids.

Fuzzy C-Means: FCM is almost similar to C-Means but separated by fuzzy membership function introduced during centroid selection. Procedure is as follows:

1. Histogram of image intensities is calculated.
2. Initialize fuzzy membership & centroids.
3. Calculate distance between centroids and update with fuzzy membership values.
4. Update centroids and fuzzy membership values.
5. Repeat steps 3 & 4 for better convergence.

Figure 4 shows original image (1st column) and c-means & FCM segmented images (3rd & 4th columns) respectively.

2.6 Edison's Mean – Shift

The EDISON system consists of edge preserving filter algorithm and applied extensively for images sharing similar foreground and background pixel intensities. Feature vectors preserving pixel intensity values along with the coordinates are computed w.r.t each pixel.

Algorithm computes average mean of feature values at local neighborhood and weights are assigned measuring the distance between each pixel and center of local neighborhood pixel. Center is shifted to the newly calculated weights and procedure repeats, hence the name mean shift [14]. Figure 5 shows original image and segmented image obtained through mean-shift.

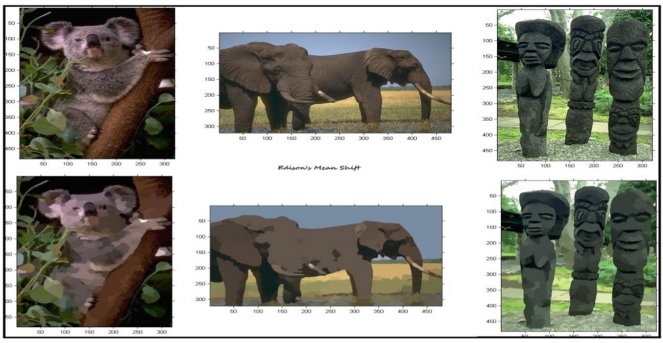


Fig. 5. 1st row: Original image (Berkeley dataset). 2nd row: Resultant image obtained using mean shift

3 Experimental Results and Discussions

All our experiments are carried out on a PC machine with an intel 5th gen (i5) 3 GHz processor and 8 GB RAM under Matlab-14 programming platform. In the following subsections we will discuss about: (i) Dataset and experimental procedure used in our experiments, (ii) Evaluation metrics (iii) Performance Analysis

3.1 Berkeley Dataset

Berkeley Segmentation Dataset (BSDS 300/500) consists of 300/500 images and is more challenging due to variety of image classes, high intensity variations, cluttered background and multiple objects [20]. Being motivated by these challenges and a fact that there are multiple ground-truth images available altogether for entire 300/500 images, we have considered this dataset to investigate the performances of above mentioned approaches.

Figure 6 shows few sample images of BSDS 300/500 dataset and its ground truth boundaries obtained from four different experts.

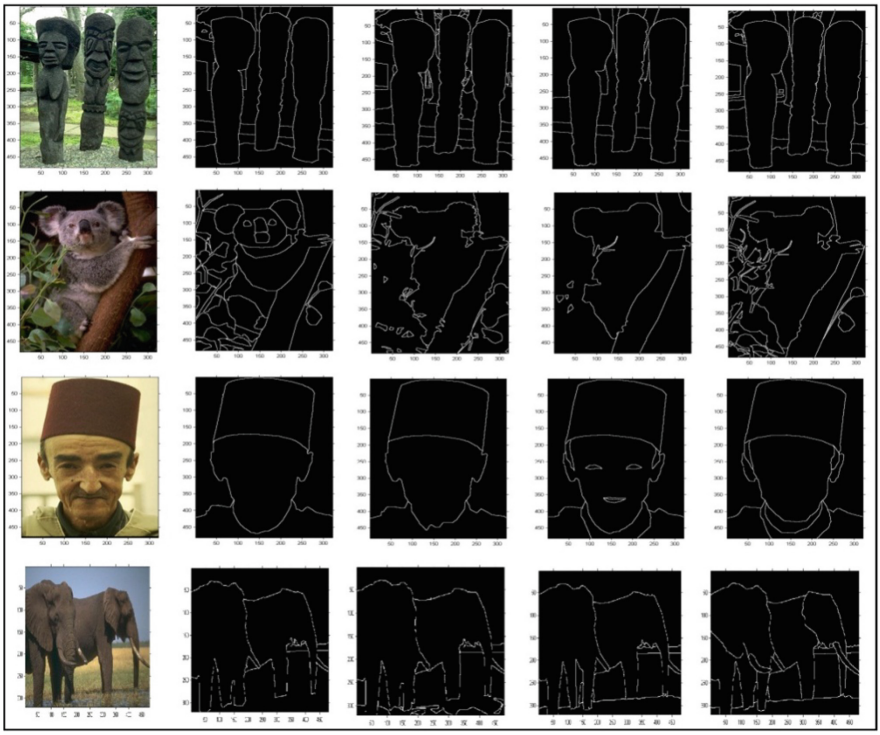


Fig. 6. Sample images of Berkeley dataset (Column – 1) and its ground-truth edges obtained from four different experts (Columns 2-5).

3.2 Evaluation Metrics

Evaluation process can be broadly classified into two broad categories of thoughts. Firstly, perceptual evaluation thought i.e. visually measuring the contribution of algorithms success in segmentation. Second, algorithms can be quantitatively measured by comparing it with appropriately derived ground-truth using some metrics.

In this paper we choose the later part of thought to evaluate the performances of segmentation approaches, edges obtained are compared with the ground-truth to measure the quality using metrics such as Jaccard & Dice coefficients, Hausdorff distance, Correlation, Mean Square Error (MSE), and Root Mean Square Error (RMSE) with following equations. Most of the metrics are un-biased and sensible in matching the boundaries with no compromise towards over/under segmentation.

Let *X* & *Y* be *(MXN)* matrices referring to edges obtained after canny edge detection on resultant images and ground-truth images respectively with m-rows & n-columns.

- Jaccard & Dice measures degree of similarity between any two data and lies in the range [0, 1] with higher values as closest match,

$$\text{Jaccard (X, Y)} = \frac{|X \cup Y| - |X \cap Y|}{|X \cup Y|} \qquad \text{Dice(X, Y)} = \frac{2||\text{XY}||}{|\text{X}| + |\text{Y}|}$$

- Hausdorff distance produces max value if X & Y possess high degree of similarity in shapes.

$$\text{H(X, Y)} = \max(\text{h(X,Y)}\&\text{h(X,Y)})$$

- Correlation in range [0, 1] defines the statistical relationship between X & Y.

$$\text{Correlation } r = \frac{\sum_m \sum_n \left(X_{mn} - \overline{X}\right)\left(Y_{mn} - \overline{Y}\right)}{\sqrt{\sum_m \sum_n \left(X_{mn} - \overline{X}\right)^2 - \sum_m \sum_n \left(Y_{mn} - \overline{Y}\right)^2}}$$ where $\overline{X}\&\overline{Y}$ are mean of *X* & *Y* respectively

- Mean Square Error & Root Mean Square Error signifies the error between two signals (X & Y) and produces floating point values in range of [0, 1], minimum value indicates better segmentation with less error.

$$\text{MSE} = \frac{\sum\sum (X - Y)^2}{m * n} \qquad \text{RMSE} = \sqrt{\frac{\sum_{i=1}^{h} (X - Y)^2}{n}}$$

3.3 Performance Analysis

Experiment of proposed algorithms is carried out on entire set of images in BSD 300/500 dataset and performance analysis is obtained using aforementioned metrics. To measure the quality of boundaries due to proposed approaches, average similarity

values computed between ground-truth and resultant segmented images by means of standard metrics are tabulated in Table 1.

Figures 7, demonstrates original images and resultant segmented edges along with its ground-truth, pbGM, pbGM2, pb2MM2, pbCanny (Row-wise). Figures 8, demonstrates resultant segmented edges along with its k-Means, C-Means, Fuzzy C-Means, Mean shift, Color space + Wavelet (Row-Wise). In regard to algorithms, depending on visual inspection it can be perceived that proposed color-space + wavelet edges give better confidence qualitatively over gradient magnitude & clustering techniques (k-means, c-means, FCM & mean shift).

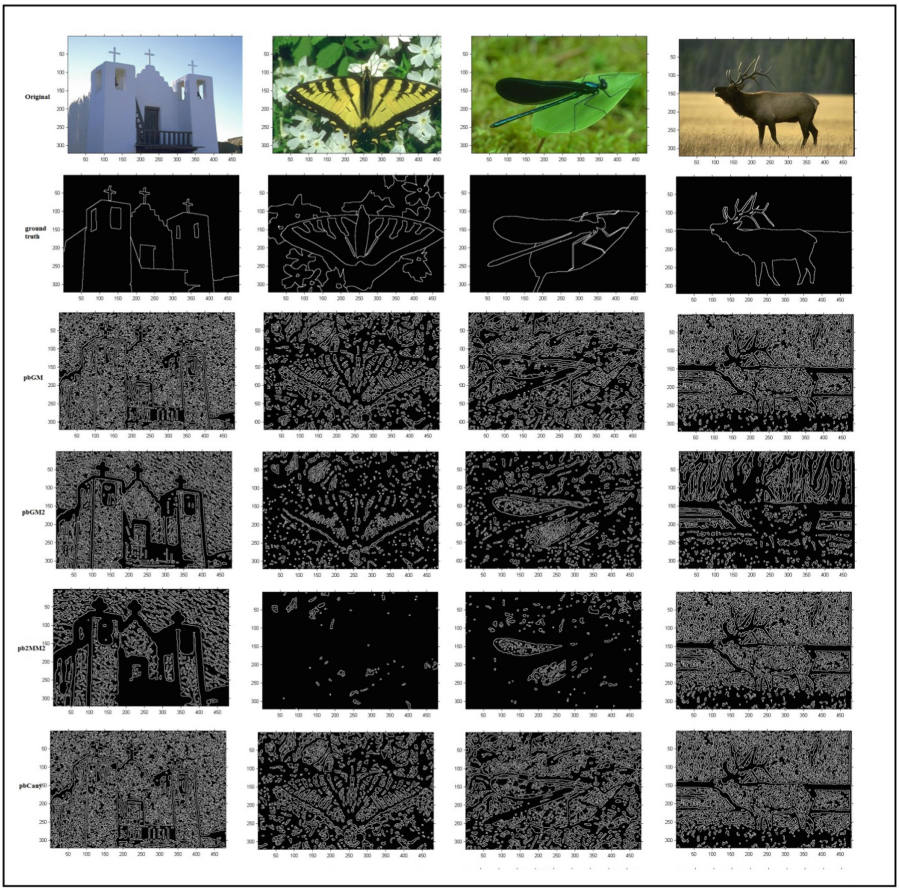


Fig. 7. Original image (Berkeley dataset) & segmented edges using: Ground truth; pbGM; pbGM2; pb2MM2; pbCanny (Row-wise).

They are also verified quantitatively and found similar compared to boundary detection & clustering techniques and briefed in next section.

From Table 1, it is evident that proposed color-space with wavelet has got significant contribution in improving the quality of segmentation. Color space selection is a prime factor in transforming image into different space so that perceptual information of objects can be preserved and better projected than regular color spaces.

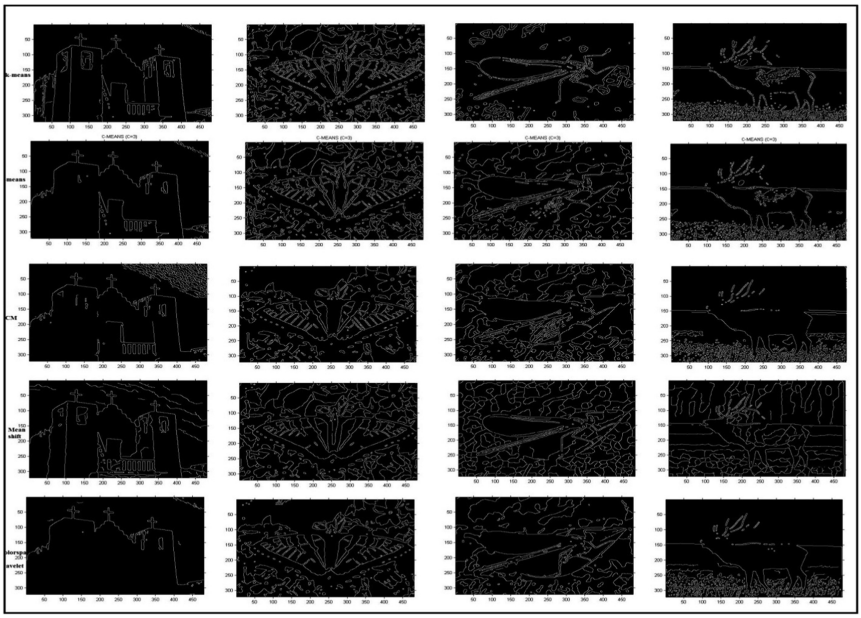


Fig. 8. Segmented edges using: k-Means; C-Means; Fuzzy C-Means; Mean shift; Color space + Wavelet (Row-Wise)

Table 1. Performance analysis

Method	Jaccard	Dice	Hausdorff	Correlation	MSE	RMSE
Edison's Mean Shift [13]	0.1150	0.2063	8.246	0.1652	0.0846	0.2808
k-Means [11]	0.2068	0.3127	10.046	0.174	0.0299	0.1815
C – Means [18]	0.2182	0.3812	9.645	0.168	0.0399	0.2015
Fuzzy C-Means [21]	0.2228	0.3128	9.762	0.1685	0.0742	0.2609
PbGM [22]	0.105	0.2786	9	0.1742	0.0654	0.1792
PbGM2 [23]	0.1046	0.2678	9.540	0.1876	0.0542	0.2036
Pb2MM2 [16]	0.176	0.2063	9.8	0.209	0.0338	0.1932
Pb Canny [24]	0.196	0.3061	10	0.212	0.0432	0.2639
Colorspace + k-means [11]	0.2768	0.3467	10.546	0.276	0.0329	0.1815
Colorspace + FCM [11]	0.4091	0.5675	13.1928	0.3146	0.0290	0.1704
Proposed Colorspace + Wavelet	0.4607	0.6043	14	0.5096	0.0137	0.0912

Note: Above tabulated results are the average similarity values computed between resultant edges (obtained due to proposed approaches) and ground-truth edges considering entire images in BSD 300/500 dataset

Proposed colorspace with wavelet reported high degree of shape similarity index (Hausdorff) along with maximum statistical dependencies (Correlation) and minimum errors, overlap ratio measured in terms of Jaccard & Dice received good appreciation with perceptually defined edges in comparison with other techniques.

4 Conclusion

The results offered in the paper concludes that metrics selected are most promising in measuring the quality of segmentation approaches and also offers close match to visual inspection/evaluation models and finally a significant evaluation. Color space models have got substantial impact during preprocessing stage leading to better image analysis and extended to segmentation. It is also carefully observed that direct application of clustering and magnitude techniques doesn't yield good results, but when applied onto images in transformed color space generates substantial edges.

The algorithms are evaluated on benchmark BSD 300/500 dataset and performance analysis is carried using efficient metrics w.r.t manual ground-truth edges. Compilation time and segmentation quality parameters give value addition to this paper. In future, this work can be used for quick and speedy evaluation of segmentation techniques in upcoming research avenues.

References

1. Fram, J.R., Deutsch, E.S.: On the quantitative evaluation of edge detection schemes and their comparison with human performances. IEEE Trans. Comput. **C-24**, 616–628 (1975)
2. Woods, R.E., Gonzalez, R.C.: Digital Image Processing. Prentice Hall, Upper Saddle River (2002)
3. Shi, J., Malik, J.: Normalized cuts and image segmentation. IEEE Trans. Pattern Anal. Mach. Intell. **22**(8), 889–905 (2000)
4. Huttenlocher, D., Felzenszwalb, P.: Image segmentation using local variation. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 98–104 (1998)
5. Yimin, Z., Wenbing, T., Hai, J.: Color image segmentation based on mean shift and normalized cuts. IEEE Trans. Syst. Man Cybern. Part B Cybern. **37**, 1382–1389 (2007)
6. Siskind, J.M., Wang, S.: Image segmentation with ratio cut. IEEE Trans. Pattern Anal. Mach. Intell. **25**, 675–690 (2003)
7. Siarry, P., Hammouche, K., Diaf, M.: A multilevel automatic thresholding method based on a genetic algorithm for a fast image segmentation. Comput. Vis. Image Underst. **109**, 163–175 (2008)
8. Diaf, M., Siarry, P., Dirami, A., Hammouche, K.: Fast multilevel thresholding for image segmentation through a multiphase level set method. Sig. Process. **93**, 139–153 (2013)
9. Wanga, Q.-Y., Yang, H.-Y., Wang, X.-Y., Zhang, X.-J.: LSSVM based image segmentation using color and texture information. J. Vis. Commun. Image R. **23**, 1095–1112 (2012)
10. Malik, J., Arbelaez, P., Bourdev, L.: Semantic segmentation using regions and parts. In: CVPR, pp. 3378–3385. IEEE (2012)

11. Mahantesh, K., Aradhya, V.N.M., Naveena, C.: An impact of complex hybrid color space in image segmentation. In: The Proceedings of 2nd International Symposium on Intelligent Informatics (ISI13), Mysore, India, vol. 235, pp. 73–82 (2013)
12. Malik, J., Maji, S.: Object detection using a max-margin hough transform. In: CVPR, pp. 1038–1045. IEEE (2009)
13. Meer, P., Comaniciu, D.: Mean shift: a robust approach toward feature space analysis. IEEE Trans. Pattern Anal. Mach. Intell. **24**(5), 603–619 (2002)
14. Jepson, A.D., Estrada, F.J.: Benchmarking image segmentation algorithms. Int. J. Comput. Vis. **85**, 167–181 (2009)
15. Blake, A., Rother, A., Brown, M., Perez, P., Torr, P.: Interactive image segmentation using an adaptive GMMRF model. In: European Conference on Computer Vision, pp. 428–441 (2004)
16. Meer, P., Comaniciu, D.: Robust analysis of feature spaces: color image segmentation. In: IEEE Computer Vision and Pattern Recognition, pp. 750–755 (1997)
17. Fowlkes, C., Martin, D., Malik, J.: Learning to detect natural image boundaries using local brightness, color and texture cues. IEEE-PAMI **26**, 530–549 (2004)
18. Mahantesh, K., Aradhya, V.N.M., Niranjan, S.K.: Coslets: a novel approach to explore object taxonomy in compressed DCT domain for large image datasets. In: El-Alfy, El.M., Thampi, S.M., Takagi, H., Piramuthu, S., Hanne, T. (eds.) Advances in Intelligent Informatics. AISC, vol. 320, pp. 39–48. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-11218-3_5
19. Mahantesh, K., Aradhya, V.N.M., Sandesh Kumar, B.V.: Benchmarking gradient magnitude techniques for image segmentation using CBIR. In: Prasath, R., Vuppala, A.K., Kathirvalavakumar, T. (eds.) MIKE 2015. LNCS (LNAI), vol. 9468, pp. 259–268. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-26832-3_25
20. Manjunath, B.S.: Image browsing in the Alexandria digital library project. D-Lib Magazine (1995). http://www.dlib.org/dlib/august95/alexandria/08manjunath.html
21. Yanga, H.-Y., Bu, J., Wanga, X.-Y., Zhanga, X.-J.: A pixel-based color image segmentation using support vector machine and fuzzy c-means. Neural Netw. **33**, 148–159 (2012)
22. Fowlkes, C., Maire, M., Arbelaez, P., Malik, J.: Using contours to detect and localize junctions in natural images. In: CVPR, pp. 1–8. IEEE (2008)
23. Fowlkes, C., Malik, J., Arbelaez, P., Maire, M.: Contour detection and hierarchical image segmentation. IEEE PAMI **33**, 898–916 (2011)
24. Yu, S.X.: Segmentation induced by scale invariance. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 444–451 (2005)

A Comprehensive Review on Automatic Diagnosis of Diabetic Maculopathy in Retinal Fundus Images

I. S. Rajesh[1(✉)], M. A. Bharathi[1], and Bharati M. Reshmi[2]

[1] Department of CSE, BMSIT & M, Bengaluru, Karnataka, India
is.rajesh081@gmail.com
[2] Department of Information Science and Engineering, BEC, Bagalkot, Karnataka, India

Abstract. Diabetic Maculopathy (DM) is one of the major problems of diabetes mellitus and it is one of the key reasons for the vision problem. It arises due to the leakage of blood from injured retinal veins. The development of DM is moderate and soundless and it is found in 10% of the world diabetic population. If diabetic maculopathy is not noticed in the underlying stage the effect this on macula is irreversible and can prompt vision loss. Therefore, screening of diabetic eye helps in finding diabetic maculopathy at the beginning stage which prevents the vision loss. In this review paper, the anatomy of the human eye and a brief overview of diabetes, diabetic retinopathy and diabetic maculopathy is presented. The literature review of various methods/techniques used for detection of DM in retinal fundus images and the performance metrics used to measure these methods are discussed in details. Issues involved in DM detection are also mentioned in this paper.

Keywords: Retinopathy (DR) · Diabetic maculopathy (DM) · Optic disc (OD) Hard exudates (HEs) · Blood vessels (BVs)

1 Introduction

The human eye is a fundamental body part associated with vision [1]. Figure 1 demonstrates the retinal anatomical structure. The lens of the human eye helps in focusing light beams onto the retina and iris decides how much light is let into the eye. Optic nerve associates the eye to the cerebrum and conveys the electrical signals framed by the retina to the visual cortex of the mind.

The retina is a sensitive layer situated at the posterior of the eye. Uncountable photoreceptors present in the retina catch the light beams and change them into electrical signals. These electrical signals drive to the brain along the optic nerve where they changed into pictures. Figure 2 exhibits the structure of the retina and its primary parts. The optic nerve is the brightest region of the retina where Blood Vessels (BVs) begin. BVs are responsible for supply of nutrition and oxygen to the retina and it must be normal for the proper working of the retina and macula. The central portion of the macula is called fovea [2].

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 410–420, 2019.
https://doi.org/10.1007/978-981-13-5758-9_35

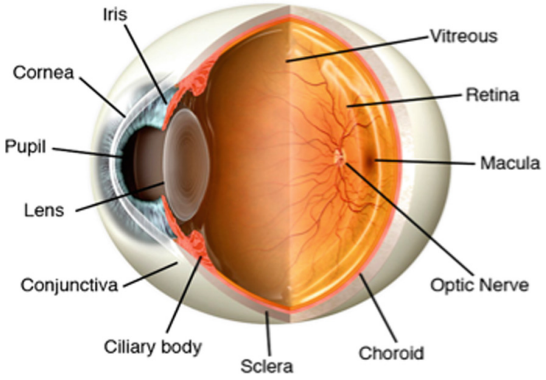


Fig. 1. Retinal anatomical structure

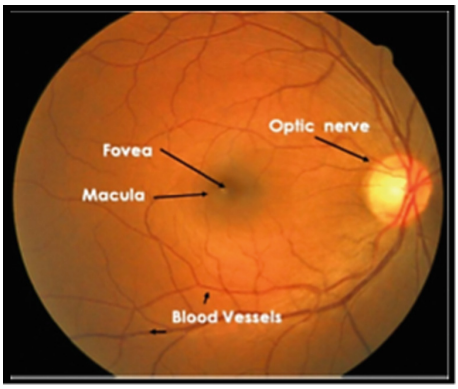


Fig. 2. Structure of retina and its main components

Optic Disc (OD): OD is one of the main anatomical structure that is visible in the color retinal fundus image. It is the most critical retinal landmark to be detected. In typical states of the retina OD is shown as a moderately round yellowish circle brighter than its surroundings. A significant part of the retinal image analysis is accurate identification of OD. OD detection is performed only on the red component because of blood vessels interfere OD detection in green the component [5]. Figure 3 shows optic disc and its appearance.

The OD is the initial point and also an access point for the major BVs. The optic disc might be round, however it is normally oval in shape with size 1.5 mm. OD is located to the nasal side of the fovea with 3 mm to 4 mm distance and 1.92 mm vertically and 1.76 mm horizontally.

Macula: The macula (also known as macula lutea) is a small elliptical shaped yellowish area found close to the focal point of the human retina. Millions of light-sensing cells in the macular region are accountable for sharp vision and it is the most delicate portion of the retina and has no BVs in its surroundings. It acts as a natural sun block because of its

ability to absorb excess blue and ultra-violet light which enters the eye. The macula location varies from person to person and in the normal retinal image, the macula has a diameter of around 1.5 mm (0.059 in), as shown in Fig. 4. When the macula is damaged, the center vision may appear dim, distorted, or dark. Once the macula is localized fovea can be determined, which helps in grading of maculopathy [3].

Fovea: Fovea is a small portion of the macula and situated at the focal point of the macular region. It is the most important component of a retinal image with 0.35 mm of diameter as shown in Fig. 4. The fovea center is situated at an approximate distance of 2 times the OD diameter from OD center. Position and distance of the fovea remains somewhat constant with respect to OD. It is a small portion having a large amount of cone cells which is responsible for sharp vision. The leakage of blood from injured BVs in the fovea region may leads to vision problem. Therefore, identification of fovea region is very important in the retinal image analysis [4].

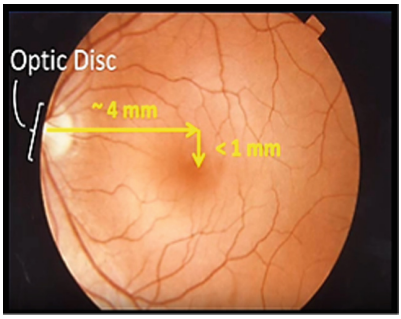


Fig. 3. Optic disk

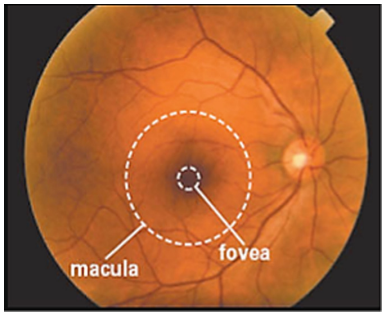


Fig. 4. Macula and fovea

This review paper is structured as follows, begins with Retinal Images Acquisition. A brief overview on Diabetic Maculopathy is explained in Sect. 3. Measures of Diagnostic Accuracy are presented in Sect. 4. An overview of existing methods to detect diabetic maculopathy is described in Sect. 5. Issues in retinal disease identification are discussed in Sect. 6. Then Sect. 6 gives research issues in DR and DM. The conclusion is presented in the Sect. 7.

2 Retinal Images Acquisition

Retinal images are captured with the assistance of fundus camera which is fundamental microscope attached with camera to photograph the inner surface of the eye for the most part retina. These retinal images can be utilized by ophthalmologists for the investigation of eye disease with the assistance of computer algorithms. Figure 5 exhibits the fundus camera.

There are many databases available for the retinal images. Few of them are listed below in the Table 1.

Fig. 5. Fundus camera

Table 1. Databases available for retinal images

Database	Link to download database
STARE	http://cecas.clemson.edu/~ahoover/stare/
DRIVE	https://www.isi.uu.nl/Research/Databases/DRIVE/
DIARETDB1	http://www.it.lut.fi/project/imageret/diaretdb1/
DIARETDB0	http://www.it.lut.fi/project/imageret/diaretdb0/
MESSIDOR	http://www.adcis.net/en/Download-Third-Party/Messidor.html
E-OPHTHA	http://www.adcis.net/en/Download-Third-Party/E-Ophtha.html
HRF	https://www5.cs.fau.de/research/data/fundus-images/

3 A Brief Overview on Diabetic Maculopathy

Diabetes is a metabolic sickness in which the individual has a high blood glucose because of less insulin creation by the pancreas cell. The pancreas cell will normally produce the exact quantity of insulin required for the cells to absorb the glucose from the blood. In patients with diabetes, the pancreas won't create insulin or pancreas may produce a little quantity of insulin. Because of this amount of glucose in the body increases. Hence, despite the fact that the blood contains a lot of glucose, the body loses its primary source of fuel due to less generation of insulin in the body.

Diabetes is categorized into three types as Type-1, Type-2 and Type-3. Diabetes caused because of an autoimmune problem is called Type-1 diabetes. Diabetes caused due to damage of insulin delivering beta cells in the pancreas is called Type-2 diabetes, which prompts no or less generation of insulin. Type-2 diabetes is caused because of malfunctioning of the beta cell, which leads to non-creation of insulin. Type-3 diabetes happens only during pregnancy, in this stage body opposes the influence of insulin delivered.

Diabetic Retinopathy (DR): Disease to the retina because of diabetes is called DR. It harms the little vein of the retina and this may cause vision problems. The DR is categorized into three stages first one is Background Diabetic Retinopathy (BDR), in this stage the veins of the retina become weak and leaky then form a small spot like structure on the retina called hemorrhages. The leaking fluid from the weak BVs in the retina leads to swelling which affect the vision. Second one is Proliferate Diabetic Retinopathy (PDR) in this stage, because of circulation problem of oxygen within the retina provokes the growing of new sensitive vessels. This phenomenon is called neovascularization. Third one is Severe Diabetic Retinopathy (SDR), here continuous growth of an unusual vessel may cause retinal detachment and gradual which leads to loss of vision.

Diabetic Maculopathy (DM): Abnormal state of the retina due to diabetic retinopathy is called Diabetic Maculopathy where exudates appear within the macular region. DM is one of the significant reasons for vision misfortune among individuals around the globe. One such explanation behind maculopathy is a direct result of diabetic macular edema where influenced retinal vein begins releasing liquid and protein onto the macula which prompts vision misfortune. The advancement of diabetic maculopathy is slow and soundless. If diabetic maculopathy is not noticed in the underlying stage the effect this on macula is irreversible and can prompt vision loss. Therefore, screening of diabetic eye helps in finding diabetic maculopathy at the beginning stage which prevents the vision loss.

The DM severity level is categorized into four stages as normal, mild, moderate and severe stage based on how close exudates appear around macular region. Three circles are drawn around the macular region with radii of 1/3 of Optic Disc Diameter (ODD), 1 ODD and 2 ODD. If the exudates are not present in a given image, then it is categorized as normal stage (Fig. 6(a)). In mild stage exudates exits outside the 2DD

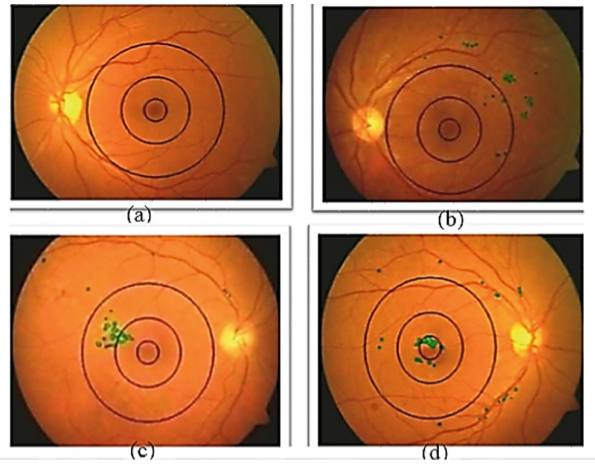


Fig. 6. Severity stages of DM (a) Normal stage (b) Mild stage (c) Moderate stage (d) Severe stage

region (Fig. 6(b)). In moderate stage exudates appears within the 2DD circle and outside the 1DD circle region (Fig. 6(c)). If the exudates appear in the 1/3 DD area, then it is categorized as a severe stage of DM (Fig. 6(d)). Compared with all stages, severe stage of maculopathy is the most vision weakening stages where vision is essentially decreased [6].

4 Measures of Diagnostic Accuracy

Diagnostic tests comprise of various types of data, for example, medical tests, medical signs and symptoms. Doctor's choice about medical action depend on diagnostic examinations, which makes the correctness of an analysis. For a given infection condition, the finest conceivable test can be chosen in view of three characteristics. They are Sensitivity, Specificity and Accuracy. How great the test is at identifying a positive infection is evaluated by sensitivity. How likely the patients without sickness can be effectively eliminated is done with the help of specificity. Accuracy measures rightness of the test.

There are four terms that are used to define accuracy, sensitivity and specificity they are True Positive (TP), True Negative (TN) and False Negative (FN), False Positive (FP). If the patient is suffering from a disease and test also indicates the same, then the outcome of the test is called TP. Likewise, if the patient is not suffering from any disease and test also demonstrates the same then test result called TN. If the patient is not suffering from any disease but test show the presence of disease, then test result is called as FP. Likewise, if the test indicates that the disease is not present but the patient is suffering from disease, then test result is called as FN. Accuracy, sensitivity and specificity are depicted in terms of TP, TN, FN and FP are shown in Eqs. 1.1, 1.2, and 1.3 correspondingly.

$$\text{Sensitivity} = \text{TP}/\left(\text{TP} + \text{FN}\right) \tag{1.1}$$

$$\text{Specificity} = \text{TN}/\left(\text{TN} + \text{FP}\right) \tag{1.2}$$

$$\text{Accuracy} = \left(\text{TN} + \text{TP}\right)/\left(\text{TN} + \text{TP} + \text{FN} + \text{FP}\right) \tag{1.3}$$

Receiver Operating Characteristics (ROC): The association between sensitivity and specificity is demonstrated by using ROC curve. It is drawn with the help of True Positive Rate (TPR) along the Y-axis and False Positive Rate (FPR) along X-axis, where TPR and FPR is calculated as shown in Eqs. 1.4 and 1.5 respectively.

$$\text{TPR} = \text{TP}/\left(\text{TP} + \text{FN}\right) \tag{1.4}$$

$$\text{FPR} = \text{FP}/\left(\text{FP} + \text{TN}\right) \tag{1.5}$$

Every single combination of FPR and FPR generates a ROC space. Both FPR and TPR decide a solitary point in the ROC space. Accordingly the area of the point in the ROC space demonstrates whether the indicative order is great or not. In a perfect

condition, a point controlled by both TPR and FPR yields organize (0, 1), or we can state that this point falls on the upper left corner of the ROC space. This point demonstrates the indicative test has a sensitivity of 100% and specificity of 100%. It is additionally called idealize grouping. Symptomatic test with half sensitivity, and half specificity can be pictured in the corner to dictate by arranging (0, 0) and organize (1, 0). Hypothetically, an arbitrary figure would give a point along this corner to corner. A point anticipated by a diagnostic test introduce in the region over the inclining speaks to a decent indicative forecast, else an awful expectation. A graphic exhibition of what explained above is demonstrated in Fig. 7.

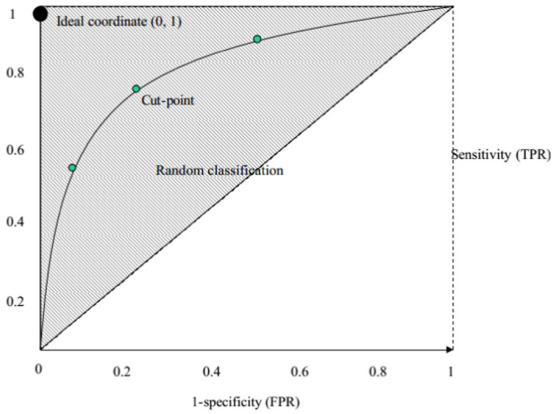


Fig. 7. Receiver Operating Characteristics (ROC) curve

5 Literature Survey

Medical imaging is experiencing an intense growth in recent years due to rise in the use of digital imaging systems for medical diagnostics. Digital imaging techniques are developed for finding retinal disease based on image processing and pattern recognition techniques. These approaches, enhanced the efficiencies of healthcare applications. Patients who are suffering from DR are at high risk to have DM, which may lead to blindness, if not detected and treated appropriately at the appropriate time. Diabetic maculopathy is an abnormal condition of the retina due to diabetic retinopathy. One such reason for DM is because of macular edema where affected blood vessel starts leaking fluid and protein onto the macula region which leads to vision loss. Therefore, screening of diabetic eye helps in finding diabetic maculopathy at the beginning stage which prevents the vision loss. The literature survey on DM detection in retinal fundus images are presented in Table 2.

Table 2. Literature survey on diabetic maculopathy detection in retinal fundus images

Author's	Methodology	Database used	Sensitivity	Specificity	Remarks
Acharya and Mookiah et al. [7]	-Hybrid system based on Radon Transform, discrete wavelet transform and discrete cosine transform -Support Vector Machine Classifier	-MESSIDOR -Private Database	Accuracy of 100% for private and 97.01% for MESSIDOR database Specificity of 99% indicating that almost zero false positive		-Method is completely automatic and does not involve any segmentation algorithms -Obtained higher accuracy with top-hat transform and DWT -System can detect Non-clinically significant macular edema (NCSME) only
Medhi and Dandapat [8]	-Intensity property -Mathematical Morphology -Horizontal Canny Edge Detector -Black Top Hat Transformation -Otsu's Threshold	DRIVE DIARETDB1 MESSIDOR HRF	100% 97.14% 95.07% 100%	100% 100% 95.45% 100%	-Removal of BVs is not required for fovea detection -Helps in estimating retina condition -Less effective in case illumination problem of the retinal image - Blurred fundus image makes identification of fovea difficult
Siddalingaswamy et al. [9]	-Histogram Thresholding, -Clustering, -Mathematical Morphology	-DIARETDB1 -Local Database (KMC)	95.6%	96.15%	-Less expensive -Provides user interface for speedy analysis -Less effective for poor quality images
Osareh et al. [10]	-Fuzzy C-Means Clustering, -Neural Network	-Local Database (Canon CR6 45 non mydriatic retinal camera)	92%	82%	-Slightly better than the standard back propagation -More expensive -Takes less time for training
Lim et al. [11]	-Watershed Transformation, -Intensity Property	-MESSIDOR	80.9%	90.2%	-Robust in OD and Exudates detection -Fails in distinguishing exudates into hard and soft exudates
Deepak et al. [12]	-By generating motion pattern	-DMED -MESSIDOR -DIARETDB0 -DIARETDB1	100%	97%	- Four freely accessible datasets were utilized for testing
Akram et al. [13]	-Gaussian Mixture Model And Support Vector Machine	-MESSIDOR	92.6%	97.8%	-Accuracy is more compared to other methods using MESSIDOR database
Ramasubramanian and Mahendran [14]	-Morphological Operation, K-Means Clustering Algorithm, -SVM Classifier, -CLAHE	-STARE -DRIVE	96% success rate		-Computational time is less compared to FCM -Offers more color information
Medhi et al. [15]	-Nick's Thresholding, -Spatial Fuzzy Level Set Algorithm	-DIARETDB0 -DIARETDB1 -DRIVE -MESSIDOR -HRF	98.6%	98.9%	-Beneficial in separating dark lesions located in macular region

(*continued*)

Table 2. (*continued*)

Author's	Methodology	Database used	Sensitivity	Specificity	Remarks
Rahim et al. [16]	-Fuzzy Filtering -Histogram Equalization, -Fuzzy Edge Detection, -K-Nearest Neighbour	-Local Database (Hospital Melaka, Malaysia)	85.9%	55.7%	-Can acts as a benchmark for the development of further retinopathy signs' -Yield more consistent screening system
Sharma et al. [17]	-Histogram Thresholding, Based On Pixel Intensities,	-DIARETDB0 -DIARETDB1 -MESSIDOR	96.2%	100%	-Grading process requires less time and less expensive -Fails to locate macula in low contrast retinal images -In future work machine learning technique can be used to improve the performance of the algorithm
Tariq and Akram et al. [18]	-Hough Transform, -Thresholding, -Mixture Model-Based (GMM) Classifier	-MESSIDOR -STARE	97.06%	98.55%	-Robust in detection of maculopathy
Punnolil et al. [19]	-Multiclass SVM -CLAHE (Contrast Limited Adaptive Histogram Equalization)	-DRIVE -DIARETDB0 -STARE	96.89%	97.15%	-Fast classification process -More efficient -Robust in detection of macula and maculopathy

From the above literature work different understandings drawn are mentioned below:

- In most of the test cases publicly available database such as DRIVE, DIARETDB1 and MESSIDOR have been used for the validation of maculopathy detection algorithms.
- The survey suggests that having a vast dataset guarantee the accuracy of the automated system.
- Most of the works studied have been carried out drawing the relationship between the macula and OD.
- In most of the cases, the green channel image used for the reason that the green plane shows the finest differentiation among the vessels and the background.
- In most of the methods morphological approach is used because of its computational efficiency.
- Diabetic maculopathy grading depends on hard exudates presence near macular region.
- In some of the works macula is localized based on blood vessel concentration near macula region.
- Hard Exudates, OD, Fovea and Macula play a major role in grading of diabetic maculopathy.

6 Issues in Detection of Diabetic Maculopathy

Analysis of color retinal fundus image becomes a difficult task because of the inconsistency of the images in terms of the retinal anatomical structures, color and the existence of specific features in different patients. Diabetic maculopathy is the major reasons for the vision problem around the globe. For appropriate identification and analysis of the DM, finding the location of the macula, fovea, OD and lesions close to the macula is vital. The identification of fovea is most important in the detection of DM. After an exhaustive survey as listed in Sect. 5, some of the difficulties in the detection of DM are discussed below:

- Fails to locate macula in low contrast and poor quality of retinal images.
- Localization of OD is not a simple task because of the boundary is not well defined in some images and the crossing blood vessels obscure several parts. Which makes macula localization difficult.
- Because of the entire OD brighter and the lesion around the OD is brighter the identification of OD becomes difficult.
- In DR hard exudates are one of the key features which may lead to DM and it is an early clinical indication. It can fluctuate in size with clear edges from small bits to substantial patches. It is extremely hard to recognize the exudates clearly, because they are small spots on the retina and has some comparative properties of the optic disc.
- Grading of diabetic maculopathy becomes difficult since hard exudates and OD appear with similar properties.
- The contrast of macula in a retinal image is quite low and sometimes it may be covered by exudates or hemorrhages. Because of this detection of the macula may fail.
- The presence of a wide variety of blood vessels makes segmentation difficult.
- Grading of diabetic maculopathy fails if hard exudates are not detected accurately.

7 Conclusion

There are numerous retinal diseases that damages the human eye. In this review paper anatomy of the human eye, retinal anatomical structures and the databases available for retinal images are defined in detail. A brief overview about diabetes, diabetic retinopathy and maculopathy is presented. In addition to this, we have discussed the existing methods/techniques for the finding of diabetic maculopathy and the performance metrics to measure these methods. Detection issues involved in maculopathy are also discussed in this paper.

The above literature review states that some methods fail in distinguishing exudates into hard and soft exudates and also in locating fovea and macula region in poor quality images. We would like to address some of the significant gaps using deep learning machine learning technique.

References

1. Acharya, U.R., Eddie, Y.K., et al.: Image Modelling of the Human Eye. Bioinformatics & Biomedical Imaging Series, 1st edn. Artech House Publication, Norwood (2008). ISBN-13: 978-1596932081, ISBN-10: 1596932082
2. Dutta, L.C.: Modern Ophthalmology, 3rd edn. JPB, Amiens (2005). ISBN-13: 978-8180614705
3. Marieb, E.N.: Human Anatomy and Physiology, 6th edn. Pearson, London (2006)
4. http://www.retinaindia.org/for-patients/retinal-disorders/diabeticretinopathy/
5. http://www.allaboutvision.com/conditions/amd.htm
6. www.maculardisease.org
7. Acharya, U.R., et al.: Automated diabetic macular edema (DME) grading system using DWT, DCT features and maculopathy index. Comput. Biol. Med. **84**, 59–68 (2017)
8. Medhi, J.P., Dandapat, S.: An effective fovea detection and automatic assessment of diabetic maculopathy in color fundus images. Comput. Biol. Med. **74**, 30–44 (2016)
9. Siddalingaswamy, P.C., Prabhu, K.: Automatic grading of diabetic maculopathy severity levels. In: International Conference on Systems in Medicine and Biology (ICSMB), pp. 331–334, December 2010
10. Osareh, A., Mirmehdi, M., et al.: Automatic recognition of exudative maculopathy using fuzzy c-means clustering and neural networks. In: Proceedings of Medical Image Understanding and Analysis Conference, vol. 3, pp. 49–52 (2001)
11. Lim, S.T., Zaki, W.M.D.W., et al.: Automatic classification of diabetic macular edema in digital fundus images. In: IEEE Colloquium on Humanities, Science and Engineering, pp. 265–269, December 2011
12. Deepak, S., Sivaswamy, J.: Automatic assessment of macular edema from color retinal images. IEEE Trans. Med. Imaging **31**, 766–776 (2012)
13. Akram, M.U., Akhtar, M., Javed, M.Y.: An automated system for the grading of diabetic maculopathy in fundus images. In: Huang, T., Zeng, Z., Li, C., Leung, C.S. (eds.) ICONIP 2012. LNCS, vol. 7666, pp. 36–43. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-34478-7_5
14. Ramasubramanian, B., Mahendran, G.: An efficient integrated approach for the detection of exudates and diabetic maculopathy in colour fundus images. Adv. Comput. **3**(5), 83 (2012)
15. Medhi, J.P., Dandapat, S.: Improved analysis of diabetic maculopathy using level set spatial fuzzy clustering. In: 2016 Twenty Second National Conference on Communication (NCC), pp. 1–6. IEEE (2016)
16. Rahim, S.S., Palade, V., Jayne, C., Holzinger, A., Shuttleworth, J.: Detection of diabetic retinopathy and maculopathy in eye fundus images using fuzzy image processing. In: Guo, Y., Friston, K., Aldo, F., Hill, S., Peng, H. (eds.) BIH 2015. LNCS (LNAI), vol. 9250, pp. 379–388. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-23344-4_37
17. Sharma, P., Nirmala, S.R., et al.: A system for grading diabetic maculopathy severity level. Netw. Model. Anal. Health Inform. Bioinform. **3**(1), 49 (2014)
18. Tariq, A., Akram, M.U., et al.: Automated detection and grading of diabetic maculopathy in digital retinal images. J. Digit. Imaging **26**(4), 803–812 (2013)
19. Punnolil, A.: A novel approach for diagnosis and severity grading of diabetic maculopathy. In: Advances in Computing, Communications and Informatics (ICACC)

Multi-image Enhancement Technique Using Max-Plus Algebra-Based Morphological Wavelet Transform

Sreekala Kannoth[1(✉)] and Halugona C. Sateesh Kumar[2]

[1] Visvesvaraya Technological University, Belgaum, India
kannothsree@gmail.com
[2] Sai Vidya Institute of Technology, Bangalore, India
hcsateesh@yahoo.com

Abstract. Image enhancement is the process of improving the quality of a digitally stored image so that it is more suitable for certain applications. Here a multiimage enhancement method is proposed which uses non linear wavelets. Though wavelet transform is a linear operation, nonlinear extensions can be made by combining it with mathematical morphology. Max-plus algebra based morphological wavelet transform is a type of non-linear wavelet. Unlike the linear wavelet, this wavelet transform uses maximum and minimum operations instead of linear analysis filters. It is used to maintain the edge information of reconstructed image. The original image is decomposed into scaled, horizontal, vertical and diagonal components using max-plus algebra based morphological wavelet transform. These components undergo bilinear interpolation and coefficients are combined by averaging. Then it undergoes inverse transformation to produce a high resolution image. This method results in less mathematical calculations than other existing methods. Performance comparison parameters are calculated for the output image and it is compared with the parameters obtained from other techniques.

Keywords: Max-plus algebra · Morphological wavelet
Mathematical morphology · SURF · Bilinear interpolation

1 Introduction

A high quality image is formed from one or more low quality images by using different image processing techniques. It is used in many applications which include text image improvement, medical image processing, facial image improvement and so on. Linear and non-linear filter are two different type of filtering techniques. Linear filter uses linear operation on individual input image and then summing the output. They destroy the edge details of an image. Non-linear filter uses non-linear operation and it helps to preserve edge details.

Discrete wavelet transform process only low frequency coefficient in each decomposition level. The high frequency coefficients are not processed in further

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 421–432, 2019.
https://doi.org/10.1007/978-981-13-5758-9_36

levels and the high frequency information in the image will be lost. The analysis filters used in linear wavelet decomposition method are linear lowpass filters and it smoothen the edges whereas the analysis filters in the morphological wavelet transform [1,11] are non-linear, and they help in preserving the edge information. These filters are constructed using the mathematical morphological operation such as erosion (minimum) or dilation (maximum). Morphological operations are done on images based on shapes. It uses a structuring element on the input image and generate an output image of the same size.

Mathematical morphological operations together with wavelet transform maintain the high frequency and low frequency information in the image. In morphological wavelet transform (MWT) [12], both the coefficients i.e., approximation coefficients and detailed coefficients of the former levels are decomposed in the succeeding levels. So the higher frequency coefficients are not lost. In this way MWT works well in reconstructing better quality image than the DWT. This method can be implemented in images and videos, for obtaining a good quality output. Non-linear class of wavelet is efficient and more simpler than other wavelets. The objective here is to apply nonlinear wavelet transform on 2D images to generate a high quality output image.

Here a new multi-image enhancement technique is proposed which is based on mathematical morphology and wavelet transform, known as max-plus algebra based morphological wavelet transform (MMT). This is considered as non linear wavelet because it uses maximum and minimum search operations in decomposition and reconstruction equations. The rest of the paper is organized as follows. Section 2 gives the details of different papers referred. Section 3 is motivation and objective of this work. Section 4 describes the proposed enhancement technique, which includes image registration technique used, max-plus algebra based morphological wavelet decomposition, interpolation technique used, inverse transform equations and different performance comparison parameters that are used for checking the performance of the enhancement method. Section 5 discusses about the output and comparison of the parameters calculated using different methods. Section 6 is conclusion and future work.

2 Literature Review

Farhan and Swarup [1] speaks about a scheme used for microgrid protection which is based on mathematical morphology. Mathematical morphology can be used for the extraction of signal component more accurately. They have used morphological Haar wavelet for detection of faults. It is a non-linear scheme used for multi-resolution signal decomposition and analysis. Here the mapping of information at different levels of the pyramid is done using analysis and synthesis operators. Lifting scheme is used for the construction of new wavelet decomposition. Performance analysis was done for islanded and grid connected mode of operation in different network topology.

Abijith et al. [2] speaks about an image watermarking method based on max-plus algebra based morphological wavelet transform. Here the string to

watermark the image is taken as input and stores in binary form. Area of the image to be taken out is defined by the length of the binary string. The selected image is converted to BMP format and RGB values of each pixel will be stored in an array. These RGB values are converted to YCbCr and embedding is done so that there is no considerable difference between original and watermarked image. This helps in identifying the source and restrict the unauthorized distribution of the image. This method is comparatively faster than the other transform based watermarking methods.

Kumaki et al. [3], proposes the idea of max-plus algebra based morphological wavelet transform(MMT) for watermarking. It is defined by nonlinear operation. Main advantages of using MMT is it gives minimum quantization error, easy to implement in hardware and it preserve the edge information. It decomposes the signal based on the scan window pattern. The research was done to verify image distortion focused on binary images. Peak Signal to Noise Ratio (PSNR) was calculated for the proposed method and it was compared with the PSNR values of the other methods like HWT, DCT and FFT.

Kumaki et al. [4] introduces the concept of adaptive max-plus algebra based wavelets and how that can be applied for single image quality improvement. It is more similar to MMT but varies in the way it selects the sampling window. This also uses the same signal decomposition equations mentioned in [3], but differs in the way they select the sampling window. Sampling window selection is done based on the features of the input image.

Liu et al. [5] gives the concept of reconstructing super resolution image by multi-scale Undecimated morphological wavelet transforms which consists of filters that uses morphological mathematical operations. Undecimated morphological wavelet (UMW) is extended in two dimensions. These transforms are multi-directional and shift-invariant. Their design is much simpler and are capable of extracting the meaningful morphological on different directions. They have also done a performance comparison analysis to show the efficiency of the method in terms of image quality and computational complexity.

Bede and Nobuhara [6] speaks about the applications of max-plus algebra based wavelet transform in image processing, which are defined using multi-resolution approach. They have done a detailed theoretical and practical studies of their proposed method. This method takes neighboring pixels in their calculation. They are computationally less complex, easy to implement and sampling window size can be varied.

Fadnavis [7] speaks about the different image interpolation schemes which are based on multi-resolution technique. Image interpolation is used to improve the quality of an image and to find out the values lying between two samples. The interpolation techniques discussed in this paper are nearest neighbor, bilinear interpolation, bicubic interpolation, basic-splines and lanczos interpolation.

Shirai et al. [8] discuss about different directional and non-directional sampling windows. Morphological wavelet transform on cellular automata processing (CAM2) gives high processing speed and the computation time depends on the number of sampling windows used. How to choose sampling window for the

images is also described. Greater potency can be achieved by use of directional sampling windows in multidirectional morphological wavelet transform.

Goutsias and Heijmans [9] gives the idea for building linear and nonlinear pyramid decomposition methods for signal synthesis and analysis. It is based on the fact that the pyramid contains number of levels, can be finite of infinite, and the information content decreases as it goes to higher levels. Implementation of higher level values is done by analysis operator, and the lower level implementation is done by synthesis operator. These two operations together gives identity operator. The paper also gives an idea about the difference between single scale and multi-scale decomposition methods.

Goutsias and Heijmans [10] speaks about the nonlinear extensions of wavelet transform and the nonlinear extensions started becoming popular with the introduction of lifting scheme. The paper introduces wavelet based mathematical morphology such as morphological Haar wavelet. Lifting scheme gives an easy way for constructing morphological wavelets. It also discusses some examples which retains local maxima in the signal over a range of values.

Heijmans and Goutsias [11] speaks about pyramids and wavelets which are the different types of multiresolution image decomposition schemes. A wavelet decomposition always involves a pyramid transform and they have illustarted both with examples using quincunx scheme. They have also discussed a nonlinear wavelet transform which uses the lifting scheme.

Srivastava et al. [12] speaks about MRI images enhancement which uses approaches in frequency and spatial domain. They have done enhancement in wavelet domain and morphological filters for edge enhancement with some denoising techniques. They have also done the comparison analysis using Contrast improvement index (CII), peak signal to noise ratio (PSNR) and average signal to noise ratio (ASNR).

Claypoole et al. [13] speaks about the different characteristics of wavelet transform constructed with lifting schemes which helps in incorporating adaptivity and nonlinear operators. They have used "update first" technique to preserve the multiresolution properties of the transform. Within that there is an algorithm to avoid predicting across edges and represent edges efficiently. The adaptive lifting transform given in the paper can be used for lossy compression and helps in reducing edge artifacts and ringing.

Zhuang [14] proposes a morphological wavelet transform based digital watermarking algorithm which helps in copyright. It has a gray image watermark to alert the copyright owner and a binary image watermark to protect the copyright. It is a simple algorithm with less computational complexity.

Shao and Zou [15] proposes a wavelet transform and mathematical morphology based edge detection algorithm. Wavelet transform helps in removing the noise whereas mathematical morphology helps in extracting smooth edges. Low frequency edge extraction is done using morphological edge detection operator and for high frequency edges wavelet modulus maxima edge extraction is used.

Xiaoping et al. [16] discuss about a set of flat structuring elements used morphological wavelet transform which is used for improving high frequency detail

coefficients. Morphological wavelet transform is used for enhancement and filtering to obtain clear fingerprint image. Original image undergo multi-scale forward morphological transformation and then use flat-structuring elements to correct detail coefficients. Finally apply multi-scale morphological reconstruction to obtain an enhanced output image. Mathematical morphology helps in maintaining the image information and eliminates unwanted components.

Huang et al. [17] discusses about the image representation using fuzzy morphological wavelet. Here the operations performed are similar to conventional mathematical morphology and can be executed much faster using parallel machines. In fuzzy mathematical morphology, the fitting is between the values zero and one. It transforms one fuzzy set to another.

Huynh-Thu and Ghanbari [18] speaks about the importance of PSNR in assessing the quality of images and videos. PSNR is a valid quality measure when video content and the codec type are not changed. But if the content is changed, then PSNR cannot be a reliable method.

Gupta and Cecil [19] gives image registration techniques using SURF and SIFT methods. They vary in the way they employ detecting features. SURF uses box filter approximation because rectangular box filters can be computed in integral images with almost constant time. SIFT slightly better in detecting more features, but it is slow. SURF is fast.

3 Motivation and Objective

There are many techniques available in literature for the enhancement of noisy and distorted images. This paper presents an image enhancement technique using nonlinear wavelets with bilinear interpolation. It is a multiimage enhancement technique where the input can be a sequence of random images of the same scene. Nonlinear wavelets used here is a max-plus algebra based morphological wavelets where the filters are built using mathematical operations such as maximum and minimum. These operation preserves both high frequency and low frequency content in the image and hence the image edge information will not be lost. It consists of less mathematical calculations compared to other methods which uses linear wavelets and hence it can be used for hardware implementation. Since the input is a sequence of images it can be used to extract details from video frames.

To summarize, the main objective of this paper is to present a fast image enhancement technique, where the input is a sequence of random images of the same scene, with the help of max-plus algebra based morphological wavelets and bilinear interpolation.

4 Proposed Method

The proposed method to obtain an enhanced high resolution image from a sequence of low resolution images is shown in Fig. 1. First the input image sequence is registered using Speeded Up Robust Features (SURF) technique. This registered image undergoes max-plus algebra based morphological wavelet

decomposition and produces four sub-bands. The decomposed coefficients undergo bilinear interpolation. Finally take the inverse Morphological wavelet transform to get the enhanced image.

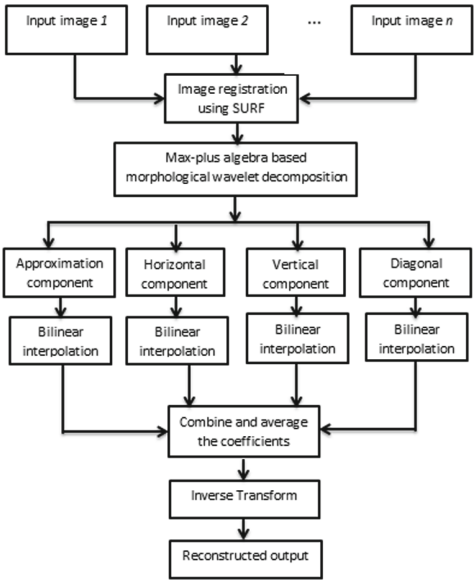


Fig. 1. Proposed method-flow diagram

4.1 Image Registration

Image registration is the technique of overlaying two or more images of the same scene which are taken at different times or by using different cameras. This process align the reference and other sensed images. Here the point matching is done using SURF [19,20]. The main concept behind SURF detector is that it is based on the approximated Hessian Matrix and builds a stack without 2:1 down sampling for upper level in the pyramid. Resulting image will have the same resolution. For filtering the stack, SURF uses a box filter approximation of second order Gaussian partial derivatives and the nearest neighbor will be taken as the key point with minimum Euclidean distance for the invariant descriptor vector.

4.2 Max-Plus Algebra Based Morphological Wavelet Decomposition

Morphological haar wavelet transform (MHWT) [9,10] decomposes low quality 2D image into approximate coefficients and detailed coefficients(vertical, horizontal and diagonal components) with the help of structuring element (B) [16]

or a window. Here a morphological wavelet transform which uses max-plus algebra concept is used for the decomposition [2,3]. MMT decomposes the image into scaled component, horizontal component, vertical component and diagonal component, and reconstruct it in accordance with the scan window pattern. For MMT processing there are three types of sampling windows, square, vertical and horizontal [3,4,6]. The process of decomposition of 2×2 pixel square sampling window [8] is given in Fig. 2.

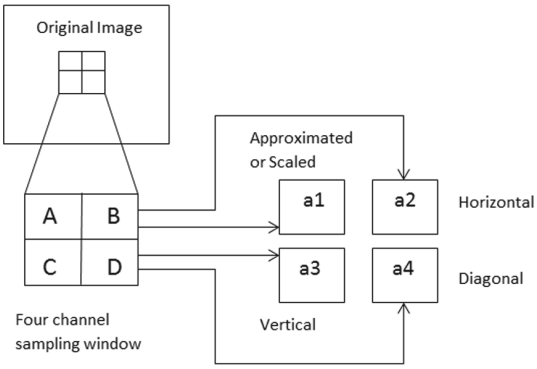


Fig. 2. Decomposition using MMT

The operation are represented with the following equation:

$$a_1 = min(A, B, C, D) \tag{1}$$

$$a_2 = B - A \tag{2}$$

$$a_3 = C - A \tag{3}$$

$$a_4 = D - A \tag{4}$$

Where a_1 represent the minimum of all the pixels values in 2×2 sampling window and is called the approximation coefficient, a_2, a_3, a_4 represents horizontal, vertical and diagonal coefficients. These operations are performed on every pixel.

4.3 Bi-linear Interpolation

Bi-linear Interpolation [7] is the resampling method used here. New pixel value is found using the distance weighted average of the four nearest pixel values. Here the two linear interpolations are performed in one direction and the other two linear interpolation are performed in perpendicular direction. The interpolation kernel is given as

$$u(d) = \begin{cases} 0, & |d| > 1 \\ 1 - |x|, & |d| < 1 \end{cases} \tag{5}$$

where d denote the distance between two interpolation points.

4.4 Reconstruction

Image reconstruction using inverse transform [3] is dine using the following Equations.

$$A = a_1 + max(-a_4, -a_2, -a_3, 0)$$

$$\begin{aligned} B &= a_1 + max(a_2 - a_3, a_2 - a_4, a_2, 0) \\ &= a_1 + max(-a_4, -a_2, -a_3, 0) + a_2 \\ &= A + a_2 \end{aligned}$$

$$\begin{aligned} C &= a_1 + max(a_3 - a_2, a_3 - a_4, a_3, 0) \\ &= a_1 + max(-a_4, -a_2, -a_3, 0) + a_3 \\ &= A + a_3 \end{aligned}$$

$$\begin{aligned} D &= a_1 + max(a_4 - a_2, a_4 - a_3, a_4, 0) \\ &= a_1 + max(-a_4, -a_2, -a_3, 0) + a_4 \\ &= A + a_4 \end{aligned}$$

4.5 Performance Comparison Parameters

Two parameters are used here to compare the effectiveness of the proposed enhancement technique with others. They are Mean Square Error (MSE) and Peak Signal to Noise Ratio (PSNR).

MSE is the sum of square of difference between the original and reconstructed images divided by number of pixels in the image.

$$MSE = \sum_{n} \frac{[I_1(l,n) - I_2(l,n)]^2}{L * N} \tag{6}$$

where $I_1(l,n)$ is the intensity value of original image and $I_2(l,n)$ is the intensity value of reconstructed image. $L * N$ is the total number of pixels in the image.

PSNR helps in identifying the quality of reconstructed image. High PSNR results in the reconstruction of high quality image. It is expressed in dB (Decibel) and is defined as the square of peak value of image divided by mean square error.

$$PSNR = 10log_{10}\frac{[R]^2}{MSE} \tag{7}$$

$R = 255$ for grayscale image.

5 Results and Discussion

5.1 Results and Analysis

A sequence of three distorted Grayscale Lena image is processed using proposed method. Figure 3 shows one of the input sequence images and the registered image. Registered image is decomposed to generate approximated and detailed (horizontal, vertical and diagonal) components which are shown in Fig. 4. These coefficients undergo bilinear interpolation and then the average of the coefficients is found. After this take the inverse transform which results in a better quality output image, shown in Fig. 5. The experiment was repeated with three image sequences of Cameraman, Barbara, Baboon and Pepper. One input image and reconstructed image for Baboon and Barbara is given in Fig. 6. Result can be improved by increasing the number of input images.

Fig. 3. Input image 1 and registered image

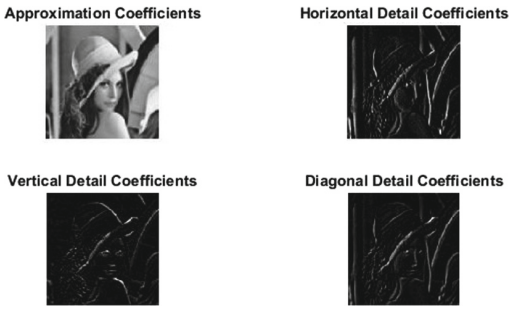


Fig. 4. Decomposed components after applying MMT

5.2 Comparative Study

The experiment was done using the proposed method for different image sequences and parameters PSNR and MSE are calculated for performance

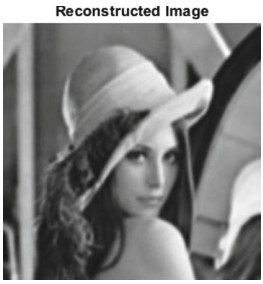


Fig. 5. Reconstructed image

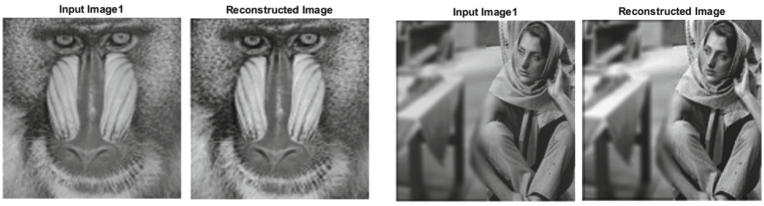


Fig. 6. Example input image 1 and reconstructed image - Baboon and Barbara

analysis [18]. The result is compared with the other methods like Normal MMT [3] and undecimated morphological wavelet [5] and the comparison is given in Tables 1 and 2.

The method is applied on distorted Lena, Pepper, Cameraman, Baboon and Barbara images. From Table 1, it is clear that average Mean square error is least for our method compared to the other related methods. Table 2 shows the values of Peak signal to noise ratio for the different images and we can see that our proposed method gives the highest PSNR values. An efficient method should result in low MSE and high PSNR values. Hence it is clear that the proposed method gives comparatively better image.

Table 1. Mean square error

Images	Proposed method	Undecimated morphological wavelet [5]	Normal MMT [3]
Lena	75.43647	90.5191	82.4526
Pepper	91.3356	91.9963	91.7622
Cameraman	156.8629	268.7326	202.1156
Baboon	63.9352	183.7385	99.6210
Barbara	76.7857	242.7170	122.2356

Table 2. Peak signal to noise ratio

Images	Proposed method	Undecimated morphological wavelet [5]	Normal MMT [3]
Lena	29.35499	28.5634	28.9687
Pepper	28.5244	28.4931	28.5041
Cameraman	26.1756	23.8376	25.0748
Baboon	30.0734	25.4888	28.1472
Barbara	29.2871	24.2798	27.2581

6 Conclusion and Future Work

Here a high quality image reconstruction technique using non linear wavelets with bilinear interpolation is implemented to generate an enhanced image from a sequence of randomly captured images of the same scene. Max-plus algebra based morphological wavelet transform helps in extracting the meaningful information by preserving the edge details with less computational complexities. From the calculated MSE and PSNR values it is seen that the proposed method gives a better output image compared to other methods. The work can be extended for enhancing the video quality.

References

1. Farhan, M.A., Swarup, S.K.: Mathematical morphology-based islanding detection for distributed generation. IET Gener. Transm. Distrib. **11**(14), 3449–3457 (2017)
2. Abijith, Sabhahit, G., Rao, N., Sequeira, P., Venugopala, P.S., Kumaki, T.: Performance of max-plus algebra-based morphological wavelet transform watermarking in android devices. In: International Conference on Wireless Communications, Signal Processing and Networking (2017)
3. Kumaki, T., Murakami, Y., Shuhei, I., Kei, N., Fujino, T., Ogura, T.: Max-plus algebra-based morphological wavelet transform watermarking for highly-parallel processing with mobile embedded processor. J. Signal Process. **16**, 547–556 (2012)
4. Kumaki, T., Tanito, Y., Tokunaga, T., Fujita, T., Ogura, T.: Adaptive multi-directional max-plus algebra-based morphological wavelet transform. J. Signal Process. **20**, 9–19 (2016). https://doi.org/10.2299/jsp.20.9
5. Liu, W., Liang, Y.S., Chen, Z.B., Zhang, J.H.: An image super resolution reconstruction algorithm based on undecimated morphological wavelet. In: IEEE International Conference on Digital Signal Processing (DSP), pp. 600–603, 21 July 2015
6. Bede, B., Nobuhara, H.: A novel max-plus algebra based wavelet transform and its applications in image processing. In: Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics San Antonio, TX, USA, October 2009
7. Fadnavis, S.: Image interpolation techniques in digital image processing: an overview. Int. J. Eng. Res. Appl. **4**(10), 70–73 (2014)
8. Shirai, S., Nakai, M., Kumaki, T., Fujita, T., Nakanishi, M., Ogura, T.: Morphological wavelet transform using multiple directional sampling windows on cellular hardware platform. In: IEEE 9th International New Circuits and systems conference, pp. 354–357, 26 June 2011

9. Goutsias, J., Heijmans, H.J.: Nonlinear multiresolution signal decomposition schemes- I. Morphological pyramids. IEEE Trans. Image Process. **9**(11), 1862–76 (2000)
10. Goutsias, J., Heijmans, H.J.: Nonlinear multiresolution signal decomposition schemes-II. Morphological wavelets. IEEE Trans. Image Process. **9**(11), 1897–913 (2000)
11. Heijmans, H.J.A.M., Goutsias, J.: Morphological pyramids and wavelets based on the quincunx lattice. In: Goutsias, J., Bloomberg, D.S. (eds.) Mathematical Morphology and its Applications to Image and Signal Processing. CIVI, vol. 18, pp. 273–281. Springer, Boston (2002). https://doi.org/10.1007/0-306-47025-X_30
12. Srivastava, A., Alankrita, R.A., Bhateja, V.: Combination of wavelet transform and morphological filtering for enhancement of magnetic resonance images. In: Snasel, V., Platos, J., El-Qawasmeh, E. (eds.) ICDIPC 2011. CCIS, vol. 188, pp. 460–474. Springer, Heidelberg (2011). https://doi.org/10.1007/978-3-642-22389-1_41
13. Claypoole, R.L., Davis, G.M., Sweldens, W., Baraniuk, R.G.: Nonlinear wavelet transforms for image coding via lifting. IEEE Trans. Image Process. **12**, 1449–1459 (2003)
14. Zhuang, L.: Multipurpose digital watermarking algorithm based on morphological wavelet transform. In: IEEE International Conference on Communications and Mobile Computing, vol. 3, pp. 396–400, 6 Jan 2009
15. Shao, K., Zou, Y.: An image edge detection algorithm based on wavelet transform and mathematical morphology. In: Wen, Z., Li, T. (eds.) Foundations of Intelligent Systems. AISC, vol. 277, pp. 485–495. Springer, Heidelberg (2014). https://doi.org/10.1007/978-3-642-54924-3_46
16. Xiaoping, J., Shijun, Y., Jing, Z., Heping, C., Naiqi, L.: Research of morphological wavelet transformation in fingerprint image preprocessing. In: IEEE 6th International Conference on Web Information Systems and Applications, 18 September 2009, pp. 41–44 (2009)
17. Huang, C.P.: Image representation using fuzzy morphological wavelet. In: INTECH. Open Access Publisher (2008)
18. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of PSNR in image/video quality assessment. Electron. Lett. **44**(13), 800–801 (2008)
19. Gupta, V.K., Ceci, K.: An analytical study of SIFT and SURF in image registration. Int. J. Eng. Innovative Technol. **3**(9), 130–134 (2014)
20. Bay, H., Tuytelaars, T., Van Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). https://doi.org/10.1007/11744023_32

Power and Area Efficient FSM with Comparison-Free Sorting Algorithm for Write-Evaluate Phase and Read-Sort Phase

T. A. S. Bhargav(✉) and E. Prabhu(✉)

Department of Electronics and Communication Engineering,
Amrita Vishwa Vidyapeetham, Amrita School of Engineering, Coimbatore, India
coolbhargavtas@gmail.com, e_prabhu@cb.amrita.edu

Abstract. In this paper, a comparison-free sorting algorithm is proposed for negative and positive elements which satisfies the conditions such as hardware complexity. The basic idea is to sort the array of input integer elements without performing any comparison related operations between the data. Sorting technique for negative and positive numbers are executed involving similar hardware. Therefore, it doesn't require any complex hardware design. This avoids any usage of memory such as SRAM or any circuitry involving complex design as compared to that of the ones used in other conventional sorting practices. Instead the proposed work utilizes basic registers to store the binary elements. FSM module is proposed with a comparison-free sorting algorithm in order to reduce hardware complexity. All the designs are coded in VHDL and verified using ModelSim SE10.4d simulator. The conventional and the proposed design are synthesized using Vivado and Synopsys DC Design Compiler (90 nm CMOS technology). From the synthesis report, it is observed that proposed FSM with comparison-free sorting algorithm has reduction in power and area compared to the conventional design.

Keywords: Sorting algorithm · Finite state machine · Write-evaluation phase Read-sort phase

1 Introduction

Sorting is a term which is widely used in the field of computer science in order to arrange the given data in a particular order. Over the decades, a wide range of sorting algorithms was introduced and are practically applicable in multiple specializations [2, 7]. Every sorting algorithm is unique in its approach. The sorting algorithms can be categorized based on a variety of parameters such as the number and the type of input elements provided or on the rate of recurrence of input elements and other parameters such as the stability, adaptability [3], usage of memory, complexity or the method of sorting such as insertion, selection, etc. Sorting can also be classified based on the manner it arranges the elements which can be either numerical or lexicographical. Sortings are widely used in many applications depending upon its requirements. Out of all sorting algorithms, some sorting algorithms have been specialized to use high-speed sorting, processing data for large computational values [4], single CPU by using sorting [5], for using of less

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 433–442, 2019.
https://doi.org/10.1007/978-981-13-5758-9_37

memory and to increase performance [6], and using RISC processor for multiple CPUs and also for big data processing as well.

There are various sorting algorithms which were introduced over the past few years off which the most popular algorithms are the simple sorting, efficient sorting, bubble sorting and distribution sorting to name a few. Each of these algorithms has an order of complexity which decides the best, average and worst behavior of the algorithm. There is rapid development in computational power for parallel processing in many GPU and CPU based handling devices wherein sorting has proven to be more efficient in managing the resources for computational power [1]. While some sorting algorithms are application specific and are best suited for specific computational methods, the remaining sorting algorithms can be implemented for any problem statements provided the algorithms meet the required specifications.

In computer science technology, sorting algorithms can be used to arrange large data values. Let A denotes n elements in the memory A1, A2 and so on An. Contents of A will be arranged either in increasing or decreasing order, A1 <= A2 <= A3 <= A4 … <= An. From the relation stated above, A comprises of n number of elements and hence, can store these elements in n factorial ways. Suitable algorithms are used for sorting these n factorial ways.

In the field of Vedic mathematics, every sutra is unique in its way of execution [10]. When it is put to practical applications, the complexity and the computation of the entire design increases. The sorting techniques on the other when used for such design, allows one to study the responses from such designs in an orderly manner.

Sorting algorithms can be classified into data size based or data size information based. Sorting techniques which rely on the data size are further categorized as external sorting and internal sorting. While internal sorting arranges the data as these are being stored in the primary memory, an external sort uses both a primary memory as well as a secondary memory wherein the former stores the current data to be sorted and the latter stores the ones that cannot occupy the former. Furthermore, sorting techniques which depend on the information of data size are again classified into comparison based sorting and non-comparison-based sorting. A comparison-based algorithm arranges the given sorting array by individually comparing the value of one element with the value of other elements. On the other hand, a non-comparison-based sorting algorithm arranges an array by sorting its value using the internal character.

In software engineering and mathematics, to place elements in a certain order to form a list, sorting algorithm is used. The most-utilized are numerical and lexicographical orders. Proficient arranging is imperative to upgrade the utilization of different calculations that require arranged records to work effectively; it is likewise frequently helpful for creating intelligible yield. All the more formally, the yield must fulfil two conditions:

1. The output is in non-diminishing order.
2. The output is a transposition of the input.

Considering all the existing sorting algorithms, the insertion sort has proved by all accounts most effective for a modest number of components. Selection sorts do not take the pre-existing sorting request. A few more complex sorting algorithms are considerably quicker on considerably larger sets.

Complexity and internal structure are some of the ways for classifying sorting algorithms. As compared to another form of sorts as explained in [8], these are some of the common sorting techniques used in information technology. While swap-based sorting refers to the comparison of adjacent elements and swapping accordingly from either left to right or right to left, merge-based sorting partitions the array of the element into two groups repeatedly till each of the groups has a single element and sorts the groups before combining it to form a sorted list. On the other hand, parallel sort: Parallelization of several sorting algorithms undergo parallelization in order to reduce the process time.

Sorting algorithms can further be classified based on other criteria such as:

- **Computational complexity**: It categorises the algorithms under three cases of complexities i.e., good, typical and worse. The number of comparisons required in good and average cases is O(n log n) while for worst case is O(square of n).
- **Stability**: Sorting stable calculation keeps up the relative requests of 3 record with approach keys while sorting of unstable calculation do not keep up the relative requests of record with meet keys.
- **Memory usage**: Temporary registers are required to store temporary values during sorting.
- **Recursion**: A few algorithms are further categorized only as repetitive or non-repetitive but remaining ones show both of the characteristics.

While comparison based implements the traditional approach of inspection and contrast of two elements at a time, quicksort is considered to be the most chosen sorting algorithm due to its better performance and popularity among the sequential sorting.

There are various reasons for the sorting to be considered as a basic problem for computer scientists.

- At some point, the need for the arrangement of data is natural in an application. For instance, with a specific end goal to prepare the client articulations, the bank needs to sort checks with check number (Fig. 3).
- Algorithms frequently utilize the arrangement technique as one of the subroutine which are a bit primitive. For instance, a program that renders graphical items which are being layered over each other require some arrangement or sorting.
- There is a wide assortment of arranging calculations and they utilize numerous strategies. In reality, numerous vital strategies are achieved through calculation outline that has been developed throughout the years.

2 Related Works

Previous work on the comparison-free sorting algorithm as mentioned in [1] makes use of several circuit components such as a one-hot decoder, register arrays, a serial shifter, a parallel counter, and many more. The functionality of the data path and the corresponding data flow depends on write evaluation and read-sort phase. Unlike the previously mentioned work wherein the author design the sorting algorithm for unsigned numbers, we are modifying the required hardware to sort for signed as well as unsigned

numbers for our proposed work. We use a similar hardware specification as mentioned above with some small modification such as the removal of the tri-state buffers, multiplexers (which could have been used as one-hot logic, [11]) and one-detector which is utilized for unsigned numbers through FSM technique while for signed numbers, an additional component is added i.e., a signed set keeping into consideration about the area and the power optimization of the proposed design. Unlike the work proposed in [9] where the entire methodology depends on the scheduled Data-Flow Graph, our work completely realizes the FSM thus obtained and provide a comparative study between the existing work and the proposed design both in using FSM technique and designs for Signed inputs. The complete circuit design is implemented under 90 nm CMOS technologies as is the work proposed in [1] (Fig. 2).

3 Methodologies

3.1 Comparison-Free Sorting Algorithm

Here we provide a brief summary of the algorithm in order to follow the methodology for our work. For instance, let's consider an input binary bus of n-bits for which the input values must be sorted. It is known from the fact that the number of inputs to be sorted are M = 2n. The sorting algorithm functions using the corresponding one-hot weight representation of the associated input element as it is unique for every element. Therefore, there exists a hot encoding representation of N-bits corresponding to every M input elements such that N = M. The conversion of the binary input element to its corresponding one-hot encoded representation is a simplified process which utilizes a traditional one-hot decoder. Utilizing such representation method ensures that every element is orthogonal to each other when projected onto an Rn linear space.

3.2 Finite State Machines

A finite-state machine (FSM) can be defined as a computational model which can be utilized for modelling different problems or algorithms and provides a constructive mapping for implementing the same through hardware or software. It is a theoretical model of the algorithm or problem statement designed in such a manner that it can be in precisely one of the finite number of states at any given time. The given FSM transitions from one proposed state to another as a response associated with some external inputs. An FSM functions only by using a finite number of states, the starting or the initial state and the conditions to be satisfied before each transition. The states and conditions are defined by the user developing the FSM.

For our proposed work, we developed an FSM as shown in Fig. 1 consisting of six states which describe the functionality of a comparison free sorting algorithm dealing with n input elements. The states utilized for the FSM have no relation with the number of input elements.

At some point, the need for the arrangement of data is natural in an application. For instance, with a specific end goal to prepare the client articulations, the bank needs to sort checks with check number.

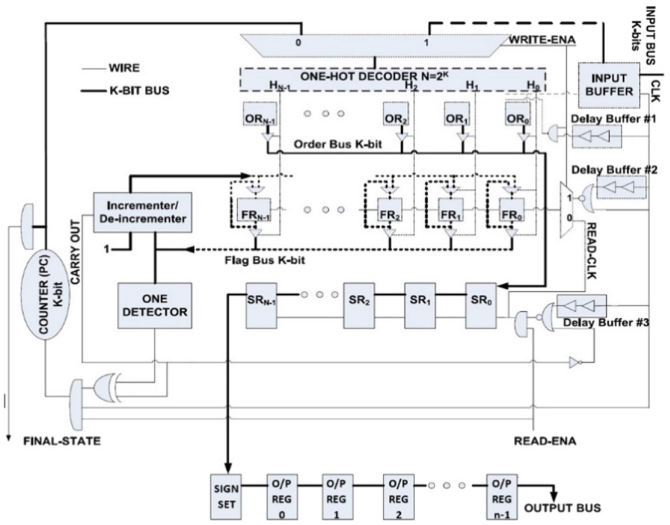


Fig. 1. Block diagram for signed read-sort phase

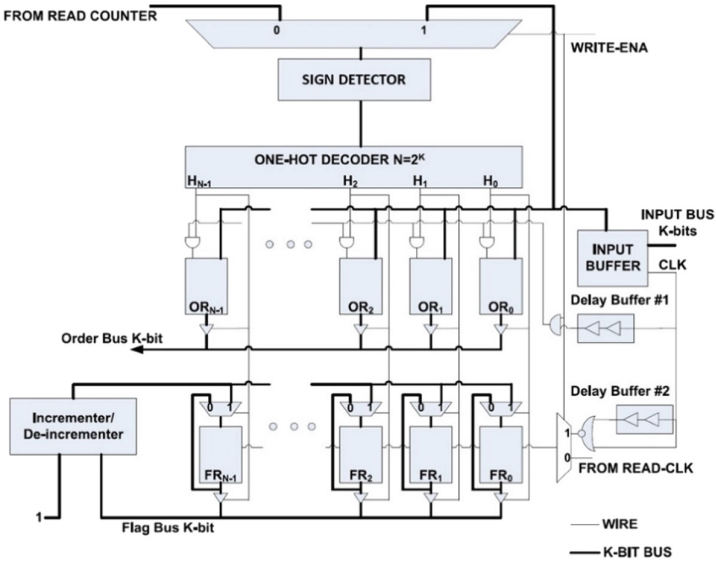


Fig. 2. Block diagram for signed write-evaluation phase

- State 1 (One hot decoder): Inputs are primarily fed into the one hot decoder and the variable 'count' is used to store the current or the present value of the input.

– Condition: If the value of *count* remains less than (2^n-1) where n is the number of bits and (2^n-1) is the maximum value of the input bus. On satisfying the condition it will allow the state to transition from state 1 to state 2.

- State 2 (Order register): it consists of a register of (2^n-1) bit length in order to primarily store the corresponding one hot encoded value of the current input value i.e., by setting the corresponding bit position of the value stored in the *count*.

– Condition: If the value of *count* remains less than (2^n-1) where n is the number of bits and $(2^n\text{-}1)$ is the maximum value of the input bus. Once the condition is satisfied it will allow the state to transition from state 2 to state 3.

- State 3 (Frequency register): The register is of (2^n-1) bit length and the main objective is to store the recurrence of an input element in its corresponding bit position. Since our system describes a comparison free sorting algorithm so as to rearrange the given input elements, unique elements are sorted out by order register. Apart from rearranging the elements based on its value, it is difficult to identify an input element whether it is repeated or not. At this point, the frequency register plays a significant role.

– Condition:

3. If the value of *count* remains less than (2^n-1) where n is the number of bits and (2^n-1) is the maximum value of the n-bit input. A state transition will occur from state 3 to state 4 on satisfying the condition.
4. If the value of *count* is equal to (2^n-1) and the current value of the frequency register (I) is equal to 0, then the FSM state will transition from state 3 to state 6.
5. If the value of *count* is equal to (2^n-1) and the current value of frequency register (I) is not equal to but rather greater than 0, then the FSM state will transition state 3 to state 5.

- State 4 (Up counter): The up-counter increments the pointer in order to pass the value of the next input element after every transition (till the value of the pointer reaches a count which is equal to the total number of inputs) which then allows a state transition from state 4 to state 1.
- State 5 (Up counter): This up-counter is responsible for incrementing the value of variable 'i' which represents bit position in frequency register provided its value is less than or equal to the total number of input elements.

– Condition: If the value of 'i' is less than a total number of inputs, then a state transition occurs from state 5 to state 3.

- State 6 (Sorted register): This register stores the value of sorted input which represents the output of the modified system. It stores the value of the input elements in an ascending order till the variable 'i' reach its maximum value.

– Condition: If the value of 'i' is less than a total number of inputs, a state transition occurs from state 6 to state 3.

4 Result and Discussion

For our simulation, we have primarily used the Synopsys DC Design Compiler tool under 90 nm CMOS technology and Vivado software. Synopsys DC Design Compiler is used for obtaining the area and power values. While Vivado software is utilized for providing the required device utilization values. The values obtained in Table 1 represents the device utilization values for both the conventional design as well as the proposed FSM module. We observed that there is a significant reduction in the number of components used such as FF, LUT, I/O and BUFG by 54.6%, 52.1%, 23.21% and 16.6% respectively for post synthesis. Similarly, the reductions are also observed in post-implementation giving percentage change of 65.26%, 64.84%, 23.21% and 16.6% respectively. Table 2 represents the comparative analysis of the conventional and the proposed FSM with respect to the parameters such as area and power. For our work, we have also observed the similar parameters using Signed inputs for write-evaluation and read-sort phase for which the device utilization values are given in Table 3 (Figs. 4, 5 and 6).

Table 1. Comparison of basic components used in existing model and proposed FSM model

Components	Availability	Post synthesis		Post implementation	
		Conventional [1]	Proposed (FSM)	Conventional [1]	Proposed (FSM)
FF	1424000	1154	524	898	312
LUT	712000	979	468	674	237
I/O	1100	1034	794	1034	794
BUFG	128	12	10	12	10

Table 2. Comparison of existing model and proposed FSM model using area and power parameters

Parameters	Conventional [1]	Proposed FSM design	Reduction (%)
Area (um^2)	114018.3015	52739.1005	53.745
Power (mW)	3.2954	1.0403	68.4318

Table 3. Device utilization values for write-evaluation and read-sort phase with signed inputs

Components	Availability	Signed write-evaluation and read-sort phase	
		Post synthesis	Post implementation
FF	1424000	879	667
LUT	712000	823	721
I/O	1100	794	794
BUFG	128	12	12

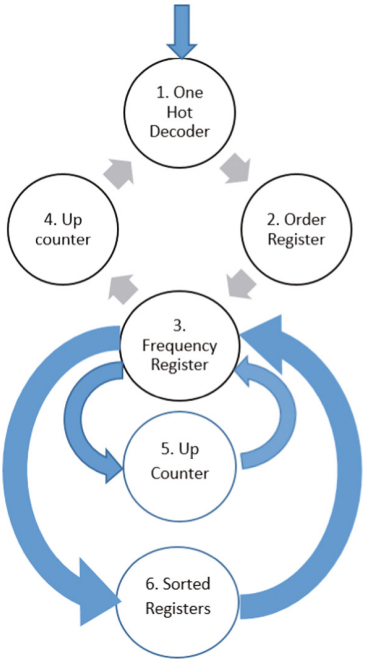


Fig. 3. Finite state machine for comparison free sorting algorithms

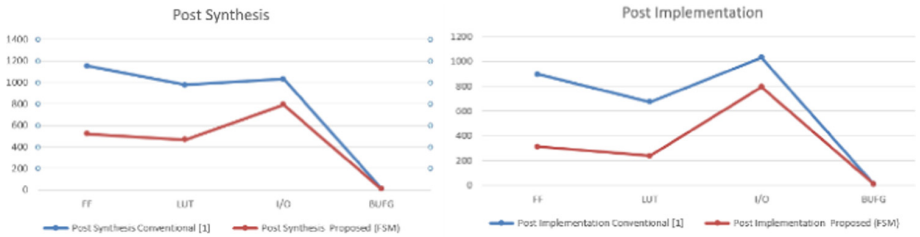


Fig. 4. Graphical representation of the data obtained in Table 1

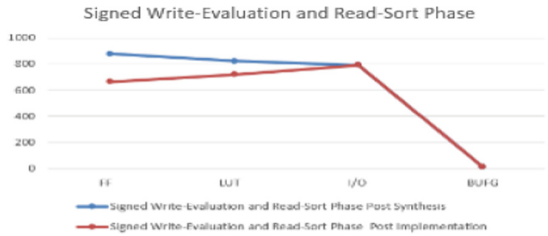


Fig. 5. Graphical representation of the available components as observed from Table 3

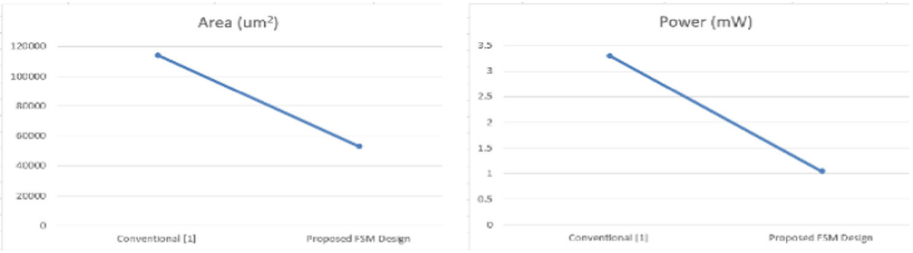


Fig. 6. Area and power comparisons of the proposed and conventional design respectively as shown in Table 2.

5 Conclusion

In this paper, we proposed an algorithm for comparison-free sorting by using FSM module. Our work combines the operation of write-evaluation and read-sort phase in a single block. The advantages of our work over the previously described implementations are various parameters as provided from our results. This technique can be utilized in order to map the random output response generated by any logical design while dealing with random inputs which can be either forced inputs by the observer or can be generated automatically. Furthermore, the design implemented for signed inputs. In the field of Design and Trust, it is a requirement for the designer to map the responses with the corresponding inputs while dealing with malicious hardware or information leakage. Any deviation of the responses from the required reaction must be recorded correctly along with its respective input. Its challenging enough to spot the affected output manually instead sorting the input and the responses will allow the designer to compare and analyses the observations.

References

1. Abdel-Hafeez, S., Gordon-Ross, A.: An efficient O(N) comparison-free sorting algorithm. IEEE Trans. Very Large Scale Integr. Syst. **25**, 1930–1942 (2017)
2. Canaan, C., Garai, M.S., Daya, M.: Popular sorting algorithms. World Appl. Program. **1**(1), 6271 (2011)
3. Busse, L.M., Chehreghani, M.H., Buhmann, J.M.: The information content in sorting algorithms. In: Proceedings IEEE International Symposium Information Theory (ISIT), pp. 2746–2750 July 2012
4. Koike, A., Sadakane, K.: A novel computational model for GPUs with application to I/O optimal sorting algorithms. In: 2014 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), pp. 614–623. IEEE, May 2014
5. Lupescu, G., Sluanschi, E.I., Tpu, N.: Using the integrated GPU to improve CPU sort performance. In: 2017 46th International Conference on Parallel Processing Workshops (ICPPW), (pp. 39–44). IEEE 2017, August
6. Sareen, P.: Comparison of sorting algorithms (on the basis of the average case). Int. J. Adv. Res. Comput. Sci. Softw. Eng. **3**(3), 522–532 (2013)

7. Lindstrom, E.E., Vitter, J.S., Wong, C.K.: Introduction–sorting. IEEE Trans. Comput. **100** (4), 293–295 (1985)
8. Khan, F.G., Khan, O.U., Montrucchio, B. Giaccone, P.: Analysis of fast parallel sorting algorithms for GPU architectures. In: 2011 Frontiers of Information Technology (FIT), pp. 173–178. IEEE, December 2011
9. Harish Ram D.S., Bhuvaneswari, M.C., Logesh, S.M.: A novel evolutionary technique for multi-objective power area and delay optimization in high level synthesis of datapaths. In: Proceedings 2011 IEEE Computer Society Annual Symposium on VLSI, ISVLSI 2011, Chennai, pp. 290–295 (2011)
10. Prabhu, E., Mangalam, H., Gokul, P.R.: A delay efficient vedic multiplier. In: Proceedings National Academy of Sciences, India, Section A: Physical Sciences (2018). https://doi.org/10.1007/s40010-017-0464-4
11. Madhukar Reddy, B., Prabhu, E.: An efficient 16-bit carry select adder with optimized power and delay. Int. J. Appl. Eng. Res. **10**(11), 27909–27916 (2015)

Technologies and Methods for 3D Reconstruction in Archaeology

Suma Dawn(✉) and Prantik Biswas

Jaypee Institute of Information Technology, Noida, India
{suma.dawn,prantik.biswas}@jiit.ac.in

Abstract. Digital Modeling of archaeological sites and the inherent cultural heritage has proved to be a tool of great importance not only for discovery of data but also for recovery and understanding of data from the archaeological site remains. The irreversible and destructive nature of excavation can be mitigated by using the techniques of 3D digital modeling using the available hardware and methodologies of computer vision and reconstruction. The widespread adoption of 3D technologies have facilitated not only timely and accurate recording methods but also storage and reusability of data that can be further used for collaborative and reconstructive tasks. This allows for preservation of the excavated objects, artefacts and landscapes and also for inter-disciplinary research. In this work we have put forth the numerous techniques that are used for 3D reconstruction of objects and artefacts, creation of virtual environments of heritage sites and underwater excavation incorporating not only unearth artefacts but also maps and location information to form a virtual and immersive environment.

Keywords: 3D modeling · Multi-resolution · Multi-layered model
Integration · Image acquisition · Feature extraction and matching
Range-sensor devises · Image-sensors · Object reconstruction · Stereo vision
Structure-from-motion · Virtual environment · Underwater object reconstruction

1 Introduction

"Archaeology" as defined in the dictionary is the study of human history and prehistory through the excavation of sites and the analysis of artefacts and other physical remains. Not only is narrative reconstruction of events and contexts a part of archaeology but also very important for archaeologists to bring the stories of the past – scenes and sequences – for better understanding of our culture, heritage and overall growth and development. Usually wax, porcelain, wooden, or plaster was used for simulation of archaeological models before the introduction and wide-spread use of computer graphics and virtual reality for showcasing of 3D models. These prototypes were useful both in terms of education as well as entertainment [12, 16] and are useful for numerous reasons- (i) efficient representation of the objects/scenes; (ii) allows for statistical and evaluative measurable; (iii) efficient modeling and existing of various analysis tools and methods; (iv) easy building of virtual models; and (v) easier transmission, manipulation, storage, and representation [8] making this an open field

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 443–453, 2019.
https://doi.org/10.1007/978-981-13-5758-9_38

for multi-disciplinary experts and these have given rise to the use of methods that are traditionally thought to be in the fields of Computer Graphics, AI, GIS, Soft Computing, etc.

Unearthing archaeological findings, their acquisition, analysis and interpretation by scientists, historians, and excavation experts were time-consuming processes. With the advent of digital technologies and their widespread usage through internet have not only reduced this process to its original minuscule part but also allowed for cross-border expertise to be at hand. Digitization has not only been proven helpful but is not considered a necessary in many spheres in archaeology. Information uncovered during field investigations, including detection, documentation, analysis, visualization and their digital storage have increased the opportunities for researchers and helped in protecting and conserving the heritages of a particular place. These technologies have intern lead to the growth of visualization and modeling of not only objects but also in depiction of monuments and even cities [8, 12, 16].

The process of digitization in archaeology has bought forth the traditionally isolated activity of excavation, preservation and study to a global front wherein the data can be easily shared and analyzed by varied personnel [23]. Modeling of archaeological contexts is achieved in numerous manners - Acoustics and optical for underwater and submerged requirements, 3D scanners, photogrammetry and satellite imagery. These help in acquisition and 3D reconstruction of object as well as intra-site modeling and virtual city building. These digital proofs, in terms of the various artefacts and excavation scenes not only remove any personal bias in the interpretation of the facts but also allows for a global increase in knowledge by providing tools for resolving any theoretical or narrative conflicts by providing datasets that are accurate, in terms of their photographic parameters, and accessible. A reconstruction model can be created for archaeological research during or after the excavation, for public presentation or for educational purposes [20]. In [19] to augment the cognitive ability of the archaeologist, promotion of an archaeological site and the monitoring of its conditions of degradation multi-disciplinary methodology were proposed.

Not only does the digitalization technologies allow for recording, but also for visualization, representation and reconstruction. These involve GIs applications, image-based and range-based modeling of the datasets, and virtual 3D aggregate modeling too. Also, these recordings are available not only during the field excavation, but also in the labs long after the process is physically over [21]. Apart from 3D reconstruction and visualization, other advantages that many expert believe are in (i) limiting fragmentation of the preserved artefacts by application of visualization, (ii) allowing for multi-disciplinary classification of the findings, (iii) searching, and querying on numerous tags, (iv) limiting subjective characterization, (v) enriching and extending the archaeological research, to name a few.

Since the complete setup cannot be streamlined to be an automated process, there are some other major problems too that the process of 3D reconstruction faces - requirement of specific hardware and technical expertise, management of 3D data, and cost factor. However, once these hurdles are sorted, the datasets then have higher longevity, accessibility, compatibility and visibility. However, the interpretation of the datasets and a need for standardization pose recent challenges too.

The focus of this paper is to present a compendium of ideas and methodologies, currently being used by researchers for enriching the archaeological findings, either on land or underwater. In Sect. 2, we have summarized the various steps used for 3D reconstruction of artefacts as well as for intra site reconstruction. In Sect. 3, we have described various methods that are traditionally used as well as in-vogue for 3D reconstruction. Section 4 concludes this paper.

2 3D Reconstruction Methodology

Archaeological excavation is inherently a destructive process. To unearth the site for study, intact earthen deposits need to be removed. Much data is lost during this process either due to the requirements of the process itself or due to the non-availability of experts for relevant interpretation or may be due to gross negligence. However, using digital recording limits this and helps in addressing the issues of excavated layers, as well as for preserving their primary contexts [2]. Also, it opens up the knowledge base, so acquired for geometric morphometrics, shape analysis, animation of 3D artefacts for public interpretations for education and entertainment [4].

The first phase of any task is dataset procurement. In the archaeological context this procurement is acquisition of artefacts, site maps, images, building plans, environment information etc. The basic process of 3D reconstruction from images is as shown in Fig. 1.

The next phase is of detection and mapping of the features in the site or artefacts. This is performed by finding features within the image sets. Relevant feature or control points are established and these are then registered or mosaic with one another to form an initial model of parse cloud point of the object or a relevant edge set of the relevant site model as the case may be. 3D reconstruction is performed by transpiring a mesh structure on top of the cloud points or edge set. After which texture mapping is done to give the look of a solidified object or site plan. Subsequently, texture mapping is performed to give the mesh structure an object or site resemblance.

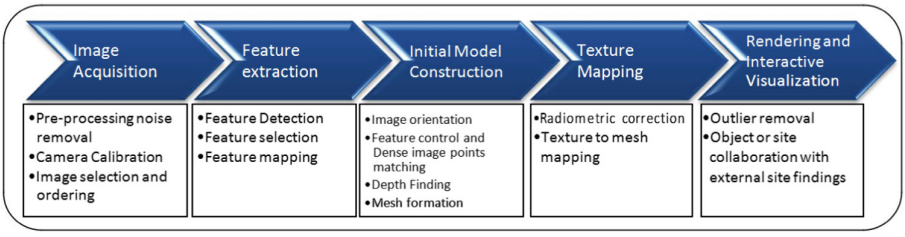


Fig. 1. Basic 3D reconstruction process

The final phase is of rendering of the textured map or object and enabling their interactive visualization. This reconstructed dataset may be used for viewing form multiple angles or may support any movement within the virtual site plan if designed properly.

3 Technologies in 3D Reconstruction for Archaeology

Archaeological recording is a fascinating and important aspect of capturing and acquiring archaeological findings. This spatial recording can be easily performed by using digital devices such as Total Station Theodolite (TST) and laser scanners, digital cameras or camcorders including 3D cameras, and Global Positioning Systems (GPS). While optical and acoustic sensors are used for unearthing underwater sites, satellite imagery helps in recording the terrestrial and aerial data, whereas the active sensors, mentioned earlier are used for capturing surface characteristics of the objects, artefacts, and sites. Many a times spatial recording involves the survey to combine rectified digital photography and real-time data along with inputs from site experts so as to ensure the quality of the input datasets. Also, to keep a track of the overall data collection, „working shots are recorded that record the excavation process itself. The datasets of the artefacts and the sites so collected are then used for feature mapping which can lead to object formation or virtual site designing, which may be extended to form a virtual reality tour for the complete project site under consideration.

3.1 Range-Sensor Devices

The type of sensors used for capturing data depends on factors such as project requirements and budget, required accuracy, site location and characteristics, object and artefact dimensions, etc. They are used in numerous fields of building and constructions, surveying and engineering, mining and excavation. Though these are generally much costlier, bulky and partial flaccidity compared to image-sensor devices, but they help in directly capturing data with higher accurate geometric details. Measurement of distance is accomplished by reading the object under survey's reflected signal's modulation pattern as is shown in Fig. 2.

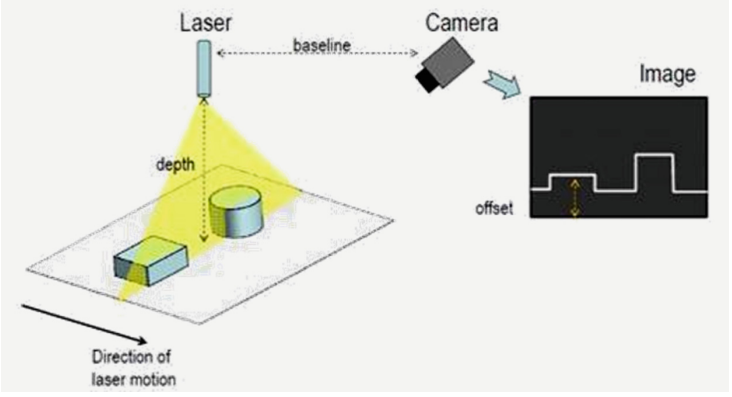


Fig. 2. System diagram for a laser scanner system [25].

Some of these range-sensor devices are elaborated as follows:

(a) Total Station Theodolite (TST): The total station is a range sensor device that uses an electronic theodolite integrated with an electronic distance meter to read slope distances from the instrument to a particular point. [13].
(b) Laser scanners: These scanners are used for collecting data in which the distance to a surface at each point in the picture is inherently imbibed so as to identify the 3D positioning of each point in the picture.
(c) LIDAR scanners: These sensors measure the distance by irradiating a target area or object with a laser and then analyzing the reflected light. These are usually employed for airborne data collection of a site and find much usage in the fields of geomatics, archaeology, geography, geology, geomorphology, seismology, forestry, remote sensing, atmospheric physics, etc. [6].
(d) Sonar sensors: Sonar sensor use the echo principle by sending out sound/acoustic energy waves underwater to locate objects and study their characteristics.
(e) Optical sensors convert light, or a change in light, into an electronic signal. The technique used for measuring the object characteristic is based on measuring an intensity change in one or more light beams or phase changes in the light beams.
(f) Time-of-flight scanners: These sensors resolve the distance between the object under consideration and the camera based on the known speed of light and by measuring the time-of-flight of a light signal between them for each point of the image as compared to bombarding the object with laser beam for data collection and is described in Fig. 3.

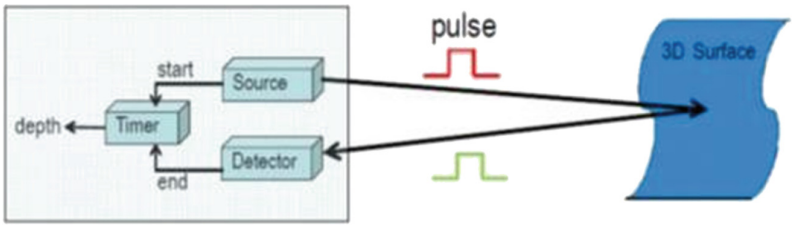


Fig. 3. System diagram for a pulsed light source TOF camera [25].

3.2 Image-Sensor Devices

Many a times, analogue and/or digital cameras of various configurations are used for data capturing from archaeological sites. These constituent the image-sensor devices in which the sensor detects and conveys the information as an image, usually of two dimensional nature. These are then used for extracting the 3D information of the objects by employing methods of stereo-imaging, etc.

3.3 Techniques for 3D Reconstruction

The aim is to construct 3-D structure(s) that may be resolved from a series of overlapping, offset images.

3.3.1 Object Reconstruction from Artefact Images

3D object reconstruction allows viewers to view the object from multiple angles, understand, assimilate and appreciate their nitty-gritty as per their own requirements without handling the objects personally or even within the confines of their own personal spaces.

Challenges

Object reconstruction is usually done by using data acquired through image-sensor devices, mostly using digital cameras of high resolution.

Methods

The methods used are required to construct the correspondence between the various images, of the same object, to project the information in a 3D space depicting the object in the virtual environment.

A. Structure from Stereo Vision

Structure from stereo vision refers to perception of depth and its subsequent usage for obtaining 3D information from, usually, two images. Just as within our eyes, calibrated cameras are used for obtaining image displace horizontally from one another, separated by a pre-defined distance as described using Fig. 4. This makes the system to view the surrounding from slightly different perspectives.

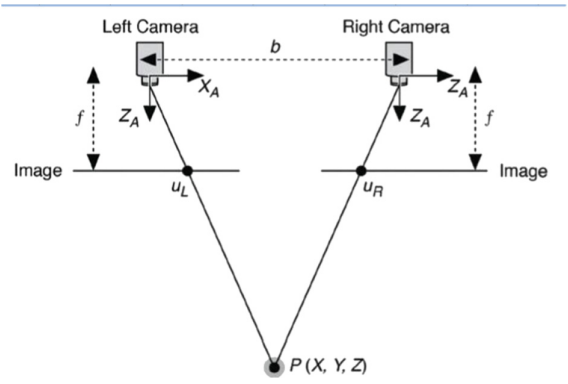


Fig. 4. A simplified stereovision system [24]

Distance between those two projected points “disparity”, is used to calculate the depth information, which is the distance between real-world point “P” and the stereo vision system.

Stereo vision from 2 images can be extended to multi-baseline stereo images that uses more than two images for depth finding and volumetric stereo that use a voxel volume to get a view-independent representation [5, 15].

B. Structure from Motion

Structure from motion refers to our cognitive understanding of an object and its environment by virtue of their motion relative to the observer. The information gathered over time, due to motion, is used for deciding the 3D data from a series of overlapping images [22].

3.3.2 3D Site Reconstruction

Reconstructing a complete site in virtual environment is a daunting task [10]. It requires not only proper excavation details, but also site maps and other historical input apart from the designing and visualization skills. This is detailed in Fig. 5 as shown below. The aim of such a reconstruction is not only to represent the archaeological entities accurately and elaborately, but also to enhance the historical events and their impact on the current land and culture [1].

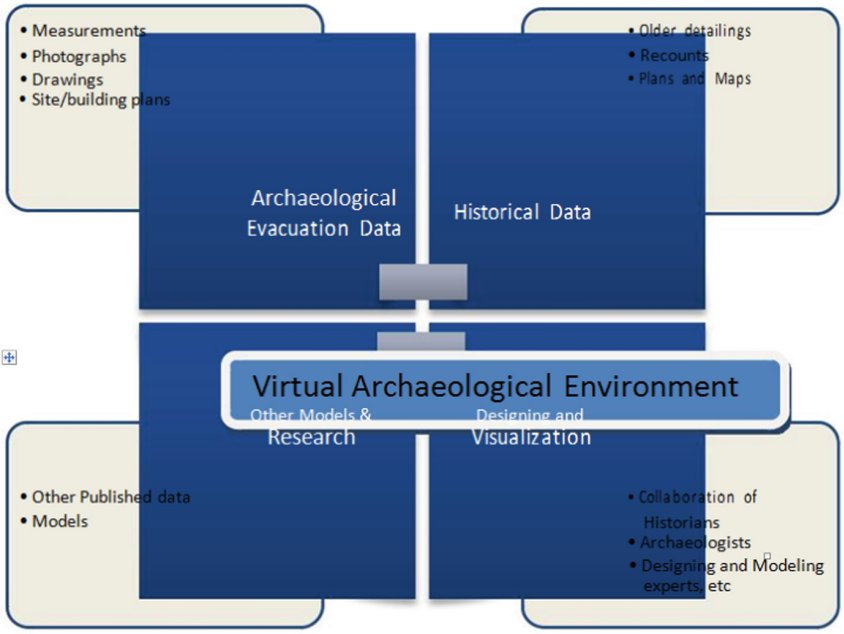


Fig. 5. Virtual Archaeological Environment Model creation requirement.

Challenges

Some of the challenges that arise are related to creating and mapping the virtual site plan to that of the original findings and integrating of the data sets from other sources into the environment. These challenges are encountered majorly in - (a) Data

acquisition from aerial and terrestrial images, and their noise removal, calibration and orientation; (b) Matching datasets to give proper correspondence; (c) Integration of data sets, their site plans with proper texturing; (d) Corrections to models; (e) integration 3D object and artefact sets within the 3D virtual environment; and (f) enabling of the walk through and interaction. Apart from the challenges related to creation of the virtual environment, there are other challenges that this virtualization might put forth include improper understanding of the complete environment, their archaeological tagging and metadata formation and finally collating data from multiple archaeological databases [1, 20].

Methods

One usual method of creating a virtual 3D archaeological site as reported in [11] is that of **Multi-resolution Modeling methodology** which was used for building the Forum in Pompeii which encompassed a large area in terms of both landscape as well as architectural requirements. The model creation employed a top-bottom style in which traditional aerial images were used for generic low-resolution viewing and enabled viewing of higher resolution geometric details by using range data and terrestrial images too depending on the level of detail required by the users. Apart from using Digital Surface Model (DSM) and underlying Digital Elevation Models (DEM), the method also used multi-point matcher to generate a dense point cloud for further usage in surface alignment, data editing and model building. This technique used the integration of different modeling mechanisms and processes like those of photogrammetry, active sensors, topographic surveying, etc. for resulting in an elaborative and fairly accurate model.

Another method involved using **Multi-layered interactive 3D model** for modeling of the interactive virtual Acropolis city [12]. This used the method of dividing the available information chronologically such that the city's growth could be visualized by inserting information chronologically and assimilating its development.

Such virtualization of ancient and archaeological findings help in making observations about the environment in whole and allow for evaluation of certain hypotheses related to archaeological concerns. Apart from this, computer reconstructions also facilitate understanding of relative features in similar situations or sites elsewhere allowing for learning to happen in multiple aspects [20]. In [12], authors have also incorporated the usage of Virtual reality within the 3D environment to develop a virtual system which can be accessed using devices such as mobiles or tablets too [18].

3.3.3 Underwater Surface Reconstruction

Computer vision algorithms like stereo triangulation and structure from motion cannot be employed directly underwater [9, 17]. Further calculations can also include methods such as point clouds for forming 3D object structure [26]. Point cloud generation utilizes the underlying methods of stereo triangulation. It is a network of points that is used to represent a 3D object. And maybe considered as its data structure.

Challenges

Underwater navigation and data acquisition for archaeological purposes is challenging not only due to the landscape but also due to water's inherent nature. Some very

common problems faced by researchers while data collection, though not restricted to, include [3, 7] – volatile and selective nature of the underwater optical effects, changes effected due to season or due to the day/night cycle, constituents of water and wastewaters. Also, the amplitude and orientation of the waves can introduce noise.

Methods

Apart from using sonar datasets, the methods of optical scanners are frequently used for underwater image acquisition as shown in Fig. 6 [3]. LED light sources are used for this purpose whose reflection is captured by the camera for getting the object immersed in the underwater environment. The image is captured in two separated channels of the light spectrum. The separation of the channels is assured by the use of optical filters. Usually near-infrared and visual spectrum of light are used.

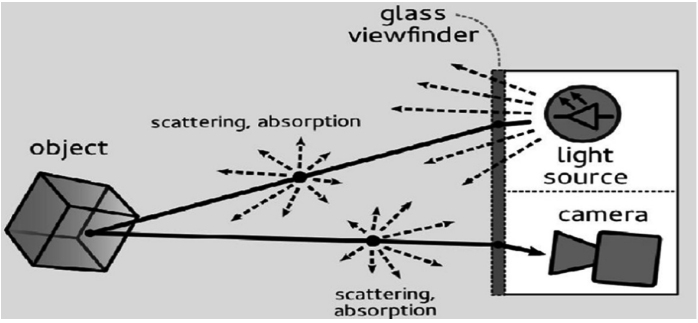


Fig. 6. Underwater image acquisition method.

Another method that has gained importance these data is using depth sensors for mapping the underwater immersed objects. Equipments such as Microsoft Kinect and Asus Xtion sensors are also in usage [7]. However, these sensors need to be held just above the water surface and cannot be immersed. Voxel hashing algorithm have been used for performing reconstruction [14] as it was found to be remarkably robust to noise even when depth image of the surface completely indistinguishable due to minor motions in the water waves. This noise is caused by the change in reflection due to the change in the angle between the water surface and the sensor's line of sight.

Due to the inherent difficulties in image acquisition, the system have a large scope for improvement mainly during the fusion and mapping phase wherein exact pixel corresponding is difficult.

4 Conclusion

In this paper we have discussed not only data acquisition methods but also technologies prevalent for development of 3D environment and 3D objects in various scenarios of landscape and terrestrial data but also from underwater excavation expeditions. Furthermore, we have presented challenges in the various techniques. These techniques

that utilize various hardware and software allow for users and researchers to achieve not only better understanding of the analytics of the architectural and structural requirements but also for interpretation of objects, artefacts, sites having unique archaeological features to be appreciated by multi-disciplinary experts. Though demonstrations of the digital reconstruction have their own fair share of drawbacks, they allow for exploration of such models for endless and non-invasive investigation and documentations too. Also, virtualization of the archaeological environment helps in addressing many socio-cultural issues and preserving of cultural heritage as well as generating public interest within such fields.

References

1. Barceló, J., Vicente, O.: Some problems in archaeological excavation 3D modelling. Bar Int. Ser. **1227**, 400–404 (2004)
2. Barrett, B.J.W.: Beyond documentation : 3D data in archaeology. In: Returning to Columbus, vol. 58, no. 4 (2014)
3. Biegański, W., Kasiński, A.: Image Acquisition in an Underwater Vision System with NIR and VIS Illumination (2014). arXiv preprint arXiv:1402.1151
4. Campana, S., Sordini, M., Rizzi, A., Remondino, F.: Geomatics techniques for the 3D documentation and visualization of archaeological building. In: The Conference, p. 151, December 2010
5. Chen, Q., Medioni, G.: A volumetric stereo matching method: application to image-based modeling. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1999, vol. 1. IEEE (1999)
6. Cracknell, A.P., Hayes, L.: Introduction to Remote Sensing, 2nd edn. Taylor and Francis, London (2007). ISBN 0-8493-9255-1. OCLC 70765252
7. Dancu, A., Fourgeaud, M., Franjcic, Z., Avetisyan, R.: Underwater reconstruction using depth sensors. In: SIGGRAPH Asia 2014 Technical Briefs, p. 2. ACM, December 2014
8. Evans, T.L.: Digital Archaeology: Bridging Method and Theory. Psychology Press, London (2006). Ferdani, D.: 3D reconstruction in archaeological analysis of medieval settlements (n.d.)
9. Foresti, G.L.: Visual inspection of sea bottom structures by an autonomous underwater vehicle. IEEE Trans. Syst. Man Cybern. Part B: Cybern. **31**(5), 691–705 (2001)
10. Frischer, B., Dakouri-Hild, A.: Reasoning in 3D: a critical appraisal of the role of 3D modelling and virtual reconstructions in archaeology. Beyond Illus.: 2D 3D Digit. Technol. Tools Discov. Archaeol. **1805**, 36–45 (2008)
11. Guidi, G., Remondino, F., Russo, M., Menna, F., Rizzi, A., Ercoli, S.: A multi-resolution methodology for the 3D modeling of large and complex archeological areas a multi-resolution methodology for the 3D modeling of large and complex. Int. J. Arch. Comput. **07** (01), 39–56 (2009)
12. Higgett, N., Baines, E., Everitt, D., Saucedo, G., Tatham, E.: Virtual Romans : virtual reconstruction of Roman Leicester (Ratae Corieltauvorum) 210AD, pp. 13–16 (2012)
13. Kavanagh, B.F., Glenn Bird, S.J.: Surveying principles and applications, 4th edn, pp. 257–264. Prentice Hall, Englewood Cliffs (1996)
14. Niessner, M., Zollhöfer, M., Izadi, S., Stamminger, M.: Real-time 3D reconstruction at scale using voxel hashing. ACM Trans. Graph. (TOG) **32**(6), 169 (2013)

15. Okutomi, M., Kanade, T.: A multiple-baseline stereo system. IEEE Trans. Pattern Anal. Mach. Intell. **15**(4), 353–363 (1993)
16. Premo, L.S.: Book: Beyond Illustration: 2D and 3D Digital Technologies as Tools for Discovery in Archaeology, vol. 1805. Archaeopress, Oxford (2008)
17. Schechner, Y.Y., Karpel, N.: Clear underwater vision. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, vol. 1, pp. I–536. IEEE, June 2004
18. Hermon, S.: Between the real and the virtual: 3D visualization in the archaeological research - expectations and prospects. Virtual Archaeol. Rev. **2**, 59–63 (2011)
19. Stanco, F., Tanasi, D.: Virtual acropolis. Digital recreation of a Sicilian Archaic sanctuary. Virtual Archaeol. Rev. **3**(5), 126–130 (2012)
20. Sylaiou, S., Patias, P.: Virtual reconstructions in archaeology and some issues for consideration. IMEROS: Annu. J. Cult. Technol. **4**(1), 180–191 (2004)
21. Tsiafaki, D., Michailidou, N.: Benefits and problems through the application of 3D technologies in archaeology: recording, visualisation, representation and reconstruction. Sci. Cult. **1**(3), 37–45 (2015). https://doi.org/10.5281/zenodo.18448
22. Westoby, M.J., Brasington, J., Glasser, N.F., Hambrey, M.J., Reynolds, J.M.: "Structure-from-Motion" photogrammetry: a low-cost, effective tool for geoscience applications. Geomorphology **179**, 300–314 (2012)
23. Zubrow, E.B.: Digital archaeology. In: Digital Archaeology. Bridging Method and Theory, pp. 10–31. London (2006)
24. A Guide to Stereovision and 3D Vision. https://www.techbriefs.com/component/content/article/ntb/features/feature-articles/14925. Accessed 27 Nov 2015
25. Science Behind 3D Vision. http://www.depthbiomechanics.co.uk/?p=102. Accessed 27 Nov 2015
26. Fan, H., Su, H., Guibas, L.J.: A point set generation network for 3D object reconstruction from a single image. In: CVPR, vol. 2, no. 4, p. 6, July 2017

Curvelet Based Thinning Algorithm

R. L. Jyothi[1(✉)] and M. Abdul Rahiman[2]

[1] College of Engineering Chengannur, University of Kerala, Thiruvananthapuram, India
jyothianil@gmail.com
[2] LBS Institute of Technology for Women, University of Kerala, Thiruvananthapuram, India

Abstract. Efficiency of a recognition algorithm depends upon availability of noiseless and classes of images with unique and differentiable features. Thinning plays an important part in building a recognition system for various types of images. It is mainly used in handwritten character recognition systems and Biometric systems. In Handwritten character recognition systems where there exist large variations within same class of images, thinning plays a vital role. Here curvelet based thinning algorithm is proposed which produces better results compared to other thinning algorithms. The capability of curvelet transform in estimation of directional features added to the strength of proposed work. Curvelet transform is combined with watershed algorithm, binarization and skeletonization to produce the final thinned output. The efficiency of the proposed work has been estimated in this work through detailed experimental analysis and comparisons with 7 efficient thinning algorithms. The measures used for comparison in this works are number of foreground pixels, PSNR (peak signal to noise ratio), MSE (mean square error), RMSE (root mean square error), reduction rate and thinness ratio.

Keywords: Curvelet transform · Watershed algorithm
Handwritten character recogntion systems · Palm leaves
Grantha script

1 Introduction

Thinning algorithm plays a significant role in building an efficient object recognition system especially biometric and hand written character recognition system (HCR). In handwritten character recognition system where there exist large variations in the same character depending upon the materials used for writing and writer's writing style, thinning is an important preprocessing step. Thinning is a morphological operation applied for structural analysis of an image. Thinning algorithms aims at producing skeletonized image of one pixel width by removing selected pixels from the image. An efficient thinning algorithm always

S. M. Thampi et al. (Eds.): SIRS 2018, CCIS 968, pp. 454–463, 2019.
https://doi.org/10.1007/978-981-13-5758-9_39

preserves overall structure of the image maintains characteristics like connectivity preservation, single pixel width preservation, prevention of excessive erosion and conservation of important attributes of an image. There exist various types of thinning algorithms but none of the existing algorithm work efficiently in all types of images. If statistical and geometrical recognition algorithm is taken into consideration none of the recognition system works efficiently without an appropriate thinning algorithm. Thinning algorithms are mainly characterised into iterative and non-iterative thinning algorithm. Iterative thinning algorithm works in a pixel by pixel manner [2]. It is classified into sequential and parallel thinning algorithm. In sequential thinning algorithm the algorithm works based on a fixed sequence. In a sequential algorithm current iteration depends upon all the iterations done till now. But parallel thinning algorithm works based on only the results produced in the previous iteration. Non-iterative algorithms does not work on pixel by pixel basis. The most common non iterative algorithms are medial transforms and distance transforms.

Here dataset used for experimental analysis are grantha characters extracted from palm leaves. Grantha script was the script used in southern part of ancient India to write sanskrit language. Example of a grantha palm leaf is shown in Fig. 1. This dataset has been chosen as there exist large variation for the same character in different palmleaves set. The variations in characters exist due to different materials used for writing. Due to these variations building an efficient recognition system for these characters are a complex task.

Eventhough there exist an ample number of algorithms for thinning purpose, none of the algorithms seems to be efficient in all types of images. Here a curvelet based thinning algorithm is proposed which works very efficiently in highly degraded grantha character images.

In Sect. 2 detailed literature survey of related work is revealed. Section 3 explains the proposed work on Curvelet Based Thinning Algorithm. In Sect. 4 detailed experimental analysis and comparison with other algorithms are demonstrated. Last Section depicts conclusion on the work.

2 Related Work

Thinning algorithms works by selective deletion of object pixels along the image until it becomes one pixel width. The implementation of thinning algorithm started with medial axis transformation [3,4].

Hilditch algorithm [5] is mainly used to thin the image edges. The algorithm checks all the pixels and decides to change a foreground pixel to background if the conditions specified below are satisfied. The first condition is that the number of nonzero pixels in the 8-neighbourhood of processing pixel P are in the range 2 to 6. The second condition specifies that the number of 0, 1 patterns in the 8-neighbourhood of processing pixel P is exactly one and the two pixel wide horizontal and vertical lines do not get completely eroded. The main problem of the method is that it works by thinning the image edges and is less concentrated on the whole image. This method always produces unwanted branches.

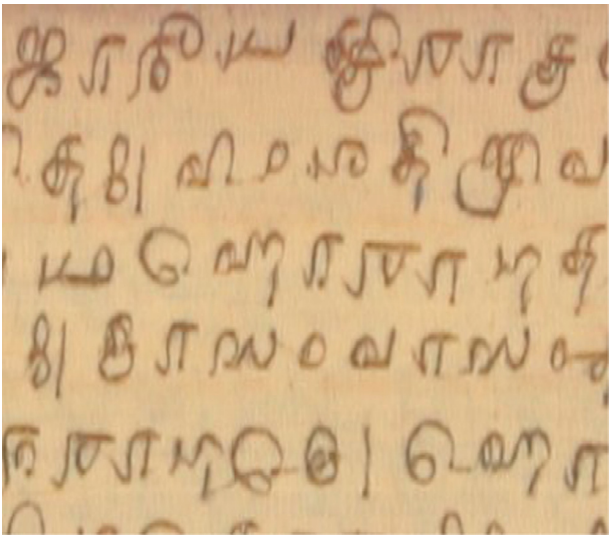


Fig. 1. Grantha palm leaf

Zhang suen (ZS) algorithm [6] is an example of parallel thinning algorithm. The algorithm is simple to implement and works fastly by performing sub iteration twice on the corresponding pixels. The principle used in first sub-iteration removes the south-east border pixels and in second sub-iteration which will removes the remaining north-west pixels. Finally it finishes the current iteration by removing all the marked pixels. This algorithm thins the digital patters to one pixel width and executes faster. The main drawback of this method is that it does not preserve 8-connectivity. Due to this it reduces long or short diagonal lines of 2-pixel width to a single dot or line consists of only two points. Another problem is that the obtained skeletons may suffer diagonal line preservation.

LW algorithm [7] proposed in order to solve the problem of ZS. The LW is a modified form of ZS. Here only condition (1) in the ZS algorithm $2 \leq N\ (Pi) \leq 6$ is replaced with $3 \leq N\ (Pi) \leq 6$ (i.e., the number of nonzero pixels in the 8-neighbourhood of processing pixel P are in the range 3 to 6). Using the modified condition, LW algorithm able to preserve 8-connectivity. So that the slant lines are retained but they are of two pixel width. In some cases algorithm ends with complete elimination of two-pixel wide horizontal and vertical lines. WHF algorithm [8] is a modified form of LW algorithm. It will resolve the problem of LW algorithm. In most cases it can produce one pixel width but sometimes it makes needless trees in the image and does not preserve 8-connectivity.

Enhanced parallel thinning algorithm [9] is an improvement over Zhang suen (ZS) algorithm. Here some additional conditions are added into each pass of the algorithm. This may helps to erase the remaining lines with two pixel width. This algorithm can extract one pixel slim lines by removing the undesirable branches in the image. It can also preserve the connectivity in the image.

In the Arabic parallel thinning algorithm [10] the input pattern is scanned from upper left corner to lower left corner in each passes. The algorithm contains 4 sub-iterations instead of 2 in the previous algorithms. Here the pixel connectivity and end point pixel preservation are checked. Since the algorithm passes through multiple iterations the presence of unwanted branches became reduced. This algorithm can gives one pixel width image but the image is not completely free from spurs.

Guo-Hall's thinning [11] is a thinning algorithm with two subiterations that works by preserving the end points and also removes the redundant pixels. Compared to ZS and LW, the algorithm produce thinner skeletons by deleting the redundant pixels. The algorithm gives results in less number of iterations but it does not preserve the structure.

Abdulla et al.'s Algorithm [12] is processed with two sub-iterations. In the first sub-iteration the image is scanned horizontally by using a $3 * 4$ sized block and then delete the horizontally isolated or adjacent points. Similarly the second sub-iteration scans the image vertically using a $4 * 3$ sized block and then delete the vertically isolated or adjacent points.

The HSCP algorithm [13] considers the information about the edges along with the values computed by placing a $3 * 3$ mask. This reduced the sub-iterations two to one as compared to ZS and LW algorithms. The algorithm computes the edge status of the current pixel and those three neighbors (S, E, and SE). Thus the condition formed by the conjunction of both edge and mask values are used to determine final thinning procedure. The algorithm does not preserve the end points but eliminates the staircase effect without eliminating the connectivity. In this algorithm 450 and 1350 inclined diagonal lines are reduced to two pixel width vertical lines.

The Hall's algorithm [14] proposed to preserve the diagonal lines in the image. Here all the marked pixels are deleted in parallel except certain pixels, which will preserve the connectivity in the image. The algorithm also preserves end points and shape of the diagonal lines.

In Rosenfeld parallel thinning algorithm [15] a subset of boundary pixels (contour pixels) of the object or a particular region is to be removed.

Nagendraprasad-Wang-Gupta [16] is a fast parallel thinning algorithm. It works very well in handwritten characters. Here the obtained skeleton is connected and due to the asymmetric nature of the algorithm some unwanted pixels are not removed. Modified NWG thinning makes the algorithm symmetric and which will eliminates the redundant pixels that are irrelevant to recognize the hand written character. In [19] a new algorithm is proposed which works by deleting successive contours of the set to be thinned while identifying pixels where disjoint parts of boundary have been mapped. In [17] a thinning algorithm is proposed to extract thin skeletons from object made of voxels. Algorithms produce curve or surface skeletons from object made of voxels. In [18] a new robust and efficient skeletonization algorithm for 3D objects based on minimum cost is introduced. In [20] a survey on skeletonization algorithms were performed and the principles, benefits and drawbacks were discussed. In [21] a one pass parallel

asymmetric thinning algorithm is proposed which has good noise resistance, produce perfect 8 connected skeleton and faster speed without severe erosion. In [22] a new algorithm for thinning is proposed and compared with K3M method for thinning, Huang method and ZS (Zhang Suen) method. The proposed method achieved superior topology preservation compared to other methods.

3 Curvelet Based Thinning Algorithm

The proposed method is Curvelet based Watershed thinning Algorithm, which improves all of above existing algorithms. The new algorithm preserves all the characteristics of a skeletonized image. They include basic requirement like connectivity, one pixel width and end point preservation. The proposed algorithm will go through various stages. Figure 2 shows the methodology of the proposed work. The first stage of the algorithm implements Curvelet Transform [1] on the binary image for extracting curvic structure of the image. The Curvelet is a multiscale representation suitable for images with curves and handles curve discontinuities well. Curvelets can reveal edges better than wavelets and is therefore well-suited for multiscale edge enhancement. Here first generation curvelet

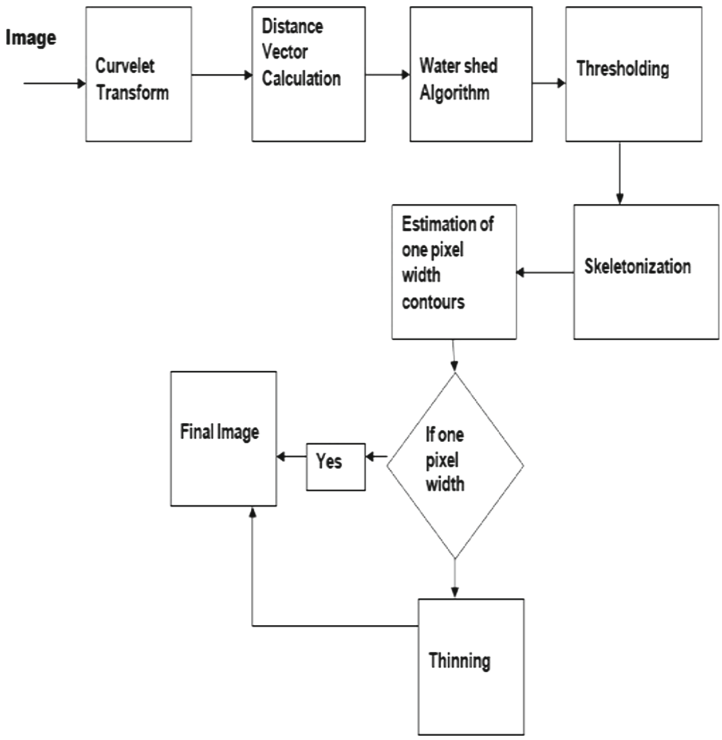


Fig. 2. Architecture: curvelet based thinning algorithm

is implemented. Curvelet transform is carried out by blockwise ridgelet analysis. Ridgelet analysis is 1-D wavelet transform of the concerned block in radon domain. During the next stage a distance vector of each pixel of image respective to the curvelet image is calculated and a new distance vector image is generated. Thus the resultant image is a restored form of input image. Next stage implements watershed algorithm on the distance vector image. The watershed image provides image contours for the skeletonization process. It will produce efficient connected boundaries than the traditional edge based techniques, which results in disconnected boundaries. The boundaries of the resulting regions always correspond to contours in the image. Next step calculates maximum intensity value from the watershed image and a thresholding operation based on this maximum value is performed. Finally the morphological thinning is performed on the resultant image by removing selected foreground pixels from the binary image and thereby producing skeletonized image with one pixel width. The deletion criteria use a $3 * 3$ neighborhood mask around the current pixel P. Thus a set of rules for deletion of foreground pixels are applied based on pixels in the neighborhood. Figure 3 shows a $3 * 3$ mask for pixel removal.

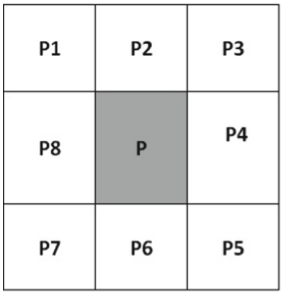


Fig. 3. 3×3 mask

Using this mask the current pixel P is deleted if the following conditions are satisfied:

1. $2 \leq N(P) \leq 6$
2. $S(P) = 1$
3. Apply the following
   (a) $P2 * P4 * P6 = 0$, $P4 * P6 * P8 = 0$; for 1^{st} sub iteration
   (b) $P2 * P6 * P8 = 0$, $P4 * P6 * P8 = 0$; for 2^{nd} sub iteration
   (c) $P2 * P4 * P8 = 0$, $P2 * P6 * P8 = 0$; for 3^{rd} sub iteration
   (d) $P2 * P4 * P6 = 0$, $P2 * P4 * P8 = 0$; for 4^{th}sub iteration
   Let N(P) be the number of nonzero pixels in the 8-neighborhood of processing pixel P in the range of 2 to 6, which ensures that no end point pixel and no isolated pixel is deleted. S(P) be the number of 0, 1 patterns in the 8-neighborhood of processing pixel P and it should be exactly one, which

provides a connectivity test, in which no image pattern is disconnected. The remaining conditions in each sub iterations perform removal of two pixel wide horizontal and vertical lines in all directions.

4 Experimental Analysis

Cropped grantha characters taken from a set of palmleaves are used for analysis purpose. The antiquated palmleaves are of 1000 years of age. The characters taken are highly degraded and exhibit high variations.

The proposed method is compared with some the existing efficient thinning algorithms. The proposed algorithm is compared with algorithms like Hilditch [5], ZS [6], LW [7], Guo-Hall [11], Abdulla's thinning algorithm [12], WHF [8] algorithm and two pass parallel algorithm [16]. The Fig. 4 shows the result of testing some of the existing efficient algorithms with the cropped Grantha Character 'Ka'. The algorithms were tested with 128 samples of each grantha character. From different analyzed results it can be concluded that Hiditch algorithm produces spaces in most of the images. Guo-Hall algorithm erases parts of slanting lines. LW algorithm does not produce 1 pixel width slanting lines in most image cases. In most cases the skeletons produced by LW algorithm are of low quality. In most image cases WHF and two pass algorithms produces branches in

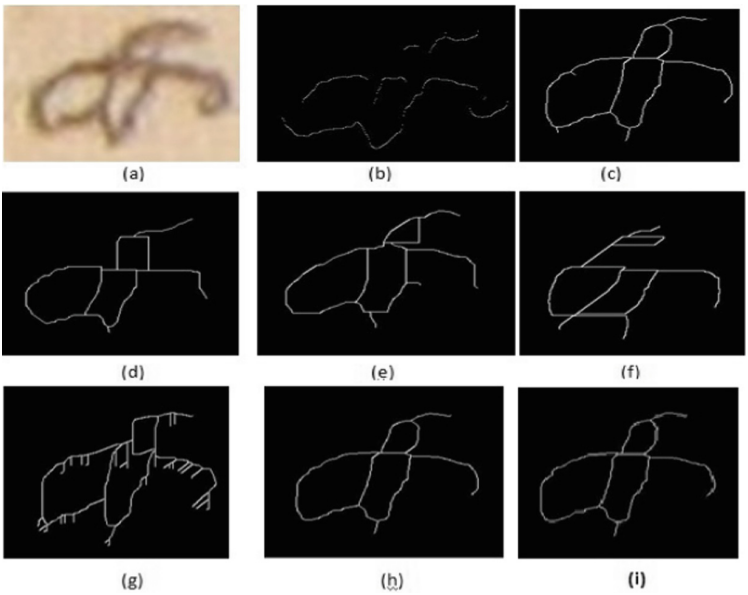


Fig. 4. Result of implementation of various algorithms on grantha character 'ka'. (a) input image (b) Abdulla's Method (c) LW Algorithm (d) Hilditch Algorithm (e) Two Pass Algorithm (f) Guo-Hall Algorithm (g) WHF algorithm (h) ZS algorithm (i) Curvelet Based proposed thinning algorithm

images. The algorithms are compared based on measures like Number of pixels, PSNR (Peak Signal to Noise Ratio), MSE (Mean Square Error), RMSE (Root Mean Square Error), reduction rate [23] and thinness ratio [23]. The calculation of reduction rate and thinness is done based on the original image while the calculation of PSNR, MSE and RMSE is done based on ground thruth images created based on degraded thinned version of original image. The mean value for Number of pixels, PSNR, MSE, RMSE, reduction rate and thinness for 128 samples of grantha character 'Ka' having an average value of 9975 for total number of foreground pixels is shown in Table 1. For a thinning algorithm to be efficient it should have minimum number of pixels, high PSNR, lower values for RMSE, MSE and with reduction rate and thinness should be nearer to 1. From the tabulated result it can be seen that the proposed algorithm produces minimum number of pixels compared to all existing algorithms, higher PSNR, lower values for RMSE, MSE and the thinness and reduction rate nearer to 1.

Table 1. Comparative result of average values of various measures for 128 samples of grantha character 'ka'

Algorithm	Number of pixels	PSNR	MSE	RMSE	Reduction rate	Thinness
Abdullas	758	51.4628	0.4643	0.6814	0.924	0.9799
Hilditch	594	54.4498	0.2334	0.4831	0.9405	0.9925
Two pass parallel	666	51.4628	0.4643	0.6814	0.9332	0.9862
Guo-Hall	696	51.4904	0.4150	0.6442	0.9302	0.9813
WHF	769	50.6165	0.5642	0.7511	0.9221	0.9779
ZS	654	51.3284	0.4789	0.6920	0.9344	0.9879
LW	650	53.4904	0.2911	0.5396	0.9348	0.9895
Proposed	562	57.8585	0.1001	0.3104	0.9437	0.9937

5 Conclusion

A thinning algorithm based on curvelet transform is simulated and compared with some of the efficient thinning algorithms. The proposed algorithm is compared with algorithms like ZS algorithm, LW algorithm, WHF algorithm, Abdullas algorithm, Guo-Hall algorithm and two pass parallel algorithm. The algorithms were evaluated based on measures like number of foreground pixels, mean square error, root mean square error and peak signal to noise ratio. It has been estimated through experimental analysis that the proposed curvelet based thinning algorithm exhibit significant performance over other algorithms based on the evaluated measures.

References

1. Ma, J., Plonka, G.: The curvelet transform. IEEE Sig. Process. Mag. **27**, 118–133 (2010)
2. Kumar, H., Kaur, P.: A comparative study of iterative thinning algorithms for BMP images. Int. J. Comput. Sci. Inf. Technol. **2**, 2375–2379 (2011)
3. Blum, H.: A transformation for extracting new descriptors of shape. In: Symposium on Models for the Perception of Speech and Visual Form. MIT Press, Cambridge (1964)
4. Palvidis, T.: Structural Pattern Recognition. Springer, Heidelberg (1977). https://doi.org/10.1007/978-3-642-88304-0
5. Hilditch, C.J.: Linear skeletons from square cupboards. In: Machine Intelligence, pp. 403–420. University Press, Edinburgh (1969)
6. Zhang, T.Y., Suen, C.Y.: A fast parallel algorithm for thinning digital patterns. Commun. ACM **27**, 236–239 (1984)
7. Lu, H.E., Wang, P.S.P.: An improved fast parallel algorithm for thinning digital patterns. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 364–367 (1985)
8. Wang, P.S.P., Hui, L., Fleming, T.: Further improved fast parallel thinning algorithm for digital patterns. In: Computer Vision Image Processing and Communication Systems and Applications, pp. 37–40 (1986)
9. Kwon, J.-S., Gi, J.-W., Kang, E.-K.: An enhanced thinning algorithm using parallel processing. In: Proceedings of 2001 International Conference on Image Processing (2002)
10. Tellache, M., Sid-Ahmed, M., Abaza, B.: Thinning algorithms for Arabic OCR. In: Proceedings of IEEE Pacific Rim Conference on Communications Computers and Signal Processing, pp. 248–251 (1993)
11. Guo, Z., Hall, R.W.: Parallel thinning with two subiteration algorithms. Commun. ACM **32**, 359–373 (1989)
12. Abdulla, W.H., Saleh, A.O.M., Morad, A.H.: A preprocessing algorithm for handwritten character recognition. Pattern Recogn. Lett. **7**, 13–18 (1988)
13. Holt, C.M., Stewart, A., Clint, M., Perrott, R.H.: An improved parallel thinning algorithm. Commun. ACM **30**, 156–160 (1987)
14. Hall, R.W.: Fast parallel thinning algorithms: parallel speed and connectivity preservation. Commun. ACM **32**, 124–131 (1989)
15. Rosenfeld, A.: Connectivity in digital images. J. ACM **17**, 146–160 (1970)
16. Nagendraprasad, M.V., Wang, P.S.P., Gupta, A.: Algorithms for thinning and Rethickening of binary images. In: Digital Signal Processing, vol 3, pp. 97–102. MIT Press (1993)
17. Couprie, M., Bertrand, G.: Asymmetric parallel 3D thinning scheme and algorithm based on isthumes. Pattern Recogn. Lett. **76**, 22–31 (2016)
18. Jin, D., Iyer, K.S., Chen, C., Hoffman, E.A., Suha, P.K.: A robust and efficient curve skeletonization algorithm for tree like objects using minimum cost paths. Pattern Recogn. Lett. **76**, 32–46 (2016)
19. Pavlidis, T.: A thinning algorithm for discrete binary images. J. Comput. Graph. Image Process. **13**, 142–157 (1980)

20. Saha, P.K., Borgefors, G., di Baja, G.S.: A survey on skeletonization algorithms and applications. Pattern Recogn. Lett. **76**, 3–12 (2016)
21. Dang, W.: A fast parallel thinning algorithm for binary image skeletonization. Int. J. High Perform. Comput. Appl. **14**, 65–81 (2000)
22. Abu-Ain, T., Abdullah, S.N.H.S., Bataineh, B., Omar, K.: A fast and efficient thinning algorithm for binary images. J. ICT Res. Appl. **7**, 205–216 (2013)
23. Gramblicka, M., Vasky, J.: Comparison of thinning algorithms for vectorization of engineering drawings. J. Theor. Appl. Inf. Technol. **94**(2), 265–275 (2016)

Author Index