H.S. Saini
R.K. Singh
K. Satish Reddy   *Editors*

Innovations in Electronics and Communication Engineering

Proceedings of the Fifth ICIECE 2016

Springer

Lecture Notes in Networks and Systems

Volume 7

The series "Lecture Notes in Networks and Systems" publishes the latest developments in Networks and Systems—quickly, informally and with high quality. Original research reported in proceedings and post-proceedings represents the core of LNNS.

Volumes published in LNNS embrace all aspects and subfields of, as well as new challenges in, Networks and Systems.

The series contains proceedings and edited volumes in systems and networks, spanning the areas of Cyber-Physical Systems, Autonomous Systems, Sensor Networks, Control Systems, Energy Systems, Automotive Systems, Biological Systems, Vehicular Networking and Connected Vehicles, Aerospace Systems, Automation, Manufacturing, Smart Grids, Nonlinear Systems, Power Systems, Robotics, Social Systems, Economic Systems and other. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution and exposure which enable both a wide and rapid dissemination of research output.

The series covers the theory, applications, and perspectives on the state of the art and future developments relevant to systems and networks, decision making, control, complex processes and related areas, as embedded in the fields of interdisciplinary and applied sciences, engineering, computer science, physics, economics, social, and life sciences, as well as the paradigms and methodologies behind them.

H.S. Saini · R.K. Singh · K. Satish Reddy
Editors

Innovations in Electronics and Communication Engineering

Proceedings of the Fifth ICIECE 2016

Springer

Editors
H.S. Saini
Guru Nanak Institutions
Ibrahimpatnam, Telangana
India

K. Satish Reddy
Guru Nanak Institutions
Ibrahimpatnam, Telangana
India

R.K. Singh
Guru Nanak Institutions
Ibrahimpatnam, Telangana
India

Printed on acid-free paper

Preface

The Fifth International Conference on Innovations in Electronics and Communication Engineering (ICIECE 2016) by Department of Electronics and Communication Engineering, Guru Nanak Institutions Technical Campus, was conducted in association with Defense Research and Development Organization and Department of Science Technology on July 8 and 9, 2016. The conference has grown exponentially over the years and has become a platform for scientists, researchers, academicians, and students to present their ideas and share their cutting-edge research in various fields of electronics and communication engineering. The focus for this year conference was "Smart Cities" and we had distinguished speakers from India and abroad who shared innovative solutions and technologies.

Over 290 papers were received of which 100 papers have been selected for Springer Conference Proceedings publication. The papers selected were presented by the authors during the conference. Parallel sessions were conducted to accommodate all the authors and to give them ample time to discuss their ideas. We were happy to note that all the authors were happy with the arrangements and encouraged us to conduct such conferences in the future as well.

We would like to thank all the participants, speakers, session chairs, committee members, reviewers, international and national board members, Guru Nanak Institutions Management, and all the people who have directly or indirectly contributed to the success of this conference. The editors would also like to thank Springer Editorial Team for their support and for publishing the papers in "Lecture Notes in Networks and Systems" series.

Ibrahimpatnam, India
<div align="right">

H.S. Saini
R.K. Singh
K. Satish Reddy
</div>

Committee

Editorial Board Members

Patrons
Sardar Tavinder Singh Kohli, Chairman, Guru Nanak Institutions
Sardar Gagandeep Singh Kohli, Vice-Chairman, Guru Nanak Institutions

Conference Chair
Dr. H.S. Saini, Managing Director, Guru Nanak Institutions

Conference Co-chairs
Prof. R.K. Singh, Associate Director, GNITC
Dr. S.V. Ranganayakulu, Dean-R&D, GNITC

Steering Committee
Dr. B. Veeranna, Director, GNITC
Dr. S. Sreenatha Reddy, Principal, GNITC
Prof. Parthsaradhy, Associate Director, GNITC
Dr. Rishi Sayal, Associate Director,GNITC
Dr. M. Narendra Kumar, Vice-Principal, GNITC

Technical Committee
Dr. Madhavi Latha, JNTU, Hyderabad
Dr. P. Chandra Sekhar, OU, Hyderabad
Prof. V. Rajanikanth, SKIT, Srikalahasti
Prof. A.S. Sai Prasad, VCE, Hyderabad
Dr. K. Jaya Sankar, IETE, Chairman, Hyderabad
Dr. Elizabeth Rani, Head - Electronics and Instrumentation, GITAM University, Visakhapatnam
Dr. V. Malleswara Rao, GITAM
Dr. K. Manjunath Achari, GITAM

Convenors
Dr. K. Satish Reddy, Professor, GNITC
Dr.K. Santhi, Professor, GNITC
Prof. M.A. Khadar Baba, HOD-ECE, GNITC
Prof. B. Kedarnath, HOD-ECE, GNITC

Committee Members
Prof. S. Maheshwara Reddy, HOD-ECE, GNITC
Prof. Hemasundara Rao
Prof. A. Mohan
Prof. K. Nagaiah
Prof. E.N.V. Purna Chandra Rao
Prof. B. Krishna Kumar
Mr. Sandeep Patil, Associate Professor
Mr. S.K. Saidulu, Associate Professor

Advisory Board-International
Prof. Libor Rufer, CIMA, Micro & Nano Systems Groups, France
Dr. Hasan Sevim, Dean, School of Engineering, Southern Illinois University, Edwardville, USA
Dr. D.D. Ebnizer, Associate Editor, JASA, USA
Dr. C. Satish Kumar, Battery Research-Engineer, Ford Motors Co., USA
Prof. Dr. Wei-Hsiang(Lewis), National Cheng Kung Univeristy, Tainan, Taiwan
Dr. Fauziah Haji Abdul Aziz, University Malaysia Sabha, Malaysia
Dr. Raj Kumar Patro, Nizwa College of Technology, Oman
Dr. Kuei-Hsien Chen, National Taiwan University, Taiwan
Dr. Bhanu Bhaskara Murthy Majmaah University, Saudi Arabia
Dr. G. Satish Reddy, Scientific Advisor to Ministry of Defense
Dr. U. Chandrasekhar, Prochancellor VELTECH University, Tamilnadu
Dr. E.S.R. Rajgopal, IISC, Bangalore
Dr. P. Subba Rangaiah, Director, VEDA IIT

Advisory Board-National
Dr. Sheela Rani, Vice Chancellor, Satyabhama University Chennai
Dr. M Mukunda Rao, Professor Emeritus, IIT, Chennai
Dr. N.V.L. Narasimha Murthy, IIT Bhuvaneswar
Dr. M. Ravi Babu, IIT Ropar
Dr. Srinivasulu Thadishetti, Principal, K.U
Dr. Mahaveer Singh, Sr. Principal Scientist, National Physical Laboratory, CSIR, New Delhi
Dr. B.N. Bhandari, JNTUH, Hyderabad
Dr. L. Pratap Reddy, JNTUH, Hyderabad
Dr. D. Srinivasa Rao, JNTUH, Hyderabad
Dr. Ajish K. Abraham, HOD-Electronics, AIISH, Mysore
Dr. P. Lakshmi Narayana, Osmania University, Hyderabad

Dr. Manoj Kumar, Principal DAVIET, Jalandhar
Dr. Pamela Chawla, Director, Institute of Engineering, MAU, H.P.
Dr. Novalina Jacob, Scientist, ADRIN, Hyderabad
Dr. Satish J. Sharma, RTM Nagpur University Nagpur
Dr. A. Bhujanga Rao, Department of Instrumentation, A.U. Visakhapatnam
Dr. N.K. Narayanan, Kannur University, Kerala
Dr. N.M. Renukappa, Professor and Scientist, JSS Technical Institutions, Mysore

Conference Committee Members

S. No	Name of the committee	Name of the faculty	
		Coordinator	Members
1.	Finance Committee	Mr. Sandeep Patil	D. Varaprasad
2.	Programme Committee	Prof. A. Mohan Dr. Sathishkumar B. Srinivas	K. Shashidar B. Anitha Abhijith Dey B. Aruna
3.	Invitation/Reception Committee	Prof. Anithaswamidas	Prof. Nagaiah-Delivery G. Kiranmayee V. Ramadevi N. Ramya Teja
4.	Registration Committee	Dr. S.J. Sukumar	M. Sirisha C. Sailaja V. Rajesh A.V. Rameshwar Roa Naveen Murali krishna
5.	MOC Committee - Inaugural/Panel Discussion	Mr. Sandeep Patil	Mrs. ManinderKaur K. Sujatha
6.	Proceeding Committee	Prof. KrishnaKumar	B. Srinivas N. Srikanth prasad
7.	Pre Conference tutorial	Md. Shoukath Ali R.P. Naraiah	K. Sathiya Surjith D. Naresh C. NaveenKumar Srinivas Nanda
8.	Panel Discussion	Dr. S.J. Sukumar	Prof. KrishnaKumar Abhijith Dey
9.	Food Committee	B. Sridhar	R. Gopinath
10.	Accommodation Committee	Prof. C. Hemasundera Rao	1. Chinna Narasimhulu 2. O.Ravinder
11.	Decoration/Seating Arrangement/ Banners	N. Srikanth prasad	
12.	PA & Audio Visual Committee	Mr. Raju	1. Electrician from maintenance
13.	VIP Committee	Prof. V. Bhagya Raju	Mr. D. Surendra rao

(continued)

S. No	Name of the committee	Name of the faculty	
		Coordinator	Members
14.	Transport Committee	Prof. P.C. Roa	1. P. Bhargav 2. G. Srinivas 3. G. Santhosh kumar
15.	Certificate/ Momento Committee	G. Kiranmayee	1. S. Swetha 2. Ramyasree 3. N.V.S. Murthy
16.	Purchase Committee	Prof. Maheshwara Reddy	1. S. Saidulu 2. V. Sai Babu
17.	IT Committee	Mr. D. Sriharsha	1. V. SaiBabu 2. IT services (2 members)
18.	Health care	Mr. Raju Ramyasree	
19.	Press Release for Conference	B. Sridhar	
20.	Conference Office	Prof. Anithaswamidas	1. Simarpreet kaur 2. Vinisha

Reviewer's List

Name of Internal Reviewers

S. No	Name	Address
1.	Prof. R.K. Singh	Associate Director
2.	Dr. S.V. Ranganayakulu	Dean/R&D, GNI
3.	Prof. M.A. Kadar baba	HOD/ECE, GNITC
4.	Dr. K. Satish reddy	Professor/ECE
5.	Dr. K. Santhi	Professor/ECE
6.	Dr. S. Sukumar	Professor/ECE
7.	Dr. Syed Jalal Ahmad	Professor/ECE
8.	Prof. Anita Swamidas	Professor/ECE
9.	Prof. CH. Hemasundara Rao	Professor/ECE
10.	Prof. N.V. Purnachandar Rao	Professor/ECE
11.	Prof. A.Mohan	Professor/ECE
12.	Mr. V. Bhagya Raju	Professor/ECE
13.	Mr. K. Shashidhar	Associate Professor/ECE
14.	Mr. B. Sridhar	Associate Professor/ECE
15.	Mrs. G. Kiran maye	Associate Professor/ECE
16.	Mr. Bachu Srinivas	Associate Professor/ECE
17.	Dr. N. Satish Kumar	Professor/ECE, GNIT
18.	Dr. R. Vinoth Kanna	HOD/EEE, GNIT

Name of External Reviewers

S. No	Name	Address
1.	Dr. S.P. Singh	HOD/ECE, MGIT, HYDERABAD
2.	Dr. B. Thiyaneswaran	Assistant professor, Sona college of Technology, Salem, Tamilnadu
3.	Dr. C. Rajasekaran	Professor/ECE, KSRCT, Namakkal, TN
4.	Dr. S. Arul Jyothi	Associate Professor, SREC, Coimbatore
5.	Dr. M. Nirmala	Associate Professor, KSRCT
6.	Dr. Somasundaram	Associate Professor
7.	Dr. P.G. Geetha	Professor/ECE, Sree Vidythanikethan Engineering college, Tirupathi
8.	Dr. M. Maheswaran	Kongu engineering college, Erode,Tamil Nadu
9.	Dr. S. Gunavathi	Associate Professor VIT, Vellur
10.	Dr. K. Venkatesh Sharma	Professor/CSE, Sri Indu college of Engineering and Technology

National and International Advisory Board List

International Advisory Board for ICIECE 2016

S. No.	Name	Affiliation	E-Mail
1.	Prof. Libor Rufer	CIMA, Micro & Nano Systems Groups, Francisis	libor.rufer@imag.fr
2.	Dr. Hasan Sevim	Dean, School of Engineering, Southern Illinois University, Edwardville, USA	sevim@engr.siu.edu
3.	Dr. D.D. Ebnizer	Associate Editor, JASA, USA	d.d.ebenezer@gmail.com
4.	Dr. C. Satish Kumar	Battery Research Engineer, Ford Motors Co., USA	satishbc@gmail.com
5.	Prof. Dr. Wei-Hsiang (Lewis)	National Cheng Kung Univeristy, Tainan, Taiwan	whlai@mail.ncku.edu.tw
6.	Dr. Fauziah Haji Abdul Aziz	University Malaysia Sabha, Malaysia	afauziah@upnm.edu.my
7.	Dr. Raj Kumar Patro	Nizwa College of Technology, Oman	rajkumar.patro@nct.edu.om
8.	Dr. Kuei-Hsien Chen	National Taiwan University, Taiwan	chenkh@pub.iams.sinica.edu.tw
9.	Dr. Bhanu Bhaskara Murthy	Majmaah University, Saudi Arabia	b.bhaskara@mu.edu.sa
10.	Dr. G. Sateesh Reddy	Scientific Advisor to Min. of Defense, India	gsreddy@rcilab.in

(continued)

S. No.	Name	Affiliation	E-Mail
11.	Dr. U. Chandrasekhar	Prochancellor VELTECH University, T.N	rapidchandra@gmail.com
12.	Dr. E.S.R. Rajgopal	Indian Institute of Science, Bangalore, Karnataka	gopal@physics.iisc.ernet.in
13.	Dr. P. Subba Rangaiah	Director, VEDA IIT, Hyderabad, Telangana	profksr@vedaiit.com

National Advisory Board for ICIECE 2016

S. No.	Name	Affiliation	E-Mail
1.	Dr. Sheela Rani	Vice Chancellor, Satyabhama University Chennai	vc@sathyabamauniversity.ac.in
2.	Dr. M. Mukunda Rao	Professor Emeritus, IIT, Chennai	mmrao@ee.iitm.ac.in
3.	Dr. N.V.L. Narasimha Murthy	Professor, IIT Bhuvaneswar	murthyn@iitbbs.ac.in
4.	Dr. M. Ravi Babu	Professor, IIT Ropar	ravi@iitrpr.ac.in
5.	Dr. Srinivasulu Thadishetti	Principal, Kakatiya University, Telangana	principal.kucet@gmail.com
6.	Dr. Mahaveer Singh	Sr. Principal Scientist, National Physical Laboratory, CSIR, New Delhi	abc..@nplindia.org
7.	Dr. B.N. Bhandari	Professor, JNTUH, Hyderabad	bnb@jntuh.ac.in
8.	Dr. L. Pratap Reddy	Professor, JNTUH, Hyderabad	prataplr@jntuh.ac.in
9.	Dr. D. Srinivasa Rao	Professor, JNTUH, Hyderabad	dsraoece@jntuh.ac.in
10.	Dr. Ajish K. Abraham	HOD-Electronics, AIISH, Mysore	ajish68@aiishmysore.in
11.	Dr. P. Lakshmi Narayana	Osmania University, Hyderabad	laxminarayana@osmania.ac.in
12.	Dr. Manoj Kumar	Principal DAVIET, Jalandhar	drmanojkumarindia@gmail.com
13.	Dr. Pamela Chawla	Director, Institute of Engineering, MAU, H.P.	pamela_chawla@yahoo.com
14.	Dr. Novalina Jacob	Scientist, ADRIN, Hyderabad	novalinajacob@gmail.com

(continued)

(continued)

S. No.	Name	Affiliation	E-Mail
15.	Dr. Satish J. Sharma	RTM Nagpur University Nagpur	sjsharma@rediffmail.com
16.	Dr. A. Bhujanga Rao	Department of Instrumentation, A.U. Visakhapatnam	dr_abrao@yahoo.co.in
17.	Dr. N.K. Narayanan	Kannur University, Kerala	nknarayanan@gmail.com
18.	Dr. N.M. Renukappa	Professor & Scientist, JSS Technical Institutions, Mysore	reunijagal@gmail.com

Contents

About the Editors

Dr. H.S. Saini is Managing Director of Guru Nanak Institutions and obtained his Ph.D. in the field of computer science. He has over 22 years of experience at University/College level in teaching UG/PG students and has guided several B.Tech., M.Tech., and Ph.D. projects. He has published/presented more than 30 high-quality research papers in international and national journals, and proceedings of International Conferences. He is the editor for Journal of Innovations in Electronics and Communication Engineering (JIECE) published by Guru Nanak Publishers. He has two books to his credit. Dr. Saini is a lover of innovation and is an advisor for NBA/ NAAC accreditation process to many Institutions in India and abroad.

Prof. R.K. Singh is Associate Director of Guru Nanak Institutions Technical Campus, is an alumnus of REC (Now MNIT, Jaipur), and did his M Tech from IIT Bombay, in the field of communication engineering. Prof. Singh has served Indian Army in the core of electronics and mechanical engineering for 20 years before hanging his uniform as Lt. Col. He has rich industrial experience as Army Officer Managing Workshops and has been teaching faculty for more than six years while in services. He started his career as teaching assistant at MNIT, Jaipur, for one year before joining the army. As professor, he has served for more than eleven years after premature retirement from the

army services. He has served as HOD, vice principal, and principal of engineering college before being approved as Associate Director of this institute.

The professor had hands-on experience on high-tech electronic equipments and has done many courses on radars and simulators. He has published many papers on microstrip antennas, VLSI, and radar signal processing in national and international conferences. He is presently pursuing his PhD from GITAM University.

Dr. K. Satish Reddy did his B.Tech (ECE) from VNR Vignana Jyothi Institute of Engineering and Technology (JNTUH) in the year 2001, Master of Science (VLSI Design) from University of Texas at Dallas, USA, in the year 2004, Master of Science (Optical Science and Engineering) from University of New Mexico Albuquerque USA, in the year 2009, and PhD (Microwave Engineering) in the year 2014. He has got thirteen national and international publications under his credit. He is Convener of International Conference on Innovations in Electronics and Communication Engineering and he is also Assistant Editor for Journal of Innovations in Electronics and Communication Engineering. He is acting as reviewer for number of national and international reputed journals. His interest areas are wireless communications and signal processing. He has guided more than 45 B.Tech and 12 M.Tech projects in his 12 years of teaching and research experience.

Part I
Devices, VLSI and Embedded System

Redesigned and Reinforced Low-Power Data Encryption Standard Using Hash Codes for High-Speed and High-Secure Applications

Thrived Dharbhashayanam, K. Manjunatha Chari and Pramod Dharbhashayanam

Abstract Data Encryption Standard was the first block cipher which was widely used for 25 years. In this paper, we propose to build a new and strong 128-bit block cipher based on DES which can be safely used wherever DES was used without compromising security. Reinforcing of DES weak areas such as key generation and Static S-boxes and also redesigning with the help of hash codes are proposed to make DES-128 strong enough to competitors. Detailed cryptanalysis is also provided on the new cipher along with other results. Xilinx XC3S500 device of Spartan family is used for hardware.

1 Introduction

In the overwhelming technological world, sending raw data is always a high threat as the Internet usage explosively expanded. Sending your data in disguised format is much needed, and cryptography deals with this area.

Encryption standards are there from ages in one or other forms. DES [1] was much popular and widely employed all over the world. Though there were several theoretical explanations which showcased DES weakness, it was succesfully brute-forced with costliest equipment in 22 h and 15 min. This kind of breaking is not always possible with the cost and time in consideration in real time. DES can

T. Dharbhashayanam (✉) · K. Manjunatha Chari · P. Dharbhashayanam
Department of ECE, GITAM School of Technology, Hyderabad, AP 502329, India
e-mail: thrived.d@gmail.com

K. Manjunatha Chari
e-mail: manjunath.gitam@rediffmail.com

P. Dharbhashayanam
e-mail: pramod.cbit@outlook.com

still be used safely for moderate secure application. But to reuse this algorithm in all applications, we need to make it to strong, secure and speed [2, 3].

DES can still be an attractive solution for power-efficient designs such as smart cards [4], RFIDs as the Industry-Standard algorithm AES is still trying to get low-power-efficient model.

This paper is an attempt towards making the current DES strong, secure and high speed using hash codes. Though there are several attempts [5–7], unfortunately none addressed the core issues of the DES algorithm.

When we scanned the DES algorithm at microscopic level [1], observed the various cryptanalysis [8] done on DES and also studied various theoretical attacks, we concluded to state the defective areas of DES algorithm are as below:

1. 64 (56 effectively)-bit key length;
2. Key generation;
3. Static S-boxes.

We propose mechanisms to answer these three issues by redesigning and reinforcing the DES algorithm such that the improved DES can be used elsewhere without any hesitation like AES.

The remaining structure of the paper goes like this. In Sect. 2, we dig out DES algorithm and identify and illustrate the weak areas. In Sect. 3, we discuss various redesign and reinforce mechanisms to resolve these three issues and to propose DES in new avatar. In Sect. 4, we detail the cryptanalysis done on the new real improved DES when compared to DES and show that how robust it is along with other results. Later, we conclude in Sect. 5.

2 DES and Its Weak Areas

2.1 Overview of DES

DES is designed in such a way that it should not be broken easily and also should sustain for longer time by withstanding all new techniques to break it. It successfully survived through the tremendously speeding span of era and stood as monopoly in encryption standards for more than 2 decades.

DES was designed to resist against all attacks by then except brute-force attack as the exhaustive search was though near to impossible with 2^{56} combinations at that time. DES was the first block cipher which powerfully implemented confusion and diffusion principle of cryptography.

At the time of DES designing differential attacks are well-known [9] and hence designed the 16 rounds iterative algorithm such that it's robust against Differential attacks. According to experiments conducted till now, 2^{47} chosen plaintexts needed to break DES in differential cryptanalysis way which is a strong performance to say. We will discuss this in detail in Sect. 4. DES takes 64-bit plaintext and 64-bit key to

give 64-bit cipher text. Hence, its block size is 64. And the algorithm is briefed as below as in [1] (Fig. 1).

The 64-bit plaintext will be given to IP box which shuffles the data and then halved to two parts. We always perform much complex operations on right half and later swap such that alternatively data are encrypted in full fashion. Before the rounds start, a non-dependent key generation algorithm will run to produce 16 keys one for each round from the parent key which is supplied. This key and right half of the round input will be given to Fiestel box or F-box which carries crux of encryption. In this box, 32-bit message will be expanded and will be mixed with key, and later in a nonlinear approach, we deduce 32 bits from the 48 bits (through substitution box or S-box) which will be sent as left part of next round. This operation called as round logic will be performed for 16 times, and at the last round, we do one more swap and one more permutation which is actually inverse of initial permutation, and the final clubbed 64-bit output is given out as ciphertext. The same process will be performed with just reverse of keys for decryption process.

The same block does both encryption and decryption with only difference of keys order. It is an excellent mechanism with which we provide efficiency and extreme security in one shot unlike Public-Key ciphers. Various shuffling structures

Fig. 1 DES algorithm
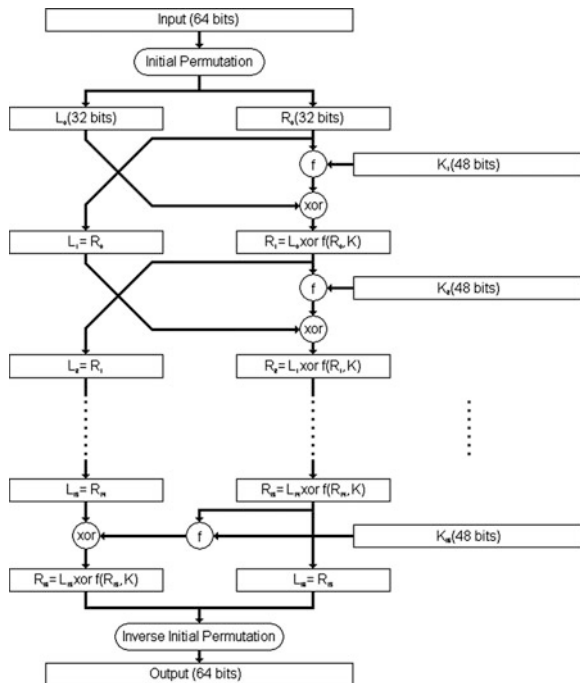
available in algorithm are confusion mechanisms which are linear structures, and the S-box is implementation of diffusion which is nonlinear structure.

2.2 DES Weak Areas

On careful study of the algorithm and various attacks happened, we would like to steer the weak area discussion in sync with attacks.

The major reason DES was doomed and lost its standard position is due to its small key size. 64-bit key at first will be stripped out to 56; hence, effective key size is 56 bits, and hence, we just need to have 2^{56} keys to do exhaust search known as brute-force attack. In 1999 and later in 2006 (cost-effective search $10,000 set-up), it was brute-forced with known plaintext and ciphertext combination to identify the hidden key. So the *key size* is main culprit of DES fall, and we state this *as the first weak area of DES.* When you consider the key generation algorithm depicted as below, if you are able to find out any round key, you can easily traverse back and find the original key as the key generation algorithm has no complex mechanism other than permutations and shifting which are very easy to reverse (Fig. 2).

Fig. 2 Existing key generation
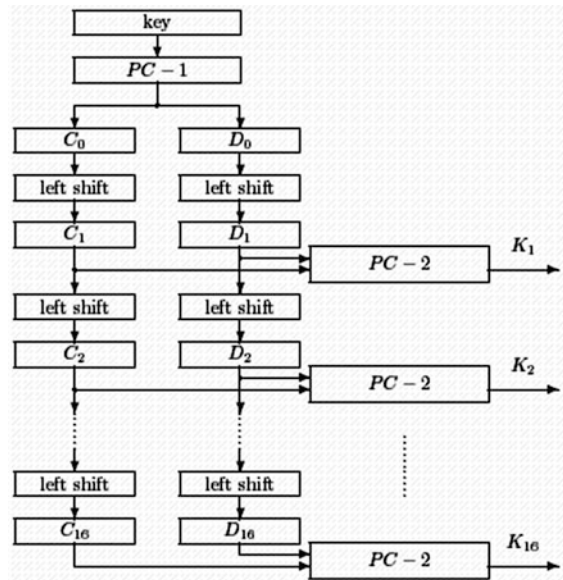
Through side-channel attacks or known plaintext and ciphertext pair attack, if we able to get one round key then traversing back is much easier to find the key. In block ciphers that too private key ciphers, the key is top secret, and if we get to know the key, we lost confidentiality. *So we would like to state that key generation area is second weak area.*

With a large set of known plaintext and ciphertext combinations, a different analysis is carried known as differential attack, where the amount in difference of plaintext and amount in difference of ciphertext will be compared to find the key. DES has shown strong resistance towards differential cryptanalysis where you would like to need minimum of 2^{47} pairs of known plaintext and ciphertexts.

Later, the same experiment carried to identify the linear approximation for the data change in single round (known as linear cryptanalysis) is performed. The key identified with 2^{43} sets of known plaintext and ciphertexts which was tremendous improvement over differential cryptanalysis. Though this seems to be speedier than brute-force attack, the known plaintext and ciphertext data collection required minimum of 10 days and 2 days to break using linear cryptanalysis.

Though this seems to be tough to break, the acceptable point is that DES has open weakness towards linear cryptanalysis which can be further improved to reduce the time to break. So the culprit is S-box but not the design of entire S-box. The nonlinear transformation of 6 bits into 4 bits is an awesome idea. We would like to retain it. The only issue is static and non-random S-tables which led to prove as root cause for linear cryptanalysis.

So the Static S-tables would be third weak area of DES.

3 Reinforcing and Redesigning DES

3.1 Redesign

We loved many parts of DES algorithm such as S-box 64-bit mappings, mixing of keys and repeated confusion and diffusion-based rounds. So we are not going to bring radical changes in DES and change its face entirely. We would like to discuss weak areas 1 and 2 in this section and 3rd weak area in subsequent section.

As we discussed in previous section, key generation algorithm is pretty simple and allows us to trace back and forth if we got a clue. This is not with the case of the other algorithms as they prevent this strongly. To reinforce this section of algorithm, we propose to introduce hash function in the key scheduling algorithm of DES. This is for the first time using the hash codes in the core algorithm of an encryption standard.

Instead of just shifting and shuffling which provides a door for hackers, we are proposing all new key generation algorithm based on SHA family [10]. We first discuss a brief detail about SHA-2 [11] and then our new key generation algorithm.

SHA-2 is extended and enforced hash code standard approved by NIST which replaced SHA-1. Though SHA-1 does not have any practical weakness, some theoretical analysis [12] is published, and henceforth, SHA-2 came into picture and replaced SHA-1. We too taking SHA-2 (SHA-224 and SHA-256 in specific).

We have done various experiments and concluded with the structure given in Fig. 3 for key generation which is efficient. The proposed key generation algorithm depicted in Fig. 3 is shown.

- We need 16 keys for each round and hence running hash for each round will impact critical path;
- Hence, we utilized SHA2 concept of producing 224 hash out of 256 (i.e. SHA-224) from the initial key;
- Each SHA-2 hash code will produce 224-bit output, i.e. four 56-bit keys ($4 \times 56 = 224$) from one 56-bit input key;
- By passing these 4 keys through PC2, we will get 4 round keys;
- By repeating this procedure for just four times followed shifting, we will get 16 48-bit keys needed for encryption/decryption;
- This way the key generation algorithm completely replaced with SHA-based structure without disturbing the original DES algorithm;
- As we are repeating SHA only for four times, we are not adding much delay to critical path;
- By employing SHA-224 instead of just shifting, we provide strong random nature and nearly impossible hacking of round key from master key or vice versa.

The next change as part of redesign is extendibility of block size. As shown in Sect. 2, *key length or block size was main reason for DES fall*. Hence, increasing the key size to 128 bits is mandatory. But at the time of invention, DES 56 bits thought as safe but proven wrong 22 years later. Same thing may happen to enhanced DES as well if we just increase the key size to 128.

Instead, we propose new extendible 64X block size structure where X is of range 1–24. For example, if X is 2, then the block size is 128 bits (which is standard now) and our enhanced DES will become cipher of 128 bits with 128-bit keys. The structure would be like below (Fig. 4).

The above structure can easily extended to larger block size up to 1024 bits. As we are using SHA-224 which can take input size max of 1024 bits, we can safely use this structure up to 1024-bit key size. Also, all cryptanalysis [13, 14] proved that number of rounds needed are greater than 4, and hence, to provide more speed and counterbalance the inclusions of hash codes into critical path, we reduced the number of rounds from 16 to 8. The improved DES in the figure is the combination of redesign and reinforcement (discussed in next section) efforts. With this redesigning of current DES, we added strong non-traceable key generation and configurable and extendable structure as well to DES such that it can easily upgraded in future as per requirements then.

Fig. 3 Proposed key generation

**** Ka and Kb are two halves from Key
**** Ma and Mb are two halves from Plaintext
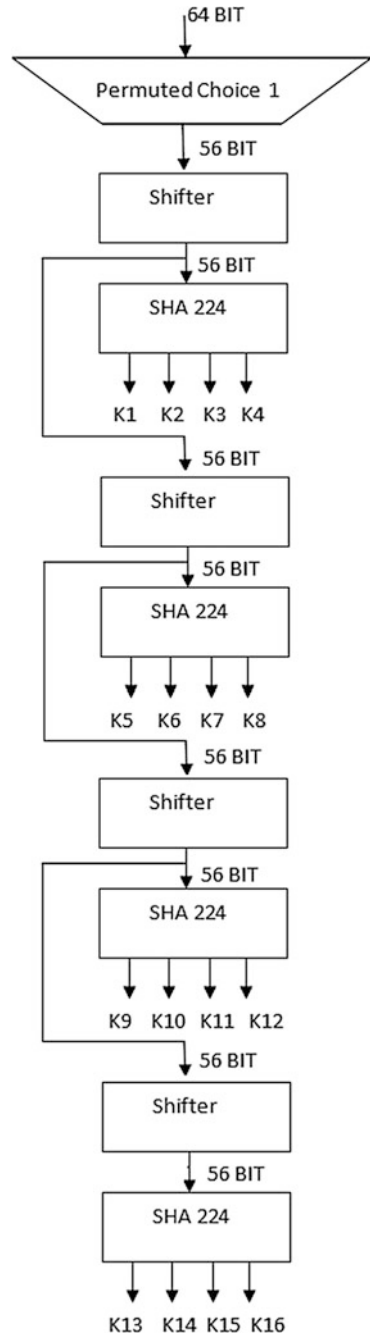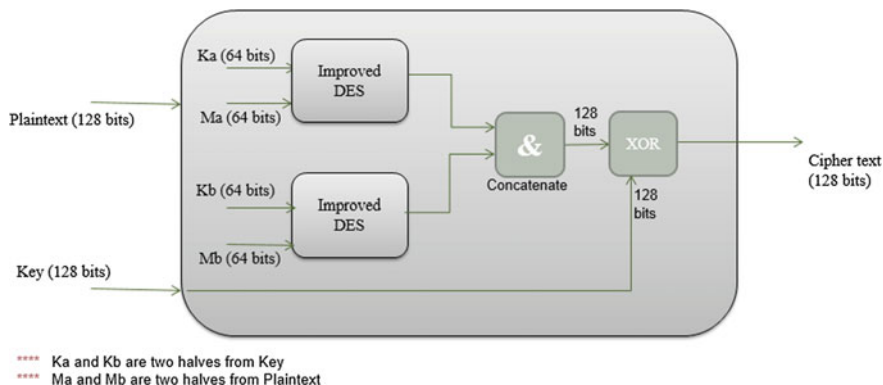
Fig. 4 Proposed top module

3.2 Reinforcement

The only section unanswered is Static S-tables. The current DES algorithm utilizes 8 standard, fixed S-tables as part of substitution. 48-bit data will be made 8 parts of 6 bits, and each part will be supplied to 8 tables to get 4-bit output from each and hence 32 bit output. This 64-bit substitution algorithm is great but deriving the 4-bit value from a static table is problematic area. Many efforts proven succesful in finding collisions in static S-tables.

These S-tables designed to stand strong against differential attacks, the same tables exposed DES weakness against linear and collision attacks. As these are only part of non-random or fixed data in the complete F box operation other than tables, we thought of reinforcing this structure by adding more random nature to it.

In a nutshell, we are not changing or redesigning the S-box algorithm, instead reinforcing it by adding *dynamic* tables in place of static and fixed tables.

How to design dynamic tables? The only inputs we have are key and message. We cannot rely on message as the input for encryption and decryption is different and the keys must be same (but in reverse order) for encryption and decryption. So the common input for both encryption and decryption is master key or original key. This time we proposed to take SHA-256 algorithm to generate dynamic tables as below.

As shown in Fig. 5, we give initial 64-bit key to SHA-256 which produces 8 S-tables as below.

- Each S-box takes 6-bit output and produces 4-bit output.
- Each S-box will have 64 numbers (value ranging from 0 to 15) in 4 rows and 16 columns, i.e. we need 64 nibbles for a table.
- So to replace fixed tables, we need 8 sets of 64 nibbles (i.e. 64 * 4 = 256 bits) which are ranging from 0 to 15.
- In other words, 8 sets of 64 nibbles. It means 8 sets of 256 bits.

Fig. 5 Dynamic S-box generation algorithm
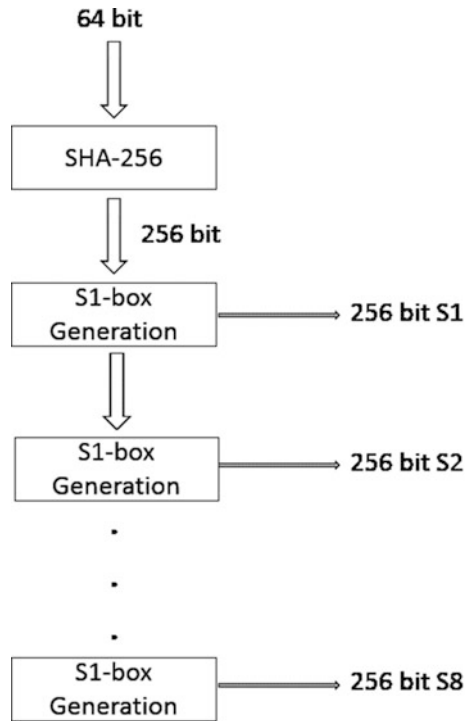
Fig. 6 S*x* box generation
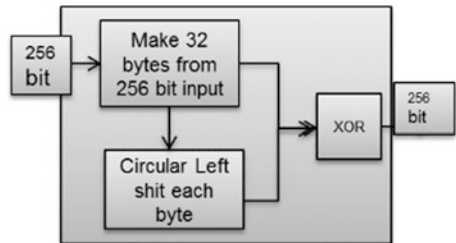
- We gave 64-bit key to SHA-256 which produces 256-bit message digest which will be employed as one S-box as shown in Fig. 5.
- By shifting this 256-bit S-box and XORing with previous 256 bit as shown in Fig. 6. We will get next S-box

Like this, we will get 8 completely dynamic key-dependent S-boxes which are random in nature. There are no more fixed tables, and if you have suspicion that key is known, then change the key all the tables will be changed.

With this redesign and reinforcements, we have added complete dynamic character to the algorithm to make it strong and robust against all kinds of attacks.

4   Cryptanalysis on Improved DES

4.1   *Differential Cryptanalysis*

Differential cryptanalysis is the cryptanalytic attempt of identifying the key with the known plaintexts with small difference. In this scheme, the role of key will be removed, and hence, with known plaintexts or their difference, one can try to identify the key.

The process for DES is explained as below as per [13].

(Legend: He → Cryptanalyst)

- In differential cryptanalysis and as shown in [14], he starts with two messages, m and m′, differing by a known difference delta-m.

That is,

$$\Delta m = m \oplus m'.$$

He considers the difference between the intermediate message halves:

$$\Delta m_i = m_i \oplus m'_i.$$

The input to S-box S_1, for example, at round i of the encipherment of message m is

$$m_i[32,1,2,3,4,5] \oplus k_{ij}[1,2,3,4,5,6],$$

and the input to S_1 at round i of the encipherment of message m' is

$$m'_i[32,1,2,3,4,5] \oplus k_{ij}[1,2,3,4,5,6].$$

From the identity $(a \oplus c) \oplus (b \oplus c) = a \oplus b$, we see that the XOR of these two inputs is

$$(m_i[32,1,2,3,4,5] \oplus k_{ij}[1,2,3,4,5,6])$$
$$\oplus (m'_i[32,1,2,3,4,5] \oplus k_{ij}[1,2,3,4,5,6])$$
$$= m_i[32,1,2,3,4,5] \oplus m'_i[32,1,2,3,4,5]$$
$$= \Delta m_i[32,1,2,3,4,5].$$

The dependence on k has disappeared.

- So the effect of key in the rounds of encryption is eliminated, and hence, the crux of cracking the key lies in the algorithm randomness in each round.
- DES randomness core is its S-box.
- So this differential cryptanalysis aims at how the S-box takes 48-bit input and how it produces 32-bit output using two known plaintexts as the values in S-boxes are constants.
- The fixed values of S-boxes helped to device the methodology of differential cryptanalysis for DES.
- As per [15], the differential cryptanalysis is effective if DES has 1–4 rounds only. As DES has 16 rounds in encryption/decryption process, the known difference of plaintext did not yield much results due to randomness built up with the help of 16 rounds.

- Still, differential cryptanalysis attack against DES requires enormous amounts of chosen plaintext. Biham and Shamir [15] estimate the amount of plaintext necessary for their attack as $2^{47.2} = 1.6 \times 10^{14}$ chosen plaintext messages (of 64 bits each).

Though DES is strongly resistant against differential cryptanalysis, we observed the zero effect of key in the process of attack on original DES and made below improvements.

- Block size is 128 bits, i.e. two cores of DES.
- S-boxes are now completely dynamic and random in nature as they are dependent on key. So key role is reintroduced into the analysis.

Below are the observations that can be deduced from above two improvements.

- Instead of one core, now two cores of DES, i.e. we doubled the complexity and time taken for differential cryptanalysis when compared to DES.
- As the S-boxes are dynamic and based on key, the existing differential crypt-analysis method on DES will not work on improved DES as the values of S-boxes are unknown in contrast to known S-box values in DES.
- The removed key role is reintroduced into the analysis to make the computation much complex and time tedious.

4.2   Linear Cryptanalysis

Linear cryptanalysis [16] is invented to reduce the complexity in differential cryptanalysis and provide practicality. Linear cryptanalysis is known plaintext–ciphertext pair-based method, and it aims at identifying the key by forming a linear approximating equation in transforming the plaintext to ciphertext.

It is proven that [7] 2^{43} known plaintext–ciphertext pairs are enough to crack DES using linear cryptanalysis which is lesser than differential cryptanalysis (2^{47}) and brute-force attack (2^{56}). Hence, it can be stated as fastest key yielding attack amongst available attacks.

The linear cryptanalysis method is trying to linearize the S-box tables by which we can break the tough nut in forming the linear approximation equation to break the key.

The principal part of linear cryptanalysis is study of S-boxes and their fixed values in DES and forming the linear approximate equation like one below.

$$P_{i_1} \oplus P_{i_2} \oplus \cdots \oplus C_{j_1} \oplus C_{j_2} \oplus \cdots = K_{k_1} \oplus K_{k_2} \oplus \cdots$$

As new improved S-boxes are completely based on key and each table is dynamically derived the existing approximation strategies will not hold. Moreover, the approximation process will be more tough as we need to include S-box

generation module as well in the equation which has irreversible hash functions, and hence, the equation never holds good.

In a nutshell, we are making all existing linear cryptanalysis methods irrelevant on new improved S-box as the source of attack is the static tables and we replaced them with key-driven tables. So the existing attacks of keyless strategy fail on new improved DES as the key is essential factor in the S-box tables.

5 Other Results

To increase the speed and reduce the space/power, we have reused SHA-224/256 single block. Moreover, we get 256 hash of initial key in key generation only, and instead of regenerating hash in S-box table's generation, we pass this hash to S-box generation process such that one round of hash preparing is saved and eliminated from critical path. Further, we reduced the number of rounds to 8 from 16 to add more speed to the algorithm.

We have increased the block size to 128 bits, i.e. double the time taking for encryption/decryption or double memory footprint/power consumption expected. Moreover, we have added SHA-224 and SHA-256 into critical path which further produces delay. So when we compare the speed, memory footprint and power consumption of improved 128-bit DES against 64-bit DES, we need to compare after doubling the original DES figures. So we are taking DES-64 as reference to ease out the comparison.

We have simulated and synthesized original DES-64 and our improved DES-128 using Xilinx ISE and XST and implemented Spartan-500E (XC3S500) development board (Table 1).

So from above results, we have provided 80% speed when compared to 128-bit original speed, 60% memory space and 80% power consumption along with secure structure against all current attacks.

On the same hardware, we have simulated and synthesized AES for 128-bit encryption and compared our model, and results are given below (Table 2).

Table 1 Result comparison

Algorithm	Speed	Memory space	Power
DES-64	1	1	1
DES-128	2	2	2
Improved DES-128	1.2	1.4	1.2

Table 2 Comparison with AES

Algorithm	Speed (in ns)	Footprint (in LUTs)
Pipelined AES	15.09	15,086
Our algorithm	11.21	12,228

6 Conclusion

In this paper, we first identified the weak areas using literature available and then proposed all new improved version of DES which is novel in its design where hash codes are used in its core to resolve existing weakness. The improvements not only answer the weaknesses but also make the existing attacks on DES obsolete. We have shown the attacks will not hold strong against new improved DES-128, and moreover, this can be extended up to 1024-bit block size and 1024-bit key size in future. The extendibility is very easy to configure as just new cores need to be added based on block size. Along with cryptanalysis and extendibility, we also made it speedier, less space taking and power efficient when compared to original DES-128. We hope our effort in making DES stronger and recapturing its wide-spread usage will eventually become true.

References

1. Data Encryption Standard, FIPS PUB 46, 1977 Jan 15. Available from NTIS; Springfield, VA 22151 USA
2. Fu L, Pan M (2009) A simplified FPGA implementation based on an improved DES algorithm. In: 3rd International conference on IEEE genetic and evolutionary computing, 2009, WGEC'09, pp 227–230
3. Wang K (2009) An encrypt and decrypt algorithm implementation on FPGA's. In: Fifth international conference on IEEE semantics, knowledge and grid, 2009, KG 2009, pp 298–301
4. Sison AM, Tanguilig BT, Gerardo BD, Byun Y-C (2011) An improved data encryption standard to secure data using smart cards. In: 2011 9th international conference on software engineering research, management and applications (SERA). doi:10.1109/SERA.2011.27
5. Patterson C (2000) High performance DES encryption in Virtex FPGA's using JBits. In: 2000 IEEE symposium on field-programmable custom computing machines, pp 113–121
6. Chueng TP, Yusoff ZM, Sha'ameri AZ (2000) Implementation of pipelined data encryption standard (DES) using Ultera CPLD. In: Proceedings of IEEE TENCON 2000, vol 3, pp 17–21
7. Pasham V, Trimberger S (2001) High-speed DES and triple DES encryptor/decryptor
8. Rouvroy G, Standaert F-X, Quisquater J-J, Legat J (2003) Efficient uses of FPGAs for implementations of DES and its experimental linear cryptanalysis. IEEE Trans Comput 52 (4):473–482
9. http://en.wikipedia.org/wiki/Data_Encryption_Standard
10. FIPS 180-1 (1996) Secure hash standard, US Department of Commerce, Washington D.C. Springer, New York
11. FIPS 180-2 (2002) Secure hash standard, US Department of Commerce, Washington D.C. Springer, New York
12. Wang X, Yin YL, Yu H (2005) Finding collisions in the full SHAI. In: Proceedings of Crypto'2005, volume 3621 of Lecture Notes in Computer Science. Springer, Berlin, pp 17–36
13. Coppersmith D (1994) The Data Encryption Standard (DES) and its strength against attacks. IBM J Res Dev 38(3):243. doi:10.1147/rd.383.0243
14. http://en.wikipedia.org/wiki/Differential_cryptanalysis

15. Biham E, Shamir A (1993) Differential cryptanalysis of the full 16-round DES. In: Lecture notes in computer science: advances in cryptology—proceedings of CRYPTO '92, Springer, Berlin, pp 487–496. See also Biham E, Shamir A (1993) Differential cryptanalysis of the data encryption standard. Springer, Berlin
16. http://en.wikipedia.org/wiki/Linear_cryptanalysis

A Review of Low-Power VLSI Technology Developments

Nakka Ravi Kumar

Abstract Ever since the invention of integrated circuits, there has been a continuous demand for high-performance, low-power, and low-area/low-cost diversified applications from a variety of consumers. This demand has been pushing the fabrication process beyond ultra-deep, sub-micron technologies such as, 32, 22, 14 nm, and so on. In this chapter, various technological aspects for low-power applications are reviewed in detail, along with the evolution of new technology, bearing in mind the PPA (power, performance, and area). Some basic reviews of components of power consumption in CMOS are also given.

Keywords Low-power VLSI · VLSI technology
CMOS power consumption · Bulk CMOS · SOI · FINFET

1 Introduction

Electronic devices and systems have entered into almost every corner of human life. Most of these devices, at present, are portable and battery operated. The evolution of the computer moved from requiring a huge powerhouse to run one device, to a small handheld battery operated device like a palm-top. All these portable electronic devices need ICs to operate under low-power consumption. Along with the lower power consumption demand, there is a demand for high performance. In addition, all these low-power and high-performance devices are demanded at low cost. The aforementioned demands have pushed the fabrication process from a micrometer level to a nanometer level. Scaling down vertical and lateral dimensions and the voltages of the MOS transistors will increase the performance of the circuit in terms of speed, decrease the overall power dissipation of the chip, and reduce the area of the chip, thereby the cost per transistor.

N. Ravi Kumar (✉)
ECE Department, Mahaveer Institute of Science and Technology, Keshavagiri Vyasapuri, Bandlaguda, R.R.District, Hyderabad 500 005, Telangana, India
e-mail: nrkumar_2002@yahoo.co.in

There are various scaling methods adopted for CMOS: constant field scaling or full scaling, constant voltage scaling, and generalized scaling [1–3].

Due to the restrictions of some material properties, such as the energy band gap, built in voltage, etc., the voltages could not be scaled down as rapidly as dimensions. Because of this, current densities and power densities will increase. An increase in power density increases heat. After a certain limit, managing heat becomes more expensive, if the power density is not addressed properly, than the actual chip fabrication cost.

In addition to this, the need of portable battery operated devices along with their reliability issues, thermal dissipation, and scalability, have demanded for low supply voltages.

CMOS is the best choice for low-power applications. The power consumption of CMOS can further be reduced to a great extent by carefully studying the parameters influencing the power consumption and then by designing the CMOS technology for power efficiency. In the text below (Sect. 2) we deal with various components of CMOS power consumption and the parameters influencing each of them.

The drive for lowering power dissipation is emphasized in Sect. 3.

Technological approaches to lowering the power consumption are described in Sect. 4. The need for low power along with high speed for certain applications has led to new technological inventions. These demands and inventions are described in Sect. 5. Section 6 concludes the chapter.

2 Components of CMOS Power Consumption

The total power consumption in a CMOS circuit consists of three components, namely, dynamic power (P_D), static power/leakage power (P_{leak}), and short circuit power (P_{SC}).

2.1 *Dynamic (Active) Power Consumption (P_D)*

This component of power consumption is due to switching activity of logic gates in a CMOS circuit. The gates will be continuously charging and discharging their load capacitances according to the logic-switching activity. This power consumption depends on power supply voltage (V_{DD}), switching frequency (f_s), and switching node capacitance (C_l). Dynamic power occurs when the device is 'ON'.

If the total number of gates in a chip are 'n' and the average switching frequency and the average switching node (load) capacitances are 'f_s' and 'C_l', respectively, then the total dynamic power consumption of the chip is given by:

$$P_D = n * f_s * 1/2 * C_l * V_{DD}^2 \qquad (1)$$

2.2 Static DC Power/Leakage Power Consumption (P_{leak})

Static power occurs when the device is 'OFF'. The MOS transistors act as switches for digital applications. An ideal switch is expected to have infinite input impedance and infinite 'OFF' state resistance.

Infinite input impedance means the gate should be perfectly isolated from the substrate. However, in practical terms when the gate oxide thickness is scaled down, there will be some leakage of current from the gate (I_{gate}) to drain due to tunneling and hot carrier injection (Fig. 1).

Infinite 'OFF' state resistance means there should not be any drain to source current when the gate voltage is below its threshold. However, there will be some diffusion current when the gate voltage is below VT. This current is called sub-threshold leakage current (I_{sub}). There is another leakage current from drain to substrate, which is called junction leakage current (I_j). One more source of leakage current is gate induced drain leakage (GIDL; I_{GIDL}) (Fig. 1).

The total leakage current (I_{leak}) is given by:

$$I_{leak} = I_{gate} + I_{sub} + I_j + I_{GIDL} \qquad (2)$$

Leakage power (P_{leak}) is given by: $P_{leak} = I_{leak} * V_{DD}$. Though leakage power is negligible compared to active power, for a long channel behavior device, it becomes comparable for devices with short channel effects. Short channel effects are the result of generalized scaling, where voltages are not scaled down as rapidly as dimensions are. As shown in Fig. 2, the power consumption due to leakage current becomes a significant factor for the total power consumption.

Fig. 1 Sources of leakage currents in an NMOS transistor
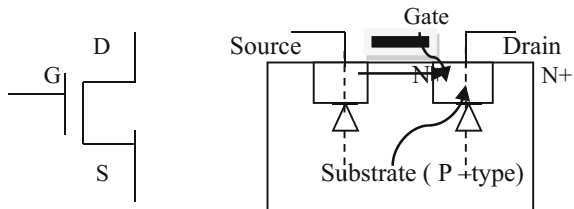
Fig. 2 Trend of dynamic and leakage power against device dimensions
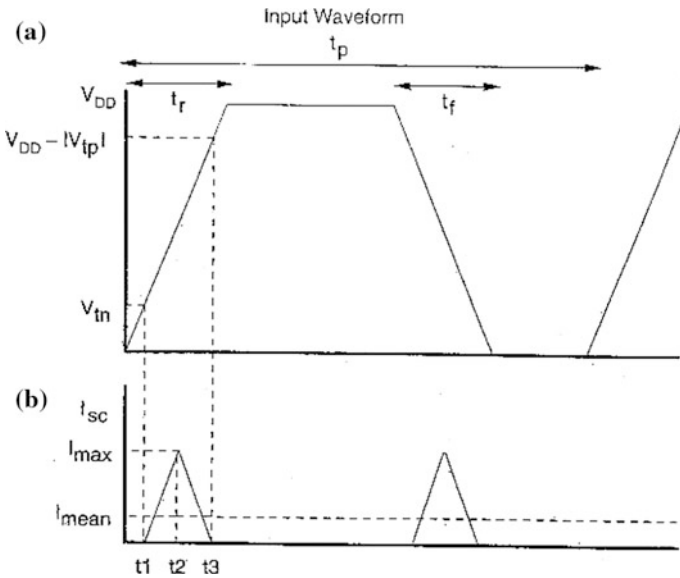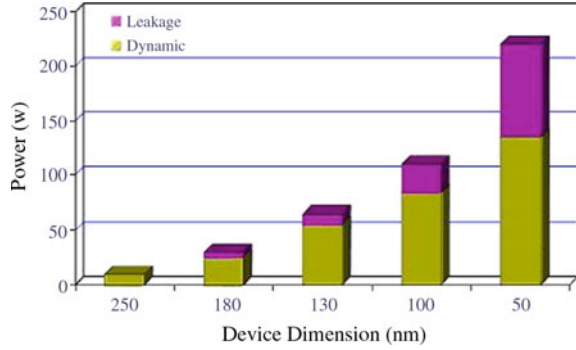
Fig. 3 **a** Input waveform; **b** Short circuit current spikes

2.3 Short Circuit Power Consumption (P_{SC})

This component of power consumption comes about due to the momentary 'ON' state of both NMOS and PMOS of a CMOS gate during transitions. Due to a finite rise time and fall time of the input waveform of a gate, both NMOS and PMOS transistors are 'ON' for short periods of time. As shown in Fig. 3, this results in spikes of short circuit currents during transitions.

The short circuit power (P_{SC}) is given by: $P_{SC} = I_{mean} * V_{DD}$. This short circuit current can be totally eliminated by making $V_{DD} < |V_{tp}| + V_{tn}$.

3   Drive to Lower Power Dissipation

Even if one can afford to pay for the power, one cannot get along with it because, as the chip size shrinks, the power dissipation per unit area increases. This will increase the temperature of the chip. The increased temperature will deteriorate the carrier mobility and V_T. Therefore, the performance of the chip will be greatly deteriorated. Moreover, as the temperature rises, the minority carrier concentration also rises. This leads to increased leakage current. As we have seen in the previous section, increased leakage current leads to increased static power dissipation. This will further increase the chip temperature and eventually lead to chip breakdown.

Though scaling of CMOS technologies reduces the overall power dissipation of a chip, the power density increases (Fig. 4) for conventional CMOS processes. The power density increases with decreasing minimum feature size of CMOS technology for generalized scaling.

The projection of power density is not far from that exhibited on a nuclear reactor surface. This has driven the requirement to invent new technologies and design methodologies for low-power dissipation density.

In addition to these thermal limitations, there is a constant demand for low-power consumption from battery operated portable devices. Low-power dissipation ensures long battery life.
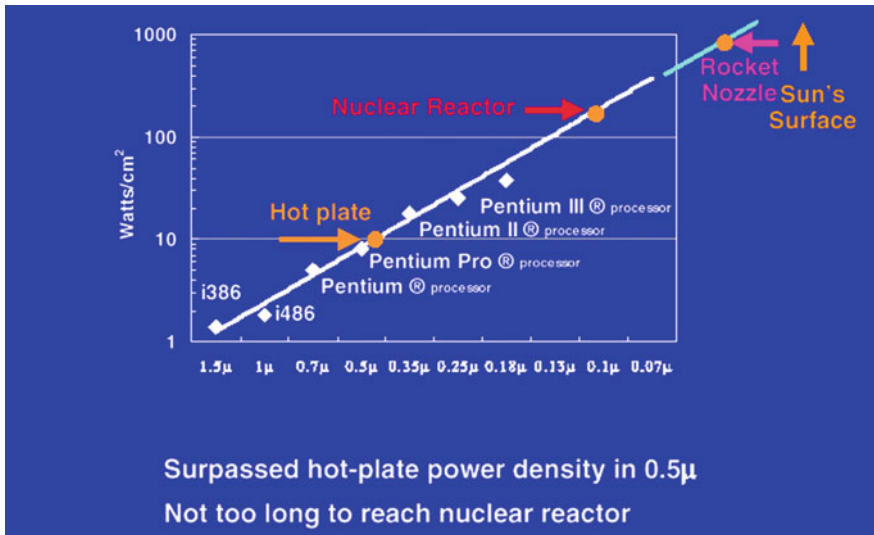
Fig. 4 Power density against feature size of CMOS technology (Courtesy of Fred Pollack, Intel CoolChips tutorial Micro-32)

4 Approaches for Lowering Power Consumption

Power consumption has to be addressed both at technological level as well as at circuit design level. For ultra-low-power applications, optimization and adoption of innovative circuit design techniques at all levels of abstraction, along with innovations in new technologies, have become mandatory.

As we have seen in Sect. 2, the overall power consumption consists of three components, from which short circuit power consumption can be eliminated.

We shall now identify methods of reducing dynamic power and leakage power.

4.1 Ways to Reduce Dynamic Power (P_D)

The dynamic power in CMOS, given by (1), depends on switching frequency (f_s), load capacitance (C_L), and power supply voltage (V_{DD}). The dynamic power can be reduced by reducing any, some or all of these parameters.

Dynamic power has a quadratic dependence on V_{DD}. Therefore, reducing V_{DD} reduces P_D drastically. However, V_{DD} cannot be reduced arbitrarily, because, it has to be compatible with the system in which the chip is assembled. However, the power supply voltage of CMOS technologies has been scaled down (as shown in Table 1) after a global agreement among major VLSI industries.

Some circuit design techniques adopt a multiple power supply. They generate a low supply voltage internally for power sensitive blocks and retain a high V_{DD} for critical path circuits [4].

Switching frequency can be reduced by using encoding and decoding techniques.

The load capacitance, C_L, constitutes gate oxide capacitance, C_{ox}, interconnect capacitance, C_I, and junction capacitance, C_J. By optimizing the gate areas and interconnect areas, load capacitance can be reduced. At the technological level, the interconnect capacitance can be reduced by low 'k' (low dielectric constant) dielectric layers underneath the metal layers. The junction capacitance can be reduced greatly in SOI and FINFET technologies.

Table 1 Typical scaling scenario of V_{DD}

Year of introduction	Technology node	Power supply (V_{DD})
1974	5 μm	10 V
1984	1 μm	5 V
1994	0.35 μm	3.5 V
2004	90 nm	1 V

4.2 Ways to Reduce Static Power

As discussed in Sect. 2.2, as the technology shrinks, static power consumption becomes comparable to dynamic power consumption.

The static power in CMOS, given by (3), depends on I_{leak} and V_{DD}. Reducing V_{DD} reduces static power. I_{leak} should also be reduced for ultra-low-power applications.

Dealing with these four components of leakage current, as given in (2), is becoming more and more of a prime concern as technology shrinks.

Sub-threshold Leakage Current (I_{sub})
Low V_{DD} requires a low threshold voltage. Leakage current increases as the threshold voltage decreases. For a particular target threshold voltage, the sub-threshold leakage current can be minimized by making the sub-threshold slope close to its minimum value. The sub-threshold slope is given by:

$$S = (\ln 10) * (kT/q) * ((C_{ox} + C_d + C_{it})/C_{ox}) \qquad (3)$$

where C_{ox} is the gate oxide capacitance; C_{it} is the interface trap capacitance; and C_d is the depletion capacitance. The minimum value of the sub-threshold slope is $(\ln 10) * (kT/q)$, which is approximately 60 mv/decade. The sub-threshold slope can be made close to its minimum value by making $C_{ox} \gg (C_d + C_{it})$.

Now, C_d can be reduced by making the substrate doping very low. However, making substrate doping low causes 'punch through'. To avoid this, new technologies, like SOI and FINFET, have been invented.

The interface trap capacitance, C_{it}, can be reduced with a good insulator–substrate interface.

Junction Leakage Current (I_j)
The substrate current generated by impact ionization at the drain substrate junction triggers a parasitic bipolar action in the MOSFET. This onset of bipolar action further increases the leakage current. So, preventing this bipolar action is essential.

Gate Leakage Current (I_{gate})
Gate leakage current consists of band-to-band tunneling current, hot carrier injection current, and gate induced drain leakage current.

As the gate dielectric thickness is reduced, gate leakage current becomes a serious problem. One possible solution is to choose high dielectric constant materials instead of silicon dioxide (SiO_2). For the same capacitance, the insulator thickness can be increased for high k dielectrics compared to SiO_2. Therefore, the electric fields are reduced and hence the leakage currents are too. Some choices of high-k dielectrics are: Al_2O_3, HfO_2, ZrO_2, Y_2O_3, etc. [1].

Gate Induced Drain Leakage Current (I_{GIDL})
This phenomenon occurs due to an accumulation of the majority carriers in the substrate under the gate near the drain. In the case of NMOS, the accumulated holes under the gate, near the drain/substrate junction, causes pre-matured impact ionization. This leakage is called I_{GIDL} [5], which can be reduced by using high-k gate dielectrics.

5 New Technologies for Ultra Low Power and High Performance

As discussed in the previous section, lowering V_{DD} reduces power consumption and also suppresses reliability problems. In conventional CMOS technology, where $POCL_3$ degenerately doped N+ poly silicon is used as a gate material for both NMOS and PMOS transistors, the threshold voltage adjustment requires a separate boron implantation. In this technology, a PMOS transistor turns out to be a buried channel device, which is also known as a compensated MOSFET. Reducing the threshold voltage of these compensated MOSFETs is difficult. Therefore, to achieve low V_T for both NMOS and PMOS, dual-poly gate technology is adopted. Here, N+ poly is used for NMOS and P+ poly is used for PMOS. The poly doping is undertaken along with a self-aligned source/drain implant [6]. However, these conventional bulk CMOS technologies are not free from sub-threshold leakage, parasitic bipolar current, punch through, DIBL, high junction capacitances, and other short channel effects. There is one further problem with this technology—the poly depletion effect [7], which becomes more prominent as the gate oxide thins.

The performance and power goals for certain applications in the advanced nodes of 40, 32 nm, etc., could not be achieved with conventional bulk CMOS processes. This has led to alternative technologies.

These new technologies are SOI and FINFET. In both of these technologies the body is made very thin so that the gate has maximum control over the channel. The threshold voltage is adjusted using mid-gap materials for gate electrodes. These technologies postpone the short channel effects to future generations. These technologies ensure not only a reduced I_{off}, due to the decreased leakage currents, but also provide increased I_{ON}, due to an increase in carrier mobility owing to very lightly doped or undoped bodies.

5.1 SOI Technology

Silicon-on-insulator (SOI) technology was invented to reduce a device's leakage current, increase its 'ON' current, and reduce its capacitance [8–10].

In SOI fabrication technology, transistors are built on a silicon layer resting on an insulating layer of SiO_2 (as shown in Fig. 5). The insulating layer sits on top of the silicon substrate.

There are different ways of manufacturing SOI wafers: SOS (silicon-on-sapphire), SIMOX (Separation by IMplanted OXygen), BESOI (bond and etch-back SOI), Smart Cut, and ELTRAN (Epitaxial Layer TRANsfer). The description of all these processes is beyond the scope of this chapter.

If the body is not very thin, it will be partially depleted. The device is then called partially depleted SOI or PDSOI. These devices suffer history effect and kinks in output characteristics due to parasitic bipolar action. These problems are avoided by
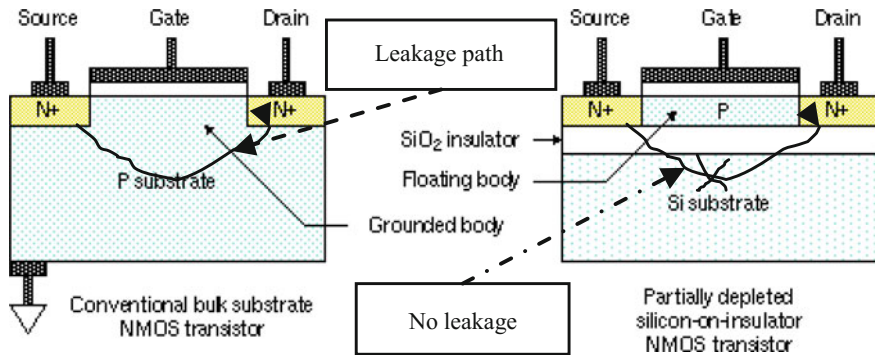
Fig. 5 A bulk NMOS transistor versus SOI NMOS

making the body and base oxide (BOX) thicknesses very thin. The body of these devices is fully depleted, and the technology is thus called fully depleted SOI or FDSOI [11]. If the body thickness is ultra-thin, the technology is called ultra-thin body SOI or UTBSOI.

The advantages of SOI over Bulk CMOS are: higher performance due to reduced capacitances and increased I_{ON}; latch-up elimination due to vertical and lateral isolation of the transistors; reduced temperature sensitivity; a removal of the need for body or well taps; reduced antenna effects; small transistor sizes saving areal extent; low-power consumption due to ultra-low-leakage currents; and reduced parasitic capacitances.

The disadvantages are: high wafer costs and self-heating.

Heating may not be a problem in ultra-low-power devices. We consider SOI technology is the best choice for RF applications.

5.2 FINFET Technology

As mentioned in the previous section, the solution to leakage problems is to make the body ultra-thin in order to have better gate control. In SOI technologies, the body is thin and planar. However, the body can also be made thin and vertical, where the gate will be on both sides, or on all sides (as shown in Fig. 6). This technology is called FINFET technology [12, 13].

The thin silicon fin can be made on a bulk silicon substrate or on a SOI substrate. As shown in Table 2, both these technologies have their own advantages and disadvantages.

INTEL uses FinFET technology down to 10-nm nodes.
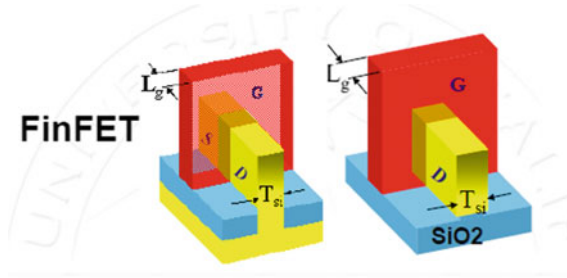
Fig. 6 Fin on bulk and fin on SOI

Table 2 Fin on bulk versus fin on SOI

Fin on bulk substrate	Fin on SOI substrate
Δ HFin (variation in Fin height) is greater ($-$)	Δ HFin is less ($+$)
Substrate is cheaper ($+$)	SOI substrate is costly ($-$)
Heat is easily dissipated ($+$)	Heat dissipation is a problem ($-$)
Parasitic BJT will exist ($-$)	No parasitic BJT ($+$)
Source/drain region epitaxy is possible ($+$)	Source/drain epitaxy is not possible ($-$)

6 Conclusion

Demand for high performance and low-power consumption has led CMOS scaling from micrometer ranges to nanometer ranges. These demands could not be met with conventional CMOS technologies and thus led to the invention of new technologies. Various components of power consumption in CMOS are discussed in detail in this chapter. The drive for low-power consumption and the methodologies adopted to reduce power consumption in CMOS are also discussed in detail. Finally, the technologies developed to achieve these goals have been reviewed.

References

1. Sze SM Physics of semiconductor devices, 3rd edn. Wiley India publications
2. Moore GE (1965) Cramming more components on integrated circuits. Electronics 38(8):19
3. Haron NZ, Hamdioui S, Why is CMOS scaling coming to an end? http://ce-publications.et.tudelft.nl/publications/510_why_is_cmos_scaling_coming_to_an_end.pdf
4. Kim K, Agrawal VD (2011) Ultra low energy CMOS using below-threshold dual-voltage supply. J Low Power Electron 7:1–11
5. Chen J et al (1987) Sub-break down leakage current in MOSFETS. IEEE Electron Device Lett EDL-8:515

6. Pandey SM, Kumar NR, Sankar GK, Pal DK, Roy JN (1999) Optimization of dual poly gate low voltage CMOS process. Proceedings of the Tenth International Work shop on the Physics of Semiconductor Devices,Vol part2, pp 871–874 ISSN: 0277786X, ISBN: 9780819436016 [0819436011]
7. Sallase JM (2000) Improved analytical modeling of polysilicon depletion in MOSFETS for circuit simulation. Solid-State Electron 44:905–912
8. Bhat KN, Silicon on insulator (SOI) devices. Lecture notes. http://ece.iisc.ernet.in/~navakant/nano/2007/Lecture23.pdf
9. Singh RK, Saxena A, Rastogi M (2011) Silicon on insulator technology review. Int J Eng Sci Emerg Tech 1(1):1–16 ©IJESET. ISSN: 2231-6604
10. Crisoloveanu S (2000) Architecture of SOI transistors: what's next? In: 2000 IEEE international SOI conference, pp 1–2
11. Sakurai T, Matsuzawa A, Douseki T (2006) Fully-depleted SOI CMOS circuits and technology for ultralow-power applications. Springer, New Jersey
12. Liu T-JK (2012) FinFET history, fundamentals and future. In: 2012 symposium on VLSI technology short course
13. Hu C, FinFET and other new transistor technologies. Lecture notes. http://microlab.berkeley.edu/text/seminars/slides/2011-8_FinFET_and_the_Concept_Behind_It.pdf

Architectures of Charge Pump
for Digital Phase Locked Loops

D.S. Rajeshwari and P.V. Rao

Abstract Varieties of charge pump architectures are discussed and simulated with constant loop parameters. DC mismatch is significantly compared for single-ended and fully differential charge pumps. The charge pumps are designed with foundry tsmc, 90 nm and 1.8 V CMOS technology. The improved versions of single-ended and differential charge pumps are explicating providing high performance. The current mismatch is achieved less than 0.01–0.2% for different architectures.

Keywords DPLL · Charge pump · Single-ended · Fully differential
Switch errors · Current mismatch

1 Introduction

Phase difference of the reference frequency and the feedback frequency implies digital signals to charge pump (CP). These signals act as functionality to charge pump as charging, discharging, and no change at the output voltage. The block diagram of CP is shown in Fig. 1. The output control voltage is processed through voltage-controlled oscillator generating the frequency signal. Digital phase locked loop (DPLL) locks when output frequency or feedback frequency signal is equal to reference frequency. Ripple on output control voltage propagates in the DPLL loop to increase the jitter [1]. While charging and discharging to phase frequency detector (PFD) output digital signals, charge pump has its errors. They are charge

D.S. Rajeshwari (✉)
Electronics Engineering, Jain University and Altran Technologies Pvt. Ltd.,
Bengaluru 560001, India
e-mail: dsrajeshwari@gmail.com

P.V. Rao
Electronics and Communication, Vignana Bharathi Institute of Technology,
Hyderabad 501301, India
e-mail: pachararao@rediffmail.com

sharing, clock feed-through, charge injection, and current mismatch. Current mismatch results are due to static and dynamic mismatches. These mismatches also correspond to leakage current. In every period, comparison of the phase error results variation in the width of UP and DOWN signals. IUP and IDN CP currents have to be same to obtain the same amount of charging and discharging for finite width of UP and DOWN pulses. Dynamic mismatch results when there is a difference in switching time of UP and DOWN. In the locked condition DPLL has small static phase error which leads to jitter [2]. Non-idealities of single-ended charge pump are shown in Fig. 2a and those of differential charge pump are shown in Fig. 2b. Differential nature will also have static and dynamic mismatches [3, 4]. As shown in Fig. 2b, the effect of non-idealities will be avoided. Ripple of differential CP is far less than single-ended CP. Differential CP has more power consumption and it is a complex circuit.

Many single-ended CPs and differential CPs are discussed in the literature [2–12]. The literature [4] exploits improved version of CP with switch-at-source and the same architecture is executed in differential CP. The literature [6] exploits significance of charge pump in DPLL loop dynamics and also about charge sharing, charge injection, and clock feed-through. In the literature [2] the different types of CP, proposed CP are discussed. Here CP is based on transmission gate and operational amplifier structure to achieve current mismatch. The literature [8] describes differential-in single-ended output and switch-at-drain and exploits comparison of CP structures. The literature [9] exploits current-steering concept for CP which is a better method to avoid non-idealities associated with PFD/CP.

In this paper, comparative study of all types of charge pump for DPLL applications is presented. Analysis, design, and simulation of these charge pump results are briefed. The simulation is performed using tsmc foundry, 90 nm technology, and 1.8 V power supply. Section 2 explains the configurations of CP, Sect. 3 deals with different types of CP, Sect. 4 provides simulations results, and Sect. 5 concludes the whole work.

Fig. 1 Block Diagram of a charge pump PLL
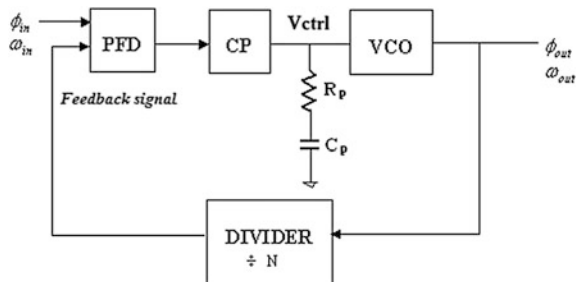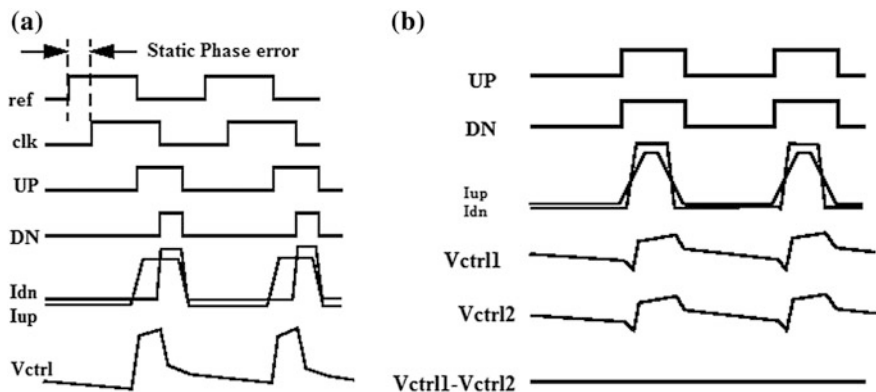
Fig. 2 Waveforms of non-idealities **a** single-ended charge pump and **b** differential charge pump

2 Charge Pump Configurations

Basically, the charge pump comprises switches and current sources. To generate two current sources of the same value, it is advisable to use a single constant current using current mirrors. This type of current mirror topology can be done in three different ways, with them differing only by the position of the switches [5, 12]. Switch-at-drain, switch-at-gate, and switch-at-source are three different configurations.

As shown in Fig. 3a switch is placed at drain of current mirrors MOS M2 and M1 and is placed near to load capacitance. Charge sharing phenomenon occurs when switch turns ON to OFF. Switch UP is initially turned ON and Vout is charged. Thus, switch UP holds channel charge [8]. As soon switch UP is turned OFF, part of stored charge in channel is shared with load capacitance and node on the other end of switch. This traditional switch-at-drain suffers seriously from charge sharing. Another phenomenon is sequentially turning ON to OFF, thereby creating high peak current. Along with charge pump UP and DN current, these high peak currents are needed to be matched. If voltage at the drain of switch could hold the same voltage level when switch is turned OFF, switch-at-drain would be a efficient architecture for high-speed DPLL performance [2].

Another type of CP configurations is switch-at-gate, as shown in Fig. 3b. Switch is placed away from the load capacitance; thus greatly charge sharing is minimized. As current mirrors *M1–M4* are always in saturation region, generating high peak current is possibly avoided. These types of CP have high power consumption. Gate of current mirrors is connected by two more transistors which are switch and another branch of current mirror. This increases capacitance at the gate and consequently reduces the switching speed.

Switch is placed at source of currents IUP and IDN as shown in Fig. 3c. Here, IUP and IDN are derived from common current reference ICH. This configuration
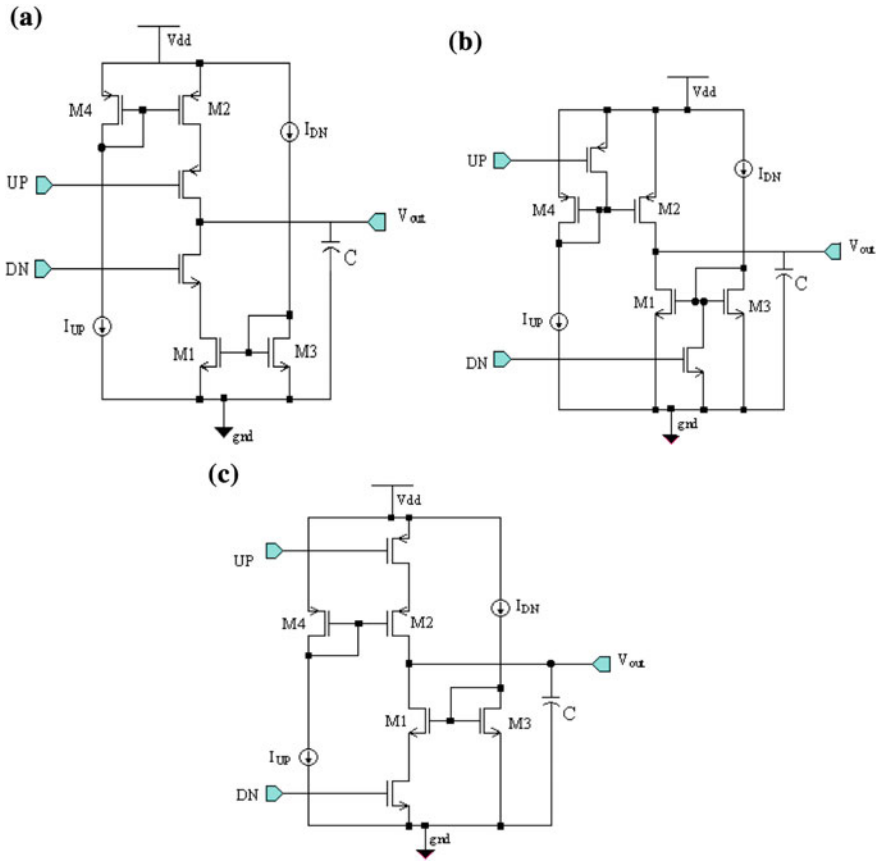
Fig. 3 Charge pump configurations **a** switch-at-drain, **b** switch-at-gate, and **c** switch-at-source

takes away the drawbacks of above types [9]. Comparing to Fig. 3a and b, both the way switch is away from the load capacitance and does not loading more capacitance at gate, thus charge sharing maximally avoided and will not affect the switching speed too.

3 Architecture of CMOS Charge Pump

Charge pump configurations are varied to improve the performance of DPLL. Extended types of charge pump are single-ended CP, differential-in single-ended-out CP, and differential-in differential-out CP.

3.1 Single-Ended CP/Differential-In-Single-Ended-Out CP

Charge pump that is shown in Fig. 4 is also single-ended charge pumps. These charge pumps are very popular for their low power consumption, and they do not demand for additional loop filters. One of the improved version of Fig. 3c is shown in Fig. 4a. It is implemented with pull-up transistors *M9* and *M10*. These transistors are added to reduce the charge coupling to the gate and help in enhancing the switching speed.

The dependence of charge injection upon the input level makes it necessary to seek methods of canceling the effect of charge injection [5]. One of the techniques is that suppressing the charge injection by using dummy switch. Figure 4b have
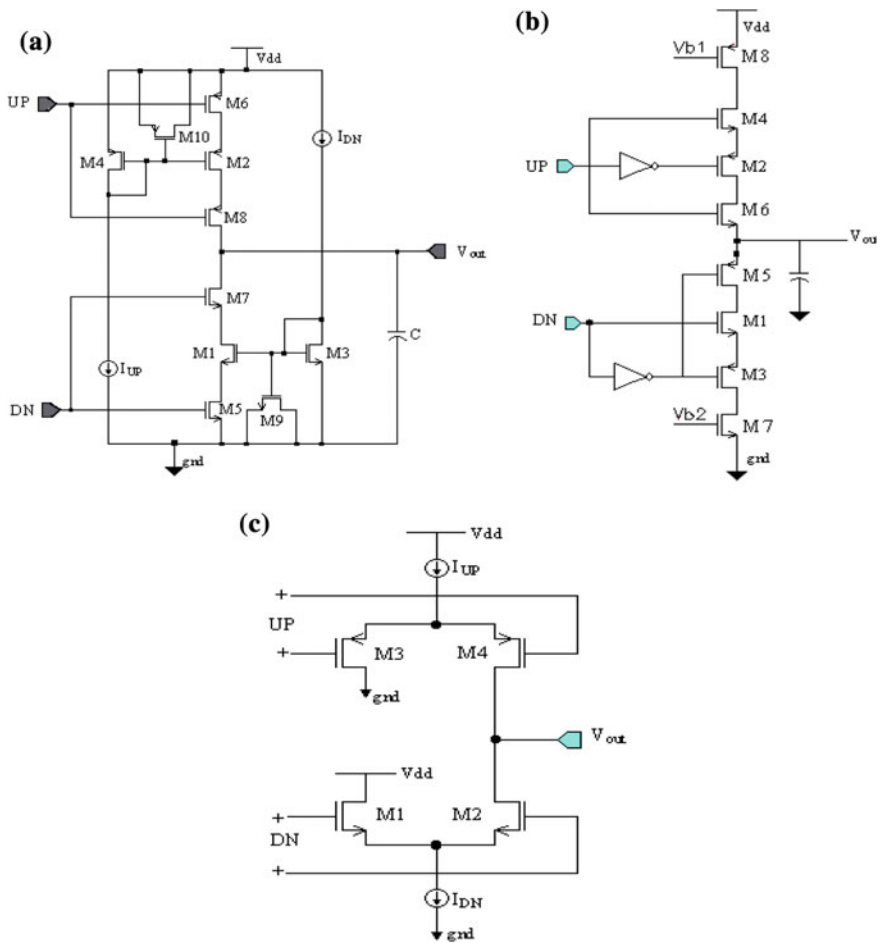
Fig. 4 Single-ended CP **a** with pull-up transistor and **b** with dummy switches, **c** differential-in single-ended output current-steering CP

dummy switch, is placed near to switch and the output load capacitor and it reduces drawback of charge sharing in contrast to Fig. 3a. Single-ended charge pumps finds application in digital clock generators.

Current-steering CP is a differential-in single-ended output and is shown in Fig. 4c. Here, when *M3* is ON, *M4* is OFF and vice versa. Current steering works on dividing the current and no node floating [10]. The current-steering concept allows improving the switching speed and clock skew and minimizing leakage current. Some variations are applied to Fig. 3a for improving the performance. Figure 5a shows the CP with an active amplifier. Whenever switch is placed at drain, the architecture significantly suffers from charge sharing as switch turns ON to OFF. Thus, voltage at the drain of switch can be maintained to the output node using an active amplifier [11]. When the switch is OFF, unity gain voltage followers uphold the output node. This architecture can be employed when parasitic capacitance around switches is as good as charge pump output capacitor. Even though current steering and active amplifier efficiently avoid the switch errors, the causes due to natural difference between NMOS and PMOS cannot be avoided. As solution charge pump using only NMOS switches with same functionality can be implemented [5, 7]. Charge pump with NMOS switches having symmetric load is shown in Fig. 5c.

3.2   Differential-In Differential-Out CP

Compared to single-ended charge pump, the fully differential charge pump has a number of advantages. The DPLL performance is insignificantly affected by the NMOS and PMOS transistor switches mismatch. Because of differential nature mismatch between NMOS transistors and between PMOS transistors is comparatively better than that between NMOS and PMOS transistors [2, 12]. The differential-in and differential-out charge pumps are fully symmetric operation. The ideal fully differential CP is shown in Fig. 6a. Thus, delay-compensated inverters will not generate offset. Output range limits the tuning range of VCO. Output range is doubled with differential CP. The two differential outputs are less sensitive to leakage current. They appear as common-mode offset.
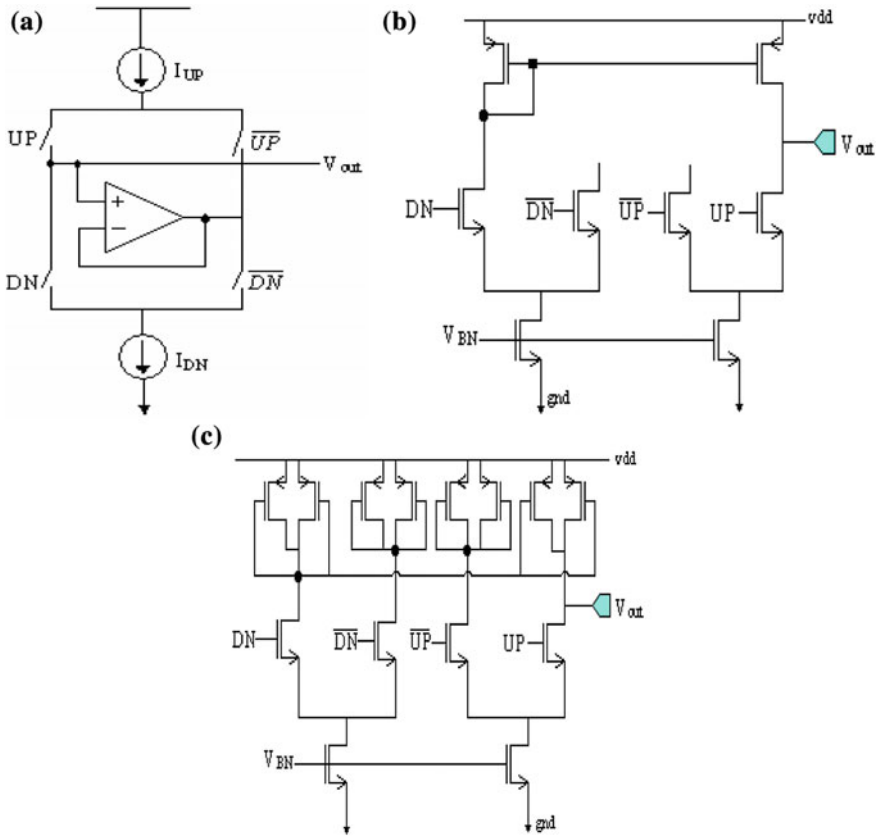
Fig. 5 Differential-in single-ended output CP **a** with active amplifier and **b** using only NMOS switches, and **c** complete schematic of the offset-canceled charge pump with symmetric loads

Differential charge pump has four switches acting on complementary UP and DN digital signal from PFD. Table 1 explains the working principle of differential CP. The same is implemented in fully differential current-steering method as shown in Fig. 6b.
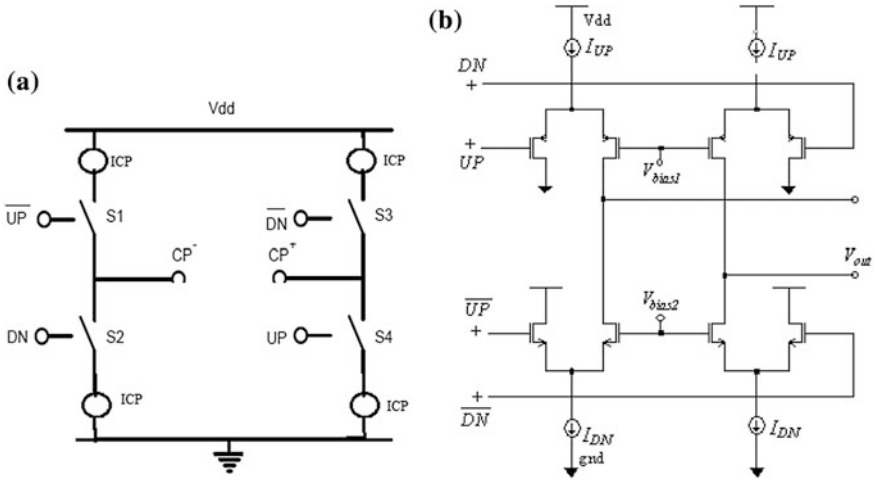
Fig. 6 **a** Ideal differential charge pump **b** Fully differential charge pump

Table 1 Functionality of differential CP

UP	DN	S1, S4	S2, S3	CP−	CP+
Low	High	OFF	ON	Decrease	Increase
High	Low	ON	OFF	Increase	Decrease

4 Simulation Results

Reference frequency (REF_CLK) and feedback frequency (FB_CLK) are applied to DPLL that finds UP and DN signals outputs of PFD and same input to the charge pump. FB_CLK is obtained as output of divider through VCO. The waveform shown in Fig. 7 explains the output of PFD and charge pump combination. The DC mismatch and output voltage range for different CP architectures simulated using the loop parameters are shown in Table 2. The CP results are shown in Table 3. Switch-at-source-based architecture avoids switch errors. Clock feed-through is signal-dependent; thus it can't be completely avoided.
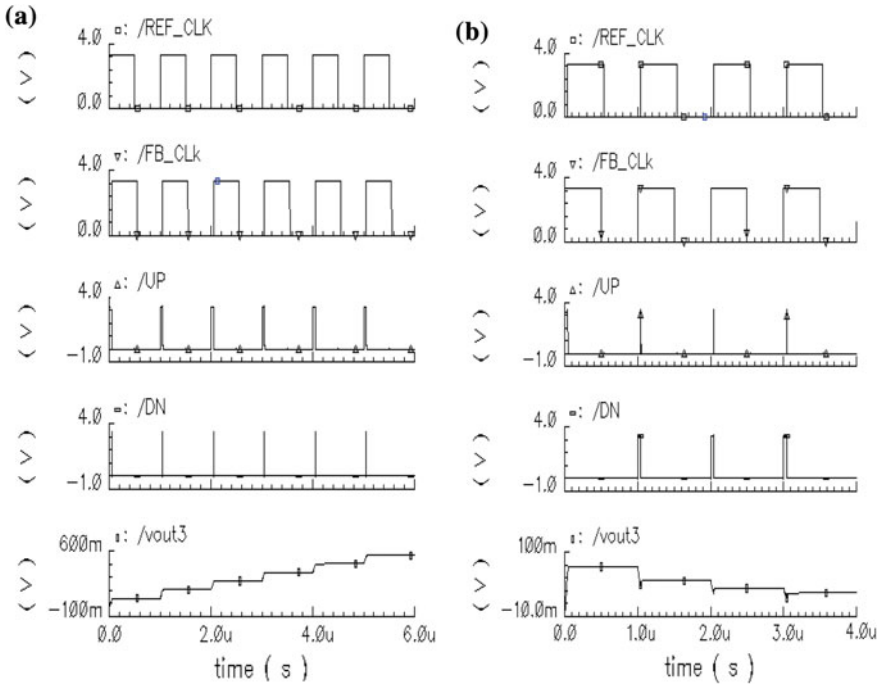
Fig. 7 Step response of PFD/CP/LF **a** REF_CLK lags FB__CLK, **b** REF_CLK lags FB_CLK

Table 2 Parameters of the VCO

Parameter	Values
Tuning range	50–100 MHz
Gain	220–300 MHz/V
Peak jitter	100–200 ps at 250 MHz

Table 3 Results of charge pumps

Charge pumps	Current mismatch (%)	Charge pump output voltage (V)
Charge pump with pull-up transistor Fig. 4a	<0.2	0.3–1.4
Charge pump with dummy switches Fig. 4b	<0.14	0.65–1.3
Differential-in single-ended output current-steering CP Fig. 4c	<0.17	0.4–1.4
Charge pump with active amplifier Fig. 5a	<0.19	0.5–1.4
CP using only NMOS switches Fig. 5c	<0.1	0.3–1.4
Fully differential charge pump Fig. 7	<0.11	0.01–1.76

5   Conclusions

Numerous architectures of single-ended and fully differential charge pumps are investigated. The non-idealities of charge pumps increase jitter in the loop. Pros and cons of CP configurations are discussed. Improved versions of these basic charge pumps are able to achieve high speed and reduced current mismatch. Differential CP has wide output range which helps in tuning VCO. Differential nature of CPs allows better immunity to both power variation and other common-mode disturbances, such as leakage current.

References

1. Rajeshwari DS Rao PV (2014) 3.3 V low power low peak jitter voltage controlled oscillator for MICS applications. In: National conference on VLSI signal processing communication and soft computing, pp. 331–336
2. Zhou J, Wang Z (2008) A high-performance CMOS charge-pimp for phase-locked loops. In: International conference on microwave and millimeter wave technology IEEE 2008, vol 2, pp. 839–842
3. Yang Z, Tang Z, Min H (2008) A fully differential charge pump with accurate current matching and rail-to-rail common-mode feedback circuit. IEEE international symposium on circuits and systems, pp. 448–451
4. Xue L, Zhang Z (2009) Differential charge pumps circuit for high speed PLL application. In: IEEE symposium on industrial electronics and applications, vol 2, pp. 885–888
5. Kalenteridis V, Papathanasiou K, Siskos S (2010) Analysis and design of charge pumps for telecommunication applications. In: International federation for information processing, pp. 43–60
6. Rajeshwari DS, Rao PV, Rajesh V (2016) Charge pump with improved high-swing cascode current source for accurate current matching in DPLL. In: Advances in intelligent systems and computing, pp. 39–47
7. Sun Y, Siek L, Song P (2007) Design of a high performance charge pump circuit for low voltage phase-locked loops. In: International symposium on integrated circuits, pp. 271–274
8. Zheng S, Li Z (2011) A novel CMOS charge pump with high performance for phase-locked loops synthesizer. In: IEEE ICCT 2011, pp. 1062–1065
9. Lee J-S, Keel M-S, Lim S-I, Kim S, Charge pump with perfect current matching characteristics in phase-locked loops. Electron Lett 6(23):1907–1908
10. Wu X-L, Chen J-N, Ke D-M, Zhang X-J (2008) Design of high-speed charge-pump in PLL. In: IEEE 2008, pp. 1–3
11. Ji Jin-Yue, Liu Hai-Qi, Li Qiang. A 1-GHz charge pump PLL frequency synthesizer for IEEE 1394b PHY. J Elect Sci Tech 10(4):319–326
12. El-Hage M, Yuan F. Architectures and design considerations of CMOS charge pumps for phase-locked loops. In: IEEE CCECE 2003, vol. 1, pp. 223–226

Performance Evaluation of 14-nm FinFET-Based Ring Counter Using BSIM-CMG Model

Bhavesh Soni, Gaurav Aryan, Ronit Solanky, Adit Patel and Rajesh Thakker

Abstract Digital circuits are the heart of any modern microprocessor or micro-controller. Because of the vast advantages over analog integrated circuits, digital circuits are superior in terms of speed, performance and power consumption. Nowadays, digital designs are made up using semiconductor elements like MOSFET. This provides high-speed performance. Since the last 40 years, integrated circuits have been improving as per the Moore's law. Reduction in size of transistors introduces several problems such as SCE and DIBL. One of the possible solution is FinFET which can mitigate the problems. Using FinFET and its different topologies, positive edge-triggered D flip-flop-based ring counter is examined at 14 nm technology. Simulation is done using HSPICE and BSIM-CMG FinFET model. By changing gate geometry and substrate, results are carried out. From the results, we can conclude that quadruple gate is better option in terms of delay, average power and current compare to tri-gate and double gate FinFET.

Keywords MOSFET (metal oxide semiconductor field effect transistor)
FinFET · SCE (short-channel effect) · BSIM-CMG (Berkeley short-channel IGFET model-common multigate)

B. Soni (✉) · G. Aryan · R. Solanky · A. Patel · R. Thakker
U.V. Patel College of Engineering, Ganpat University, Mehsana, India
e-mail: bhavesh.soni@ganpatuniversity.ac.in

G. Aryan
e-mail: gaurav.aryan1993@gmail.com

R. Solanky
e-mail: ronitsolanky@gmail.com

A. Patel
e-mail: aditkpatel58@gmail.com

R. Thakker
e-mail: rathakker2008@gmail.com

1 Introduction

As the number of transistors per chip is increasing at an exponential rate, in early 1960s, Gorden Moore gave a prediction on the growth rate of chip complexity. This prediction is called as the Moore's law which states that "the number of transistors per chip would quadruple every 3 years [1]". The semiconductor firms have been following this prediction for the last 40 years. It is not every time possible to follow the prediction as limitations and many other problems that will not allow to follow. This law gave birth to the International Technology Roadmap for Semiconductors (ITRS) organization. Every year, ITRS publishes a report for semiconductor companies and this serves as the benchmark for them. Moore's law can be seen graphically in Fig. 1.

MOSFET is one of the most important semiconductor device having three terminals which are source, gate and drain. The basic principle of MOSFET is that by using the gate voltage, we can control the current between source and drain. The basic structure of MOSFET is shown in Fig. 2. To fulfill the Moore's law, it is required to scale down the MOSFET to integrate more number of transistors on a single chip. It is found that after certain μm and nm scaling range MOSFET performance is degraded. The problems encountered are termed as short-channel effects (SCE). There are many problems associated with SCE.

A MOSFET is called a short-channel device if the channel length (L_{eff}) is approximately equal to the depletion region thickness (x_d) of the drain and source [4]. Since, drain and source are doped with n-type and substrate is of p-type, depletion region does exist. So when the channel length between drain and source $L_{eff} \approx x_d$, device is called short-channel device.

Fig. 1 Transistor number per chip over time [2]
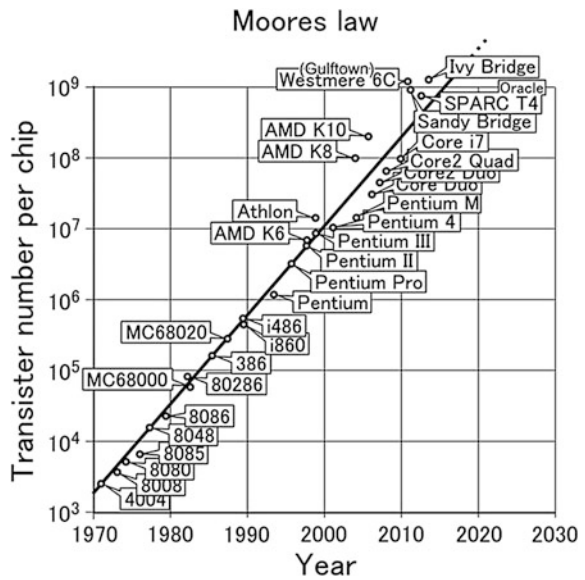
Fig. 2 MOSFET structure
[3]
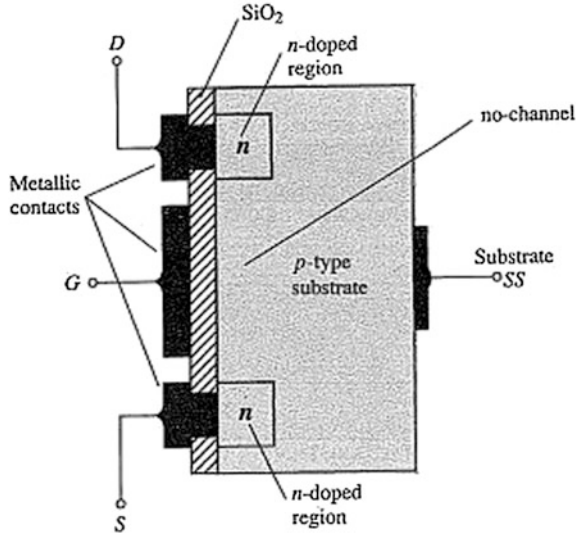
Fig. 3 Tri-gate FinFET [5]
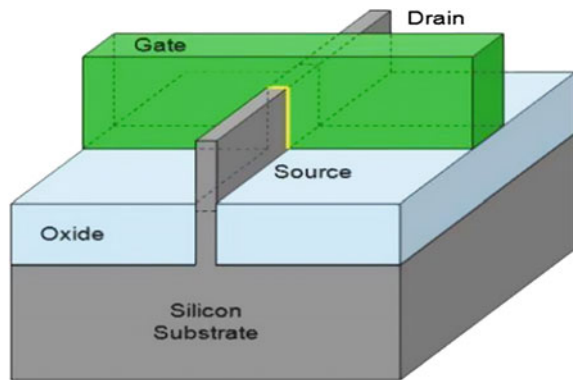
To counter the problems with MOSFETs, FinFET is the possible solution. FinFET structure is shown in Fig. 3. It contains vertical fin, which acts as a channel. This vertical fin allows more area of channel to be covered by the gate. Hence, control over the channel is more in the case of FinFET which offers less SCE.

So, as stated earlier that MOSFET is not a good choice at lower technologies (∼<25–30 nm), we can replace it with FinFET. Since many digital circuits are based on the CMOS technology, we can use FinFET to make an inverter.

Based upon the coverage area of gate over the channel, there are different FinFETs are available. Figure 3 shows tri-gate FinFET in which gate is covering three sides of channel (fin). Accordingly, there are three gates, namely back gate, front gate and top gate. Other combinations are also possible such as double gate (DG) FinFET, Ω-shaped FinFET and cylindrical gate FinFET. Figure 4 shows certain combinations. Different configurations give different advantages.

Fig. 4 Different gate
structures [1]
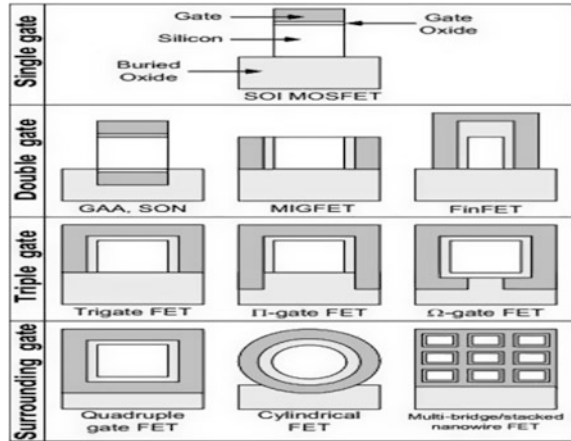
Table 1 Parameters list

Parameters	Value
Gate length, L_g	14 nm
Thickness of Fin, TFIN	15 nm
Height of Fin, HFIN	30 nm
Substrate selector, BULKMOD	0 or 1
Structure selector, GEOMOD	0–2
Number of Fins, NFIN	Variable
Supply voltage, VDD	0.7 V

In this paper, Positive edge-triggered D flip-flop-based ring counter is simulated. One of the most basic elements in digital circuits, it is important to make them work as efficiently as possible in lower technologies.

2 Methodology

Synopsis HSPICE is taken as a simulation tool, whereas BSIM-CMG model is taken for the MOSFET model. HSPICE K-2015.06-SP1 and BSIM-CMG 110.0.0 versions are used for HSPICE and MOSFET model, respectively. BSIM-CMG contains different model's parameters. Some of them are shown in Table 1. By changing these parameters accordingly, we can improve the performance.

For the D flip-flop-based ring counter, gate length is kept at 14 nm. The effective channel width is the function of number of fins (NFIN), thickness of fin (TFIN) and height of the fin (HFIN). The relation can be given as [6, 7]:

$$W_{eff} = NFIN \times (TFIN + 2HFIN) \tag{1}$$

Fig. 5 Positive edge triggered D flip-flop [8]
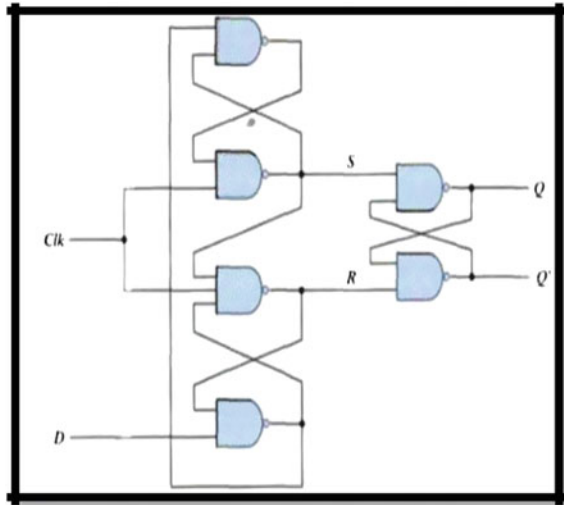
TFIN and HFIN are kept at 12 and 26 nm, respectively, for both nmos and pmos. By changing the NFIN, we can change the W_{eff}. It is important as we need to keep $(W/L)_p \approx 2.5(W/L)_n$ [4].

By selecting proper W_{eff} of nmos and pmos, we can improve the delay of the circuit. Using logical effort, we can properly size the nmos and pmos. In this paper, by selecting the proper NFIN for pmos and nmos, BULKMOD and GEOMOD are changed. BULKMOD = 0 for multigate on SOI substrate and BULKMOD = 1 for multigate on bulk substrate. GEOMOD = 0 for double gate, 1 for triple gate and 2 for quadruple gate. For all these conditions, average power, average current, rise and fall time of output and delay are measured for NAND gate and for D flip-flop (Fig. 5).

3 Simulation Results

Transient analysis is done for 4-bit ring counter. The transient time of 1 p to 10 ns is applied.

We can use D flip-flop for the design of the 4-bit ring counter. Total of 104 transistors are needed to implement 4-bit ring counter. It is shown in Fig. 6. HSPICE allows us to use NAND gate and D flip-flop as a subcircuit. Ring counter is made using subcircuit concept in HSPICE (Fig. 7).

The waveforms can be shown using Custom Waveview. The results of positive edge Triggered D flip-flop are shown in Fig. 8.

The truth table of the D Flip-flop can be verified from the waveform. One can verify the same graph. The simulated results are shown in Table 2.

Fig. 6 4-Bit ring counter schematic using D flip-flop
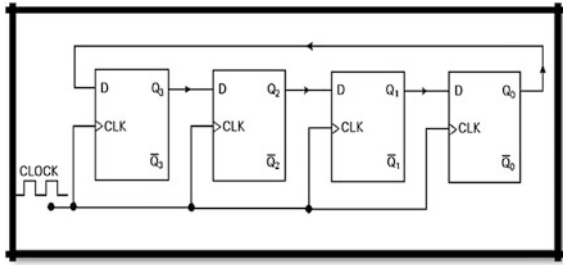
Fig. 7 State diagram and truth-table of ring counter

Q_1	Q_2	Q_3	Q_4	After clock pulse
1	0	0	0	0
0	1	0	0	1
0	0	1	0	2
0	0	0	1	3
1	0	0	0	4
0	1	0	0	5
0	0	1	0	6
0	0	0	1	7

(a) State diagram   (b) Sequence table
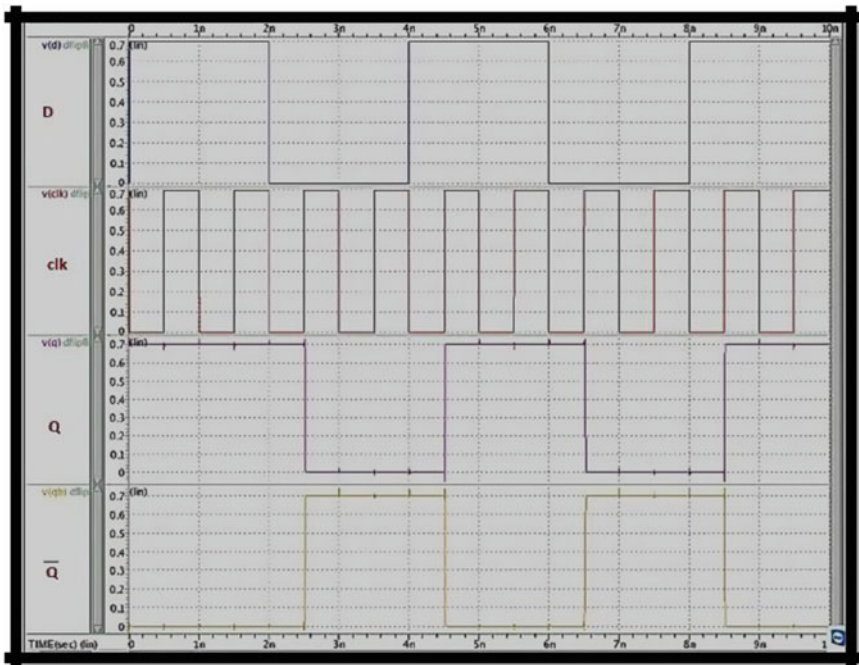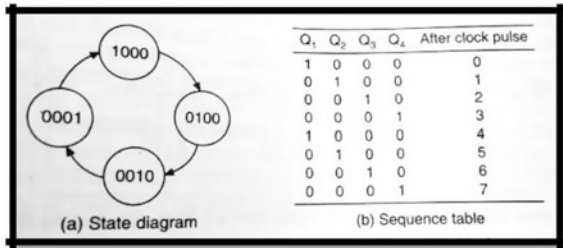
Fig. 8 Positive edge-triggered D flip-flop waveforms [9]

Table 2 Simulated results for 4-bit ring counter

BSIM-CMG parameters		Ring counter		
GEOMOD	BULKMOD	Avg. power (μW)	Avg. current (μA)	Delay (ns)
0	0	4.5218	6.4597	1.9902
1		1.9642	2.8060	1.99
2		1.7717	2.5310	1.9897
0	1	4.5178	6.4540	1.9902
1		1.9919	1.99	2.8456
2		1.7883	2.5548	1.9896
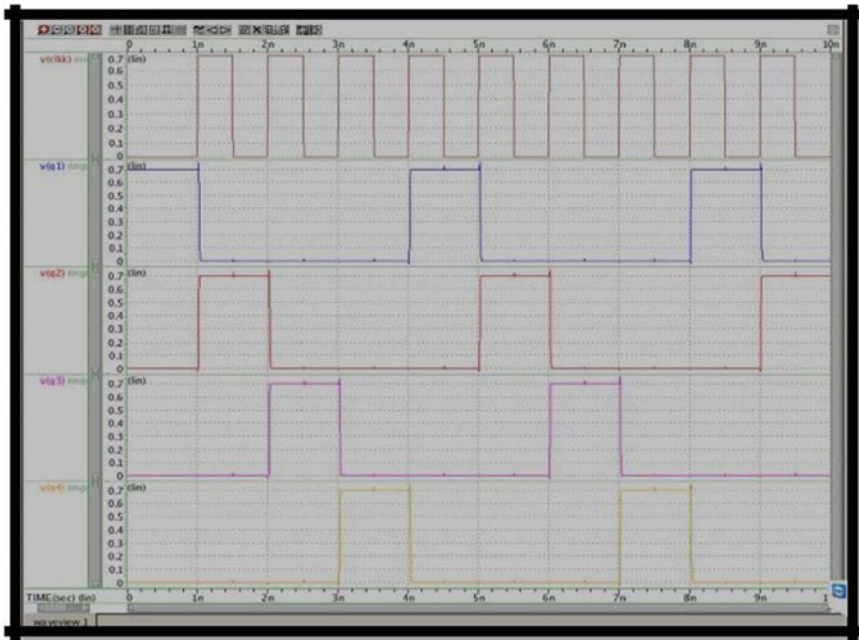
Fig. 9 Waveform for the 4-bit ring counter

As we can see from the graph that at the positive edge of clock, Q changes its state according to input D. Otherwise, it retains the input D. $\overline{Q}$ is the inverted output (Fig. 9).

From the figure, we can see the output wave form of the 4-bit ring counter, initially we have present the Q3 to logic 1 and all other bits are set to logic 0. For every positive edge, clock of the 1 bit is shifted right according to state diagram as shown in figure. So you can verify output waveform from its state diagram.

From the Table 2, we can see that there is no more difference in results for different values of BULKMOD i.e. for SOI or Bulk substrate, there are no much differences. The delay is approximately equal for both BULKMOD value. We can

see that GEOMOD has major impact over the circuit in terms of average power and current. For (GEOMOD = 2), we are getting the best results. Gate all around covers all sides of fin and hence provides better control over the channel. So power dissipation is less because of less leakage current. Similarly, Triple-gate FinFET is better choice over the double gate FinFET.

D flip-flop is used as a subcircuit of 4-bit ring counter, and the number of transistors is more compare to D flip-flop which leads to more current and more power dissipation. We can see from the Table 2 that for quadruple gate on bulk subtract has the least delay, power and average current. From the results we can see that, as the GEOMOD value increases, gate control increases and we get improved results of average power dissipation and average current d. So the gate all around is the best structure in the terms of power, delay and current.

4 Conclusion

Positive edge-triggered D flip-flop-based 4-bit ring counter is examined. BSIM-CMG model is used as a FinFET model. Different gate topologies and substrates are examined for ring counter. 14 nm technology and its related parameters are used for the simulation. We can conclude from the Table 2 that quadruple gate (GEOMOD = 2) is the best choice for ring counter because of the low power, low average current and less delay. Quadruple gate with bulk substrate is better in terms of delay but provides more power and current because of more number of transistors. NFIN does not provide much impact over delay, but it changes the rise time and fall time of the output. Also increase in NFIN will, increase the current and ultimately increases the power dissipation. So, triple-gate FinFET is better in terms of average power and current.

References

1. Colinge JP (2008) FinFET and other multigate transistors. Springer, Berlin
2. Breaking Moore's law [Online]. http://betanews.com. Accessed April 2016
3. Boylestad Electronic devices and circuit theory, 9th edn
4. Kang S-M, Leblebici Y CMOS digital integrated circuits: analysis and design, 3rd edn. Tata McGraw-Hill
5. Bohr M Standards 22 nm—3D trigate transistors presentation. Intel Corporation
6. Lim W, Chin H, Lim C, Tan M (2014) Performance evaluation of 14 nm FinFET-based 6T SRAM cell functionality for DC and transient circuit analysis. J Nanomater 2014
7. Binti A, Tahrim A, Chin H, Lim C, Tan M (2015) Design and performance analysis of 1-bit FinFET full adder cells for subthreshold region at 16 nm process technology. J Nanomater 2015

8. http://electronics.stackexchange.com. Accessed April 2016
9. Morris Mano M, Ciletti MD Digital design. Pearson, Prentice Hall
10. Patil N, Martin C, Oruklu E (2014) Performance evaluation of multi-gate FETs using the BSIM-CMG model. In: IEEE 2014
11. Debajit B, Niraj JK (2014) FinFETs: from devices to architectures, vol 2014. Hindawi Publication
12. Mishra P, Muttreja A, Niraj JK (2011) FinFET circuit design. Nano Electron Circuit Des
13. HSPICE® user guide: basic simulation and analysis, version K-2015.06, June 2015
14. HSPICE® reference manual: MOSFET models, version K-2015.06, June 2015

Theoretical Investigations on the Thermal Effects of VCSEL and Its Impact on the Frequency Response of Multimode Fiber Optic Link

Murali Krishna Karunakaran and Ganesh Madhan Muthu

Abstract The temperature dependence of Vertical Cavity Surface Emitting Laser (VCSEL)'s frequency response and its impact on a multimode fiber optic link is investigated for Radio-over-fiber applications. The frequency response of a 863 nm, bottom emitting VCSEL is evaluated in the temperature range of 20–120 °C. The multimode fiber link bandwidth is determined under different VCSEL temperatures and link lengths. The maximum 3-dB bandwidth is found to be 2.46 GHz for a length of 0.2 km at a bias current and operating temperature of 10 mA and 100 °C, respectively. However, the link bandwidth reduces to 0.37 GHz for a length of 1 km at 100 °C. This study helps to identify the optimum MMF length at different VCSEL temperatures.

Keywords VCSEL · Frequency response · Radio over fiber · Thermal effects
Multimode fiber · Link simulation · Bandwidth

1 Introduction

The market for Vertical Cavity Surface Emitting Lasers (VCSEL) is expected to increase many folds in the coming years, as they find increased application in the fields of optical fiber data transmission, analog broadband signal transmission, gesture recognition, data storage, and chip scale atomic clocks [1]. Due to the development of long wavelength, higher bandwidth, lower cost, wavelength tunability, and low drive currents, VCSELs have become potential candidates for passive optical access networks. The operating temperature for VCSELs not only depends on the ambient, but also increases with bias current, based on the device

M.K. Karunakaran (✉) · G.M. Muthu
Department of Electronics Engineering, Madras Institute of Technology,
Anna University, Chennai, Tamil Nadu 600044, India
e-mail: kmuralikrishnamk@gmail.com

G.M. Muthu
e-mail: mganesh@annauniv.edu

profile. The effects of temperature result in thermal lensing and lead to a significant change in the lasing wavelength, threshold current, optical output power, and mode spectra [2]. Many literatures have dealt with modeling the laser diodes with thermal effects. The temperature effect on static and dynamic characteristics of multilongitudinal mode laser diode had been investigated [3]. An improved transmission line model to study the thermal effects in semiconductor laser diode was reported [4]. The effect of temperature-dependent variations in distributed-feedback laser (DFB) output and the single-mode fiber-induced dispersion on the performance of digital optical link were analyzed [5]. The performance of thin oxide aperture 850 nm VCSEL incorporating thermal effects and complete optical link for gigabit application was investigated [6].

As high-speed optical transmission requires a large modulation bandwidth, factors affecting the frequency response have to be analyzed. The electrical, thermal, and optical issues form the extrinsic factors, while intrinsic limits are based on resonance frequency, damping, and K-factor. Modeling of VCSEL along with parasitics can be mapped to a first-order low-pass filter transfer function. A three-pole filter function includes relaxation–oscillation frequency (f_R), intrinsic damping (γ), and parasitic roll-off (f_{par}) [7]. These parameters can be tuned in the VCSEL design to optimize the modulation bandwidth. The intrinsic high-speed properties of the VCSEL depend strongly on the photon lifetime. The small signal modulation response for increasing currents at room temperature and 85 °C was carried out for a 850 nm VCSEL transmitter [8]. An optimized cavity-gain detuning and advanced 850 nm VCSEL design, operating in the range of 20–90 °C, shows a maximum modulation bandwidth of 13 GHz with modulation current efficiency factor (MCEF) beyond 10 GHz/mA$^{1/2}$ [9]. In longer wavelength VCSELs also, small signal modulation analysis was carried out. In 1550 nm VCSEL, record-high modulation bandwidths beyond 17 GHz at 20 °C and 11 GHz at 80 °C, with proper electro-optical and thermal design are reported [10]. A temperature-dependent small signal analysis for different currents was carried out on a 980 nm VCSEL operating between 25 and 120 °C. It was observed that single-mode devices are limited by damping and multimode devices are restricted by thermal effects [11]. The small signal transfer function analysis of single-mode VCSEL model at 25 and 50 °C shows that the resonance frequencies increase with bias, with the magnitudes of the peaks eventually decreasing [12]. For a bottom emitting laser with pad, the 3-dB modulation bandwidth of over 7 GHz and average threshold current of 0.7 mA was reported [13]. A 3.1-μm-small aperture VCSEL has a maximum 3-dB bandwidth of 15.2 GHz at 2.1 mA bias current. In this model, a bandwidth of 18.2 GHz was achieved for no parasitic effect and 45 GHz intrinsic bandwidth due to gain compression [14]. The modulation response of 863 nm VCSEL for different temperature was simulated using symbolically defined devices (SDD) model in ADS software [15].

It is well known that multimode fibers (MMF) are suited for short range connectivity than single-mode fiber (SMF), when operated at 850 and 1300 nm. They provide near zero absorption but level of scattering decreases at longer wavelengths [16].

The speed of an optical communication system depends not only on the information carrying capacity of fiber, but also on the rate at which the information can be modulated at the transmitter and the speed of response of photodetector at the receiver. From the literature, it is observed that frequency response of VCSEL has been thoroughly investigated. However, the temperature-dependent response and the total VCSEL–MMF link have not been studied extensively. This becomes important in VCSEL-based radio-over-fiber (analog) links, where uncooled, low-cost transmitters are to be used. In this paper, thermal effects on VCSEL are modeled using rate equations and the frequency response is determined. The combined effects of VCSEL's operating temperature and MMF dispersion are analyzed to evaluate the bandwidth of the overall link. The 3-dB bandwidth of the link at various VCSEL operating temperatures is calculated, based on MATLAB simulation.

2 Link Modeling

An optical link comprising of VCSEL diode, MM fiber, and photo detector is modeled to determine the overall frequency response. Hence, the task of obtaining individual frequency responses has to be carried out. As wide band photo detectors are commonly used, its impact is not included in the link model and the cascaded effect of laser and fiber responses in the frequency domain is investigated. Both thermal effects of VCSEL and dispersion, and attenuation effects of MMF are considered in the study.

2.1 VCSEL Diode

An 863 nm bottom emitting VCSEL with a 16-μm aperture diameter is used for the link simulation. The operating temperature is varied from 20 to 120 °C. The VCSEL rate equations incorporating thermal effects are given as [17]

$$\frac{\partial N}{\mathrm{d}t} = \frac{\eta_i(I - I_{\mathrm{off}}(T))}{q} - \frac{N}{\tau_n} - \frac{G_{\mathrm{o}}(N - N_{\mathrm{o}})S}{1 + \varepsilon S} \ . \tag{1}$$

$$\frac{\partial S}{\mathrm{d}t} = -\frac{S}{\tau_p} + \frac{\beta N}{\tau_n} + \frac{G_{\mathrm{o}}(N - N_{\mathrm{o}})S}{1 + \varepsilon S} \ . \tag{2}$$

$$T = T_{\mathrm{o}} + (IV - P_{\mathrm{o}})R_{\mathrm{th}} - T_{\mathrm{th}}\frac{\partial T}{\partial t}. \tag{3}$$

where the term I_{off} represents the offset current which is determined by the following expression,

Table 1 Parameters for the VCSEL laser diode model [17, 18]

Parameter	Value
Injection efficiency, η_i	1
Spontaneous emission coupling coefficient, β	1×10^{-6}
Carrier recombination lifetime, τ_n	5 ns
Output coupling efficiency, k	2.6×10^{-8} W
Gain coefficient, G_o	1.6×10^4 s^{-1}
Carrier transparency number, N_o	1.94×10^7
Photon lifetime, τ_p	2.28 ps
1st Temperature coefficient, a_o	1.246×10^{-3} A
2nd Temperature coefficient, a_1	-2.545×10^{-5} A/K
3rd Temperature coefficient, a_2	2.908×10^{-7} A/K^2
4th Temperature coefficient, a_3	-2.531×10^{-10} A/K^3
5th Temperature coefficient, a_4	1.022×10^{-12} A/K^4
Thermal impedance, R_{th}	2.6 °C/mW
Thermal time constant, T_{th}	1 μs
Gain compression factor, ε	5×10^{-7}

$$I_{\text{off}}(T) = a_0 + a_1 T + a_2 T^2 + a_3 T^3 + a_4 T^4. \tag{4}$$

The coefficients a_0–a_4 are obtained during parameter extraction. 'N' and 'S' refer to carrier number and photon number in the cavity. The laser output power is evaluated using the following equation and it is given as [17].

$$P_o = kS \tag{5}$$

where k is the output coupling efficiency.

The VCSEL parameters used in the rate equations are tabulated in Table 1.

2.2 Multimode Fiber (MMF)

Multimode fiber (MMF) can be equivalently represented as a simple low-pass filter and modeled using a transfer function approach. The MMF used in the simulation has a 50 μm core diameter, whose functionality as filter can be evaluated using impulse and frequency responses.

The impulse response of a MMF can be generalized as a probability density function of a normal distribution and is given as [19]

$$h_{\text{mmf}}(t) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(t-\tau)^2}{2\sigma^2}\right) \tag{6}$$

where τ is the delay of the channel; σ is the standard deviation of the impulse response.

The frequency response of the fiber is given by,

$$H_{\text{mmf}}(\omega) = \exp\left(\frac{-\omega^2\sigma^2}{2} - j\omega\tau\right)$$ (7)

where $H_{\text{mmf}}(\omega)$ is the Fourier transform of the impulse response of Eq. (6).

The fiber parameters used in this work are similar to that of Yuen et al. [19].

The cascaded effect of the VCSEL and MMF models can be represented in frequency domain using convolution theorem [20],

$$F[h_1(t) * h_2(t)] = [H_1(f)] \cdot [H_2(f)] = [H_2(f)] \cdot [H_1(f)].$$ (8)

Thus, the frequency-domain response of the overall link is the product of the frequency responses of the individual systems.

3 Simulation Results and Discussion

3.1 VCSEL Diode DC Characteristics

The rate equations for the VCSEL provided by Eqs. (1)–(5) are solved using an ODE solver available in MATLAB software, for the temperature range of 20–120 °C. Figure 1 denotes the power–current (P–I) characteristics of the diode, which illustrate a linear variation of optical power and then followed by a roll-off,
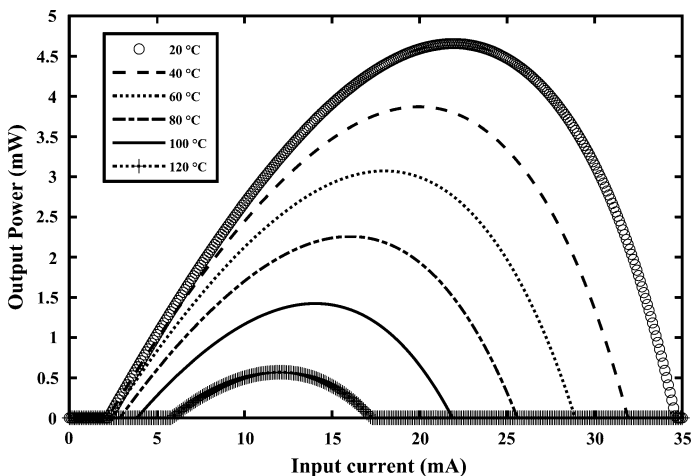
Fig. 1 P–I curves of VCSEL diode

after certain maximum value. These results are in accordance with the work done by
Mena et al. [17], for the same structure, thereby validating our simulation.

3.2 VCSEL Diode Small Signal Characteristics

The VCSEL is DC biased with a current of 6 mA and RF current with peak-to-peak
value of 2 mA is applied. The analyses are carried out in the frequency range of
10 MHz–10 GHz, for different bias currents in the range of 6–12 mA. As the bias
current increases, optical power rises and bandwidth also improves in the linear
regime. For comparing the response curves at different biases, they are normalized
by the low frequency value, calculated at 10 MHz [12]. Figure 2a, b illustrate the
simulation results, for 20 and 100 °C, which indicate that both resonant frequency
and modulation bandwidth decreases as operating temperature increases. This is in
accordance with literature [17]. As the optical power output at 120 °C is less, the
simulation is carried out for 20–100 °C.

3.3 MMF Impulse and Frequency Response

The MMF link length is varied from 200 m to 1 km, which is quite optimum range
for MMF in RoF applications. The impulse response shows that as the fiber length
increases, attenuation and dispersive effects become predominant. The graph is
normalized to unity for comparison. The frequency response is obtained by taking
Fourier transform of the fiber's impulse response. The 3-dB bandwidth decreases as
the length of the fiber is increased, which is logical. Figure 3 explains the distance
effect in both time and frequency domains.
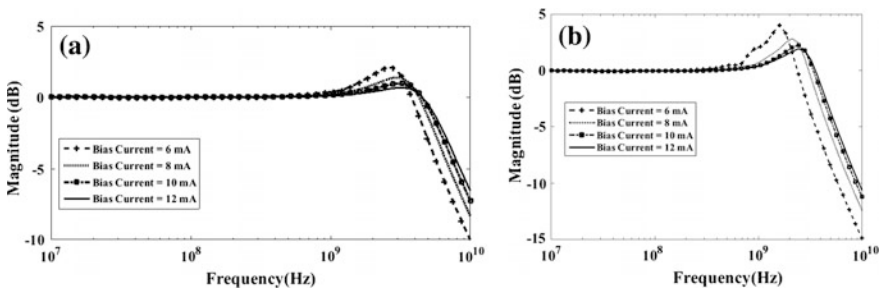
Fig. 2 Small signal modulation response of VCSEL at **a** 20 °C, **b** 100 °C
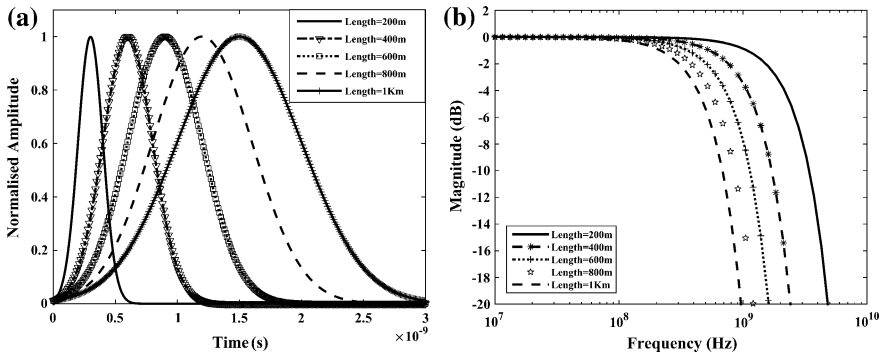
Fig. 3 MMF **a** impulse response, **b** frequency response

3.4 Frequency Response of VCSEL–MMF Link

The combined effect of VCSEL temperature and MMF dispersion plays a crucial role in the overall frequency response. The VCSEL is biased at 8 mA and the temperature is fixed as 20 °C. The individual responses of the laser, MMF, and the combined response are plotted in Fig. 4. It is observed that, at lower link lengths, the total bandwidth is dictated by the laser's response (Fig. 4a). However, for larger link lengths, the fiber dispersion becomes predominant factor affecting the bandwidth, which is illustrated in Fig. 4c.

The simulation is repeated for a maximum operating temperature of the VCSEL, namely 100 °C, biased at 8 mA. The results shown in Fig. 5 are quite similar to that of Fig. 4, except the effect of temperature is more pronounced in the link, compared to 20 °C case. This has a significant effect and tends to increase the 3-dB bandwidth of the overall link.

The simulation outcome for laser response alone is shown in Fig. 6, which indicates an increase in 3-dB bandwidth for increasing bias current. However, the modulation bandwidth decreases with increasing temperature. For 20 °C, the 3-dB bandwidth found to be 6.85 GHz at 12 mA, while it is 5.85 GHz at 8 mA. Also for 100 °C, the 3-dB bandwidths are 4.69 GHz, 3.93 GHz for 12 mA, and 8 mA bias currents, respectively.

The following Fig. 7 shows the frequency response of VCSEL–MMF link, when the laser diode is operated at 20 °C and biased with 10 and 12 mA. The link length is fixed as 200 m. The response of fiber dominates laser response in the cascaded effect. Hence, the link response is similar to that of the fiber.

Figure 8 shows the frequency response of VCSEL–MMF link, as the laser diode is operated at 20 and 100 °C, biased at 10 mA. The link length is fixed as 200 m. The response of fiber provides lesser bandwidth than the laser. Hence, the link response is similar to that of fiber, but the effect of laser operating temperature decides the deviation from the fiber response. The laser response at 100 °C provides
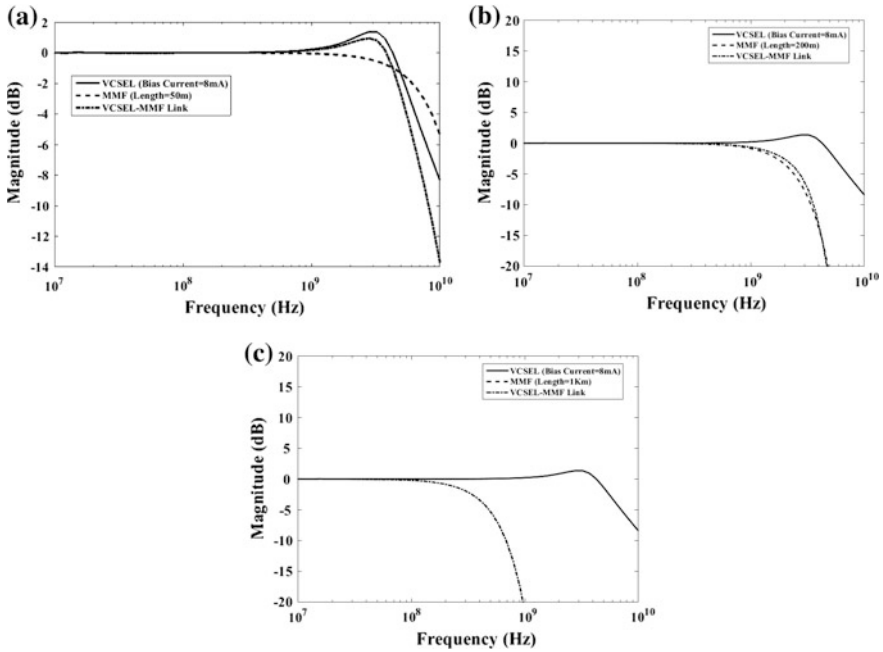
Fig. 4 Frequency response of optical link with VCSEL biased at 8 mA and 20 °C: **a** length—0.05 km, **b** length—0.20 km and **c** length—1 km

a significant resonant peak compared to 20 °C which increases the overall 3-dB link bandwidth.

Table 2 provides details about the maximum bandwidth of VCSEL–MMF analog link, at different temperatures and lengths. It is observed that even at higher bias currents, the overall link bandwidth is not enhanced. This is due to response of the MMF. However, the presence of significant resonance peak at lower bias current also slightly improves the overall bandwidth. Further, higher operating temperature also provides a hump in the frequency response, due to increased threshold. This in fact aids in improving the link bandwidth.

In conventional electronic amplifiers, inductive peaking technique provides a resonance peak near the 3-dB cutoff frequency of the original amplifier, thereby increasing its bandwidth. Similarly, at higher operating temperatures (100 °C), VCSEL provides a significant resonance peak, as shown in Fig. 8. If the bias point of the device is near the threshold, the resonance peak is predominant. This behavior depends on the bias current and the threshold current of the device, which is temperature dependent. As the bias point moves away from the threshold (or) bias point fixed as several multiples of threshold, the resonance peak disappears. Thus, the 3-dB bandwidth calculated is more for larger bias current than near threshold for a particular temperature case. Also, while operating at different
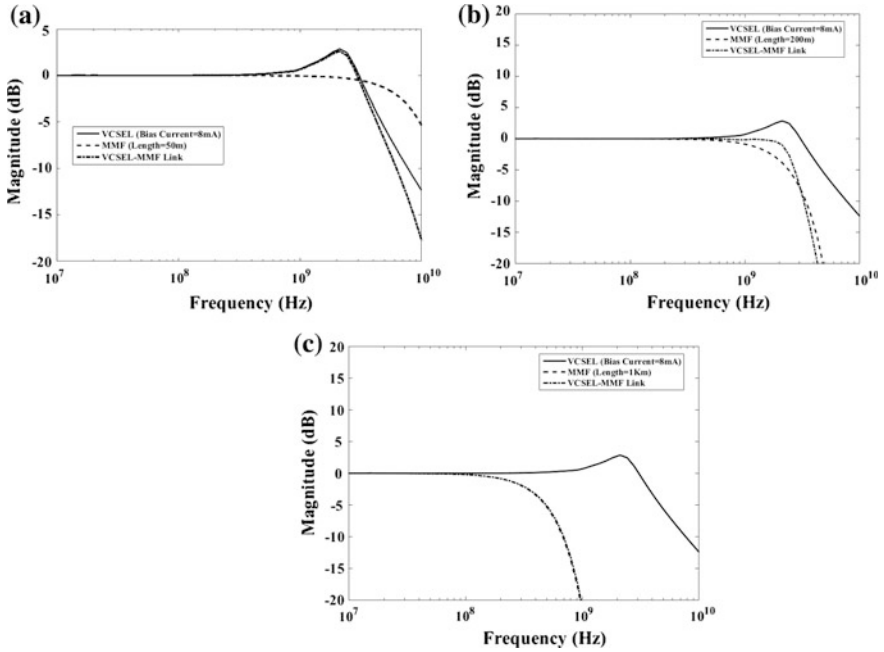
Fig. 5 Frequency response of optical link with VCSEL biased at 8 mA and 100 °C: **a** length—0.05 km, **b** length—0.20 km and **c** length—1 km
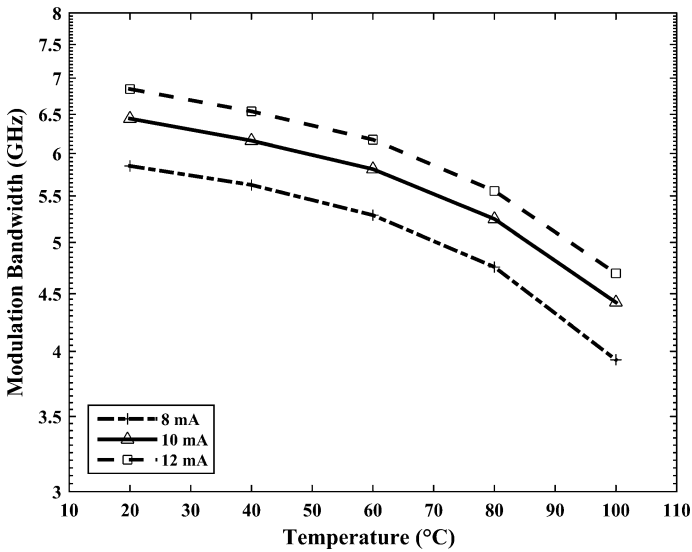
Fig. 6 Modulation bandwidth versus temperature
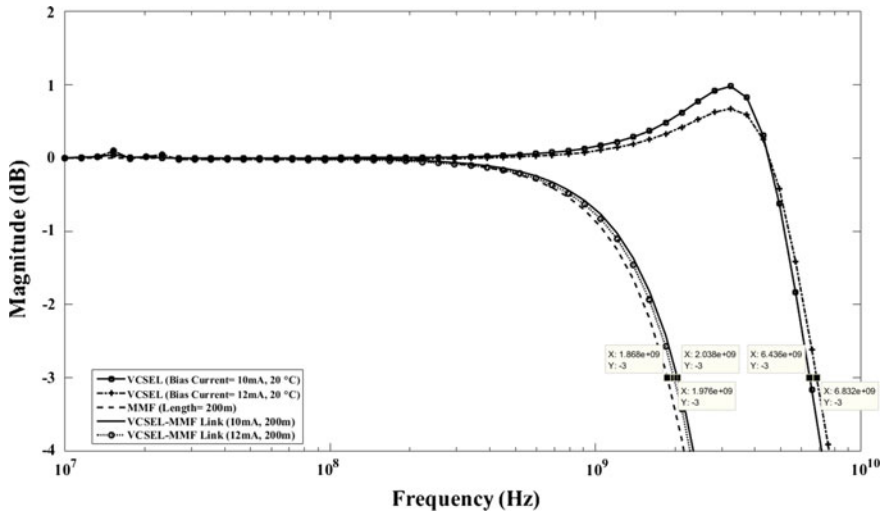
Fig. 7 Frequency response of VCSEL, MMF, and optical link at 20 °C biased at 10 and 12 mA and link length of 0.2 km
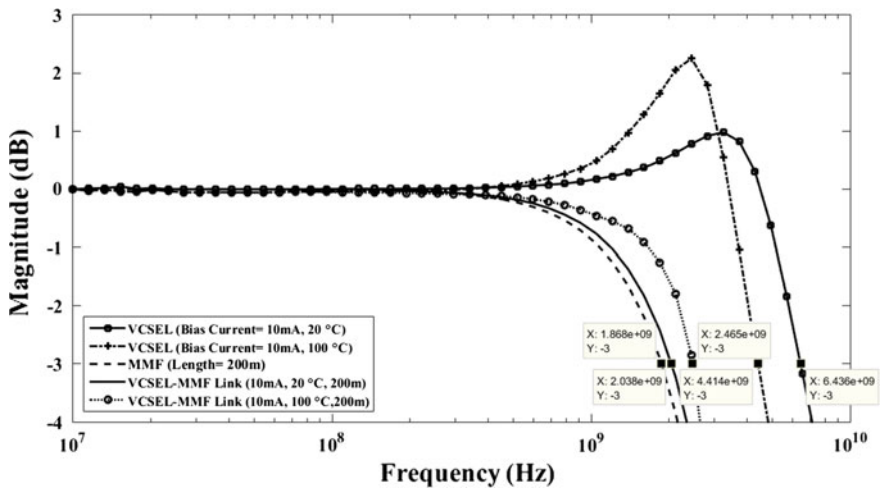
Fig. 8 Frequency response of VCSEL, MMF, and optical link at 20 and 100 °C biased at 10 mA and link length of 0.2 km

temperatures for the same bias current to VCSEL, the bandwidth depends on the operating current with respect to the threshold. Hence, when the bias current is chosen far away from threshold for lower temperatures, it becomes closer for the higher temperature case, with respect to the former. In this case, the laser response

Table 2 Link bandwidth

Distance (m)	Temperature (°C)	Optical link bandwidth (GHz)	
		I_b = 10 mA	I_b = 12 mA
200	20	2.04	1.98
	60	2.14	2.06
	100	2.47	2.39
600	20	0.63	0.62
	60	0.63	0.63
	100	0.64	0.63
1000	20	0.37	0.37
	60	0.37	0.37
	100	0.37	0.37

of higher temperature shows a significant resonance peak and when combined with MMF response, it provides an increase in 3-dB bandwidth than the lower operating temperature.

4   Conclusion

In this paper, optical link for RoF application comprising of a 863 nm bottom emitting VCSEL diode and MMF is simulated. While considering laser diode alone, increasing the operating temperature decreases the modulation bandwidth. Further, when bias current is increased within the particular operating temperature, bandwidth also increases. However, when cascading the laser response with fiber response, the thermal effect plays major role in slightly improving the optical link bandwidth except at longer distance where dispersion effect dominates. For a VCSEL bias current of 10 mA and operating temperature at 100 °C, and fiber length of 0.2 km, the 3-dB bandwidth is found to be 2.46 GHz. However, the link bandwidth reduces to 0.37 GHz for a length of 1 km with same laser operating conditions.

References

1. VCSEL market growing at 21.9% from $781.6 m in 2015 to $2.1bn in 2020. http://www.semiconductor-today.com/news_items/2016/apr/bcc_070416.shtml
2. Osinski M, Nakwaski W (1994) Thermal effects in vertical-cavity surface-emitting lasers. Int J High Speed Electron Syst 5(4):667–730

3. Ganesh Madhan M, Vaya PR, Gunasekaran N (2001) An unified approach to study the thermal dynamics in multi longitudinal mode semiconductor lasers. Fiber Integr Opt 20 (2):159–170

4. Ganesh Madhan M, Neelakandan R (2008) An improved transmission line laser model for multimode laser diodes incorporating thermal effects. Opt Quantum Electron 40:535–550

5. Koussalya B, Ganesh Madhan M (2009) Simulation of thermal effects in laser diode and its impact on high speed fiber optic link. J High Speed Netw 17(4):175–184

6. Murali Krishna K, Ganesh Madhan M (2016) Performance analysis of a low cost VCSEL transmitter based multimode fiber optic link for gigabit ethernet application. In: 5th IEEE international conference on communication and signal processing, pp 0227–0231

7. Hofmann W, Müller M, Nadtochiy A, Meltzer C, Mutig A, Böhm G, Rosskopf J, Bimberg D, Amann MC, Chang-Hasnain C (2009) 22-Gb/s long wavelength VCSELs. Opt Express 17 (20):17547–17554

8. Kuchta DM, Rylyakov AV, Schow CL, Proesel JE, Baks CW, Westbergh P, Gustavsson JS, Larsson A (2015) A 50 Gb/s NRZ modulated 850 nm VCSEL transmitter operating error free to 90 °C. J Lightwave Technol 33(4):802–810

9. Nazaruk DE, Blokhin SA, Maleev NA, Bobrov MA, Kuzmenkov AG, Vasil'ev AP, Gladyshev AG, Pavlov MM, Blokhin, AA, Kulagina MM, Vashanova KA, Zadiranov YM, Fefelov AG, Ustinov VM (2014) Single-mode temperature and polarization-stable high-speed 850 nm vertical cavity surface emitting lasers. In: 16th Russian youth conference on physics and astronomy, IOP Publishing Ltd., pp 1–6.

10. Müller M, Gründl T, Horn M, Nagel, RD, Wiedmeier W, Rönneberg E, Böhm G, Amann MC (2010) Small-signal analysis of high-temperature stable 1550 nm high-speed VCSELs. In: 6th joint symposium on opto- and microelectronic devices and circuits, pp 1–4

11. Mutig A, Fiol G, Pötschke K, Moser P, Arsenijevic D, Shchukin VA, Ledentsov NN, Mikhrin SS, Krestnikov IL, Livshits DA, Kovsh AR, Hopfer F, Bimberg D (2009) Temperature-dependent small-signal analysis of high-speed high temperature stable 980-nm VCSELs. IEEE J Sel Top Quantum Electron 15(3):679–686

12. Mena PV, Morikuni JJ, Kang SM, Harton AV, Wyatt KW (1999) A comprehensive circuit-level model of vertical-cavity surface-emitting lasers. J Lightwave Technol 17 (12):2612–2632

13. Dutta AK, Kosaka H, Kurihara K, Sugimoto Y, Kasahara K (1998) High-speed VCSEL of modulation bandwidth over 7.0 GHz and its application to 100 m PCF datalink. J Lightwave Technol 16(5):870–875

14. Thibeault BJ, Bertilsson K, Hegblom ER, Strzelecka E, Flyod PD, Naone R, Coldren LA (1997) High-speed characteristics of low-optical loss oxide-apertured vertical-cavity lasers. IEEE Photonics Technol Lett 9(1):11–13

15. Tian X, Wang Z, Gao J (2006) Thermal analysis based on rate-equation model for VCSELs. In: Proceedings on SPIE 6352, optoelectronic materials and devices, 63523L, pp 1–4

16. Facts—yes, you do need to read this, https://www.garlandtechnology.com/blog/fiber-facts-yes-you-do-need-to-read-this

17. Mena PV, Morikuni JJ, Kang SM, Harton AV, Wyatt KW (1999) A simple rate-equation-based thermal VCSEL model. J Lightwave Technol 17:865–872

18. Qi C, Shi X, Wang G (2011) High-order circuit-level thermal model of vertical-cavity surface-emitting lasers. IET Optoelectron 5(1):19–27

19. Yuen R, Fernando XN, Krishnan S (2004) Radio over multimode fiber for wireless access. In: Canadian conference on electrical and computer engineering, vol 3, pp 1715–1718

20. Jeruchim MC, Balaban P, Sam Shanmugan K (2002) Simulation of communication systems modeling, methodology and techniques, 2nd edn. Kluwer, Dordrecht

Design and Modeling of Quadcopter for Elephant Monitoring Using Aerial Image Processing Approach

S.J. Sugumar and V. Sai Babu

Abstract Human elephant conflict is a major problem in villages located near the forests. They are the huge threat to life and property of the dwellers. Till now, many traditional methods are employed to monitor the elephants along with some technologies are also deployed. One such is the camera, it is not efficient to use permanent cameras since the placement of cameras is critical. We propose an alternate way using quad copters for surveillance of elephants using cameras, thanks to their agility and size. It is propelled by four high rpm motors attached to the end of four arms with adjacent arms perpendicular to each other. The motion of the drone is controlled by varying the speed of individual motors. The quad copter is deployed for the surveillance of elephants at a height of 100–300 m in a 2 km span depending on the flora characteristics and terrain in elephant prone regions of the forest. The image captured from the top view is transmitted to the receiver and is processed by converting into HSV and extracting the hue model using structural element function. BLOB analysis is performed to detect the presence of elephants. An early warning is sent which allows the forest executives to alert the local people and to take necessary actions to divert the elephants.

Keywords HEC · Elephant · Quad copter · HSV · BLOB

1 Introduction

In recent times the intrusion of elephants into human habitat has increased drastically. This is mainly due to the loss of the habitat of the elephants. Due to this intrusion, there is loss of life and property of the local residents and also the life of the elephants [1]. It is also our prime responsibility to conserve the Elephant

S.J. Sugumar (✉) · V. Sai Babu
ECE, GNITC, Ibrahimpatnam, Telangana, India
e-mail: sjsgnitc@gmail.com

V. Sai Babu
e-mail: saivoruchu@gmail.com

Fig. 1 Elephant intrusion for the past 6 years in Coimbatore
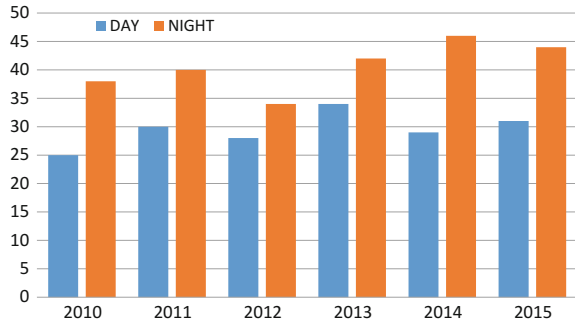
population which is decreasing at an alarming rate. In Africa and Asia, elephant habitat is being replaced by agriculture—both by small-scale farmers, international agribusiness such as palm oil, industrial growth, educational instructions, infrastructural growths etc. Elephants are being squeezed into smaller areas. As a result, elephants frequently raid and destroy crops which are planted near their corridors. This makes elephants to come closer to human living areas which lead to human conflict (HEC). Due to HEC, both elephants and human loss is increasing every year. Fierce competition for living space has resulted in human suffering, a dramatic loss of forest cover, and reduced Asian elephant numbers to between 25,600 and 32,750 animals in the wild. Asian elephant populations are highly fragmented, with more than 1000 individuals in a contiguous area, greatly decreasing their chances for survival. Most of the national parks and reserves where elephants reside are too small to accommodate viable elephant populations. The conversion of forest areas to agricultural use also leads to serious elephant–human conflicts. Human and elephant deaths due to conflicts during the period 1999–2015 in Coimbatore Forest Division is analyzed and based on secondary data collected from the Forest Department we could find out the damages caused due to various conflict issues. The human causalities between 2010 and 2015 alone attributed 59% of overall deaths. Most of the human deaths (67%) were recorded in outside of the forest areas in forest borders. Totally 133 elephant deaths were recorded from 1999 to 2015 [2] (Fig. 1).

2 Related Work

Vermeulen et al. [3] has proposed an unmanned aircraft to survey the elephant. The unmanned aircraft has been flown through a certain path in the forest at about 100 m of altitude. The camera which is installed in the aircraft captures images automatically and stores in the SD card. Number of elephants detected has been observed from the stored images after the flight has returned to the base station. Sugumar and Jayaparvathy [4] installed Cameras on tower or trees which capture the images of the intruding elephant. The captured image is transmitted to the base

station via RF network. The received image is processed by comparing it with the stored database images of elephants. Luukkonen [5] explains the mathematical modelling and control of a quadcopter. The mathematical model of quadcopter dynamics is presented and the differential equations are derived from the Newton–Euler and the Euler–Lagrange equations. The model is verified by simulating the flight of a quadcopter with Matlab. Bendale and Karwankar [6] describes a method for tracking moving objects from a sequence of video frame. The required object is then identified by placing a bounding box around the object. In this paper we propose a method of aerial survey to detect the intrusion of wild animals and to alert the local people and also the forest rangers. An aerial drone quadcopter is deployed in this work and it is made to fly around a known path and by using the camera attached to it, aerial photos of the elephant are taken and transmitted. The images after being received are then processed to detect the presence of elephants.

3 Methodology

The model of the proposed work is shown in Fig. 2 and explained below. Quadcopter consists of four brushless direct current (BLDC) motors individually controlled by four electronic speed controllers (ESC). ESCs take input from battery and according to the PWM input obtained from the microcontroller, sends alternating voltage to the motor windings. Embedded controller gets input from the Inertial Measurement Unit (IMU) sensor that consists of accelerometer and gyroscope and compares with the threshold value. According to the error signal computed, it uses the manually tuned Proportional and Integral values to correct the error.

Fig. 2 Block diagram of the project
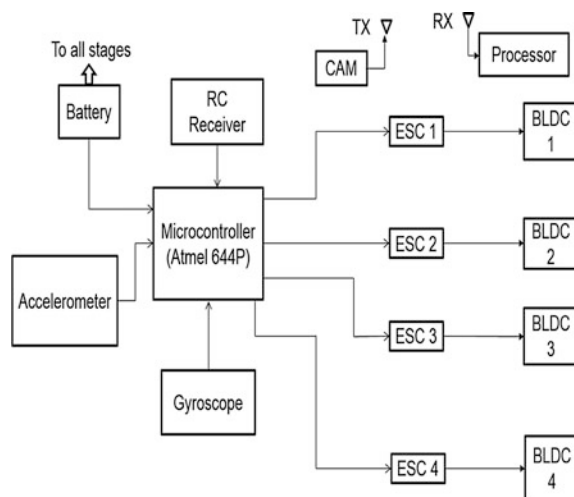
This is a vital part of the device since stability is a critical feature for aerial applications. The alternating voltage in the stator then, according to electrome-chanical energy conversion principle, is converted into rotational motion. The direction of the quadcopter is controlled by varying the speed of individual motors. The control signal is given manually which is wirelessly transmitted to the receiver connected to the microcontroller. A camera that is connected to the quadcopter takes pictures and transmits them continuously to the base station (a personal computer). Using suitable image processing techniques [6] in MATLAB for ele-phant detection. The video recorded on the camera mounted on the quadcopter is wirelessly transmitted to the base station. This video is given as input to PC. Elephant Image Segmentation involves filtering out of the required portion of the object from the image.

4   Quadcopter Modelling

The quadcopter is basically governed by three motions namely pitch, roll and yaw about the axis x, y and z respectively can be modelled as a function of thrust. There exists 6 degrees of freedom controlled by 4 independent rotor speeds [7]. When all the motors rotate at same rpm, the quadcopter lifts in vertical direction. The pitch is forward and backward motion of the quadcopter. When the speed of motors M2 and M3 are increased above motors M1 and M4, the quadcopter flies in forward direction and vice versa for backward direction. The roll is the left and right motion of the quadcopter. When the speed of motors M2 and M4 is increased above the motors M1 and M3, the quadcopter flies to the left and vice versa for right direction. Yaw motion is the rotation of the quadcopter about its own axis. When the speed of motors M3 and M4 are increased above the speed of motors M1 and M2, the quadcopter rotates in clockwise direction vice versa for the quadcopter to rotate in counter clockwise direction [8]. The mechanical parameters can be evaluated in inertial frame (frame of observer) and body frame (frame of quadcopter). In order to switch between inertial and body frame, rotational matrix should be used. Rotational matrix for roll motion about x axis,

$$R_x(\Phi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & c\Phi & sn\Phi \\ 0 & -sn\Phi & c\Phi \end{pmatrix} \tag{1}$$

Rotational matrix for pitch angle about y axis,

$$R_Y(\theta) = \begin{pmatrix} c\theta & 0 & -sn\theta \\ 0 & 1 & 0 \\ sn\theta & 0 & c\theta \end{pmatrix} \tag{2}$$

Rotational matrix for yaw angle about z axis,

$$R_Z(\psi) = \begin{pmatrix} c\psi & sn\psi & 0 \\ -sn\psi & c\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3}$$

The rotational matrix for the complete model can be obtained by multiplying the rotational matrix along the individual axis.

$$R_{xyz}^{x_a x_b x_c} = \begin{pmatrix} c\theta * c\psi - sn\Phi * sn\theta * sn\psi & -c\theta * sn\psi & c\psi * sn\theta + c\theta * sn\Phi * sn\psi \\ c\Phi * sn\psi + sn\Phi * sn\theta * c\psi & c\Phi * c\psi & sn\theta * sn\psi - sn\Phi * c\theta * c\psi \\ -c\Phi * sn\theta & sn\Phi & c\Phi * c\theta \end{pmatrix} \tag{4}$$

This is the rotational matrix of the quadcopter. Thrust is resolved into individual axis and represented in the matrix form, since the thrust is produced only in z axis,

$$\overrightarrow{T_{xyz}} = \begin{pmatrix} c\psi * sn\theta + c\theta * sn\Phi * sn\psi \\ sn\theta * sn\psi - sn\Phi * c\theta * c\psi \\ c\Phi * c\theta \end{pmatrix} * T \tag{5}$$

Acceleration in individual axis is derived from the above equation.

$$\overrightarrow{\ddot{r}_x} = \frac{(c\psi * sn\theta + c\theta * sn\Phi * sn\psi) * T - C_d * \rho * S * V_x^2}{m} \tag{6}$$

$$\overrightarrow{\ddot{r}_y} = \frac{(sn\theta * sn\psi - sn\Phi * c\theta * c\psi) * T - C_d * \rho * S * V_y^2}{m} \tag{7}$$

$$\overrightarrow{\ddot{r}_y} = \frac{(c\Phi * c\theta) - m * g - C_d * \rho * S * V_z^2}{m} \tag{8}$$

From the quadcopter modelling [9], the various motions of quadcopter and the requirements to achieve those motions can be understood. The mathematical model that governs the motion of quadcopter is derived and explained. Maximum flight time can be calculated by dividing total capacity of the battery by the current drawn by four motors. Maximum flight time (in hours) = (battery capacity)/Current drawn by motor * No. of motors = (4200 mAh/1000)/(9A × 4 motors) = 0.116 h.

$$\text{Maximum flight time (in min)} = 0.116 \times 60 = 7\, \text{min.}$$

5 Quadcopter Force and Velocity About Each Axis

Force produced by the quadcopter to climb at given maximum RPM [10] is given by

$$F_{\text{climb}} = F_{\text{thrust}} - F_{\text{gravity}} - F_{\text{drag}} \tag{9}$$

$$F_{\text{thrust}} = 14.7585 \text{ N} \tag{10}$$

$$F_{\text{gravity}} = m * g = 12.2625 \text{ N} \tag{11}$$

$$F_{\text{drag}} = \frac{\rho}{2} * c_d * S * V^2 = 2.43 \text{ N} \tag{12}$$

$$F_{\text{climb}} = 0.0641 \text{ N}$$

$$F_{\text{forward}} = \sqrt{1 - \left(\frac{mg}{T}\right)^2} * T - \frac{\rho}{2} * c_d * S * V^2 = 5.78 \text{ N} \tag{13}$$

Velocity about horizontal and vertical axis is given by,

$$v_{\text{ver}} = \sqrt{2 * \frac{(T - mg)}{\rho * C_D * S}} = 5.25 \text{ m/s} \ . \tag{14}$$

$$v_{\text{Hor}} = \sqrt{\frac{2 * \left(\left(1 - \left(\frac{mg}{T}\right)^2\right)^{0.5}\right) * T}{\rho * C_D * S}} = 9.51 \text{ m/s.} \tag{15}$$

$$\text{Maximum Pitch Angle} \quad \alpha = \arcsin\left(\frac{mg}{T}\right) \alpha = 56.18^{\circ} \tag{16}$$

The various components used to build this project and their weight distribution are shown in the table. The quadcopter is designed for 1250 g and the actual value falls under it.

6   Results

The mechanical structure of the quadcopter is made of aluminum that consists of base plate and arms. Arms are fitted to base plate by screws and ends are cut and filed to fit the motors. The adjacent arms are perpendicular to each other in order to achieve symmetry. The center of gravity is brought as close as possible to the center of the quadcopter in order to obtain better stability. Battery and microcontroller are placed above the frame at the exact center. Camera is mounted at the bottom of the quadcopter frame. The picture of elephant is taken from the camera mounted on quadcopter at Peraiyur, Virudhunagar district in Tamil Nadu. This is used as the test image. The algorithm is implemented and the results are shown using this image in

Fig. 3 Hardware testing of captured image from the quadcopter
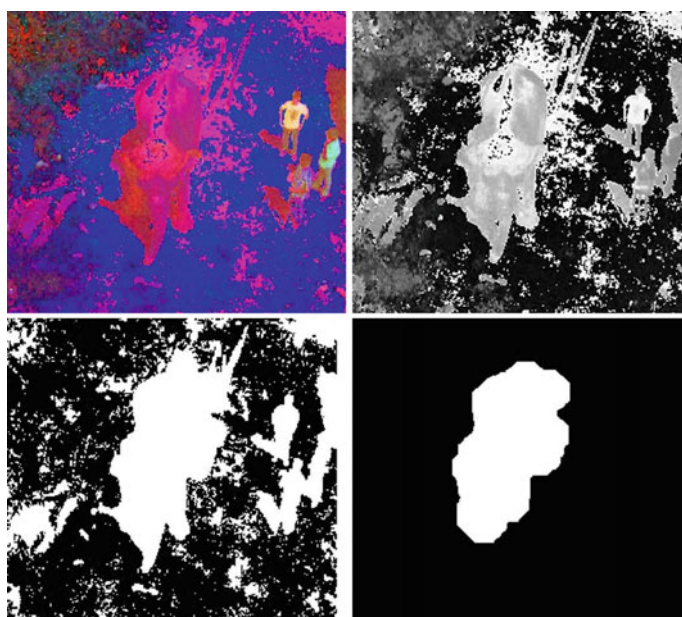
Fig. 4 Hardware testing of captured image from the quadcopter

Fig. 3. The input colour image is converted into a binary image based on a threshold value. Input image with luminance greater than the threshold value have the value 1 (white) and replaces all other pixels in the image with the value 0 (black). The binary image is then processed using 'imfill' function which fills holes in the binary image. Then by using strel function the elephant portion in the image is filtered out and the rest of the areas are nullified and is made black. The filtered image is then subjected to BLOB analysis which is shown in Fig. 4. Using 'regionprops' function the centroid and the areas of the each blobs and the number of blobs in each image can be found out. A rectangular box called bounding box which is of the area of each blob is placed around each blob to indicate the presence of elephants in the image.

In the forest border areas elephants wonder as groups. The model developed were also tested with images of elephant groups which are taken from the quadcopter aerial camera (Fig. 5).

Apart from elephants, bison also has the same structure when viewed from the top. Because it has a dual colour tone (both black and brown) it can be filtered out by 'strel' function and it will not be detected.
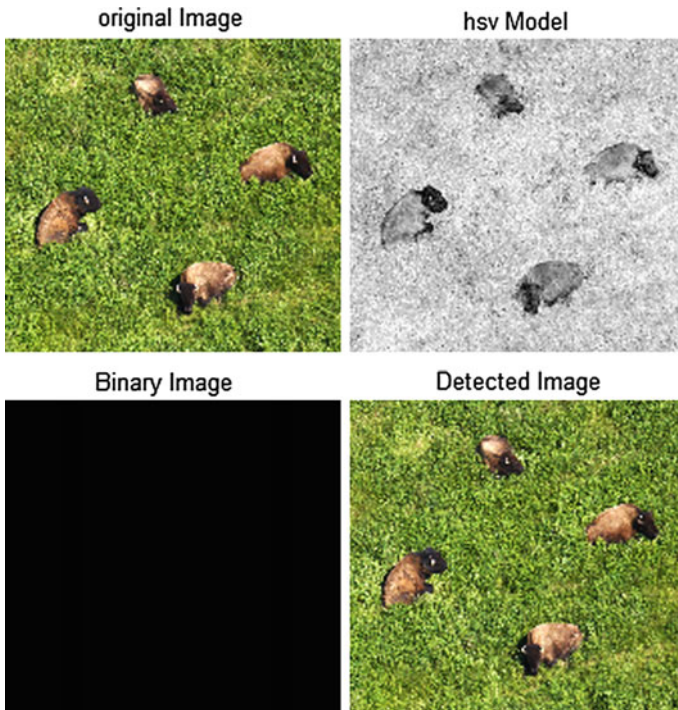
Fig. 5 Testing with other bison image

7 Conclusion

According to the statistics, the elephant ventures into human habitat in an unpredictable way. This quadcopter model we developed can be deployed for this purpose. It can be carried anywhere and made to fly in any desired place. The camera covers a wide range when flying above. Rugged terrains, inaccessible paths are never a hindrance to deploy a quadcopter. Image processing is an efficient way for this issue. The unique shape of the elephant makes the algorithm simple and effective. Apart from detecting, the total number of elephants at a particular place can also be obtained so that it helps in taking a better decision regarding protective measures. Also, the direction in which the elephant proceeds can also be known from the received information. There is hardly any miscommunication since the only animal that closely resembles elephant is bison and it is clearly distinguished by the algorithm.

References

1. Lenin J, Sukumar R (2011) Action plan for the mitigation of elephant–human conflict in India. Final report to the U.S. Fish and Wildlife Service. Asian Nature Conservation Foundation, Bangalore
2. http://www.conservationindia.org/gallery/human-elephant-interaction
3. Vermeulen C, Lejeune P, Lisein J, Sawadogo P, Bouche P (2013) Unmanned aerial survey of elephants. PLoS ONE 8(2):1–7
4. Sugumar SJ, Jayaparvathy R (2014) An improved real time image detection system for elephant intrusion along the forest border areas. Sci World J vol 2014
5. Luukkonen T (2011) Modelling and control of quadcopter. In: Independent research project in applied mathematics, Espoo, 22 Aug 2011
6. Bendale BC, Karwankar AR (2012) Moving object tracking in video using MATLAB. IJECSCSE 2(1): 5
7. Lebedev A (2007) Design and implementation of a 6DOF control system for an autonomous quadcopter. Julius Maximilian University of Würzburg, Würzburg
8. Nair AM, Talreja MA, Bhaskar P (2014) Design solution for flight control of a quadcopter. IJERT 3(5). ISSN: 2278-0181
9. Elruby AY, El-khatib MM, El-amary NH, Hashad AI (2012) Dynamic modeling and control of quadrotor vehicle. In: Proceedings of the 15th international AMME Conference. Vol 29
10. Hoffmann GM, Huang H, Waslander SL, Tomlin CJ (2007) Quadrotor helicopter flight dynamics and control. In: Proceedings of AIAA guidance, navigation and control conference and exhibit

Online Detection of Subclinical Mastitis Using Electrical Conductivity

Kalpak Shahane, V.D. Bachuwar and Pooja P. Gundewar

Abstract Bovine mastitis of a cow causes a huge loss of milk in dairy industries. These losses result in decreased yield of milk products such as cheese, curd. Early detection can lead to complete remedy. There are two types of mastitis: subclinical mastitis and clinical mastitis. Subclinical mastitis is challenging because we cannot see any visible changes by naked eye in the milk or the udder. General detection method for subclinical mastitis is counting somatic cells in milk. The aim of our work is to develop cost-effective solution for early detection of subclinical mastitis. This paper contains method for online detection of subclinical mastitis. Mastitis detection is done by checking electrical conductivity of milk analyzing conductivity with standard reading. Using Internet of Things (IoT), data will be stored and used for diagnosis of mastitis and related treatment will be given to cow.

Keywords Bovine mastitis · Somatic cells count · Udder · Online detection
Offline detection · Internet of Things · Inflammation

1 Introduction

Milk production of India was around 140 million tonnes in 2013–2014 on the basis of result shown in [1]. Milk production in 2030 would be more than 200 million tonnes. Milk is essential part of food and is an vital item in food table in India. It is also major source of earnings for millions of farmers. Any situation leading to drop in milk quantity and quality will therefore have a strong impact on source of

K. Shahane (✉) · P.P. Gundewar
Department of Electronics and Telecommunication, MITCOE, Pune, India
e-mail: kalpakshahane@gmail.com

P.P. Gundewar
e-mail: pooja.gundewar@mitcoe.edu.in

V.D. Bachuwar
School of Physical Science, Solapur University, Solapur, India
e-mail: Vdbachuwar@sus.ac.in

revenue and household nutrition [2]. One of the key causes for low yield and reduced quality of milk is mastitis, which is listed at top of the diseases causing considerable loss to dairy farmers.

The modern trend of revolution from traditional dairying to commercial dairying where hundreds to thousands of animals are raised together requires development of proper measures to control mastitis, especially subclinical mastitis. The contrary impact of subclinical mastitis on small production systems is also very high. Mastitis management not only has high significance to limit the losses related to low production, but it causes human health-related issues [3]. Due to difficulties in detecting subclinical mastitis, the milk arrives at the milk procurement system, the pathogens and toxinsin in milk may reflect in health problems. To boost farmers to lower level of mastitis, numerous dairies offer exceptional payment programs for milk restricted level of mastitis and decrease in payment for milk with higher mastitis level [4].

The most common method to detect mastitis is somatic cell count (SCC), as cow's internal immune system fights inflammation by increasing SCC. Mastitis modifies milk by increasing cell count in milk. Somatic cells consist of blood cells such as lymphocytes, macrophages, and epithelial cells from the inner udder wall. There is ample amount of evidence [5–8] of direct correlation between SCC and the degree of udder inflammation. Luís Carlos Vinhas Ítavo et al. proposed SCC level measured normal is less than 200.000 cells/ml of milk, SCC of more than 200.00 cells/mL of milk, suggesting animal could be infected with subclinical mastitis [3, 9].

Electrical conductivity (EC) has been used as indicator trait over the last decade for detection of mastitis. The growth of subclinical mastitis is going with a rise in the concentrations of sodium and potassium chloride salts and lactose [6, 10], which immediately lowers its electrical conductivity. Typically, EC of milk seems to be between 4.0 and 6.0 mS/cm at 25 °C [10, 11].

In this paper, a platinum probe sensor-based electrical conductivity measurement system with new concept of IoT system which detects subclinical mastitis is explained.

2   Electrical Conductivity Measurement

For analytical purpose, electrical conductivity is a very important property of milk. In general, electrical conductivity can be calculated by

$$\sigma = \frac{\text{Kcell}}{\text{Rx}} \tag{1}$$

where Rx (Ω) is the resistance of the solution and Kcell is geometric cell constant. A platinum probe electrode has been used because of its higher accuracy and precision. Electrical conductivity sensors are nonpolarized; therefore, measurements

Fig. 1 Platinum probe
conductivity sensor

have to be taken under a.c. regime. That is why we measure a frequency-dependent complex impedance $Z(f)$ which contains real resistance and complex capacitance because of several parasitics caused by the physical design of sensor.

To avoid such a polarization effect, measurements have to be performed at higher frequencies [12]. Figure 1 shows platinum probe conductivity sensor with cell constant $k = 1$.

To eliminate electrode polarization effect, electrical conductivity measurements must be performed on >2 kHz frequency on which conductivity readings are stable and accurate.

When conductivity measurements come into picture, we have to take care of temperature because the term conductivity is itself temperature dependent because when temperature increases, ions' mobility increases that will result in conductivity deviation and we have to compensate that deviation. It is calculated that rising temperature by $1°$ conductivity increases by 2–3% [13].

3 Block Diagram

The system block diagram is divided into few parts, frequency generation, current to voltage conversion, and AC to DC converter (Fig. 2).

An opamp-based relaxation oscillator is used for 2.4 kHz squarewave frequency generation. The output of oscillator is step downed to 1 V as if voltage is increased more than 1.5 V probe may get damaged. That AC excitation is fed to sensors. Sensor readings change with change in conductivity readings. The output of sensor is current so that I–V converter is used to convert it to a voltage as microcontroller can sense changes in voltage.

Because of AC excitation, the output of I–V converter is still AC. A rectifier converts it to pulsating DC value. Filter is used to filter out AC contents.
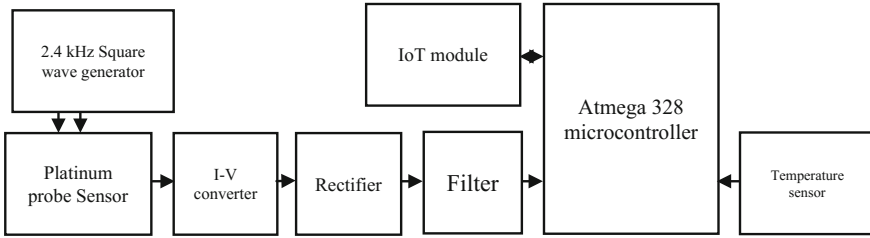
Fig. 2 Block diagram of electrical conductivity meter

Atmega328 microcontroller is used for sensing the voltage and converting it to millisiemens conductivity readings. 10-bit ADC of Atmega328 is sufficient to provide accurate output.

The Atmega328 microcontroller coded using Arduino IDE software which has a built in compiler. Figure 3 shows subclinical mastitis detection flowchart. Firstly, the ADC, LCD, and GPIO pins are initialized. Then, temperature pins are selected and the ADC0 is started, storing the result in a variable. Next, conductivity pin is selected and ADC1 is started. Voltage to conductivity conversion is done using calibration curve. Subclinically affected cow may not always show an increased conductivity from the infected quarter and the within-milking difference between smaller quarter reading and higher conductivity quarter reading can detect disease. Therefore, a combination of the EC values and ΔEC detects abnormal milk as shown in Fig. 3. Distinctive threshold values of conductivity are as follows: EC = 6.2 mS, ΔEC = 0.5 mS [13].

Lastly, all the data of all milk animals will be stored over the Internet using the concept of IoT with Wi-fi IC using AT commands.

4 Experimental Results

Calibration procedure is completed using diluted KCl solutions with EC close to EC of milk. Solutions' ECs are measured using commercial conductivity meter (equiptronics conductivity meter) and the output voltage of the system and obtained a calibration curve of Fig. 4.

The nonlinearity error referred is about 1%. The straight line equation is calculated using 2 point slope and executed in the Atmega328 microcontroller to obtain the EC.

Microcontroller programming is done and mastitis detection decided as listed in Table 1.

System was verified with different samples to check precision and reproducibility.
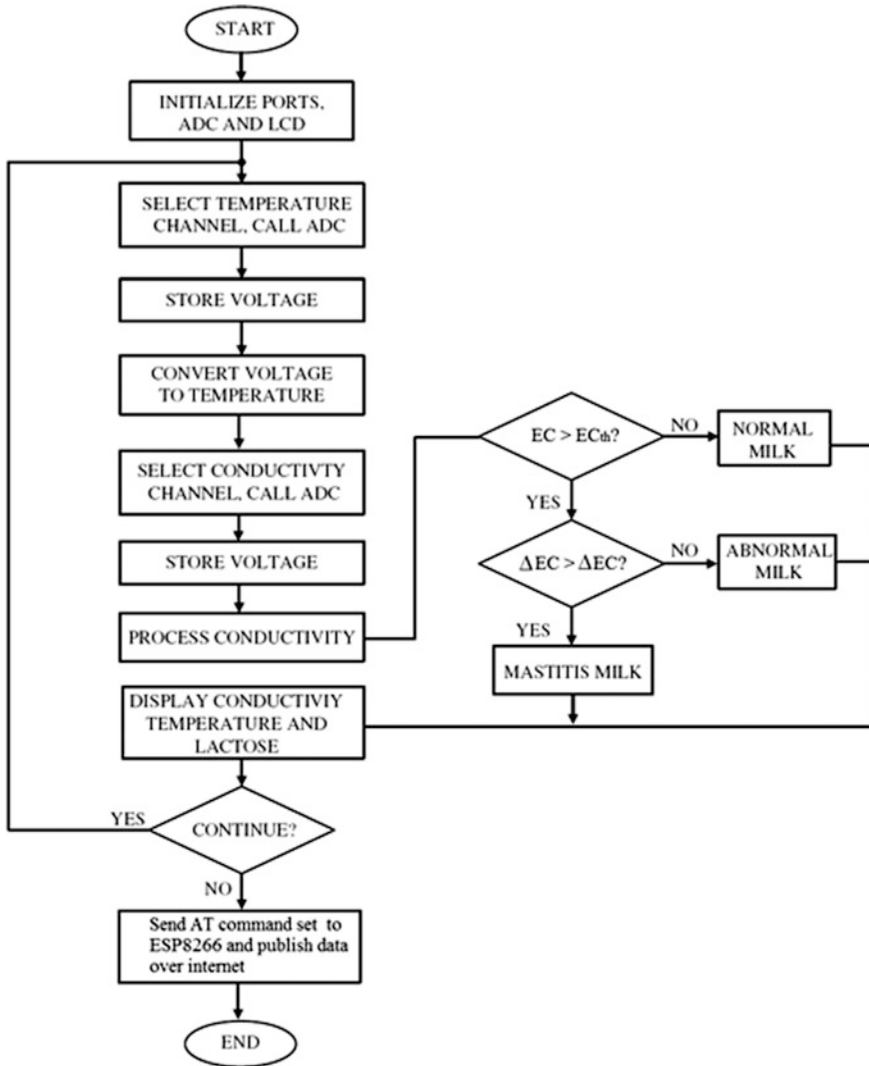
Fig. 3 Flowchart for subclinical mastitis detection

$$y = 0.0093x + 0.0187$$
$$R^2 = 1$$
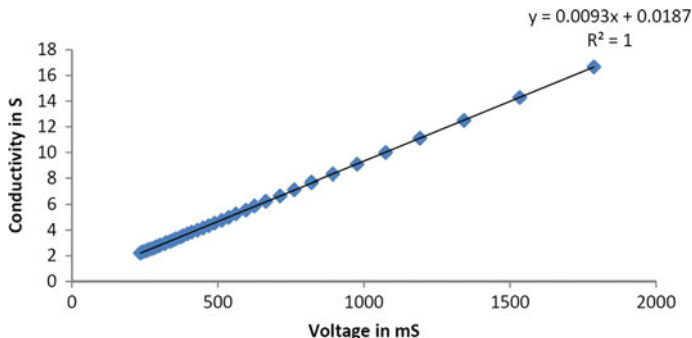
Fig. 4 Calibration curve for different values of EC

Table 1 Assessment range in mS of milk by absolute EC and differential EC

EC (mS/cm)	Differential EC (mS/cm)	Evaluation
>6.2	>0.5	Normal
>6.2	<0.5	Mastitis
<6.2	>0.5	Subclinical mastitis
<6.2	<0.5	Subclinical mastitis

5 Conclusion

Electrical conductivity method can be implemented by unskilled cattleman. The system is capable of quickly detecting mastitis level by performing EC check with temperature compensation. If we use difference between quarter (with highest and lowest conductivity) method, accuracy of subclinical mastitis detection considerably increases. Electrical conductivity test results in quick online cow-sided checking.

Number of methods are exist for detection of mastitis, in order to screen udder health performance. Some methods are exceptionally precise and accurate, but they are very costly and unavailable online field testing. This paper put forward a cost system for helping cattlemen to detect health of animal and the milk quality based on the measure of electrical conductivity at dairy farm itself. Based on EC values and ΔEC values abnormal milk is detected. The system has great sensitivity.

An EC result indirectly relates to SCC so many other factors such as environmental stress, late lactation, number of lactation, old age, and change in diet may results in decreased accuracy [3].

Effective online test for the detection of subclinical mastitis is provided, and using Internet of Things, data will be stored. Stored data will be used according to the severity of disease.

References

1. https://community.data.gov.in/milk-production-in-india/
2. Mastitis management in dairy animals, policy paper 61. National Academy of Agricultural Sciences, New Delhi, September 2013
3. Favry V (2004) Design and development of a novel electronic sensor for detecting mastitis based on conductance. Master of Science thesis, Dublin City University
4. Puri BR, Parkash S (1963) Electrical conductivity of milk. Indian J Dairy Sci 16:47
5. Viguier C, Arora S, Gilmartin N, Welbeck K, O'Kennedy R (2009) Mastitis detection: current trends and future perspectives. Trends Biotechnol 27(8):486–493
6. Watts L (1988) Etiological agents of bovine mastitis. Vet Microbiol 16(1):41
7. Nielen M, Schukken YH, Brand A, Haring S (1995) Com parison of analysis technique for on line detection of clinical mastitis. J Dairy Sci 78:1050
8. Dohoo IR, Meek AH (1982) Somatic cell counts in bovine milk. Can Vet J 23:119
9. Ítavo LCV, dos Santos GT (2001) Milk quality and subclinical mastitis detection through somatic cells counting. Acta Scientiarum Maringá 23(4):1065–1068
10. Wong NR (ed) (1988) Physical properties of milk. In: Fundamentals in dairy chemistry. Van Nostrand Reinhold, New York, p 409
11. Ferrero FJ, Valledor M, Campo JC (2014) Screening method for early detection of mastitis in cows. Measurement 47:855–860
12. Ferrara E, Callegaro L, Durbiano F (2000) Optimal frequency range for the measurement of AC conductivity in aqueous solutions. In: Proceedings of the 17th IEEE instrumentation and measurement technology conference, vol 2, pp 775–779
13. Maron SH, Prutton CF (1972) Principles of physical chemistry, 4th edn. IBH Publishing, Oxford, pp 423–425

Home Surveillance System
Using Internet of Things

Niraj Ukunde and Virendra V. Shete

Abstract In recent years, lots of research have been done in computer vision domain. Video surveillance in real-time scenario, especially for humans, like tracking and behavior analysis is one of the most active research topics in computer vision and artificial intelligence in present situation. The main focus is providing the low-cost and efficient video surveillance system for home application and can have wide scope in other areas such as elevator monitoring and server room monitoring. As the traditional surveillance system requires huge storage capacity and consumes lot of network bandwidth, hence it becomes necessary to provide solution for such design issues. This paper presents the implementation of motion detection algorithm for live camera streaming and thus allows analyzing incoming image stream, and recognizing any movement occurs in the area and if so then triggers video recorder to save the information-contained video clip.

Keywords Computer vision · Motion detection · Video surveillance

1 Introduction

In present condition, the home security is major issue which is to be dealt by every person. There are several security systems available such as CCTV, IP cameras, and different alarm systems. But most of them are not smart enough to ensure security. Hence, the aim of this project is to replace such traditional surveillance system which requires human operators by intelligent system. At present situation, the most widely used digital network transmission for video monitoring system, which provides the advantages of unlimited distance control, flexible extension, etc., is becoming a new standard of security systems. Internet of Things is one of the most trending research

N. Ukunde (✉) · V.V. Shete
MIT College of Engineering, Pune, India
e-mail: neeraju247@gmail.com

V.V. Shete
e-mail: virendra.shete@mitcoe.edu.in

domains by which the devices can communicate with each other rather than the location of their placement. The IoT significantly uses low-cost computing devices where there is less energy consumption and limited impact to the environment. The scope of Internet of Things is so wide that it can be embedded with computer vision. Hence, Internet of Things can be merged with computer vision so as to provide real-time surveillance system. The IOT has revolutionized capabilities to change everyday life of human and can help in many areas such as surveillance systems. Nowadays, consumer cameras become more inexpensive so many other intelligent function features can be included into the camera-based systems like face detection, and hence, such cameras are more widely used in surveillance systems. Video surveillance systems are useful for the safety of people in home as well as facilitate to control the entrance in the home. To monitor such activities, the key function is to analyze the human behavior. Hence, the subject for this paper is to evaluate efficient algorithm so that can be used in home surveillance systems.

2   Related Studies

Cloud-based video surveillance is becoming a new trend which has been proposed and implemented recently. The improvement in the cloud technologies has created lots of opportunities in surveillance domain [1]. Over the past years, many techniques have been suggested for surveillance systems. In traditional video surveillance systems, the cost and efficiency are determined by number of cameras that have been used, but this cost could be reduced if used with cheap wireless sensors along with IP cameras [2]. This design consists of many wireless sensors and few cameras. The sensors keep track of moving objects and report to their sink nodes, and these sink uses deployed IP cameras to capture and record the present situation and thus can be monitored remotely. PIR sensors are the cheapest sensor available for detecting the presence of human, and hence, it can be implemented to develop intelligent video monitoring system [3]. In this process, when any presence is detected by the PIR sensor, the processor starts recording the video through a camera till the sensor sends the signal of presence and then the video clip is uploaded to server. The PIR sensor has its disadvantages related to its range; hence, it can be replaced by using any motion detection algorithm. Motion detection is an important part for analyzing the scenes captured by the camera. One of the known motion detection algorithms can be implemented known as background subtraction that can provide better result [4].

3   Motion Detection Methodology

Background subtraction is a very fundamental motion detection algorithm used for detecting the moving objects as well as object tracking which is a required application for the surveillance system. In this application, any conventional digital

camera is used to sample test images which are then used to perform background subtraction [5]. The foreground estimation is done by comparing test images with background image where the constant image pixels are discarded and the pixels with changed properties are meant to be the foreground image. The motion detection algorithm includes the first phase of segmenting a test image in which moving objects denoted as foreground image are distinguished from background image [6]. One of the simplest method to achieve this implementation is when camera starts capturing images then take first image as background and afterward the frames captured at time t onward can be represented as $G(t)$ will be compared with background image which is represented as B. Now, by using arithmetic calculation of image subtraction procedure of computer vision implemented for each pixel of $G(t)$, the moving objects can be differentiated. Now, consider $X[G(t)]$ be the pixel value which is to be subtracted from next corresponding pixel at the same position at next time interval $t1$ on background image represented by $X[B]$. Hence, the equation we get can be written as follows:

$$X[F(t)] = X[I(t)] - X[B] \tag{1}$$

Let us assume the background image is at time interval t and threshold is used to improve subtraction on difference image. The resultant difference in image is represented by the changes in the pixel property between two successive frames. While using a background image, we are removing the background, and hence, using this approach, it will work where moving objects are the foreground pixels and static pixels represent the background image [7].

$$kX[F(t)] = X[f(t+1)]k > \text{Threshold} \tag{2}$$

4   System Architecture

The system mainly focuses on the intelligent way of data collection where it will have wired or wireless Internet connectivity to do so. In other way, we can say that the advanced version of surveillance system will fulfill the needs of clients in terms of monitoring services. The surveillance system introduced in this paper uploads the video on cloud server when video clip is captured by the system as well as a short message will alert the owner for respective intrusion happening in the monitoring area. Hence, in short, the proposed security system will provide the features such as the alert message through GSM module local video storage and cloud storage (Fig. 1).

To achieve the objective of the system, the standard hardware is being chosen. The raspberry Pi has a promising performance in terms of the image processing, and it is provided with the RTOS named as Rasbian OS. The Rasbian OS is a derivative of most stable Linux operating system; hence, it provides the stable platform to
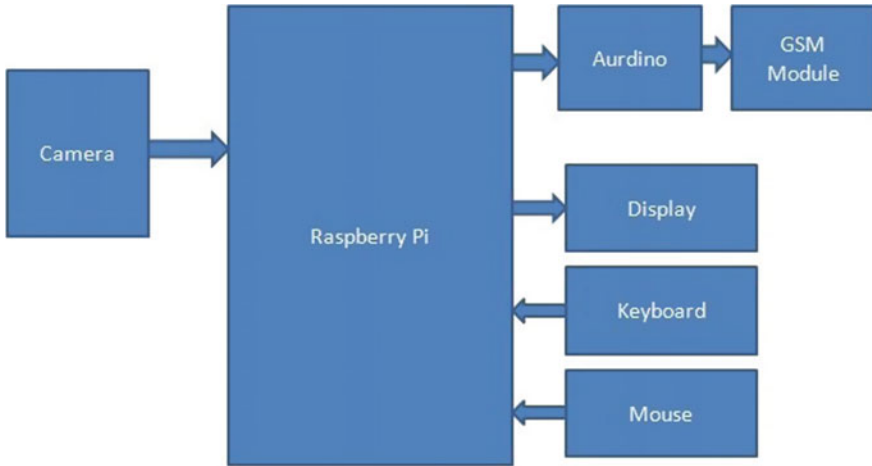
Fig. 1 Block diagram

implement our algorithm in C++ and Python. Open CV libraries are available for image processing purpose.

5 Software Flow

In this section, the overall flow for surveillance system will be discussed as follows. The digital camera is a primary input device which will provide the video frames to the processor unit. As soon as the system turned on the algorithm running on the system will start to analyze incoming video stream for any motion in the monitoring area. The execution of algorithm started when continuous image stream is captured, and thus, the stream is segmented into the frames. The background modeling uses new video frame after every predefined iteration to calculate and update background model. These background models provide statistical information about the entire background image. Background modeling is achieved by recursive or non-recursive techniques. Thresholding step is required to decide the final value of pixel. The foreground image contains the information about pixels whose values are significantly changing. Such groups of pixels showing similar pixel value properties are considered as moving object. Finally, data validation examines the group of moving object and provides final foreground image. If any motion is detected, then messaging module will get activated and the short message will be sent to the owner. At the same moment, the video capture module will start recording the video stream, and after specific time, the video will get uploaded to the server automatically. The same process will be repeated till the motion is detected by the system (Fig. 2).
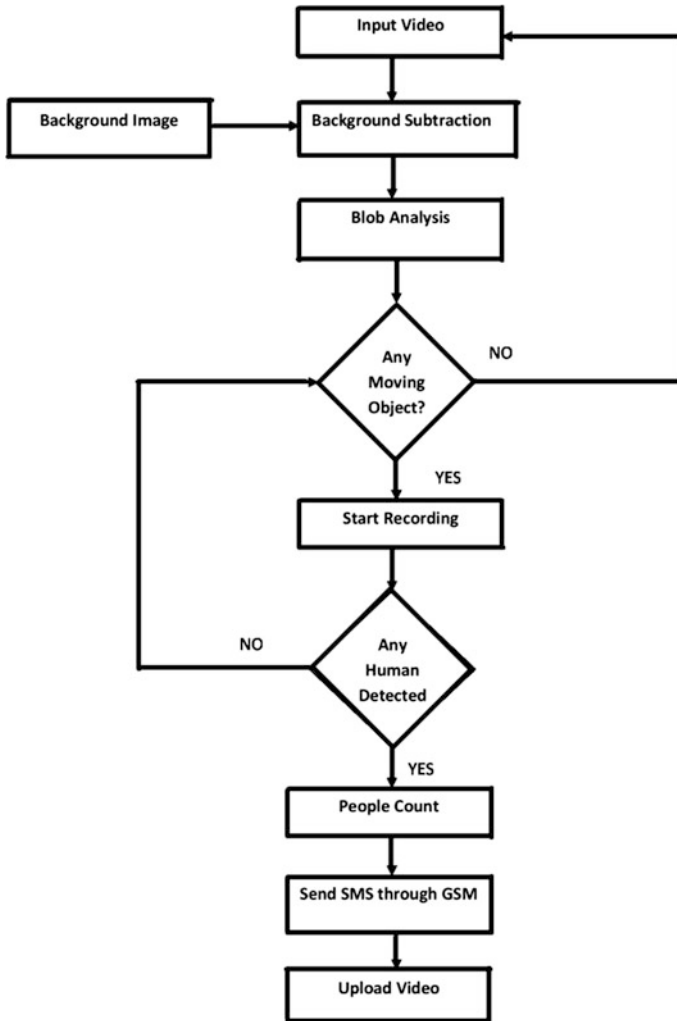
Fig. 2 Software flow

6 Experiments and Results

Following figures shows the screenshots of the result. When any movement happens, the system will analyze the incoming video stream and store the only video clip which contains the necessary information about movements. The video clips are encoded in .avi format and then can be uploaded to cloud or sent via email to the owner. These videos can be watched from any multimedia devices which have decent Internet connection such as PC or mobile phones. The latency remains fairly low for resolution of 640×360. As we are storing an only clip which contains the
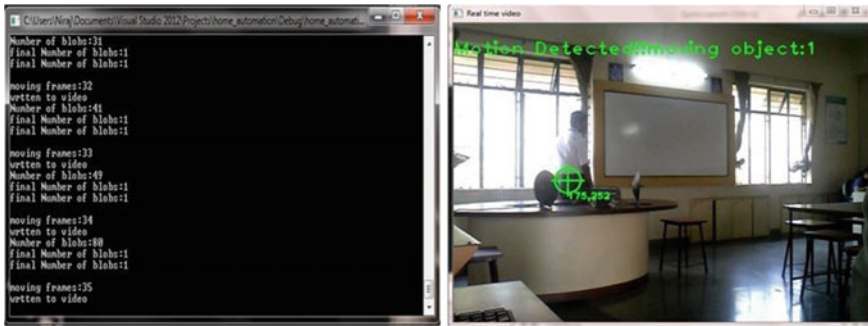
Fig. 3 Final output of system

movement, it significantly reduces the storage requirements. The output produced by the system is shown in Fig. 3.

7   Conclusion and Future Work

Video surveillance system is one of the important aspects of daily life for every human. This project is an approach for the design of intelligent video surveillance system based on motion detection algorithm implemented on Raspberry Pi. This still requires some improvements to make final design version of surveillance system. In future plan, we plan to improve the motion detection algorithm to adapt real-time scenario toward more accuracy, because motion detection algorithm depends on threshold value and it needs to adjust to improve its accuracy.

References

1. Biao S, Yuan T, Bingyin Z (2014) Design and evaluation of remote video surveillance system on private cloud. In: Biometrics and security technologies (ISBAST). IEEE
2. Chen W-T, Chen P-Y, Lee W-S, Huang C-F (2008) Design and implementation of a real time video surveillance system with wireless sensor networks. In: Vehicular technology conference, VTC spring 2008. IEEE
3. Birmohan S, Dalwinder S, Gurwinder S, Sharma N, Sibbal V (2014) Motion detection for video surveillance. In: International conference on signal propagation and computer technology. IEEE
4. Nguyen HQ, Loan TTK, Mao BD, Huh EN (2015) Low cost real time system monitoring using raspberry Pi. In: ICUFN. IEEE
5. Ansari AN, Sedkyl M, Sharma N, Tyagi A (2015) An internet of things approach for motion detection using raspberry Pi. In: International conference on intelligent computing and internet of things (ICIT). IEEE

6. Priyadarshani S, Dhanalakshmi S (2014) Foreground object motion detection by background subtraction and signalling using GSM. In: ICICES. IEEE
7. Kechao W, Xiangmin R, Zhifei W, Zongfu J, Jingwei Y (2011) Design and implementation of embedded network video monitoring terminal. In: IEEE conference on computer science and automation engineering (CSAE). IEEE

Design and Analysis of VCO
for Parameter Optimization

Virendra Shete and Anshuman

Abstract The following paper provides an introduction to analysis and design of a complementary cross-coupled LC-VCO as used in RF transmitter/receiver circuits. Different design parameters have been discussed along with their associated equations, to allow for basic on-paper calculations. The values of corresponding components are used in verification of the design via CAD simulation. The main goal is to synthesize a low-power low-noise design with a wide bandwidth of operation (tuning range).

Keywords Complementary cross-coupled LC-VCO · Phase noise · Tuning range

1   Introduction

Even though an LC-VCO is not area efficient, it is always favored in RFIC design due to its superior phase noise and jitter performance at high frequency. The performance of a VCO is determined by the quality factor (Q) of its L–C tank. Usually, on-chip spiral inductors are used for this purpose, having poor Q of around 3–5 at 2.5 GHz [2]. This affects the phase noise performance of the LC-VCO. In order to achieve low-phase noise, the inductor can be made off-chip.

Another performance parameter is the tuning range. As the tuning range of the VCO increases, it becomes difficult to meet the specifications such as low-phase noise and low-power consumption simultaneously. Increase in tuning range requires larger tuning constant (K_{VCO}) and DC power as compared to narrowband

V. Shete (✉) · Anshuman
Department of Electronics and Telecommunication, MIT College of Engineering,
Pune, India
e-mail: Virendra.shete@mitcoe.edu.in

Anshuman
e-mail: anshumaniitan@yahoo.com

VCO. However, larger K_{VCO} leads to poor phase noise response and consumes a large amount of DC power to start up the oscillation [3]. Hence, there is a trade-off between specifications of power consumption, tuning range, and phase noise in wideband LC-VCO design.

2 LC Oscillator Operation

Below figure shows a schematic of cross-coupled LC-VCO. The LC tank is responsible for determining the frequency of oscillation of the VCO and forming the drain loads. MOSFETs M1 and M2 are responsible for providing negative impedance of the value −1/gm (Fig. 1).

The associated resistance, R, of the inductor will be [3]:

$$R_L = \frac{2\pi \cdot f \cdot L}{Q_U} \tag{1}$$

In order for the VCO to oscillate,

$$\frac{1}{gm} \geq \frac{2\pi \cdot f \cdot L}{Q_U} \tag{2}$$

where

$$gm = \frac{I_D}{(Vgs - Vt)}$$

Fig. 1 Complementary cross-coupled LC-VCO [1]
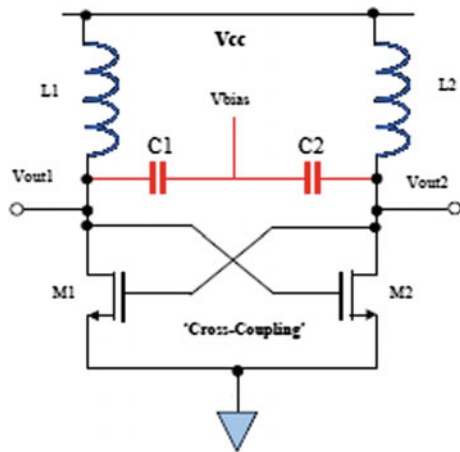
Therefore,

$$\frac{(Vgs - Vt)}{I_D} \geq \frac{2\pi \cdot f \cdot L}{Q} \tag{3}$$

Solving for I_D, we get:

$$I_D = \frac{Q(Vgs - Vt)}{2\pi \cdot f \cdot L} \tag{4}$$

In order to ensure oscillation start-up, safety factor of 2 is considered as follows:

$$\frac{2}{g_m} > Rp = \frac{2\pi \cdot f \cdot L}{Q} \tag{5}$$

3 Phase Noise Calculation

According to Lesson's equation [4],

$$L(\Delta f, K_{VCO}) = 10 \log\left\{ \left(\frac{f_0}{2Q\Delta f}\right)^2 \left[\frac{FKT}{2P_o}\left(1 + \frac{f_c}{\Delta f}\right)\right] + \left(\frac{K_{VCO}V_n}{2k_{LC}\Delta f}\right) \right\} \tag{6}$$

where Δf is offset frequency from carrier frequency f_0; P_O is the output power; V_n is the noise voltage (or control voltage); F describes the thermal noise and flicker noise of the transistor; K_{LC} is a constant associated with L and C values of tank.

It is clear from the above expression that higher the quality factor (Q) of the tank, higher will be the output power (P_O). The F parameter is dependent on the size of the MOSFETs. Larger the device size, lower will be its value. The equivalent resistance (R_{ON}) of the MOS devices also affects the thermal noise, so a large (W/L) ratio of the band switch should be chosen in order to lower the noise.

Increasing the bias current at fixed V_{DD} can improve the phase noise of the VCO. However, since the output swing is limited by VDD, increasing the bias current can only improve the phase noise up to a certain level [1]. So, the minimum phase noise will occur at the current- and voltage-limited regime. Thus, a trade-off should be made between phase noise and power consumption in designing a wideband VCO.

4   Frequency of Oscillations

The LC tank of a VCO is responsible for generating oscillations, provided the active circuit (MOS transistors) cancels out the tank losses in order to maintain oscillations by generating negative conductance equivalent to the tank losses. Additionally, voltage gain should have zero phase shift so that the output follows the input precisely.

Selecting a suitable circuit topology is the first step in designing a VCO. For this paper, a complementary cross-coupled VCO is used. The oscillation frequency of an LC tank as shown in Fig. 2 can be given by

$$w_O = \frac{1}{\sqrt{LC}}\sqrt{1 - \frac{R^2 C}{L}} \tag{7}$$

Above equation can only be applied to a VCO if MOSFETs and varactor have no losses of their own, which is implausible. MOSFETs have many parasitics that affect the performance of the VCO (Fig. 3).

Cgdo and Cgs in the above figure are overlap parasitic capacitances. These can combine with the tank capacitance and degrade the oscillation frequency. Since these are fixed capacitances and cannot be reduced by any means, it is important to reduce tank's capacitance in order to maintain high frequency of oscillations. r_{ds} is not a real resistance but is associated with the channel length of MOSFET. It also affects the quality factor of the LC tank. So, after considering the above parasitic components, (7) can be rewritten as follows [3]:

$$w_O = \frac{1}{\sqrt{L(C + C_{gs} + 4C_{gdo})}}\sqrt{1 - \frac{R^2(C + C_{gs} + 4C_{gdo})}{L}} \tag{8}$$

This will also change the start-up gm condition required for oscillation:

$$g_m \geq \frac{1}{r_{ds}} + \frac{R(C + C_{gs} + 4C_{gdo})}{L} \tag{9}$$

This equation proves that the required value of gm required to start the oscillation is directly proportional to the size of the MOSFET.

Fig. 2   LC tank
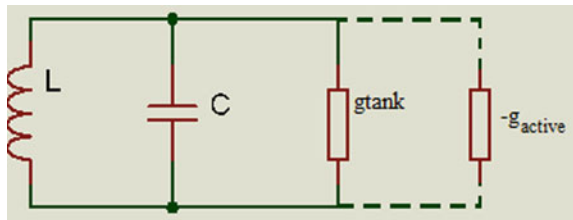
Fig. 3 MOS parasitics [3]
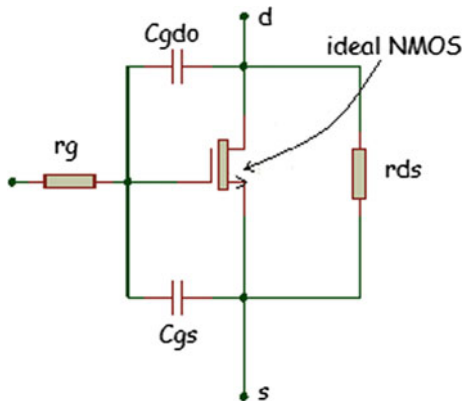
5 VCO Design Steps

A few design specifications to be considered before beginning the design are as follows [3]:

(a) DC power dissipation (max) = $V_{supply} \cdot I_{bias}$
(b) Minimum output voltage = V_{tank}
(c) Tuning Range $= \frac{w_{max} - w_{min}}{w_o} \times 100\%$
(d) Transconductance (g_m)
(e) Frequency of operation (w_o)
(f) Design area
(g) Phase noise

The first step should be setting the bias current according to expected maximum power consumption.

$$I_{bias} = \frac{P_{dc(max)}}{V_{supply}} \tag{10}$$

Next, quality factor (q_L) of the inductor should be determined at required frequency w_o,

$$Q_L = \frac{w_o L}{R} \tag{11}$$

Determine the minimum voltage of the LC tank required for the design.

$$V_{tank} = I_{bias} R_p = I_{bias} w_O L Q_L \tag{12}$$

Calculate the capacitance of the LC tank with w_O as the center frequency.

$$w_O = \frac{1}{\sqrt{LC}}\sqrt{1 - \frac{R^2 C}{L}} \tag{13}$$

Solving for C, we get:

$$C = \frac{L}{w_O^2 L^2 + R^2} \tag{14}$$

Calculate minimum required g_m for each MOSFET, given a nominal closed loop gain $\alpha_{\min} > 1$

$$g_m = \alpha_{\min}\left[\frac{1}{r_{ds}} + \frac{R(C + C_{gs} + 4C_{gdo})}{L}\right] \tag{15}$$

6 Experimental Setup

The proposed design consists of NMOS, PMOS cross-coupled pair, and LC tank as fundamental blocks. Supply voltage and bias current were selected according to power constraint. Values of L and C were selected according to expected bandwidth of operation. C6 was used to tune within the required tuning range. Output capacitors C1 and C2 were used as buffers to isolate the LC tank from load variations. Component specification is as mentioned below (Fig. 4; Table 1)
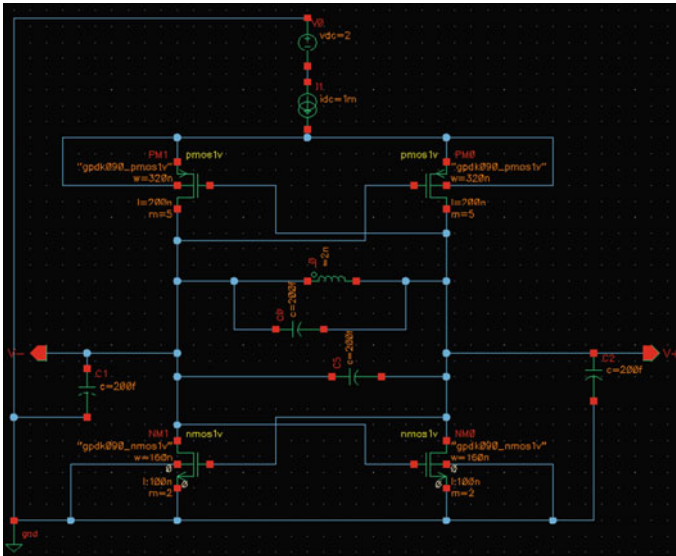
Fig. 4 LC-VCO schematic using cadence tool

Table 1 Component specification

Supply voltage	1.8 V
Bias current	1 mA
PMOS	$w = 320$ nm, $l = 200$ nm
NMOS	$w = 160$ nm, $l = 100$ nm
LC tank	$L = 100$ pH, $C = 50$ fF
Load capacitance	180 fF

Fig. 5 Output frequency plot

7 Result

This paper presented a VCO with wide tuning range. The structure achieved a tuning range of 30.5–40.4 GHz. The TR achieved by this method was 31%. Phase noise as recorded after simulation was −105 dBc/Hz. The output voltage peak-to-peak swing was recorded 0–4.4 V. This VCO technique provides a reasonable option for millimeter wave transceiver circuits and also promises further modifications as the technology scales down (Fig. 5).

References

1. Rhea RW (1995) Oscillator design and computer simulation. Noble Publishing. ISBN 1-8849-32-30-4
2. Silver JP (2009) LC oscillator tutorial. RF, RFIC and microwave design. www.rfic.co.uk
3. Hamel JS (2005) LC tank voltage controlled oscillator tutorial. UW ASIC Analog Group
4. Chiou H-K et al (2008) Design formula for band-switching capacitor array in wide tuning range low-phase-noise LC-VCO. Microelectron J 39:1687–1692

Controlling Power System of Earth Station Using Embedded System

Sunil V. Bagade and V.V. Shete

Abstract We have used two methods for earth station power controller is designed by Cortex M3 by using LPC 1768 board and FPGA board. Different loads having different power requirements and different fault conditions occurred. Earth station having 800 W inverter and having two failures named hard failure and soft failure. We have used solar energy as a main source of earth station and diesel generator for failure conditions. Using neural network, we developed algorithm in embedded C as per different types of failure, battery voltage, and solar voltage.

Keywords Earth station · Cortex M3 · FPGA · Neural network · LPC 1768 Solar energy · Power controller

1 Introduction

Earth station has been used for many purposes such as military, communication, weather forecast report, mobile communication, and satellite communication, hence earth station needs to be powered 24×7 h working, and failure in power supply is not acceptable in any condition. Earth station normally based in rural areas or away from city to avoid disturbance of other signals. They have placed on mountain area or rural area. Earth station needs an 800 W inverter including all its loads. It is difficult to supply 800 W inverter from the electricity which used for general purpose applications, because they are not dedicated for specific load and might have chances of failure. So that for dedicated load like earth station has their own power supply, which is totally independent from other application and dedicated for earth station. Generally, natural energy has been converted to electrical energy to

S.V. Bagade (✉)
VLSI and Embedded System, MIT College of Engineering, Pune, India
e-mail: sunilbagade88@gmail.com

V.V. Shete
ECE Department, MIT College of Engineering, Pune, India
e-mail: virendra.shete@mitcoe.edu.in

supply earth station as per surrounding climate and as per resources awailable wind energy; water energy or solar energy has been used. In our project, we have used solar energy as primary source to supply earth station and diesel generator as secondary source. Diesel generator has used to avoid failure conditions.

2 Earth Station Power System Block Diagram

As we have discussed in introduction part, for we have used solar panel for earth station as primary source, to utilized proper solar energy we have used battery bank to store solar energy and this energy can be utilized under different circumstances as follows.

Assume at day time, where maximum sun light is available, we have used 55 W solar panel which is more than sufficient to 12 V-3 A load and to charge battery simultaneously. Battery continuously charged until day light is sufficient to produce more than 36 W from solar panel, phenomenon of over charging battery will discuss latter.

Under different circumstances when sun light is not enough to charge the battery and drive load simultaneously, it will try to charge battery slowly but discharging rate of battery is more than charging rate due to load consumption is more than battery charging rate. At remote sites, PV systems are most effective and storage batteries are used as a backup plan. During pick time, where solar system delivered maximum energy so PV system can give a maximum result, path is divided in two ways one is used for direct power to the earth system and other for charging the batteries. In this architecture, we have used controller and inverter so that controller controls the power from PV system and battery banks. In both case, we are using DC supply.

According to load consumption, battery ratings and solar panel wattage have been decided in our project under small prototype of earth station, and it consumes maximum 30 W including controller wattage and other load still we have used 150 W inverter powered by battery for future expansion. Hence battery needs a approximate 12 V-15 A for full load condition, and for 30 W load 12 V-8 A battery is sufficient.

In this prototype, we have used satellite dish antenna and TV tuner as earth station load and c.f.l. lamp as a class room load. In actual earth station. many loads are present rather than antenna such as telephone lines, cooling fans, and fax (Fig. 1).

In this architecture, we have used controller and inverter so that controller controls the power from PV system and battery banks. In both case, we are using DC supply.

According to load consumption, battery ratings and solar panel wattage have been decided in our project under small prototype of earth station, and it consumes maximum 30 W including controller wattage and other load still we have used 150 W inverter powered by battery for future expansion. Hence, battery needs a
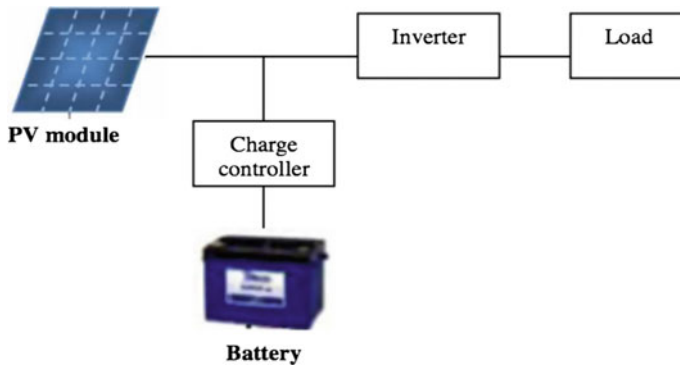
Fig. 1 Block diagram of earth station power system

approximate 12 V 15 A for full load condition, and for 30 W load 12 V 8 A battery is sufficient.

In this prototype, we have used satellite dish antenna and TV tuner as earth station load and c.f.l. lamp as a class room load. In actual earth station, many loads are present rather than antenna such as telephone lines, cooling fans, and fax machine. We have shown only class room and earth station load in this prototype.

3 Earth Station Power System

3.1 Earth Station Power System

The 430 W power required for earth station communication system, 30 W power required for telephone terminal, 100 W for FAX and graphics, and 35 W for audio converter. Different inverters for different types of load cause power loss in a system so considering all these loads and future expansion, 800 W inverter is intended to total earth station system. 48 V dc is directly taken out from battery banks for cooling fans. Here we have taken satellite dish antenna and set top box with required 12 V and 2 A approximate 24 W maximum power, TV tuner needs 5 V dc and 600 mA supplies required where as c.f.l. lamp need 11 W power. To fulfill this load requirement we have used two solar panels having capacity of 5 W and battery delivers a power 12 V and 7.5 A h, our load are TV tuner required 5 V dc and 600 mA approximately 3 W and Satellite dish antenna and Set top Box needs 12 V and 2 A rating approximate 24 W and LPC 1768 needs 1 W total output load required 29 W, hence we have used battery 12 V dc 7 A h which delivers a 29 W power for approximately 3 h and hence two 5 W and one 45 W solar panels total 55 W solar panels are used to keep battery charge.

3.2 System Controller

Controller phenomenon: As per various battery statuses, charging voltage and solar panel voltage system perform various functions. We have considered three status at 90, 40, and 10% of battery charging voltage, when battery charging voltage shows 90%, to avoid overcharging of battery controller switch off two panels out of three, hence charging rate of battery slows down. On other hand, if all panels were failed to charge battery and battery charging voltage gets below 40%, then Diesel generator ON and it charges battery.

3.2.1 Design of Failure

Controller is based on two types of failures "soft" and "hard".

(a) *Soft failure*: A "soft" failure occurs when all solar panels are not sufficient to charge battery or it produces lesser current than required load current, hence at particular charging voltage of battery (40% battery charging voltage) diesel generator gets on to charge battery.
(b) *Hard failure*: A "hard" failure does not involve automated controlling unit. It manually switches load on diesel generator. This condition occurs when controller is failed or battery get damaged due to overcharging or exceeds duty cycle limits. At the time of hard failure, defects in controller and/or in battery gets identified and repaired.

3.3 Controller Flow Chart and Programming

3.3.1 Controller Flow Chart

As we are working on prototype, some changes have been done according to solar panels and battery charging levels. From Fig. 2, when PV arrays or solar panels deliver a full power, it can drive a load without any battery support, if power of solar panel does not match with load requirement, then controller comes in picture and checks its various battery charge levels and according to that battery gets charged by solar panels or diesel generator, and according to sunlight radiation solar panels are selected for charging to prevent excess charging of battery. If charging of battery is <40%, then it will ask for all three panels for charging or else ask for diesel generator if all panels are not able to charge battery. Similarly, charging status of battery checks at 90 and 100% according to that panels are selected to charge battery and after 100% charging of battery stops to prevent overcharging of battery. Here, soft failure is handled by the controller but in case of critical damage
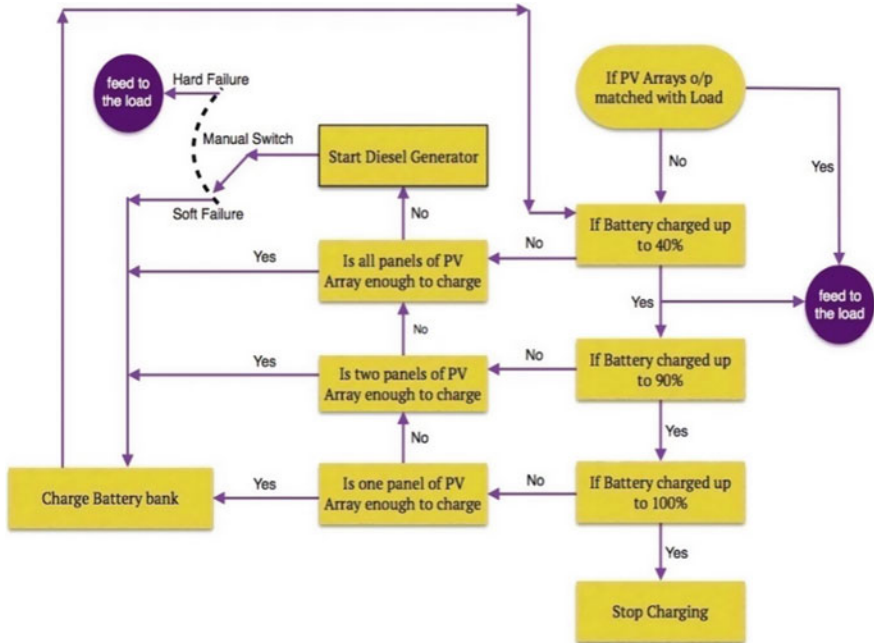
Fig. 2 Flow chart of system controller

Table 1 Logical states of switching in solar panel and diesel generator.

Battery voltage	Below 40%	Between 40 and 90%	Above 90%
Panel switching	1	1	0
D.G. status	1	0	0

in controller or a battery then diesel generator can be on manually called hard failure.

3.3.2 Programming Logic

Our results are totally dependent on battery voltage and solar panel current. We have used 12 V 15 A h battery, and to decide voltage levels of battery we divided battery levels in percentage; 100% charge battery shows more than 12.6 V, while 90% and 40% charge battery shows 12.5 and 11.9 V accordingly. While battery gets discharged fully, then expected battery voltage is 10.5 V. There are two different results to show warnings at the output and another to switching relays at output to switch panels and diesel generator (Table 1).

Controller inputs are battery voltage and panel current, as per different voltage of battery panel 3 and diesel generator get switch OFF/ON. There is another condition, as per solar current diesel generator gets OFF. We have also decided SOC of battery.

- Condition for panel switching, panel 3 is OFF if battery level more than 90%, and below 90% all panels are ON.
- Condition for diesel generator switching, diesel generator is ON when battery level below 40% and it will OFF only when solar panel is providing enough current to charge the battery or else battery is charged up to 90%.
  Condition for state of charge (SOC) battery, normally when battery discharged fully it shows 10.5 V, if battery shows below 8.5 V, then there is a problem in battery (Table 2).

3.3.3 Neural Network for Coding and Finding Values for Different Failure Conditions

Input X1 is different SOC conditions of battery; according to different SOC levels, output values are decided. We have two output values $O1$ and $O2$, $O1$ shows solar panel switching and $O2$ shows diesel generator switching. $w1$, $w2$, $w3$, $w4$, $w5$, and $w6$ are weights between the neutrons. H.L. is the hidden layer between the input and output layer (Fig. 3).

Table 2 Logical states of switching in diesel generator for diffrent voltage levels.

Solar panel 3 current	Below 13 V	Above 13 V
D.G. status	1	0
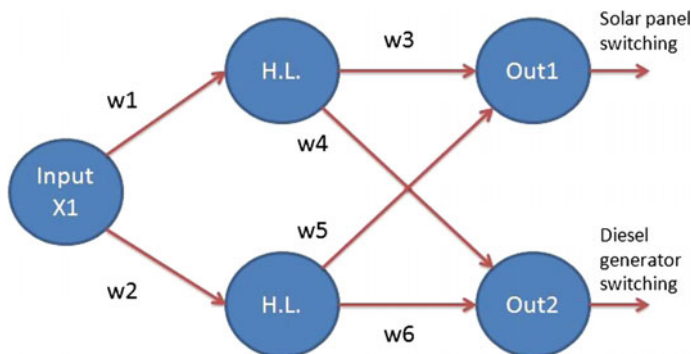
Fig. 3 Neural network for proposed system

From this neural network, we have calculated variable input values of $O1$ and $O2$. According to that, controller takes design for switching panels and diesel generator.

If we solve this, we have two equations as follows:

$$O1 = X1w1w3 + X1w2w5 \tag{1}$$

$$O2 = X1w1w4 + X1w2w6 \tag{2}$$

We assumed initial weights $w1$ and $w2$ 0.5 value then equation becomes. From Table 1, we take output $O1 = 1$, $O2 = 0$ for solar panel switching condition and $O1 = 1$, $O2 = 1$ for diesel generator switching condition. If battery SOC is more than 90% then solar panel is OFF else ON irrespective with any conditions, hence we taken value $X1 = 90$ and assumed for logic 1 minimum value is 0.75 and logic 0 maximum value is 0.25 substitute these values in Eqs. (1) and (2) and solved. We have got values $w3 + w5$ and $w4 + w6$. Same procedure is done for finding diesel generator switching; also, considering $X1 = 40$ and solving equations, we have found values of $w3 + w5$ and $w4 + w6$. Taking average of both values solved equation again we have resultant values for $O1$ and $O2$ for solar panel is more than 1.0 and for diesel generator more than 0.45. For different level of SOC if $O1$ and $O2$ equation gives different values, according to those values controller has programmed and takes design.

4   Conclusion and Results

In this paper, earth station power controller has designed experimental results that show that switching between solar panels and diesel generator handles soft and hard failure nicely, and we have demonstrated power consumption by load and power outcome from solar panels are matched by applying load as dish TV and TV tuner to receive channels having frequency in MHz. LPC 1768 controller is utilized properly for controller application; it consumes low power and design cost is also low. Signals taken from analog side are taken out in MATLAB and shown in following graphs; same outcome and results can be taken from FPGA controller by using HDL language but it is difficult to design logic circuit every time if state variables and states are increased for complex operation.

We have successfully designed battery charger circuit, while charging sources are giving low voltage across battery to charge it. We have taken out 12.56 V reference point to generate X1 and X2 input values (Fig. 4).
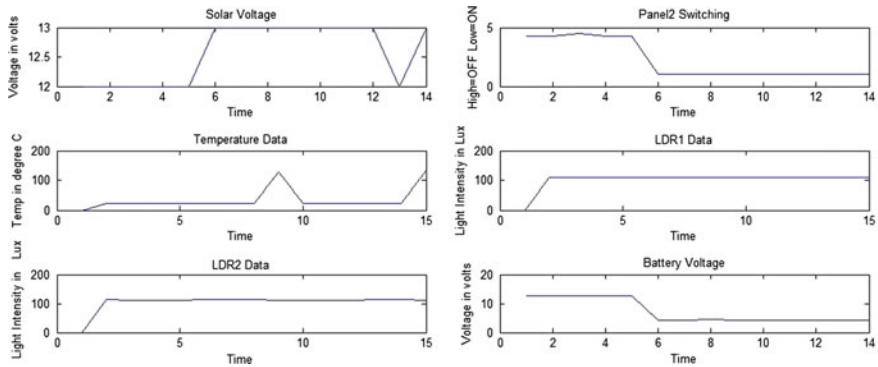
Fig. 4 Output wave forms for battery SOC and according to change in solar voltage and panel switching

References

1. Himaya T, Kouzuma S, Lanakubo T, Orteyer TH (2015) Evaluation of neural network based real time maximum power tracking controller for PV system. IEEE Power Energy Soc
2. El-Madany HT, Fahmy FH, El-Rahman NMA, Dorrah HT (2014) Design of FPGA based neural network controller for earth station power system. TELKOMNIKA 10(2):281–290
3. Petchjatuporn P, Ngamkham W, Khaehintung N (2005) A solar-powered battery charger with neural network maximum power point tracking implemented on a low-cost PIC-microcontroller. In: TENCON 2005, IEEE region 10 Nov 2005

Simulation and Control Model of Integration PV into Utility Grid Using MATLAB

Manaullah, Haroon Ashfaq and Ibrahim Imbayah Khalefah Imbayah

Abstract In this chapter, a simulation and control model of integration PV into utility grid using MATLAB, and the effectiveness of the I controller in the system under different levels of solar irradiation are presented. An MPPT module with I control developed in this work is felicitous for power applications program; however, the problem of integration of PV is that it relies heavily on weather conditions. So, there is a necessity for developing control techniques for grid integration PV system including a method for voltage and current control that stabilises the voltage and current. An MPPT algorithm using DC/DC converter (Boost converter) is applied to make PV arrays to work at maximum power point. Then, the system behaviour and performance are studied. The system stability is also considered when there is a change in solar irradiation or a fault in the system. This chapter addresses that the proposed I controller has a good performance.

Keywords Photovoltaic (PV) · Boost converter DC/DC · I controller
Maximum power point tracking (MPPT) MATLAB/SIMULINK

1 Introduction

The world is looking for alternative energy methods from renewable energy for clean and cheap energy that will get rid of the dependence on costly oil and also gas. The generating electric power by using solar cells is effective technique nowadays. The various local voltage control approaches utilising PV storage systems. These design

Manaullah (✉) · H. Ashfaq · I.I.K. Imbayah
Department of Electrical Engineering, Faculty of Engineering & Technology,
Jamia Millia Islamia, New Delhi 110025, India
e-mail: manaullah@jmi.ac.in

H. Ashfaq
e-mail: hashfaq@jmi.ac.in

I.I.K. Imbayah
e-mail: ibrahim.alzayani@gmail.com

centre on adding a voltage control ability to self-consumption plans by a set of voltage following battery charging, automatic reactive power equipment like PV power reducing [1]. PV technology provides an extra focus on the advantages of sun-powered vitality being perfect, abundant, environment-friendly, and inexhaustible. The PV framework can be considered as the proficient system to give the power to some remote off-lattice areas; furthermore, it can be financially savvy when contrasted with the expense of running electrical cables from the current matrix. Actually, there are a lot of effects that almost change the output power of the PV system, such as ambient temperature, shading, and relative humidity. Because of the environmental factors and change of solar irradiation, the PV system output power is a stochastic system [2]. The control technique is to integrate the distributed generation (DG) resources to the electrical power network. The suggested control, it can be compensated for active, reactive, and harmonic load current components during connection of DG links to the grid [3]. The researchers are interested in the generation of electric power from renewable energy, and the most focused in the solar energy. Solar energy (PV) can be used today in many applications now because of where advantages such as pollution-free can be maintained. The high cost of oil and gas has weakened the economies of several countries, since solar PV is capable of generating large amounts of clean and chip electrical energy [4].

The solar PV is an essential technique of different sustainable power sources. It is widely used all over the world because it is cheap and simple. Solar energy is assumed to play an important part in the coming days, especially in this period most of the developing countries in the world focus in this area. So the generating of electric power by using solar cells is an effective technique nowadays [5]. In this chapter, they give some ideas about a single-phase five-level photovoltaic (PV) inverter scheme for PV systems connected to the grid with a new pulse-width-modulated (PWM) control design [6]. The chapter presented a model of photovoltaic suitable for use, the ability to upgrade various designs. Dealing in later with a sound analytical PV cell model and work in this paper modelling clearly appears under SIMULINK.

The excitement of the model of the three entrances to contain the radiation, temperature, and voltage design. Who is running simulations and the results in terms of peak power and analyse the current [7]. The reliability of the system when the PV plant is linked to the grid is strengthened when the power factor, the utilisation of safety, and grid synchronisation capabilities functions are improved. Development in the efficiency of photovoltaic cells and low price makes them widely accepted and concentration in the entire world and not to rely on non-renewable energy sources, the most important advantages we have got electrical energy through clean renewable energy. The power plants in recent times need to increase their energy output, need appropriate control methods to resolve

the issues associated with the partial shading phenomena and different orientation of the PV modules closer to the solar [8].

This chapter has been referenced in the prediction model for PV out of the current, and this model depends on the cascade-forward, back propagation artificial neural network (CFNN) with two inputs and one output [9].

The experimental data are used for the PV system 1.4 kW for 2 years in this search, for this purpose an experimental database of the solar power output, module temperature, air and solar irradiance data has been used. The control approach is able to reduce the reactive power beside it can control the power quality required by a voltage sag/swell for the stability of the grid. This approach does not require any change in the hardware when compared with the employ of existing strategy [10]. In the given scheme, LLC resonant DC/DC converts is applied to get the maximum power and boost the PV array voltage.

The evolution of the grid linked large-scale power plants in light of the fact that the expanding entrance of renewable vitality. The control technique strategy can do the PQC, as well as it can remunerate the receptive force requested by a voltage drop/swell for grid adjustment. The photovoltaic (PV) systems were designed to aim, till now, to get the superior power from the PV and integrated into the utility network. The evolution of the grid linked large-scale power plants in light of the fact that the expanding entrance of renewable vitality.

The control technique strategy can do the PQC, as well as it can remunerate the receptive force requested by a voltage drop/swell for grid adjustment. The control procedure does not require any outside equipment alteration when contrasted and the utilisation of traditional current control [11]. The power plants these days are required to expand their vitality generation, needing suitable control methodologies to support take care of the issues identified with the fractional shading wonders and the distinctive introduction of the PV modules towards the sun [12].

2 Modelling PV Connected to Grid

It is a detailed model for a group of 100 kW connected to the 25 kV through a boost DC–DC converter. And it is performed maximum power point tracking (MPPT) in batch conversion by SIMULINK model using the P controller. Is the average model of network connected PV, a group of 200 kW connected to 25 kV network via two DC–DC converters boosts unit my control. It is implemented controller maximum power point tracking at the "confusion and control" technology through MATLAB function [13].

$$I = I_{\text{pv,cell}} - I_{0,\text{cell}}\left[\exp\left(\frac{qV}{\alpha KT}\right) - 1\right] \tag{1}$$

The simple PV cell does not express the *I–V* relationship of an actual PV array. Cells linked in parallel rise the current, and cells linked in series produce higher output voltages.

The actual arrays are built by a lot of linked PV cells, and the notice for the characteristics of the ends of the photovoltaic array needs the implying of such parameters to the main equation.

$$I = I_{PV} - I_0 \left[\exp \left(\frac{V + R_S I}{V_t \alpha} \right) - 1 \right] - 1 \frac{V + R_S I}{R_P} \tag{2}$$

We assume I_{sc} and I_{PV} ordinarily used in the PV device modelling as in working designs, the parallel resistance is high but series resistance is low. The diode saturation current is presented by:

$$I_0 = \frac{I_{SC,n} + K_{I\Delta_T}}{\exp \left(\frac{V_{OC,n} + K_V \Delta_T}{\alpha V_t} \right) - 1} \tag{3}$$

From the equation, the model is simplified, and the model error is equal to zero near of the open-circuit voltages, and as the results, at other regions of the *I–V* plot:

$$I_{PV} = \left(I_{PV,n} + K_{I\Delta_T} \right) \frac{G}{G_n} \tag{4}$$

The R_S and R_P the relationship is an unknown of (2) may be calculated by setting $p_{max,m} = p_{max,e}$ and solving the fulfilment equation for R_S, as given:

$$P_{max,m} = V_{mp} \left\{ I_{PV} - I_0 \left[\exp \left(\frac{q}{KT} \frac{V_{mp} + R_S I_{mp}}{\alpha N_S} \right) - 1 \right] - \frac{V_{mp} + R_S I_{mp}}{R_P} \right\} = P_{max,e} \tag{5}$$

$$R_P = \frac{V_{mp} + I_{mp} R_S}{\left\{ V_{mp} I_{PV} - V_{mp} I_0 \exp \left[\frac{(V_{mp} + R_S I_{mp})}{N_S \alpha} \frac{q}{KT} \right] + V_{mp} I_0 - P_{max,e} \right\}} \tag{6}$$

The means of Eq. 6 that for any value of R_p there will ba a value Of R_S that makes the mathematical *I–V* curve cross the experimental (I_{mp}, V_{mp}) point. The aim is to find the value of R_P and R_S for the best model solution.

$$I_{PV,n} = \frac{R_P + R_S}{R_P} I_{SC,n} \tag{7}$$

The initial conditions can be set as:

$$R_S = 0, \quad R_{P,min} = \frac{V_{mp}}{I_{sc,n\_} - I_{mp}} - \frac{V_{oc,n} - V_{mp}}{I_{mp}} \tag{8}$$

From Eq. 8, the minimum value of R_P is given by relationship between the maximum power points and the short circuit.

3 Control Approach of Grid-Connected PV System

For the control approach of grid-connected PV, MPPT theory is applied. The boost converter increase DC voltage from 273.5 to 500 V. This component utilises an MPPT system which automatically change the duty cycle so as to produce the necessary voltage to obtain maximum power.

In this chapter, I controller is good control and simulation model of PV is introduced giving an obvious view of the proper control I, and performance of any actual system with change time during the day can be used to investigate the performance of the PV generation system. The control is realised and examined in MATLAB/SIMULINK program.

3.1 MPPT Control Algorithm

The MPPT algorithms are required because PV arrays have a relationship between current and voltage is nonlinear, and there is a singular point where it provides the maximum power. This point relies on the temperature of the panels and on the irradiance terms. Both terms change during the day and are also different rely on the season of the year. Moreover, irradiation changes fast due to changing atmospheric terms such as cloudy terms. It is necessary to follow the MPP accurately under all reasonable conditions, therefore that the maximum available power is always obtained.

These key responses are maximum overshoot that should be low, rise time that should be low, and also the settling time that should be low. The increased time, steady state error, settling time, and overshoot, these four parameters will tell us that how to change the parameters of the controller. I controller has the transfer function:

$$dP/dv = 0 \quad \text{where} \quad P = V * I$$
$$d(V * I)/dV = I + V * dI/dV = 0$$
$$dI/dV = -I/V$$

The I controller minimises the error ($dI/dV + I/V$).

4   Simulation Results

The PI and I controller of integration of PV system into a grid has been created and examined by means of MATLAB/SIMULINK using the power systems block set. The selected simulation parameters of the solar equivalent circuit and MPPT are given in Table 1.

As shown in Fig. 1, the power is stable and gives good response, but the voltage is varied and disturbed. So that the I controller is only applied to solve the problem of voltage variation without affecting the power quality as depicted in Fig. 2. It is noted that the duty cycle in I controller is not highly varying as compared to PI controller.

Table 1 Parameters of the system

Parameter	Value	Parameter	Value
No. of cell	96	R_p	999.51
V_{oc}	64	R_l	5
I_{sc}	5.96	L	5 mH
V_{mp}	54.7	C_3	1200 MF
I_{mp}	5.58	K_p	2
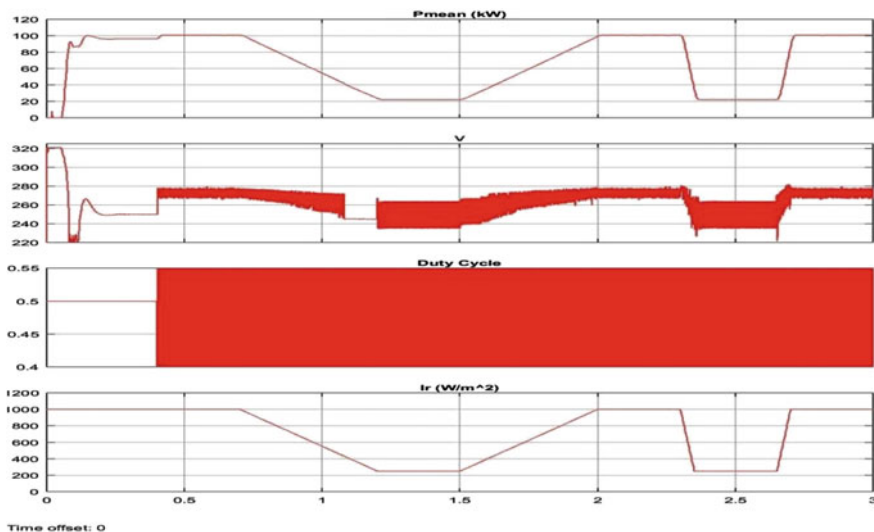
Fig. 1 PI controller
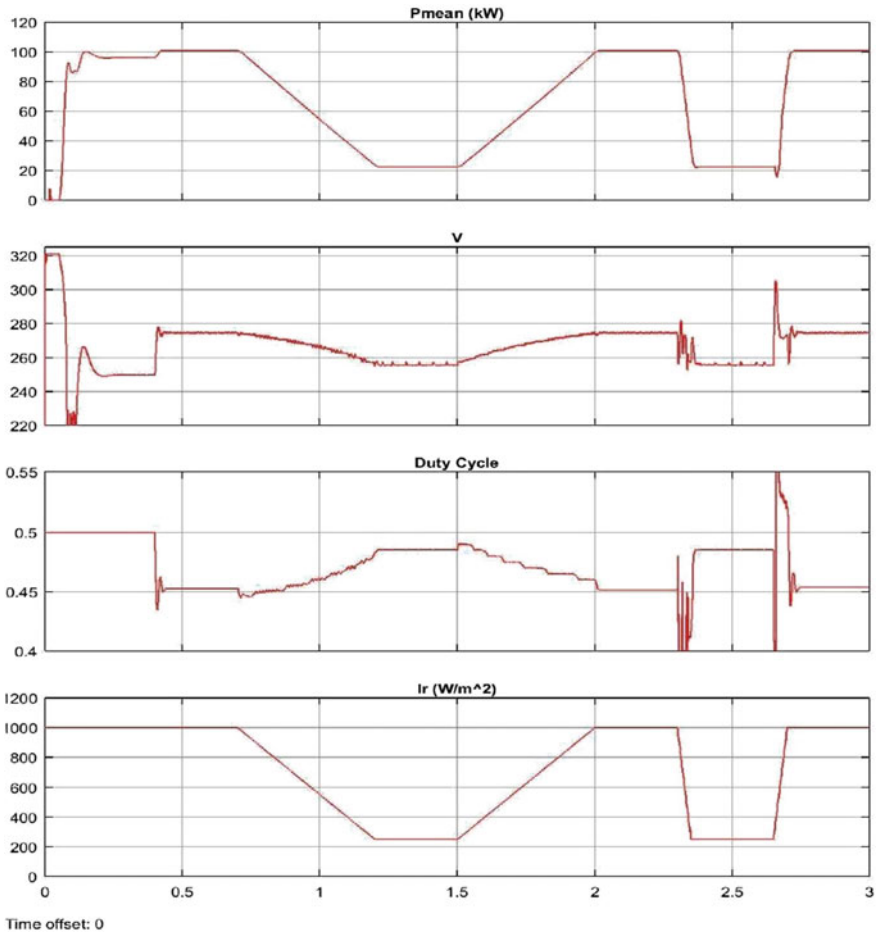
Fig. 2 I controller

5 Conclusion

The aim of this chapter is to simulate and control model of integration PV into utility grid using MATLAB. This model is based on the basic circuit of the solar cell PV system integrated into the grid the effects of parameters such as solar radiation surround. Environmental and physical parameters such as the cell temperature and solar radiation. The boost DC/DC converter is connected to PV system and the controller system is based on the maximum power point tracking (MPPT) with I controller ensure the maximum power in case of fluctuation in the weather, which is then integrated into the AC utility grid by DC/AC inverter.

References

1. Von Appen J, Stetz T, Braun M, Schmiegel A (2014) Local voltage control strategies for PV storage systems in distribution grids. IEEE Trans Smart Grid 5(2):1002–1009
2. Sumathi S, Ashok Kumar L, Surekha P (2015) Solar PV, and wind energy conversion systems. Springer, Cham
3. Pouresmaeil E, Miguel-Espinar C, Massot-Campos M, Montesinos-Miracle D, Gomis-Bellmunt O (2013) A control technique for integration of DG units to the electrical networks. IEEE Trans Ind Electron 60(7):2881–2893
4. Jiang Y, Abu Qahouq JA (2011) Study and evaluation of load current based MPPT control for PV solar systems. In: 2011 IEEE energy conversion congress and exposition (ECCE), 17–22 Sept 2011, pp 205–210
5. Ameen AM, Pasupuleti J, Khatib T, Elmenreich W, Kazem HA (2015) Modeling and characterization of a photovoltaic array based on actual performance using cascade-forward back propagation artificial neural network. J Sol Energy Eng 137(4):041010
6. Shu-Hung H, Tse KK, Hui SYR, Mok CM, HO MT (2003) A novel maximum power point tracking technique for solar panels using a SEPIC or Cuk converter. IEEE Trans Power Electron 18(3):717–724
7. Acakpovi A, Ben Hagan E (2013) Novel photovoltaic module modeling using Matlab/Simulink. Int J Comput Appl 83(16):27–32
8. Romero-Cadaval E, Spagnuolo G, Garcia Franquelo L, Ramos-Paja CA, Suntio T, Xiao WM (2013) Grid-connected photovoltaic generation plants: components and operation. IEEE Ind Electron 7(3):6–20
9. Bae Y, Vu TK, Kim RY (2013) Implemental control strategy for grid stabilization of grid-connected PV system based on german grid code in symmetrical low-to-medium voltage network. IEEE Trans Energy Convers 28(3):619–631
10. Buccella C, Cecati C, Latafat H (2015) Multi-string grid-connected PV system with LLC resonant DC/DC converter. Intell Ind Syst 1:37–49
11. Bae Y, Vu T-K, Kim R-Y (2013) Implemental control strategy for grid stabilization of grid-connected PV system based on german grid code in symmetrical low-to-medium voltage network. IEEE Trans Energy Convers 28(3):619–631
12. Romero-Cadaval E, Spagnuolo G, Franquelo LG, Ramos-Paja CA, Suntio T, Xiao WM (2013) Grid-connected photovoltaic generation plants: components and operation. IEEE Ind Electron 7(3):6–20
13. Giroux P (2012) Grid-connected PV array. Matlab Central, 2012. [Online]. www.mathworks.com/Matlabcentral/fileexchange34752-grid-connected-pv-array

Quantum Cost Optimization of Reversible Adder/Subtractor Using a Novel Reversible Gate

B.P. Bhuvana and V.S. Kanchana Bhaaskaran

Abstract The reversible logic has been gaining increased attention due to its characteristic feature of incurring low-power dissipation in digital circuits. It plays a major role in quantum computing. This paper suggests a novel 4 * 4 reversible gate called BKAS gate. One of the most interesting characteristic features of BKAS is that it can be functioned individually as a reversible full adder, as well as a full subtractor. This feature allows it to be used in the design of complex adder architectures with reduced garbage outputs, quantum cost, and ancilla input. The proposed BKAS gate is validated using a comparison with the prevailing reversible gates in the literature.

Keywords Reversible computing · Quantum cost · Quantum computing
Quantum cellular automata · Low-power adder · Low-power subtractor

1 Introduction

During the past few years, more researchers have concentrated working in the field of reversible logic due to its low power consumption in applications involving lower technology nodes. The reversible logic has an important role in the field of optical computing, quantum computation, cryptography, quantum dot cellular automata, DNA computing, digital signal processing, and low-power CMOS design.

Landauer [1] in 1960 proposed that every single bit in irreversible computation produces KTln2 joules of heat energy. Here, K and T represent the Boltzmann constant and absolute temperature, respectively. This loss occurs because of information loss during the course of irreversible computation. This was supported by Bennet [2] who stated that the energy consumed is zero when the input can be uniquely regained back from the output. The reversible circuits have no loss in information and the computation is said to be reversible only if it consists of the

B.P. Bhuvana · V.S. Kanchana Bhaaskaran (✉)
School of Electronics Engineering, VIT University, Chennai, India
e-mail: vskanchana@gmail.com

reversible gates which depicts the fact that no logic states are left unexploited. Some of the important constraints in reversible computing are as follows: (i) The fan-outs are eliminated, (ii) minimization of constant inputs, (iii) quantum cost to be a minimum, (iv) minimization of garbage outputs, and (v) minimization of the gate level or latency. Optimization of all these constraints provides an enhanced reversible circuit.

Adders and subtractors are the primary components in the digital system. A reversible adder and subtractor can be designed using a single reversible gate, which has not been referred much in the literature. References [3, 4] depict various methodologies for the design of adder and subtractor using one or more reversible gates. In this paper, a novel reversible adder/subtractor gate called BKAS is designed with a single reversible gate. It has a smaller amount of constant inputs and garbage output when compared with the existing techniques. The quantum cost, constant or ancilla inputs, and garbage outputs are estimated for the new BKAS gate. An optimized reversible 4-bit binary adder, a 4-bit BCD (Binary Coded Decimal) adder, and a 4-bit binary subtractor have been designed using the proposed gate and presented in this paper. The designed adders using BKAS gate are enhanced for their comparative quantum cost, constant inputs, gate count, and garbage outputs for justifiable comparisons.

This paper is systematized as explained. Section 2 presents the novel reversible BKAS gate and Sect. 3 shows reversible adder architectures. Section 4 describes the results and discussion, and Sect. 5 states the conclusion of the paper.

2 The Proposed BKAS 4 × 4 Reversible Gate

In this paper, a novel reversible gate, called BKAS as depicted in Fig. 1, is proposed. The width of the BKAS gate is 4, i.e., it is a 4 * 4 reversible gate. The truth table of the proposed gate is shown in Table 1. The input vectors are represented by A, B, C, and D. Outputs are defined as P, Q, R, and S. The output bits are defined by $P = A$, $Q = B \oplus D$, $R = A \oplus B \oplus C$, and $S = ((A \oplus C) \oplus D) \cdot B \oplus (A \oplus D) \cdot C$. The unique feature of the proposed gate is that it can uniquely act as a full subtractor as well as full adder. Quantum cost of proposed BKAS gate is 6, incurring 6Δ delay. Quantum representation (QR) and graphical representation (GR) of the proposed BKAS are shown in Figs. 2 and 3. The proposed BKAS gate is capable of operating like a NAND gate which is a universal gate capable of implementing all logical function.

Fig. 1 Proposed novel reversible BKAS gate
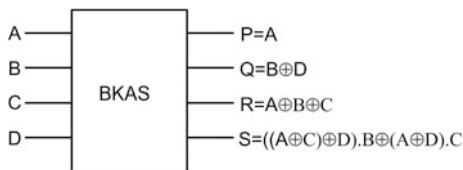
Table 1 Truth table of proposed BKAS gate

A B C D	P Q R S	A B C D	P Q R S
0 0 0 0	0 0 0 0	1 0 0 0	1 0 1 0
0 0 0 1	0 1 0 0	1 0 0 1	1 1 1 0
0 0 1 0	0 0 1 0	1 0 1 0	1 0 0 1
0 0 1 1	0 1 1 1	1 0 1 1	1 1 0 0
0 1 0 0	0 1 1 0	1 1 0 0	1 1 0 1
0 1 0 1	0 0 1 1	1 1 0 1	1 0 0 0
0 1 1 0	0 1 0 1	1 1 1 0	1 1 1 1
0 1 1 1	0 0 0 1	1 1 1 1	1 0 1 1

Fig. 2 QR of the reversible BKAS gate
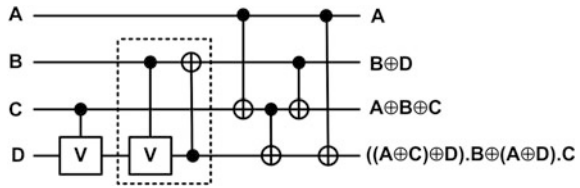
Fig. 3 GR of reversible BKAS gate
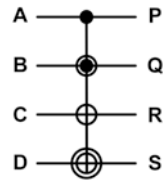
3 Design of Adder Architectures Using BKAS Reversible Gate

3.1 4-Bit Binary Adder

The binary 4-bit adder or ripple carry adder (RCA) is the most popular adder architecture which uses very simple combinational logic for the sum of two numbers. A n-bit RCA utilizes n number of full adders to derive the *sum* and *carry* outputs. The BKAS gate can be utilized to propose a full adder with 1 constant input and 2 garbage outputs [5]. Figure 4 shows the quantum representation (QR) of 4-bit ripple carry adder (RCA). The quantum cost of this 4-bit RCA is 24. It has 4 ancilla inputs and 8 garbage outputs.

Fig. 4 Binary adder
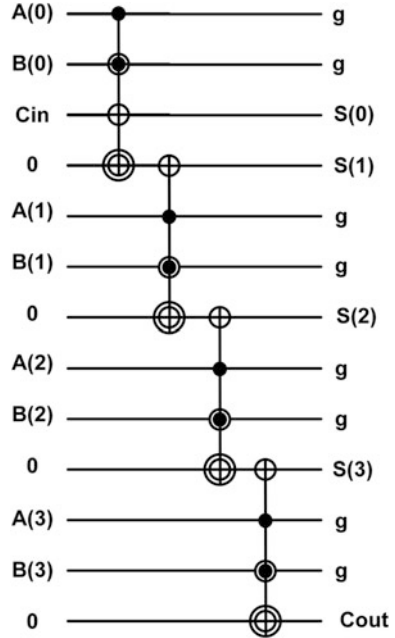
3.2   4-Bit BCD Adder

The binary coded decimal (BCD) adder is a form of binary adder that detects the sum output till 9. Any sum bit that is above 9 will be detected and corrected by the correction logic [6]. Conventional BCD adder comprises of three parts, viz. 4-bit parallel adder, overflow detection unit, and the logic for correction. When the carry produced by 4-bit parallel adder is 0, then the sum generated is considered as the final sum of 4-bit BCD adder. When the carry produced is equal to one, then the correction logic adds a 0110. For 4-bit adder, 4 BKAS reversible gate is used, overflow detection logic used one NCT gate and Toffoli gate, and the correction logic utilizes Peres gate, BKAS gate, and Feynman gate. Figure 5 shows quantum depiction of BCD adder.

3.3   4-Bit Binary Subtractor

A subtractor for 4-bit binary subtracts 2 binary numbers and produces the corresponding *difference* and *borrow* bits. An *n*-bit binary subtractor utilizes *n* full subtractors to derive the difference and borrow outputs. The BKAS gate can be utilized to implement a full subtractor with 1 constant or ancilla input and 2 outputs

Fig. 5 BCD adder
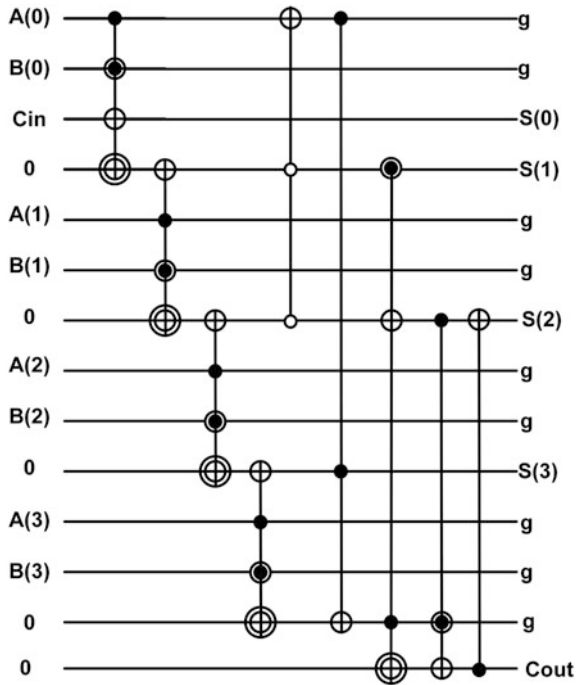
for garbage. Figure 6 shows the quantum depiction of 4-bit subtractor. The quantum cost of this 4-bit binary subtractor is 24. It has 4 ancilla inputs and 8 garbage outputs [7].

4 Results and Inferences

The efficiency of the proposed BKAS gate has been proved by comparison with the 4-bit binary adder and 4-bit binary coded decimal adder. The comparisons are made based on the total count of reversible gates employed in the design, the garbage outputs quantity, constant inputs, and the quantum cost of the design.

Comparisons have been made against existing reversible gates. Table 2 depicts the comparative analysis of 4-bit binary adder. The BKAS reversible gate has a quantum cost of 24, which is much less than the methodologies [5–7] with quantum cost 48, 56, and 32, respectively. Table 3 shows the comparison of an N-bit adder with reversibility characteristics, as a percentage of improvements vide the previous methodologies.

Table 4 presents the comparison of 4-bit binary coded decimal (BCD) adder. Proposed BCD adder design with BKAS gate has quantum cost as 46, 10 garbage outputs and 6 ancilla inputs. Thapliyal and Ranganathan [6] and Nagamani et al. [7]
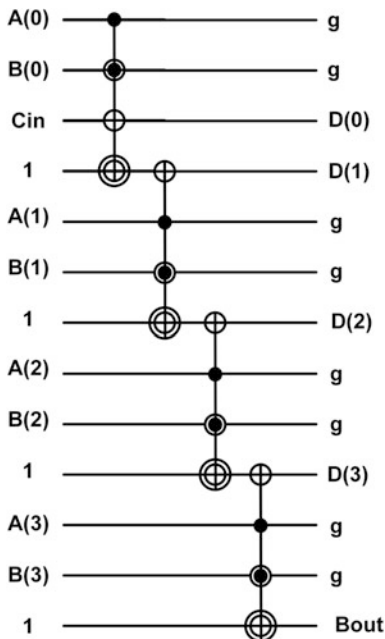
Fig. 6 Binary subtractor

Table 2 Comparison analysis of 4-bit binary adder

Design	Quantum cost	# Constant inputs	# Garbage outputs
[5]	48	4	0
[6]	56	0	0
[7]	32	4	8
Proposed	24	4	8

Table 3 Comparative analysis of N-bit binary adder with reversibility

# Bits	[5]	[6]	[7]	Proposed	Improvement (%) vide [5]	Improvement (%) vide [6]	Improvement (%) vide [7]
4	48	56	32	24	50.00	57.14	25
8	96	114	64	48	50.00	57.89	25
16	192	234	128	96	50.00	58.97	25
32	384	474	256	192	50.00	59.49	25
64	768	954	512	384	50.00	59.75	25
128	1536	1914	1024	768	50.00	59.87	25

Table 4 Comparison analysis of 4-bit binary coded decimal adder

Design	Quantum cost	# Constant inputs	# Garbage outputs
[5]	113	13	10
[6]	70	4	3
[7]	56	7	7
Proposed	46	6	10

Table 5 Comparative analysis of an N-bit BCD adder with reversibility

Bits	[5]	[6]	[7]	Proposed	Improvement (%) vide [5]	Improvement (%) vide [6]	Improvement (%) vide [7]
4	113	70	56	46	59.29	34.29	17.86
8	226	140	112	92	59.29	34.29	17.86
16	452	280	224	184	59.29	34.29	17.86
32	904	560	448	368	59.29	34.29	17.86
64	1808	1120	896	736	59.29	34.29	17.86
128	3616	2240	1792	1472	59.29	34.29	17.86

Table 6 Comparison analysis of full subtractor

Design	Quantum cost	# Constant inputs	# Garbage outputs
[5]	48	16	0
[6]	58	1	3
[8]	68	1	1
Proposed	24	4	8

Table 7 Comparative analysis of an N-bit binary subtractor with reversible characteristics

Bits	[5]	[6]	[8]	Proposed	Improvement (%) vide [5]	Improvement (%) vide [6]	Improvement (%) vide [8]
4	48	58	68	24	50.00	58.62	64.71
8	96	116	136	48	50.00	58.62	64.71
16	192	232	272	96	50.00	58.62	64.71
32	384	464	544	192	50.00	58.62	64.71
64	768	928	1088	384	50.00	58.62	64.71
128	1536	1856	2176	768	50.00	58.62	64.71

have very less ancilla or constant inputs and garbage outputs, however with very high cost of 70 and 56, respectively. Table 5 shows relative performance investigation of N-bit reversible binary coded decimal adder with percentage of improvement with respect to previous methodologies [5–7]. Table 6 presents the comparison of 4-bit binary subtractor. Proposed binary subtractor has quantum cost as 24, whereas [5, 6, 8] incur 48, 58, and 68, respectively. Table 7 portrays the comparison of N-bit reversible binary circuit with a percentage of improvement with respect to previous methodologies [5, 6, 8].

5 Conclusion

This paper proposes a novel reversible logic gate called BKAS with a distinguishable feature of individually acting as a reversible full adder and a full subtractor that is totally reversible. The efficiency of the BKAS has been validated by the use of 50, 57, and 25% a lesser quantum cost, respectively, as compared to the existing counterparts in binary adder. Furthermore, the quantum cost of BCD adder designed using BKAS is reduced by 59, 34, and 17% when compared with existing techniques. Binary subtractor designed using BKAS has 50, 58, and 64% lesser quantum cost with the existing counterparts.

References

1. Landauer R (1961) Irreversibility and heat generation in the computing process. IBM J Res Dev 5(3):183–191
2. Bennet CH (1973) Logical reversibility of computation. IBM J Res Dev 17:525–532
3. Emam MT, Elsayed LAA (2010) Reversible full adder/subtractor. In: 2010 XIth international workshop on symbolic and numerical methods, modeling and applications to circuit design (SM2ACD)
4. Bhuvana BP, Kanchana Bhaaskaran VS (2016) Design of reversible adders using a novel reversible BKG gate. In: Presented in ICIIECS-2016, to be published in IEEE conference proceedings
5. Nagamani AN, Ashwin S, Agrawal VK (2015) Design of optimized reversible binary and BCD adders. In: IEEE international conference on VLSI systems, architecture, technology and applications (VLSI-SATA), pp 1–5
6. Thapliyal H, Ranganathan N (2013) Design of efficient reversible logic-based binary and BCD adder circuits. ACM J Emerg Technol Comput Syst 9(3):17
7. Nagamani AN, Reddy NJ, Agrawal VK (2016) Design and analysis of reversible binary and BCD adders. In: Microelectronics, electromagnetics and telecommunications, pp 741–753
8. Cuccaro SA, Draper TG, Kutin SA, Moulton DP (2004) A new quantum ripple-carry addition circuit. arXiv preprint quant-ph/0410184

A Dual-Rail Delay-Insensitive IEEE-754 Single-Precision Null Convention Floating Point Multiplier for Low-Power Applications

J. Sudhakar, Y. Alekhya and K.S. Syamala

Abstract Due to plenty of intrinsic overheads in synchronous circuit design, asynchronous designs have drawn consideration in the Electronic Design Automation Industry. In an asynchronous logic design methodologies, Null Convention Logic is the best delay-insensitive logic as it has many advantages such as inherent robustness, power consumption, and modular reusability. For digital signal processing and computational dynamic range applications, floating point multiplication is a critical part with high precision and low power. Null cycle reduction technique and fine grain pipelining can be applied to the Null Convention Logic floating point multiplier to increase throughput. In this paper, Multiplier using Asynchronous Delay-Insensitive single precision NCL Floating Point Multiplier with IEEE-754 standard architecture is proposed and its performance is compared with different topologies in terms of various metrics such as delay, area, power consumption, and percentage of energy savings.

1 Introduction

The synchronous circuits have dominated semiconductor industry for the past two decades. Excessive clock skew, clock noise, increase in power consumption, and timing closure effort of synchronous circuits have led the way to the asynchronous world of VLSI [1, 2]. Null Convention Logic is an asynchronous delay-insensitive methodology that requires low power, least noise, and low electromagnetic interference (EMI) and allows modular reusability of components compared to synchronous countercircuits [3]. High dynamic range applications such as fast

J. Sudhakar (✉) · Y. Alekhya · K.S. Syamala
Electronics & Communication Engineering Department,
Vignan's Institute of Engineering for Women, Visakhapatnam, AP, India
e-mail: J.Sudhakar@gmail.com; sudhakar.jyo@gmail.com

Fourier transform require an efficient hardware to support floating point data. The proposed NCL is targeted to perform multiplication of floating point numbers and to dissipate lower power when compared to its equivalent synchronous floating point multiplier [4, 5]. The performance attributes of NCL floating point multiplier are analyzed in terms of power, delay, and percentage of energy savings.

2 Prior Work

Delay insensitivity, hysteresis, and input completeness are distinct advantages of NCL circuits [3]. Delay insensitive is achieved through dual rail or quad rail. A dual-rail signal D has two wires D0 and D1, whose values are from the set (DATA0, DATA1, NULL) as illustrated in Table 1. DATA0 and DATA1 represent Boolean logic levels 0 and 1, respectively. NULL represents empty set, a state when DATA is not available [1]. The mutually exclusive emphasizing that they cannot be asserted simultaneously. If assigned, it is called an illegal state [6, 7].

Threshold NCL gates with hysteresis state-holding capability are constructed to realize the NCL circuits. A basic threshold gate, specified as THmn gate in Fig. 1, has n inputs and 1 output. At least 'm' of the inputs must assert before the output will become asserted, and all inputs have to be deasserted before the output becomes deasserted. Input completeness illustrates that all outputs must not transit from NULL to DATA or DATA to NULL until all inputs have transited from NULL to DATA or DATA to NULL [8, 9].

The NCL modules, designed using threshold gates, are sandwiched between the delay-insensitive (DI) registers to realize input complete NCL system [10]. The flow of DATA and NULL wave fronts is controlled by the request and acknowledge signals.

Table 1 Dual-rail signal representation

STATE	DATA0	DATA1	NULL	ILLEGAL
D0	1	0	0	1
D1	0	1	0	1

Fig. 1 THmn threshold gate
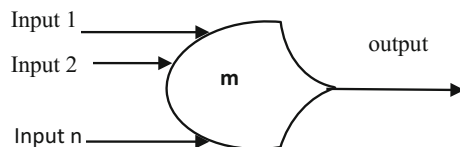
Table 2 Characteristic parameters of IEEE-754 formats

Precision	Sign (bits)	Exponent (bit)	Mantissa (bit)	Total length (bits)	Decimal digits of precision
Single	1	8	23	32	>6
Single-extended	1	≥ 11	≥ 32	≥ 44	>9
Double	1	11	52	64	>15
Double-extended	1	≥ 15	≥ 64	≥ 80	>19

IEEE-754 standard specifies four classifications to represent floating point values [11]:

1. Single precision (32 bit),
2. Double precision (64 bit),
3. Single-extended precision (≥ 43 bit, not commonly used), and
4. Double-extended precision (≥ 79 bit, usually implemented with 80 bits).

Table 2 explains the basic constituent part of the single- and double-precision formats. As shown in Table 2, the floating point number has three basic components: the sign, exponent, and mantissa. '0' at MSB denotes a positive number, and '1' denotes a negative number. The n-bit exponent field needs to characterize both positive and negative exponent values. To achieve this, a bias equal to $2^{n-1} - 1$ is added to the actual exponent in order to attain the stored exponent [12, 13].

3 Null Convention Logic Floating Point Multiplier

The fundamental components of the NCL floating point multiplier were designed using NCL combinational logic design concepts [14, 15]. NCL XOR gate, NCL ripple carry adder, NCL ripple borrow subtractor, NCL mantissa multiplier, NCL exponent incrementer (IE), and NCL intermediate product (IP) shifter are the NCL components that were needed to realize the NCL floating point multiplier [16]. The components were integrated and sandwiched between the delay-insensitive (DI) registers as shown in Fig. 2.

The NCL floating point multiplier comprises of two DI register banks (RB), one at both the input and the output. 66-bitDI RB1 receives the two inputs as normalized numbers. 55-bit DI RB2 outputs the normalized result of NCL floating point multiplier [13]. NCL XOR gate acts on the sign bits of the two inputs to produce sign bit of the product (sign). Dual-rail XOR gate comprises of two numbers of TH24compx0 threshold gates in order to produce dual-rail outputs sign0 and sign1.
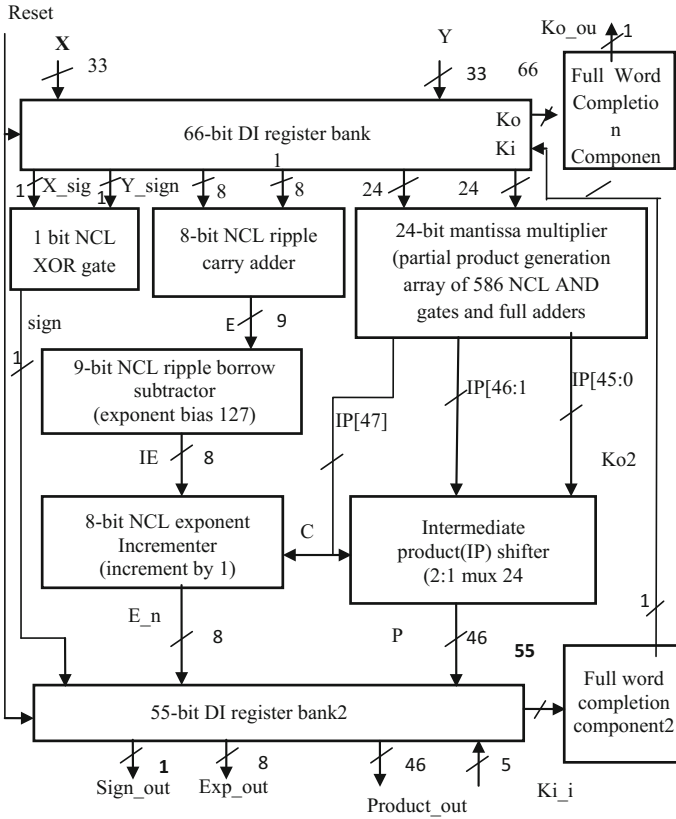
Fig. 2 Architecture of NCL floating point multiplier

Table 3 Comparison between various techniques in terms of performance metrics

Performance metric	Combinational FPM	Synchronous FPM	Booth's FPM	Wallace tree FPM	NCL FPM
Power (μW)	0.431	0.397	0.384	0.381	0.312
Delay (nS)	65	64	63	62	58
PDP (fJ)	28.015	25.408	24.192	23.622	18.096
Energy savings (%)	–	9.1	9.2	15.6	35.4

NCL significant multiplier performs multiplication of the 24 input significant bits, X [23:0] and Y [23:0]. It comprises of 576 NCL AND gates and NCL full adders. The significant bits of the inputs act on the NCL AND gates to produce bits that are given to the NCL full adders to produce the 48-bit product, IP [47:0].

4 Results and Discussions

We designed a 32-bit single-precision floating point multiplier using NCL asynchronous design style. The functionality of each block is described with VHDL. The VHDL description of NCL FPM is written at structural level [4]. It was simulated on Xilinx ISE simulator using an exhaustive VHDL test bench that generates 233×233 possible pairs of input operands [15]. Power analysis is performed using Xpower Analyzer, and the results obtained were presented in Table 3.

As presented in Table 3, Null Convention Logic design approach provides better results in terms of power and delay. We achieved 35.4% energy savings using Null Convention Logic.

5 Conclusion and Future Work

We have designed, simulated, and synthesized an IEEE-754 single-precision NCL FPM. The gate-level structural model of the proposed NCL FPM dissipated much less power than its synchronous counterpart. Hence, it can be used as a reusable library component in NCL-based digital signal processing applications that demand low power and high precision. The future work is to optimize the proposed design for higher throughput and lower power. NULL cycle reduction technique and fine grain pipelining can be applied to the NCL FPM to increase the throughput. We can reap better results with Return To One (RTO) protocol and Multi Threshold approaches.

References

1. Smith SC, Di J (2009) Designing asynchronous circuits using NULL convention logic. Synth Lect Digit Circuits Syst 4(1):1–96
2. Fant KM (2005) Logically determined design: clockless system design with null convention logic. Wiley, Hoboken
3. Bandapati SK, Smith SC (2007) Design and characterization of NULL convention arithmetic logic units. Microelectron Eng 84(2):280
4. Parsan FA, Smith SC (2012) CMOS implementation of static threshold gates with hysteresis. In: Proceedings of the IEEE/IFIP 20th international conference on VLSI and system-on-chip (VLSI-SoC '12), pp 41–45
5. Song Z, Smith SC (2013) Implementation of a fast fourier transform processor in NULL convention logic. In: Proceedings of the international conference on computer design, pp 10–16
6. Kotipalli S, Kim Y-B, Choi M (2014) Asynchronous advanced encryption standard hardware with random noise injection for improved side-channel attack resistance. J Electr Comput Eng 2014(837572):13–15
7. Smith SC (2007) Design of an FPGA logic element for implementing asynchronous NULL convention logic circuits. IEEE Trans Very Large Scale Integr VLSI Syst 15(6):672–683

8. Reese RB, Smith SC, Thornton MA (2012) Uncle: an RTL approach to asynchronous design. In: Proceedings of the 18th IEEE international symposium on asynchronous circuits and systems (ASYNC '12), Lyngby, pp 65–72
9. Bandapati SK, Smith SC, Choi M (2003) Design and characterization of NULL convention self-timed multipliers. IEEE Des Test Comput 20(6):26–36
10. Smith CM, Smith SC (2010) Comparison of NULL convention booth multipliers. In: Proceedings of the international conference on computer design (CDES '10), pp 3–9
11. Sankar R, Kadiyala V, Bonam R (2007) Implementation of static and semi-static versions of a bit-wise pipelined dual-rail NCL 2s complement multiplier. In: Proceedings of the IEEE region 5 technical conference (TPS '07), pp 228–233
12. Men L, Di J (2014) An asynchronous finite impulse response filter design for digital signal processing circuit. In: Proceedings of the IEEE 57th international midwest symposium on circuits and systems (MWSCAS '14), College Station, pp 25–28
13. Al-Ashrafy M, Salem A, Anis W (2011) An efficient implementation of floating point multiplier. In: Proceedings of the Saudi international electronics, communications and photonics conference (SIECPC '11), pp 1–5
14. Parhi KK (2007) VLSI digital signal processing systems-design and implementation. Wiley, New York
15. Smith SC http://www.ndsu.edu/pubweb/scotsmit/CCLIasync.html
16. Weste N, Harris D (2011) CMOS VLSI design: a circuits and systems perspective, 4th edn. Addison-Wesley, Boston

Part II
Signal and Image Processing

Retinal Diagnosis Exploitation Image Process Algorithms

B. Srilatha and V. Malleswara Rao

Abstract Vision is the most vital factor in human life. thus we'd like to avoid wasting our vision. that may be done by extracting retinal options. The membrane of human eyes that affects the membrane and retina construction in several ways in which. Latest technological advances within the image process helps to extract the attention diseases supported the study of feature extractions. In our projected, system we tend to used four algorithms for extracting the feature extraction. The initial step is to capture the retinal image exploitation digital anatomical structure camera. Consequent step is that the pre-processing stage, we tend to use advanced median filter to get rid of the unwanted distortions or noise with in the image. Future is the feature extractions method that is dole out on the pre-processed retinal image. The four extractions are blood vessels, exudates, small aneurysms and optic disk. The algorithms used are kirsch edge detection, modified fuzzy clustering, morphological distance based algorithm, and watershed algorithm. Here the four abnormalities are called as eye diseases. Supporting the output results of those four extractions, we discover the severity of the unwellness as gentle, moderate or severely affected. And eventually will do the treatment in early stage and that we can save our vision.

1 Introduction

Illness in membrane resulted from special diseases is verified by special photographs from membrane that are obtained by exploitation optic imaging called structure. The World Health Organization (WHO) has calculated the number of

B. Srilatha (✉) · V. Malleswara Rao
Bengaluru, India
e-mail: srilatha.ece@sairamce.edu.in

persons with eye diseases inside the globe would increase hugely from 100 thirty 5 million in 1995 to a few hundred million in 2025 [1]. Thus, on cut back the unhealthy sound effects relating to the method input image were pre-processed by a filter, and so the retinal structure image is assessed into a number of primary components such as red channel (R), inexperienced channel (G) and blue channel (B) [2]. The inexperienced channel is highly reactive to the blood vessels. The improved median filter [3] is used to induce obviate salt and pepper sound from the image. Thus the output is that the raised image a grip is academic degree sudden change inside the brightness (gray scale level) of the pixel. The removal of blood vessels in membrane pictures is a very important step in designation and medical aid of diabetic retina. For unhealthy persons, the diameter of the blood vessels might disagree or otherwise there is a break for growth of latest vessels once connected with traditional person's blood vessels [4]. The blood vessels get enveloped for diabetic patients, and it gets narrower for eye disease unhealthy patients. Exudates are first signs of membrane unwellness. Automatic transude detection would be useful for diabetic retinopathy screening method. The mix of outcome of yellow objects with terribly sharp fringe is employed to work out the exudates. The Fuzzy C-Means (FCM) gathering could be a well-known cluster technique for image partition, and several techniques are performed for exudates observation; however they need to defect. Poor quality pictures have an effect on the separation results of bright and dark injury exploitation thresholding and exudates feature extraction. Microaneurysms the first clinical abnormality to be noticed at intervals the attention for diabetic retinopathy they are red lesions. Red lesions square measure the first clinically evident injury indicating diabetic retinopathy. Therefore, their detection is important for a pre-screening system [5]. The purpose is that the shining half at intervals the standard image that will be seen as polish, spherical or vertically memory device is that the region where blood vessels and optic tract fiber enter to the membrane of human eyes. it is the shining a section of the traditional body part photos. Optic disk is the important part in the retinal image [6], and in our planned System watershed algorithmic program is employed for the optic discs detection and therefore the severity of the attention diseases is examined. Supporting the output results of those four extractions, we discover the severity of the illness as gentle, moderate or severely affected. And eventually do the treatment in early stage and that we can save our vision.

2 Materials and Methods

All the pictures employed in this paper are acquired from the govt. hospital real time patients. There are one hundred and ten retinal colour bodily structure pictures with an vary size of four hundred $\times$ 600 $\times$ three pixels, at the side of the cranial nerve congress derived by two specialists.

3 Proposed System

The proposed system consists of four stages. First stage is collecting the images of patients by fundus camera. This will be raw images and contains full of noise. This noise will be detected using filters. Second stage is pre-processing where improved median filter is used for the removal of error caused while taking the image and to reduce the noise. Third stage is the extraction of features of eye such as blood vessel, exudates, microaneurysms and optic discs using image processing algorithms. They are kirsch edge detection algorithm, fuzzy clustering algorithm, morphological distance-based algorithm and watershed algorithm, respectively, and finally fourth stage is comparison between blood sample results. The proposed system for detection of eye diseases is illustrated in Fig. 1.

3.1 I-Stage

Collect real-time patient images from the hospitals. These images are taken by fundus camera. This will be full of noise and raw images. That noise can be reduced in the second stage using filters.

3.2 II-Stage

The pre-processing step removes distinction because of image gain, similar to inhomogen. The illumination techniques like morphological operations area unit attempt to the input image. the subsequent session's the improved median filters area unit employed in pre-processing stage during this paper.
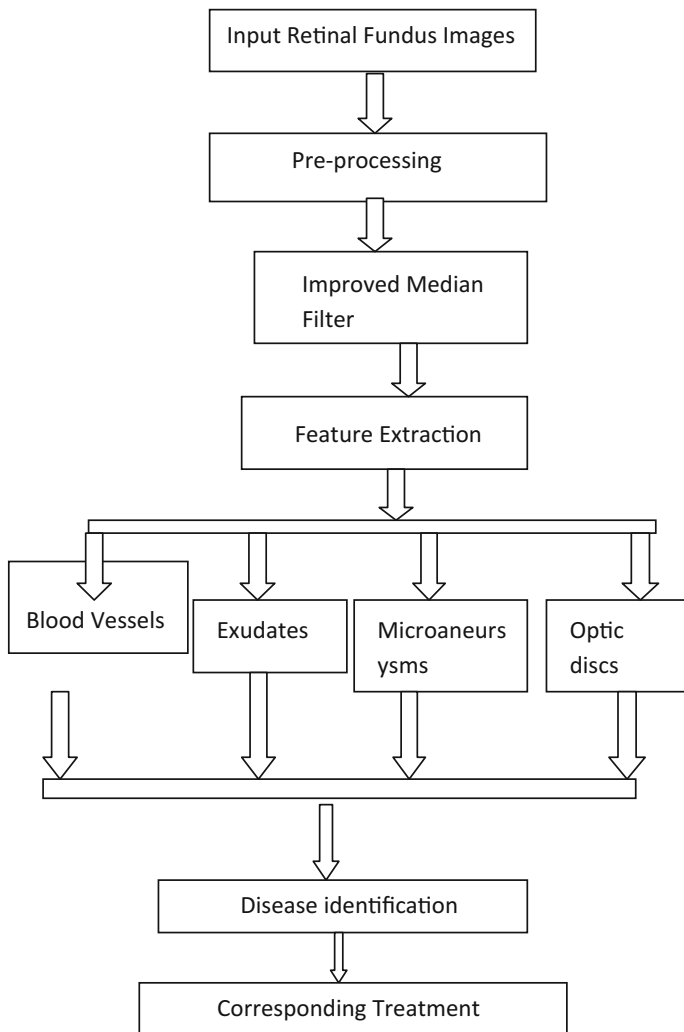
Fig. 1 Proposed system for eye diseases detection

3.3 Advanced Median Filter

To remove salt and pepper noise from the corrupted image with more algorithms is used. The output of all the filters is compared with improved median filter for the given input retinal images and also compared the performance of image with signal-to-nose quantitative relation. We used our proposed filter i.e. advanced median filter is giving best results compared to alternative filters. The algorithmic program is given below.

Step 1: Contemplate a digital input complex body part image and pad it with zeros;

Step 2: Contemplate a 3×3 matrix;

Step 3: Type the matrix in ascending order and calculate v0, pmid, pmax and pmin;

Step4: pmin < p(x, y) < pmax and pmin > 0, pmax < 255
If the condition is happy, then it is uncorrupted image, contemplate following 3×3 matrix and continue otherwise; and

Step 5: pmin < p(x, y) < pmax and 0 < pmed < 255.

If the condition is happy, then replace p(x, y) with pmid, otherwise realize vd and replace Vdmax by p(x, y).

3.4 III-Stage-Blood Vessels Feature Extraction

The observation of blood vessels from the retinal pictures could be a tedious method. Kirsch algorithm is used to find the blood vessels effectively. Authentic detection of blood vessels from membrane is associated with main job in pc-aided identification. Since the blood vessels area unit scattered in several directions, morphology process with multipath arranges components area unit to take away the vessel from the retinal pictures.

In our planned rule, cider's model is utilized for detecting the blood vessels from the retinal photographs. For edge detection, the sting photographs (i.e. detected edges) are also thought to be the house gradient. The potable gradient operator is chosen to extract the contour of the issue, and the potable edge detection uses eight filters (i.e. eight masks for connected eight main directions) that square measure applied to given image to seek out edges. Except the outer row and the outer column, every constituent and its eight neighbourhoods. Every constituent has eight outputs. Also, the foremost output of the eight templates taken in the given position. this will be made public as a result of the sting magnitude. The path of edge is made public by the connected mask. The final gradient is the summation of the improved edges by considering all ways that for RGB channel rather than any single channel alone. Here, varied directional enhanced photos square measure bestowed. This feature extraction is compared with all the algorithms. Finally, our planned rule gives correct results.

3.5 Exudates Extraction

Exudates are tiny yellow white patches with different shapes. Exudates area unit one amongst the first occurring lesions. Exudates area unit collections of lipoid and

macromolecule within the membrane. Usually, they are bright, reflective, white or cream colour lesions. They show raised vessel permeableness and a connected risk of retinal oedema. They are a marker of fluid assortment within the membrane. It is wont to highlight salient regions, extracts relevant options and eventually classifies those regions.

3.6 Fuzzy Cluster Algorithm

Fuzzy cluster is an associate in nursing overlapping cluster algorithmic rule, where each purpose may belong to tons of or two clusters with altogether completely different degrees of membership. The similarity is made public by the gap of the choices vector to the cluster middle. The cluster middle is updated until the excellence between adjacent objective performs is close to zero or a lot of, however, a predefined very little constant. The existed method is not extracting the exudates up to the best level. Once the short FCM is stopped, the IFCM algorithmic rule follows it with the values for the prototypes and membership values obtained from the short FCM algorithmic rule. The IFCM algorithmic rule then iteratively updates its priority chance, membership and centroids with these values. Once the IFCM algorithmic rule has converged, another defuzzification methodology takes place so as to convert the fuzzy segregation matrix to a crisp segregation matrix that is segmentation. Therefore, the IFCM algorithmic rule is given as follows:

Step 1: Set the cluster centroids vi convert to the histogram of the image;
Step 2: Compute the histogram;
Step 3: Compute the membership function;
Step 4: Compute the cluster centroids;
Step 5: Go to Step 3 and repeat until convergence;
Step 6: Compute the a priori probability with the obtained results of membership function and centroids;
Step 7: Recomputed the membership function and cluster centroids and the probabilities;
Step 8: If the algorithm is convergent, go to Step 9; otherwise, go to Step 6; and
Step 9: Image distribution after defuzzification, and then a region labelling procedure is performed.

This feature extraction is compared with all the algorithms. Finally, our proposed algorithm gives correct results.

3.7 Microaneurysms Feature Extraction

Microaneurysms on the membrane seem as tiny red dots of most size to be less than the diameter of the main optic veins. The recognition of microaneurysms is crucial in the operation of diabetic retinopathy grading, since it forms the idea of deciding whether or not associate image of a patient's eye ought to be treated healthy or not. Microaneurysms area unit little saccular pocket caused by native distension of capillary walls and seem as little red dots. Their walls area unit skinny and rupture simply to cause haemorrhages. To sight visible small aneurysms in tissue layer victimization size and form machine-driven small aneurysms detection and diabetic retinopathy grading and Hough transforms area unit gift however the morphological distance based mostly algorithmic rule for microaneurysms is economical and also the steps concerned as shown below.

Step 1: The pre-processing step filters the image, will increase the distinction and performs a shading correction so as to balance the non-uniform illumination across the image. The diameter-closing step may be a mathematical morphological transformation that fills all told the black dots with diameters smaller than λ.

Step 2: Once acting such closing transformation, the grey scale worth of the filled-in dots is on top of within the raised pre processed image, whereas the vessels and alternative parts stay nearly unaffected. The black top-hat step uses size and forms criteria to isolate the black elements contrasted against the background.

Step 3: The black hat remodel is that the results of the distinction between the photographs obtained by the scale closing and pre-handling steps. The above step identifies all parts with in the black region of the image and that area gives microaneurysms region.

Step 4: Finally, a K-nearest neighbours (k-NN) classifier is employed for classification. It uses the properties calculated for the candidates to search out them as either true µA or false positives supported the training set within the little information. The classifier classifies the correct microaneurysms by taking into consideration factors like size, contrast, shape and colour. Then, truth microaneurysms area unit is detected.

3.8 Optic Disc Extraction

The blind spot is the brightest half within the retina; the image is filtered so as to eliminate giant grey level distinction at intervals, the appendage region. The vessels area unit crammed implementing a straightforward closing operation. Classical watershed transformation is registered to the gradient to look at contours of the blind spot. The image is filtered so as to eliminate giant grey level distinction at
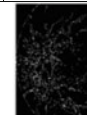
intervals, the appendage region. The vessels area unit crammed implementing a straightforward Closing operation. Pure watershed transformation is applied to the gradient to look at contours of the optic discs. during this algorithmic rule there area unit 2 steps:

Step 1: Sorting the pixels w.r.t. increasing grey worth, for unambiguous access to pixels at an exact grey level;

Step 2: A flooding step, continued level by level and starting from the minima. The implementation uses initial|a primary} in first out queue of elements, that is, a _rst-in-_rst-out system on it the next operations are going to be performed: _foadd(p; queue) adds element p at the highest of the queue, chief money handler remove(queue) returns and removes the erstwhile a part of the queue,foinit(queue) initializes associate degree empty queue, and chief money handler empty(queue) is also a ensure returns true if the queue is obvious and false otherwise. Flooding the graph using a breadth-_rst rule as follows:

Step 3: At intervals the flooding step, all nodes with grey level h unit erstwhile given the label mask. Then those nodes that have labeled neighbours from the previous iteration unit inserted at intervals the queue, and from these pixels geodesic inuence zones unit propagated at intervals the set of covert pixels. If a element is adjacent to two or extra di_erent basins, it's marked as a watershed node by the label wshed. If the element can exclusively be reached from nodes that have the same label, the node is incorporate with the corresponding basin. Pixels at the highest still have the value mask belongs to a gaggle of recent minima at level h, whose connected parts get a fresh label. As shown in [52], the time quality of rule four. 1 is linear at intervals the vary of pixels of the input image.

4   Simulation and Result

We enforced our projected methodology victimization mat science laboratory. The advanced median filter is enforced for American state noising of extremely corrupted pictures and edge bar. The brandy edge detection algorithmic rule works well for the photographs having clear distinguish between the foreground and background, since the retinal blood instrumentation may be thought about as needed foreground instruction for complex body part pictures, and brandy algorithmic rule may be effectively registered. Exudates area is one in all the foremost necessary issue of diabetic retinopathy. The red lesion is detected by the morphological distance-based mostly on algorithms. The important part in the retinal image is optic disk that could be a spherical oval form and circular region. From all

the photographs, say that projected algorithms are giving the higher results than alternative algorithms. Supporting the output results of those four extractions, we discover the severity of the unwellness as delicate, moderate or severely affected.

S No	Patient name	Patient I/P image	Eye diseases				Disease severity
			Blood vessels	Exudates	microaneary sms	Optic disk	
1	Marry						Mild
2	VenkateswaraRao						Moderate
3	Venkataswaralu						Normal
4	NagarajuKishore						Sever

5 Conclusion

Retinal pictures play different role in many applications such as sickness diagnosing and human identification. The segmental blood vessels are used for diagnosing diseases such as diabetic, eye disease and blind dot. In our proposed method the retinal image is the input to the improved median filter is applied for preprocessing stage. The stimulation result shows that the improved median filter algorithmic rule will had best with relationship between the consequences of noise reduction and time quality of algorithmic rule, and the brandy edge detection algorithmic rule will set and reset the edge to get the foremost applicable edge of the image for the retentive image details higher. Fuzzy clustering is used to extract the silent features of the retinal image effectively. The intense circular region from wherever the blood vessels arise is termed the optic disc. The simulation results show that the studied

technique is applied to differing kinds of image and supply terribly satisfying results. The conclusions of each segmentation and sweetening steps show that our technique effectively detects the skinny blood vessels, exudates, microaneurysms and optic discs.

FUTURE IMPROVEMENT

This work determines the presence of non-proliferative diabetic retinopathy in a patient by making an attempt technique of digital image changing on bodily structure pictures taken by the use of medical image camera by medical personnel in the hospital. During this work, we have got investigated and recommended a laptop-based system to establish traditional, non-proliferative diabetic retinopathy. The kirsch's operator can distinguish the blood vessels; however, the output vessels detected have a lot of breadth than the initial blood vessels, and thus sweetening is needed during this operator and also the extraction of exudates by the employment of fuzzy agglomeration technique solely the delicate and balanced stage of the polygenic disease is acknowledged and also the third stage cannot be verified. When the microaneurysms detection by the morphological distance is based mostly on algorithmic rule, the third stage is set. The foremost tough drawback of optic disc extraction is to find the region of interest. Since the situation of optical disk is treated because the landmark for the analysis and identification of disease and blood vessels in retinal pictures even supposing by currently some progress has been achieved there area unit still remaining challenges and directions for additional analysis, cherish removing completely different options and developing higher classification algorithms and integration of classifiers to relinquish higher performance and scale back the classification errors.

References

1. Jeyasri K, Subhatra P, Annaram I (2013) Detection of blood vessels for illness diagnosis. JARCSSE, pp 6–12
2. Xu L, Lua S (2010) A novel technique for vessel detection from retinal images. Research paper, pp 1–10
3. Peng Q, You X, Zhou L, Cheung YM (2010) Retinal blood vessels segmentation mistreatment the radial projection and supervised classification. In: International conference on pattern recognition, pp 1489–1492
4. Oloumi F, Rangayyan RM, Ells AL (2014) Measurement of vessel dimension in retinal structure footage of preterm infants with and disease. In: IEEE paper, pp 1–3
5. Pradeep MV, Shaik F, Abdulrahim B (2015) Extraction of blood vessels by combining Dennis Gabor filter and generalized linear mode. Int J Adv Technol Eng Sci 846–855
6. Kalashree S, Sowmya KS (2015) A novel algorithmic rule for exudates detection exploitation MATLAB. In: IEEE paper, pp 276–280
7. Khan MI, Sheik H, Mansuri AM, Soni P (2011) A review of retinal vessel segmentation techniques and algorithms. IEEE Pap 2(5):1140–1144
8. Asad AH, Azar AT (2013) An improved insect colony system for retinal vessel segmentation. In: Federate conference on applied science and knowledge systems, pp 199–205

9. Yadao P, Naval S, Maheswari A (2014) Blood vessel segmentation in retinal structure photos exploitation matched filter. In: National conference for faculty children in EEE, pp 58–61
10. Chaudari S, Chaterjee S, Katz N, Nelson M (2000) Detection of blood vessels in retinal photos exploitation second matched filter. IEEE Pap 8(3):263–269
11. Deshmukh SJ, Patil SB (2015) Vessel extraction from retinal footage by exploitation matched filter and figuring of gaussian function. In: IJIREEICE IEEE paper, pp 18–22
12. Vermer KA, Vos FM, Lemij HG, Vossepol AM (2010) A model primarily based technique for vessel detection. In: IEEE paper, pp 209–219
13. Singh D, Birmohan Singh D (2014) A new morphology primarily based approach for vessel segmentation in retinal images. In: IEEE paper, pp 8–13
14. Silkar S, Maity SP (2014) Extraction of retinal blood vessels exploitation curveletransform and fuzzy cluster method. In: Twenty second international conference on pattern recognition, pp 3392–3397
15. Bhadauria HS, Bisht SS, Singh A (2013) Vessel extraction from retinal images. IOSR J ECE 79–82
16. Kalaivani M, Jeyalakshmi MS, Aparna V (2012) Extraction of retinal blood vessels exploitation curve let transform and Kirsh templates. In: IJETAE IEEE paper, pp 360–363
17. Godwin Premi MS (2015) A novel approach for retinal blood vessels extraction and exudates segmentation. In: Analysis article, pp 792–797
18. Akila T, Kavitha G (2014) Detection and classification of onerous exudates in retinal structure footage exploitation cluster and random forest methods. In: National conference on computing and communication, pp 25–29
19. Maheswari MS, Punnolil A (2014) A novel approach for retinal lesion detection diabetic retinopathy images. In: International conference on innovations in engineering and technology, pp 1109–1114
20. Franklin SW, Rajan SE (2013) Diagnosis of diabetic retinopathy by using image method technique to watch exudates in retinal images. In: IEEE paper, pp 601–609
21. Kayal D, Banerjee S (2014) A new dynamic thresholding primarily based technique for detection of onerous exudates in digital retinal structure images. In: International conference on signal method and integrated network, pp 141–144
22. Hindu, Patil PMN (2014) Detection and classification of exudates exploitation k means cluster and color retinal images. In: Fifth international conference on signal and image method
23. Luangruagrang W, Kulkasem P, Rasmequan S, Rodtook A (2014) Automatic exudates detection in retinal footage exploitation economical integrated approaches. In: IEEE paper, pp 1–4
24. Murugha R, Reebakorah, Fathima SN, Haritha TV Microaneurysms detection ways that in retinal footage exploitation mathematical morphology. In: International J Adv Bailiwick Technol IEEE Pap, pp 120–128

Epoch Extraction Using Hilbert–Huang Transform for Identification of Closed Glottis Interval

Dhanalakshmi Samiappan, A. Jaba Deva Krupa and R. Monika

Abstract Epoch extraction helps in speech enhancement and multispeaker separation from a speech. But it is a challenging task due to time-varying characteristics of the source and the system. Epoch sequence is useful to manipulate prosody in speech synthesis applications. Accurate estimation of epochs helps in characterizing voice quality features. This chapter aims at developing an extraction algorithm independent of the characteristics of vocal tract system. It improves the accuracy of epochs extracted and pitch detected from speech signal. For feature detection, we propose a robust framework derived from Hilbert–Huang transform of speech signal. The intrinsic mode functions (IMF) sharply identify instantaneous frequencies as function of time. The proposed technique guarantees accurate pitch estimation because of better decorrelating nature of HHT compared with DCT and DFTs. The results are simulated for an input speech signal taken from NOISEX-92 database. The simulated results show that the proposed algorithm outperforms the existing methods.

Keywords Epoch · Pitch · Harmonic product spectrum
Zero-frequency resonators · Empirical mode decomposition (EMD)
Intrinsic mode factor (IMF) · Hilbert–Huang transform (HHT)

1 Introduction

Speech processing finds application in several areas like speech recognition, speaker recognition, speech enhancement, voice analysis. The main objective of speech signal processing is to extract the information of the time-varying characteristics of the speech production system. The information is represented in the form of parameters or features derived from the signal. Knowledge at different levels, such as acoustic–phonetic, prosody, lexical, syntactic, is used to interpret the message in the speech signal from the sequence of parameter or feature vectors [1, 2].

D. Samiappan (✉) · A. Jaba Deva Krupa · R. Monika
ECE Department, SRM University, Kattankulathur, Tamilnadu, India
e-mail: dhanalakshmi.s@ktr.srmuniv.ac.in; sdhanalakshmi2004@gmail.com

Epoch is the instant of significant excitation of vocal tract system within the glottal pulse. The location of epochs has many applications like accurate determination of fundamental frequency, determination of characteristics of voice sources useful in text to speech synthesis, speech enhancement [3] and extracting speaker specific information by using the excitation features around the epochs.

Another important parameter involved in speech processing is pitch, the rate of vibration of vocal folds. The quasi-periodic property of speech makes the pitch estimation task challenging. The area of the pitch estimation has attracted a lot of researchers resulting in number of algorithms that include both time domain and frequency domain techniques. However, in the case of noisy speech, the performance degradation is significantly high [4–6].

In this chapter, the features are calculated by the technique called empirical mode decomposition (EMD). It will decompose the speech signal into several intrinsic mode factors (IMF). EMD is introduced in combination with the Hilbert–Huang transform (HHT) [7], and it will analyze nonlinear and non-stationary time series.

This chapter is organized as follows. Section 2 elaborates about the review of the epoch extraction methods. Section 3 discusses the different algorithms for epoch extraction and pitch detection. Section 4 focuses on comparative results of feature extraction with recent methods in literature. Section 5 concludes the chapter.

2 Review of Epoch Extraction Methods

Most epoch extraction methods rely on the error signal derived from the speech waveform after removing the predictable portion (second-order correlations). In this section, we review the existing algorithms for pitch estimation and epoch extraction. Reliable pitch estimation is difficult for the noise-corrupted speech. In [8], Feng Huang et al. proposed a method for pitch estimation based on accumulation of spectral peaks over consecutive time frames. Popular time domain technique for pitch estimation uses SIFT algorithm based on LPC analysis and ACF (autocorrelation) function [9, 10]. Several frequency domain techniques exist of which Harmonic product spectrum derived from DFT [11, 12] is claimed to be a robust algorithm [13, 14].

In this chapter, we propose a technique which derives HPS from DCT. This technique exploits for better decorrelating property of DCT than DFT which guarantees pitch estimation at lower order, thereby reducing the computation cost.

Another approach for epoch extraction techniques depends on linear prediction (LP) analysis or exploits the periodicity property of the signal in the adjacent cycles. These algorithms are tested on a clean speech, and their robustness in presence of additive noise is studied. In this chapter, we exploit the impulse-like characteristics of excitation. The analysis is confined to a single frequency in our technique. The recent method of EMD of HHT is used to decompose the mixed signal into specific physical source rather than other frequency transformation.

3 Methodology of Epoch Extraction

Several methods are for estimating the instants of glottal closure from speech signal. These methods includes

1. Epoch extraction using zero-frequency resonators
2. Pitch detection using DCT-HPS
3. Epoch extraction based on Hilbert–Huang transform.

It should be noted that the methods placed in one category could also belong to another, depending on the interpretation of the method.

3.1 Zero-Frequency Resonators

For extraction of epoch from the given speech signal, initial step is to design a zero-frequency resonator. The resonator is designed on the fact that its output will have the information on the discontinuities due to impulse-like excitation. A double-filtering process is adopted to reduce the effects of all high pass resonances. The two cascade resonators are provided with sharper roll-off so that the large values of filtered output will override the impulse sequence discontinuities.

The characteristics of discontinuities can be extracted using the deviation of local mean from the filtered output. This can be done by subtracting the local mean taken for every 10 ms from the filtered output which we call as zero frequency-filtered signal. The following steps are to be followed

1. To remove the time-varying low-frequency bias, the speech signal s(t) is differenced.
2. $x[n]$ is then passed through zero-frequency resonator twice.
3. The trend in $y_2[n]$ can be removed by subtracting the average over 10 ms at each sample resulting in zero frequency-filtered signal.

The filtered signal clearly shows rapid changes around the positive zero crossing instants which can be taken as epochs.

Figure 1 shows the flowchart for epoch estimation using zero-frequency resonators. Successive integration process results in the extraction of the epochs. Differencing and averaging is performed to improve the accuracy.

3.2 Pitch Detection Using DCT-HPS

In this method, pitch detection is done using the harmonic product spectrum derived from the DCT. Figure 2 shows the flowchart for detecting the pitch using DCT-HPS for a given speech data. The proposed pitch detection process is of threefold—pre-processing, DCT and HPS extraction from DCT.

Fig. 1 Epoch estimation
using 0-Hz resonator
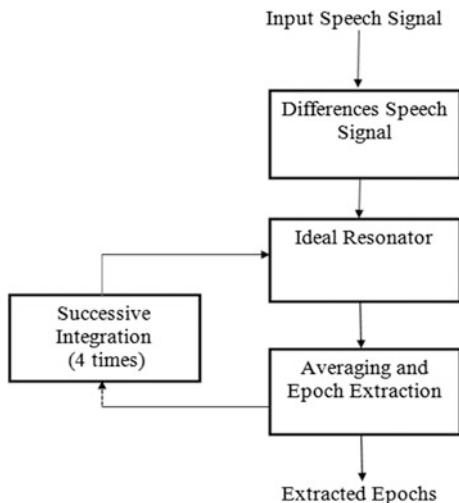
Fig. 2 Flowchart for pitch
detection using DCT-HPS
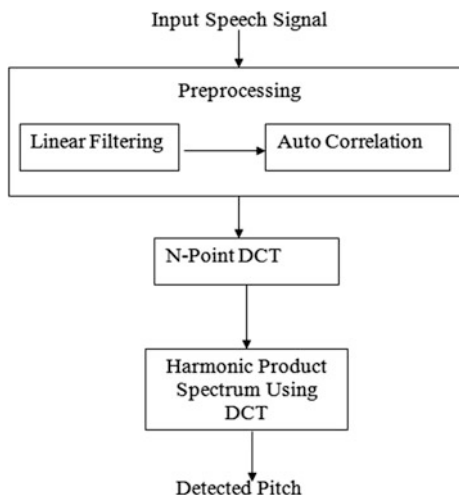
The first step of pitch detection process involves pre-processing the input speech signal. Experimental results show that pitch estimated without any pre-processing techniques yields 65% accuracy at high SNRs and drastically reduces for the techniques involving pre-processing. Analysis is done by pre-processing the noise-corrupted data using techniques such as linear filtering, autocorrelation, linear filtering followed by autocorrelation. It is found that linear filtering and then determining autocorrelation of input speech signal perform better compared to other stated pre-processing techniques. The proposed method of deriving HPS from DCT performs better than the traditional method of using DFT because of the fact that the former decorrelates the signal well causing the peaks corresponding to fundamental

frequency and higher harmonics sharper than the later. Also the required number of compressed spectra otherwise called as the order using DCT-HPS is observed to be less than DFT-HPS which guarantees reduced computational cost.

3.3 Epoch Extraction Using Hilbert–Huang Transform

The main objective of EMD is to shift the original non-stationary signal into stationary signal. With this, Huang used the EMD to decompose the multicomponent signal into monocomponent signal, which are referred to as intrinsic mode functions (IMFs). The EMD is for developing the signal into several IMFs [15]. Once the features are obtained, the instantaneous frequency of each signal will be determined. The instantaneous frequencies are symmetric with respect to local zero and zero crossing rate. The conditions are as follows:

1. In whole series of data, the number of zero crossing and number of local extremes must either be equal or differ at most by one.
2. At any point of signal, the mean value of local minima and mean value of local maxima are zero.

According to these conditions, an overview of EMD can be described in following steps. Initially we find out that all local extrema connect all values with lower and upper values, covering the entire signal between them. Then difference is computed as $h(t)$. Consider, $h(t)$ is a first component and temporary signal, the process will continue until the condition satisfies. Thus, the first IMF is obtained [16]. Find the residue by subtracting the first IMF and original signal. We can then find out remaining IMF by following the above-mentioned steps. However, satisfying the second condition of IMF in practical applications is almost impossible. Hence, the threshold value is utilized for this case. The index can be calculated as follows:

$$SD_{ik} = \frac{\sum_{t=0}^{T} \left| h_{i(k-1)}(t) - h_{i(k)}(t) \right|^2}{\sum_{t=0}^{T} h_{i(k-1)}^2(t)} \tag{1}$$

where, $h_{i(k)}$ is kth iteration of ith IMF. It noted, the numerator is equal to mean $m_{i(k)}$ i.e., $h_{i(k)}^{(t)} = h_{i(k-1)}^{(t)} - m_{i(k)}^{(t)}$. This means the ratio of $m_{i(k)}^{(t)}$ to $h_{i(k-1)}^{(t)}$ is SD_{ik}. The second SD_{ik} condition should be smaller than the assigned threshold.

In general, the threshold is assigned in range as 0.2–0.3 [15]. If $SD_{ik} < 0.3$, the first condition IMF is calculate and then iteration for ith IMF stops, then we get new IMF.

The procedure for EMD of signal is explained as follows. During the process of EMD, the cubic spline is used to generate upper and lower envelopes. Let we take original signal $X(t)$ and temporary as $Temp(t) = X(t)$.

Step 1: Find the lower and upper envelope $L(t)$ and $U(t)$, respectively. Calculate mean $m(t) = [U(t) + L(t)]/2$. Component $h(t)$ is calculated by using the formula

$$h(t) = \text{Temp}(t) - m(t) \tag{2}$$

Step 2: Check if the value $h(t)$ satisfies condition. If it satisfies, $\text{imf}_1(t) = h(t)$ and proceed to next step or else assume $h(t) = \text{Temp}(t)$ and go back to step 1.

Step 3: Calculate residue $r_1(t)$ by using the formula

$$r_1(t) = \text{Temp}(t) - \text{imf}_1(t) \tag{3}$$

Assign $r_1(t) = X(t)$ and repeat steps 1 and 2 to get second IMF.

Step 4: Repeat Step 3 for remaining IMF.

$$r_n(t) = r_{n-1}(t) - \text{imf}_n(t) \quad n = 2, 3, 4, \ldots \tag{4}$$

This will be completed when $r_n(t)$ is a monotone function. After EMD procedure, $X(t)$ is obtained as

$$X(t) = \sum_{i=1}^{n} \text{imf}_i(t) + r_n(t) \tag{5}$$

4 Experimental Results

Experiments are performed using Matlab R2012a. The input noisy and the noiseless speech signals are taken from NOISEX-92 database. In Figs. 3 and 4 (a) is the input speech signal, (b) obtained differenced signal after removing the low-frequency bias
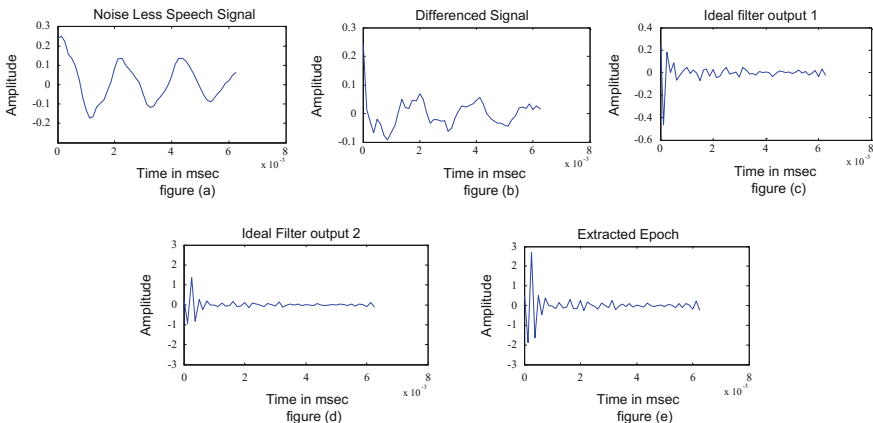
Fig. 3 Segmented noise less speech signal
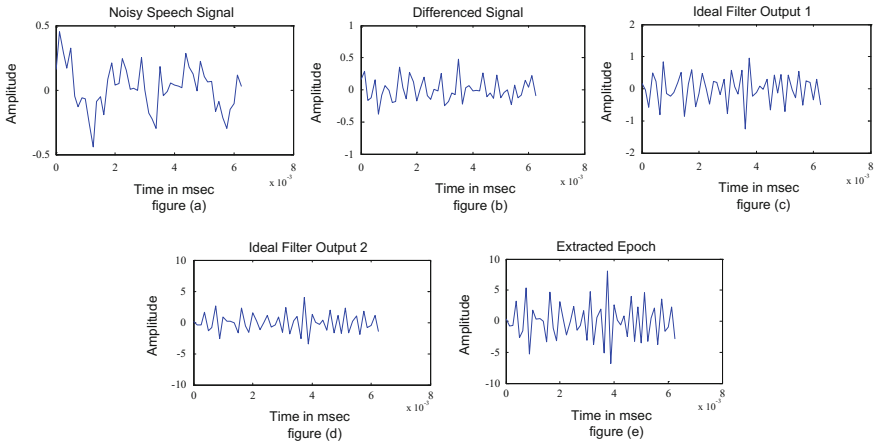
Fig. 4 Segmented noisy speech signal
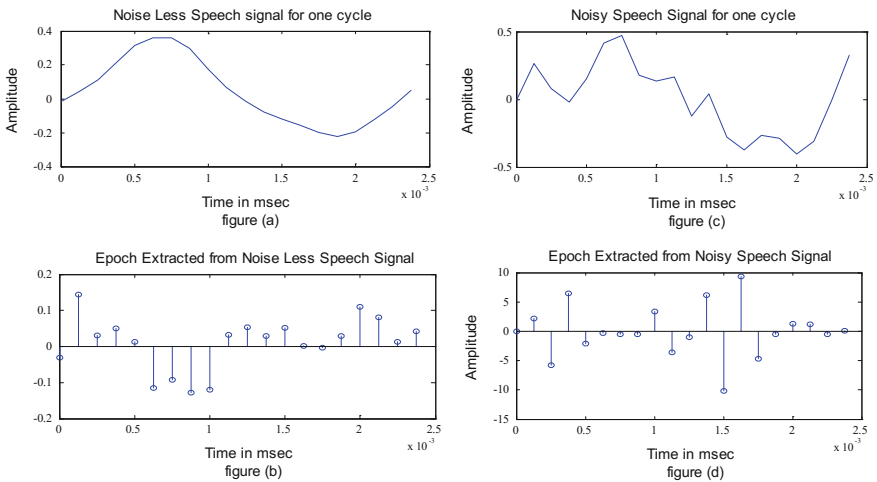
Fig. 5 Comparison of epoch extracted from noise less and noisy speech signal

in the input, (c) and (d) the resonator filter output with high frequency amplitude being integrated and low frequency amplitude being suppressed and (e) epoch extracted from the given input. On comparing the epoch extracted for a noisy and noiseless signal, we observe from Fig. 5 that for a one larynx cycle, ZFF results are found to be same in terms of identification rate (IDR), miss rate (MR), false alarm rate (FAR) and differ in terms of amplitude factor (AF) as shown.

We use the same speech segment for pitch estimation. Figures 6 and 7 show the outputs obtained for pitch detection using DFT-HPS and DCT-HPS techniques. (a) The input speech signal, (b) low pass filtered output, (c) autocorrelated filtered
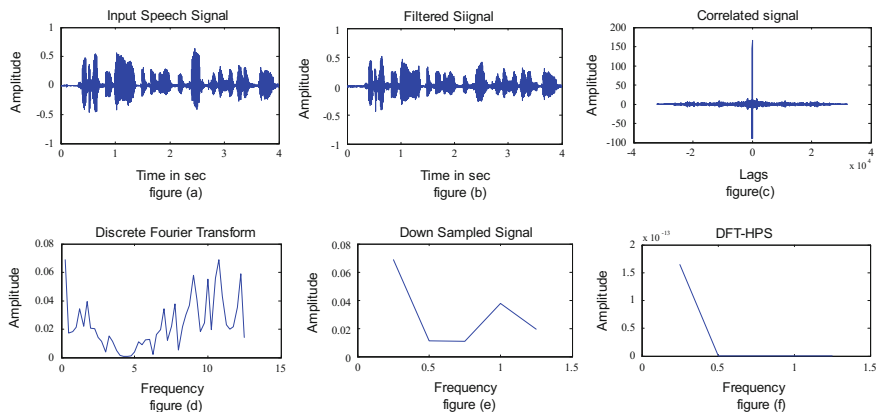
Fig. 6 Outputs obtained using DFT-HPS
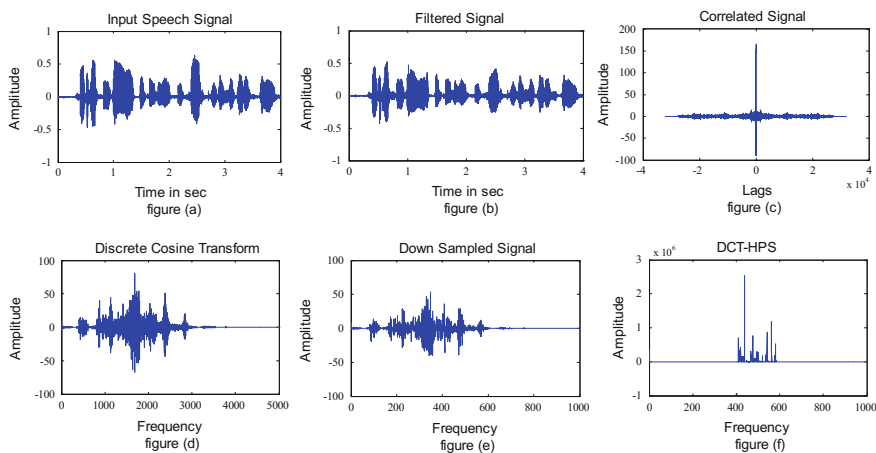
Fig. 7 Outputs obtained using DCT-HPS

output for the removal of noise and periodicity maintenance, (d) transform outputs (DFT/DCT), (e) down sampled transform output and (f) HPS derived from the respective transforms.

In Fig. 8, we see the comparison made between the HPS derived using DFT and DCT. For the given speech signal, the pitch detected using the DCT-HPS is accurate which is evident from the presence of narrow peaks. Here, the peak corresponding to fundamental frequency is the pitch which is not present in output obtained using DFT-HPS. Figure 9 shows the empirical mode decomposition of the speech signal, in which a single speech signal is decomposed into eight intrinsic mode functions. It is easy and more accurate to find the pitch from the different intrinsic modes of speech signal rather the speech signal itself.
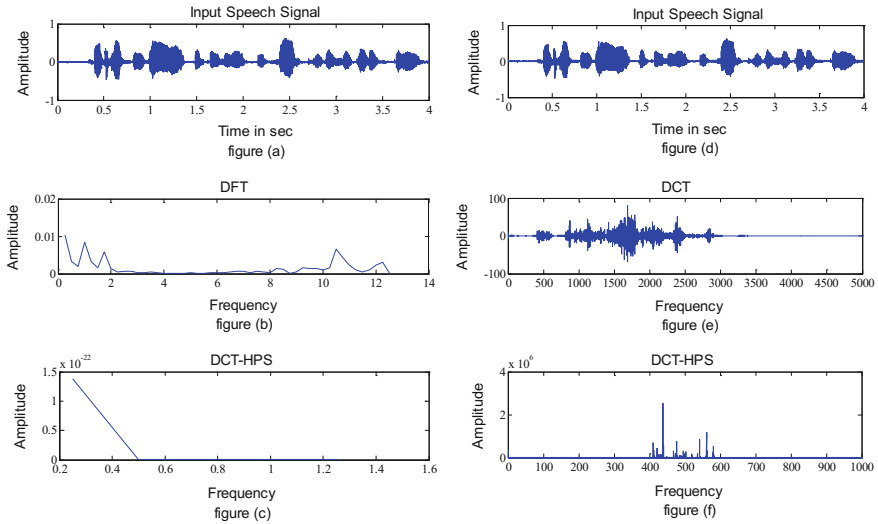
Fig. 8 Comparison between DFT-HPS and DCT-HPS
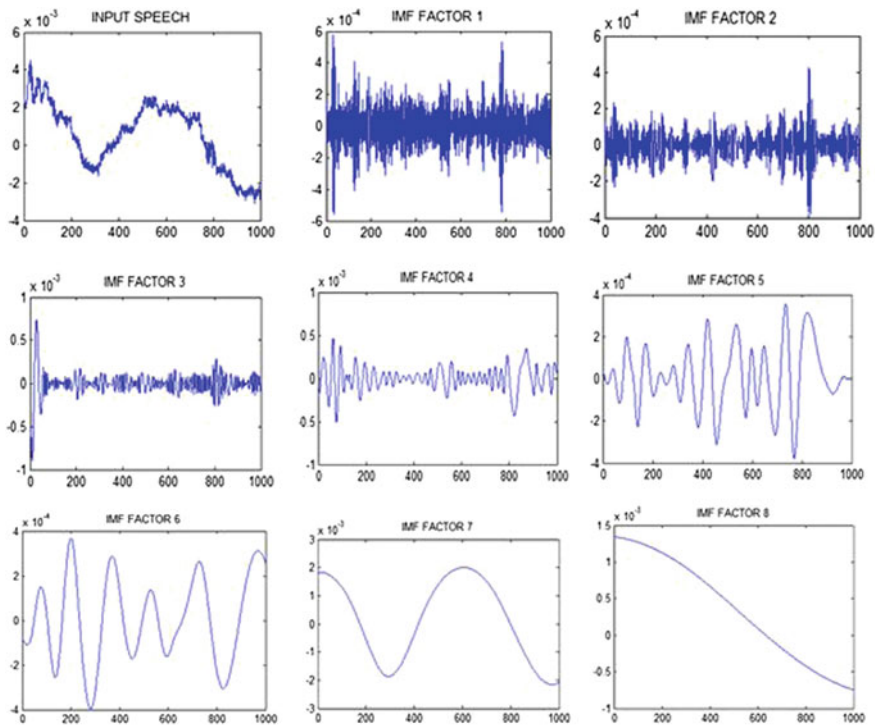
Fig. 9 Empirical mode decomposition of speech signal

5 Conclusion

In this chapter, we proposed an efficient algorithm for epoch extraction and pitch detection. The epoch extraction method is independent of characteristics of vocal tract system and uses a zero-frequency resonator. The proposed method is tested for both noisy and noiseless segmented speech signals, and its performance is found to be same except for the amplitude factor as discussed in results. This assures that the proposed method performs satisfactorily even in noisy environment. The proposed Hilbert–Huang transform-based technique is observed to give accurate pitch estimation because of its better decorrelating property. On comparing with state-of-the-art DFT and DCT-HPS-based technique, our proposal gives accurate pitch estimation, and it is also robust to noise.

References

1. Kounoudes A, Naylor PA, Brookes M (2012) The DYPSA algorithm for estimation of glottal closure instants in voiced speech. In: IEEE conference on acoustic speech signal processing, Vol 1, pp 349–352
2. Ananthapadmanabha TV, Ramakrishnan AG (2011) Epoch extraction based on integrated linear prediction residual using plosion index. IEEE Trans Audio Speech Lang Process 21 (12):2471–2480
3. Yegnanarayana B, Murty PS (2010) Extraction of vocal-tract system characteristics from speech signals. IEEE Trans Speech Audio Process 8(3):267–281
4. Yegnanarayana B, Gangashetty SV (2011) Epoch-based analysis of speech signals. In: IEEE conference on Indian academy of sciences, Vol 36, Part 5, pp 651–697
5. Hu G, Wang D (2010) A tandem algorithm for pitch estimation and voiced speech segregation. IEEE Trans Audio Speech Lang Process 18(8):2067–2079
6. Höge H, Siemens AG (2011) Basic parameters in speech processing the need for evaluation. In: Corporate Technology, IEEE Transactions on Audio, Speech and Language Processing USA
7. Molla MKI, Hirose K (2007) Single-mixture audio source separation by subspace decomposition of Hilbert spectrum. IEEE Trans Acoust Speech Signal Process 15(3):893–900
8. Huang F, Lee T (2013) Pitch estimation in noisy speech using accumulated peak spectrum and sparse estimation technique. IEEE Trans Audio Speech Lang Process 21(1):99–109
9. Markel J (2010) The sift algorithm for fundamental frequency estimation. IEEE Trans Audio Electroacoust AU-20:367–377
10. Rao KR, Yip P (1990) Discrete cosine transform: algorithms, advantages, applications. Academic Press Professional Inc, San Diego, CA, USA
11. de la Cuadra P, Master A, Sapp C (2001) Efficient pitch detection techniques for interactive music. In: International Computer Music Conference, Havana
12. Drugman T et al (2011) Detection of glottal closure instants from speech signals: a quantitative review. IEEE Trans Audio Speech Lang Process 20(3):994–1006
13. Rao KS, Yegnanarayana B (2006) Prosody modification using instants of significant excitation. IEEE Trans Audio Speech Lang Process 14(3):972–980
14. Rabiner LR et al (1976) A comparative performance study of several pitch detection algorithms. IEEE Trans Acoust Speech Signal Process ASSP-24(5):399–417

15. Nagarajan T, Sripriya N (2013) Pitch estimation using harmonic product spectrum derived from DFT. In: Audio speech and language processing, TENCON2013 - 2013 IEEE Region 10 Conference (31194)
16. Shimamura T, Kobayashi H (2001) Weighted autocorrelation for pitch extraction of noisy speech. IEEE Trans Speech Audio Process 9(7):727–730

Hilbert–Huang Transform and Its Variants in Engineering Data Analytics: State of the Art and Research Challenges

Mariselvam Ammasi Krishnan and Dhanalakshmi Samiappan

Abstract Data analysis is the elemental pursuits in scientific modern research, but the accessibility of inspection techniques downgrades it to data processing. Hilbert–Huang transform is perhaps the most notable development during last decade. HHT is physically based on the development of instantaneous frequency whose idea is relevant to non-stationary and nonlinear signals. Complexification of the signal by means of the Hilbert transform to portray the signal as far as the regulated amplitude and the related instantaneous frequencies that seem to speak to both entomb waves and intrawaves. Hilbert–Huang transform is used in the field of biomedicine, chemistry, financial, ocean engineering, speech processing, astro-particle physics, detection and localization of damage, error detection in analog and mixed-signal circuits, analysis of the landslides, temperature, environmental time series, and masking signals. But traditional EMD method has scope for improvement in mode-mixing problem and illusive IMF components. Variants of HHT are used to avoid these problems.

Keywords Empirical mode decomposition (EMD)
Intrinsic mode function (IMF) · Ensemble EMD
Hilbert–Huang transform (HHT) · Orthognal HHT · Improved HHT

1 Hilbert–Huang Transform

1.1 Theoretical Background

The HHT is an EMD-based data analysis method. It can deliver significant representations of data from nonlinear and non-stationary processes. Most of the

M.A. Krishnan (✉) · D. Samiappan
ECE Department, SRM University, Kattankulathur, Tamilnadu, India
e-mail: mariselvamak@gmail.com

D. Samiappan
e-mail: sdhanalakshmi2004@gmail.com

information investigation techniques depend on linear and stationary presumptions. Recently, new methods have been instigated to examine non-stationary and non-linear data. For example, wavelet analysis and the Wagner–Ville distribution were drawn plans of linear but non-stationary data. Furthermore, miscellaneous nonlinear time arrangement investigation mechanism was delineated for nonlinear but stationary and deterministic frameworks. Lamentably, in most legitimate frameworks, either common or even human-made ones, the information is well on the way to be both nonlinear and non-stationary [1, 26].

For analyzing such nonlinear and non-stationary systems, Huang in 1998 designed the HHT which consists couple of steps: Empirical Mode Decomposition (EMD) and Hilbert spectral analysis (HSA). This process is viable for time–frequency–energy representations [1].

1.2   The Empirical Mode Decomposition Method (Sifting Process)

Each IMF, linear or nonlinear, represents a simple oscillation, which will have the similar number or difference at most one between extrema and zero-crossings. Moreover, the oscillation will likewise be proportional concerning the mean esteem. At any given time, the data might have diverse existing together modes of oscillation, one superimposing on the others. The result is the last entangled information (Fig. 1).
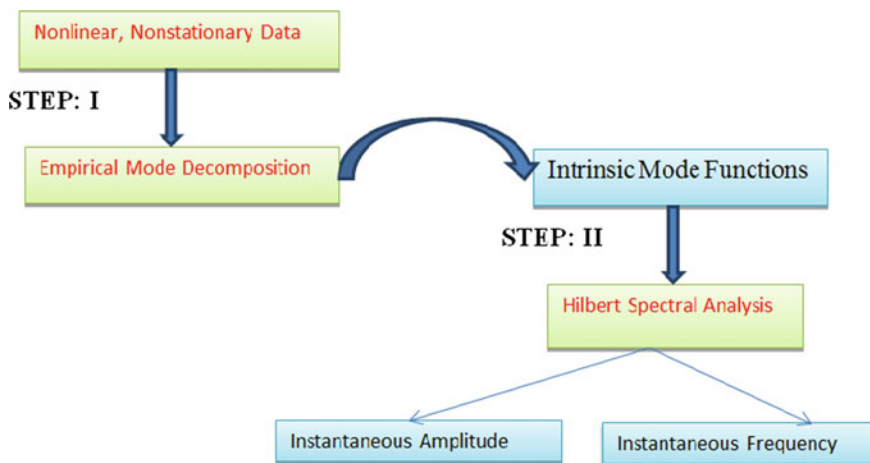
Fig. 1 Steps followed in HHT

The distinction between the data and m_1 is the first component, and h_1 is given by

$$h_1 = x(t) - m_1. \tag{1}$$

In the successive shifting process, h_1 can be treated only as a proto-IMF in the last step, and then, h_1 is considered as data following the previous step.

$$h_{11} = h_1 - m_{11}. \tag{2}$$

h_{1k} becomes an IMF; then,

$$h_{1k} = h_{1(k-1)} - m_{1k}. \tag{3}$$

$$c_1 = h_{1k}. \tag{4}$$

c_1 ought to contain the finest scale or the most limited period segment of the signal. It takes after that c_1 can be isolated from whatever is left of the data by

$$r_1 = x(t) - c_1. \tag{5}$$

Summing up the above equation

$$x(t) = \sum_{j=1}^{n} c_j + r_n. \tag{6}$$

The criterion to stop the sifting of IMFs, there are two steps: (1) the normalized squared difference between two progressive sifting operations is characterized as

$$SD_K = \frac{\sum_{t=0}^{T} |h_{k-1}(t) - h_k(t)|^2}{\sum_{t=0}^{T} h_{k-1}^2}. \tag{7}$$

If this squared difference is not exactly a preset esteem, the sifting process will stop (2). The sifting process will stop simply after S back-to-back times, when the quantities of zero intersections and extrema remain the same and are equivalent or vary at most by one. The number S is preset.

Hilbert transform

$$H[x(t)] = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau. \tag{8}$$

PV—the singular integral's principal value. The analytic signal of the Hilbert transform is defined as

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)}. \qquad (9)$$

where

$$a(t) = \sqrt{x^2 + y^2}. \qquad (10)$$

$$\theta(t) = \arctan\left(\frac{y}{x}\right). \qquad (11)$$

$$\omega = \frac{d\theta}{dt}. \qquad (12)$$

The basis of Fourier, Wavelet, and Hilbert transform is given as a priori for Fourier transform and Wavelet transform and is adaptive for Hilbert transform. While considering the frequency in Fourier transform, the convolution is global uncertainty and the same convolution is applied for Wavelet comprising of regional uncertainty, whereas in Hilbert transform we use differentiation of local certainty. The presentation is energy–frequency in Fourier, whereas in Wavelet and Hilbert transforms, it is energy–time–frequency. Nonlinearity, feature extraction, and non-stationary do not exist for FT and WT but exist for HHT (Fig. 2).

Comparison of Fourier, Wavelet, and Hilbert–Huang transforms is show below in Table 1.
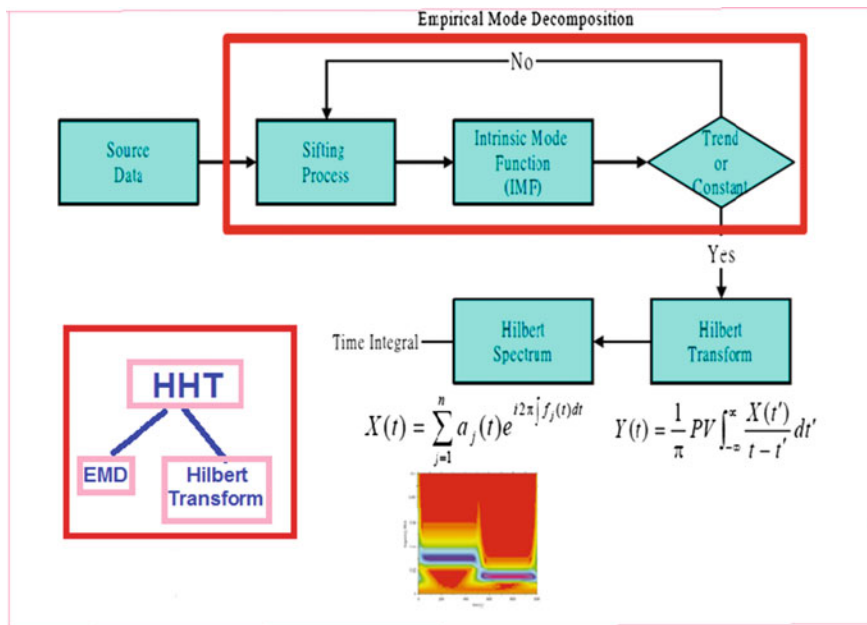
Fig. 2 Block diagram of HHT

Table 1 Comparison of Fourier, wavelet, and Hilbert–Huang transforms [1, 2]

	Fourier	Wavelet	HHT
Basis	A priori	A priori	Adaptive
Frequency	Convolution: global uncertainty	Convolution: regional uncertainty	Differentiation: local, certainty
Presentation	Energy frequency	Energy–time–frequency	Energy–time–frequency
Nonlinear	No	No	Yes
Non-stationary	No	No	Yes
Feature extraction	No	Discrete: no; continuous yes	Yes
Theoretical base	Theory complete	Theory complete	Empirical

2 HHT Variants

2.1 Variant 1 (Ensemble EMD)

In traditional EMD method mixing of signal modes fluctuates the processing of the signals. Mode mixing sometimes fails to restore the original signal [4, 5]. Short-time Fourier transform (STFT), Winger–Ville distribution, and Wavelet transform (WT) are the strategies which lack adaptivity. To overcome the mode-mixing problems in traditional EMD method, ensemble empirical mode decomposition (EEMD) can be presented in the signal, by which the integrity of signal's mapping in the different regions can be observed. To accomplish the EEMD process, it must be fed with white noise to the original noise.

The steps of ensemble EMD methods are (1) adding white noise with the unique signal and (2) decomposing the data (white noise and original signal).

The message signal is used in EMD, and along with the message signal, white noise is included to the original message signal in EEMD. In both the decompositions, 3 IMFs and 1 residue and 6 IMFs and 1 residue are used, respectively. The IMF decrement is not in order in EMD, whereas in EEMD, the IMF decrement is in appropriate order. The accuracy in EMD is comparatively less than EEMD. Hilbert spectrum analysis and marginal spectrum harmonic components and features are less in EMD when compared to EEMD.

The following table shows the comparison of EMD and EEMD which is stated above.

From Table 2, it concludes that EEMD method is best for signal processing.

Table 2 Comparison of EMD and EEMD [3]

Characteristics	EMD	EEMD
Signals	Message signals	Message signals and white noise
Decompositions	3 IMFs and 1 residue	6 IMFs and 1 residue
IMF decrement	Not in order	Proper order
Accuracy	Comparatively less	More
Hilbert spectrum analysis	Less	Decomposition is more explicit, clear, and regular
Marginal spectrum harmonic components and features	Less	More

2.2 Variant 2 (Orthogonal Hilbert–Huang Transform (OHHT))

While empirical mode decomposition (EMD) is used for earthquake motion energy changing, EMD cannot conserve energy and the end effects are spine interpolation. To avoid energy leakage, orthogonal Hilbert–Huang transform (OHHT) is used in earthquake time history process.

OHHT

By using EMD to message signal, initial n IMF components $c'_i(t)\{i = 1, 2, 3, \ldots n\}$ and final residue $r_n(t)$. Final IMF component can be used as initial component. So $c_1(t)$ can be considered as $c_n(t)$. $c_1(t)$ is called as initial step orthogonal IMF component of message signal. In order to obtain the next orthogonal IMF component of message signal $x(t)$, entire $c_1(t)$ components must remove from c_{n-1}

$$C_2(t) = c_{n-1}(t) - \beta_{21}c_1(t). \tag{13}$$

β_{21} is called as the orthogonal coefficient and $c_2(t)$ second orthogonal component. To build the orthogonality between first and second components and to get β_{21}, multiply $c_1(t)$ to both side and make time integral in above-mentioned equation (Table 3).

Table 3 Comparison of HHT and orthogonal HHT [6]

	HHT	Orthogonal HHT
Orthogonal coefficient	–	$\beta_{21} = \frac{\int_0^T c_{n-1}(t)c_1(t)\mathrm{d}t}{\int_0^T c_1^2\mathrm{d}t}$
Message signal	$x(t) = \sum_{j=1}^n c_j + r_n$	$x(t) = \sum_{j=1}^n c_j^* + r_n(t) = \sum_{j=1}^n a_j^* c_j(t) + r_n(t)$

Table 4 Comparison of EMD and IHHT [8]

Characteristics	EMD	EMD–DMS (IHHT)
Input signals	$x(t)$	$x_+(t) = x(t) + s_{mj}(t)$ $X_-(t) = x(t) - s_{mj}(t)$
To obtain IMFs	Explain in part I	Explain in part I
IMF	$c_1(t)$	$C_j(t) = (c_+(t)+c_-(t))/2$
Residue	Explain in part I	Explain in part I
Sifting process	EMD sifting process	Fm, 1 is a key parameter (sampling theorem $f_m <= f_s/2$)
Basis	Adaptive	Less adaptive (non-stationary signal)

2.3 Variant 3 (Improved Hilbert–Huang Transform (IHHT))

EMD cannot disparate the ingredients whose frequencies lie inside the identical octave [7]. To revamp the performance of the EMD, masking signals are used.

Masking signals are utilized to take care of the mode-mixing issues (Table 4).
EMD–DMS algorithm
The masking signal can be composed as

$$S_m(t) = a_m \cos 2\pi f_m t. \tag{14}$$

The masking group signals are signified as $S_{m1j}(t)$. It is denoted as dyadic masking signals if

$$f_{m,j+1} = f_{m,j}/2. \tag{15}$$

$f_m,$ $_j$ represents the frequency of the masking signals.

The additive white Gaussian noise in EMD is much narrower and has more overlaps compared to EMD–DMS. The real-world data collected from a compressor are split down into IMF 3 and IMF 4 subbands in EMD, and in EMD–DMS, IMF 5 is subbanded, which is less in EMD and more in EMD–DMS.

Comparison of EMD and EMD–DMS can be shown below (Table 5)

The much more effective way to enhance the EMD performance is by using dyadic masking signal. It should be taken into consideration that EMD–DMS is less adaptive than EMD, and in this aspect, there is a scope for further improvement.

Table 5 Comparison of EMD and EMD–DMS [8]

S. no	I/P data	EMD	EMD–DMS
1	Additive white Gaussian noise	Comparatively IMF 1 not much narrower	IMF 1 is much narrower
		More over lap in IMF	Less overlap in IMF
2	Real-world data collected from a compressor	It is broken up into IMF 3 and IMF 4 subbands	Main frequency component data transformed into IMF 5 subband
		Less IMF	More IMF
3	Additive white Gaussian noise and real-world data collected from a compressor	It is broken up into IMF 3 and IMF 4 subbands	Main frequency component transformed into IMF 5 subband

3 Applications of Hilbert–Huang Transform

3.1 Case Study 1 (Analysis of Temperature Time Series)

Global warming is a very pivotal area, and it leads the human community into empty. Global warming is getting maximum conversion because in the recent centuries the global average temperature has been increased from 0.56 to 0.92 °C [9]. Zhao considered the fangzian area which is situated in the south of Shanghai to make understanding in the temperature variation and temperature increment. Tang et al revealed that the climate warning in China ranges from 0.2 to 0.8 °C in the recent 100 years. The industries in rural side are considerably less compared to those of the urban side.

The multiscaling structures of the wavering made the investigation of the temperature time arrangement inappropriate. Multiscaling structures based on the EMD can be applied for water wave investigation [10], earthquake wave examination, and climate studies [11]. Information gathered from Shanghai meteorological administration in the fengxian district has monthly mean temperature variation from 1959 to 2012. The month-to-month mean information is gotten from the midpoints of day by day 2, 8, 14, 20 o'clock information, and days in a month. The mean is taken, and the month-to-month normal temperature (1959–2012) was recorded by the Shanghai observatory (648 months).

After applying HHT on monthly average temperature, it generated the IMFs. The IMF1 demonstrates the high recurrence wavering, which will depict fengzian temperature fluctuation in about a time period of 12 months. IMF2 is higher variation in the average timescale. IMF 6 is change in trend [12].

3.2 Case Study 2 (Detection and Localization of Damage in Aluminum Plates)

Structure damage may bring about a lot of support and well-being issues. In order to avoid such problems, many methods have been followed.

The demerits of Lamb wave experiments are confounded due to the multimodal nature and dispersive and then finding the TOF is troublesome.

The merits of single-mode guided waves are simple to analyze relatively, and it travels long distance with little dispersion.

Wavelet transform is useful in identifying the difference between the references and damage states, and it suffers in non-adaptive.

Artificial neural networks are helpful to identify the location of damages [13, 14], the drawback is to find damage, and localization requires large number of data.

The Lamb wave is convoluted due to their multimodal nature and dispersive then finding the TOF is difficult. A single-mode guided wave is simple to analyze relatively and travel a long distance with little dispersion.

The Wavelet transform has a mode shape contrast among references and harm states, but suffers in non-adaptive and Wavelet transform. Artificial neural networks help to identify the location of damages but, to find damage and localization, requires large number of data (Table 6).

Utilization of the Hilbert transform with EMD is the best approach to find out damage and location.

Lamb wave—made up of a superposition for longitudinal what's more shear modes are made available in thin plates and their propagation characteristics vary with section point, excitation, and basic geometry. A Lamb mode can be either symmetric or antisymmetric.

Ultra callous transducer transmits a pulse under those structures which is to be integrated. Reflected wave is detected by the receiver (aluminum plates). Lamb

Table 6 The merits and demerits of experiments

S. no	Experiments	Merits	Demerits
1	Lamb wave	–	Entangled due to their dispersive and multimodal nature, finding the TOF is troublesome
2	Single-mode guided wave	Simple to analyze relatively, travel long distance with little dispersion	–
3	Wavelet transform	Mode shape contrast among references and harm states	It suffers in non-adaptive and wavelet transform
4	Artificial neural networks	To identify the location of damages [13, 14]	To find damage and localization requires large number of data

Table 7 Time domain and frequency domain functions

Time domain	Frequency domain
Energy distributions $E_k = \sum_{k-L+1}^{k} (x[m])^2$	Figure of merit $\mathrm{DI_{FoM}} = \mathrm{Max}\left[\sum_{i=1}^{L}\left(\frac{\log(X(i))}{\log(Y(i))} - 1\right)^2\right]$
Root mean square deviation (RMSD) $\mathrm{DI_{RMSD}} = \sqrt{\frac{\sum_{i=1}^{L}(y(i)-x(i))^2}{\sum_{i=1}^{L}x(i)^2}}$	Spectral energy $E = \sum_{n=1}^{L}|X(k)|^2$ $\mathrm{DI_{spec}} = E_{\mathrm{NF}} - E_{\mathrm{f}}$
Correlation $\mathrm{DI_{corr}} = 1 - \dfrac{\sum_{i=1}^{L}(x_i-\mu_x)(y_i-\mu_y)}{\sqrt{\sum_{i=1}^{L}(x_i-\mu_x)^2}\cdot\sqrt{\sum_{i=1}^{L}(y_i-\mu_y)^2}}$	

wave based on damage identifications is used. PWAS (Piezoelectric wafer-active sensors) disks are fitted with aluminum plates.

The damage in plate is a diffracting substance which will create distinctive proliferation modes and different velocities. The parameter which changes to different degrees in signals is caught prior, and then, afterward, the nearness of harm in the structure is called as damage index (DI).

To develop DIs in time and frequency domain, equations are given below

The energy distribution in time domain is $\sum_{k-L+1}^{k} (x[m])^2$, while the figure of merit in frequency domain is $\mathrm{DI_{FoM}} = \mathrm{Max}\left[\sum_{i=1}^{L}\left(\frac{\log(X(i))}{\log(Y(i))} - 1\right)^2\right]$. The root mean square deviation (RMSD) is $\mathrm{DI_{RMSD}} = \sqrt{\frac{\sum_{i=i}^{L}(y(i)-x(i))^2}{\sum_{i=1}^{L}x(i)^2}}$, while the spectral energy in frequency domain is $E = \sum_{n=1}^{L}|X(k)|^2$ and $\mathrm{DI_{spec}} = E_{\mathrm{NF}} - E_{\mathrm{f}}$ (Table 7).

The correlation in time domain is

$$DI_{corr} = 1 - \frac{\sum_{i=1}^{L}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{L}(x_i - \mu_x)^2}\cdot\sqrt{\sum_{i=1}^{L}(y_i - \mu_y)^2}}$$

HHT is used to identify modal frequency in plate structure and fault diagnosis [15]. Mathematical expression [16]

$$S(t) = S_{\mathrm{d}}(t) + S_{\mathrm{br}}(t) + S_{\mathrm{dr}}(t). \tag{16}$$

$$S_{\mathrm{d}}(t) = A_1 \sin(\omega_1 t + \theta_1). \tag{17}$$

$$S_{\mathrm{br}}(t) = A_2 \sin(\omega_2 t + \theta_2) + A_3 \sin(\omega_3 t + \theta_3) \tag{18}$$

$$S_{dr}(t) = A_4 \sin(\omega_4 t + \theta_4) + A_3 \sin(\omega_5 t + \theta_4) \qquad (19)$$

$S(t)$—received signal; $S_d(t)$—direct transmitted signal; $S_{br}(t)$—reflection from auxiliary limits; $S_{dr}(t)$—reflections from defected place.

Defect localization

By using Cartesian coordinates, time to echo from damage is

$$t_{ij} = t_{off} + d_{ij}/v. \qquad (20)$$

t_{ij}, time of entry of reverberate at receiver i when transmitter j is energized; t_{off}, offset time; v, the rate at which an object covers distance.

$$d_{ij} = (d_{ri} + d_{tj}). \qquad (21)$$

$$d_{ri}\sqrt{(x_{ir} - x)^2 + (y_{ir} - y)^2}. \qquad (22)$$

$$d_{tj}\sqrt{(x - x_{jt})^2 + (y - y_{jt})^2}. \qquad (23)$$

A factual classifier in view of components removed from received signal is utilized to recognize the likely harmed structures.

3.3 Case Study 3 (Environmental Time Series)

Geosciences, particularly in the field of marine environment, and time series are nonlinear and non-stationary. The Wavelet transform is utilized for time–frequency examination method to manage non-stationery signals. In addition, Fourier transform also selects one set of data among the given stack of data and predicts the parameters of the same. Picking the wrong premise work enormously increases the quantity of terms required for fitting into the time arrangement.

Wavelet transform is acceptable, and sinusoidal function and non-acceptance function are more complex and orthogonal waveform.

The comparison of FT and WT is given in an info graphic format.

The acceptable function in Fourier transform is a sinusoidal basis function, while the non-acceptable function is a non-sinusoidal function. The acceptable function for Wavelet transform is a sinusoidal function, while the non-acceptable function is more complex and orthogonal waveform (Table 8).

Table 8 Comparison of FT and WT

S. no	Transform	Acceptable function	Non-acceptable function
1	Fourier transform (FT)	Sinusoidal basis function	Non-sinusoidal function
2	Wavelet transform (WT)	Sinusoidal function	More complex and orthogonal waveform

EMD is an adaptable process. It is feasible for nonlinear and non-stationary as told earlier.

Nevertheless, it has two disadvantages

1. The number of IMFs cannot be predicted
2. It is also computationally expensive

Acoustics Doppler current profiler is used for bottom (40 m) temperature, and current is recorded each 10 min from July 21, 2011, to January 19, 2012, in north coast of Reunion Island. This paper mainly focuses on the four temperature time series which is measured at four distinct stations. The classical spectral analysis estimates good performance when the system is linear and when information is either intermittent or stationary.

TDIC (Time-dependent intrinsic correlation)

Correlation accepts that the factors must be stationary and linear. For non-stationary time arrangement, correlation might be modified and mutilated. Providing local temporal information is limited, and it cannot differentiate the main cycle from noise [17].

3.4   Case Study 4 (Fault Detection in Analog and Mixed-Signal Circuits)

Fault test of simple and blended signal is the most imperative methods for outline and execution. Simple and blended signal circuits are mind-boggling if there should be an occurrence of fault test [18–20]. Error detection method should be sensitive to parametric fault detection such as data storage, noise reduction, and amplifying the distinction of reaction information. There are the three spaces used in signature extraction: (1) time domain [21], (2) frequency domain [22], and (3) time–frequency domain [23, 24]. Signature variation Wavelet transform (WT) and subband separating method are very useful because signature variations include both time and frequency, but there must be prior knowledge before choosing a mother wavelet in WT for the band focus area ahead for the subband strategy. The signature extraction method is used to avoid these problems, and HHT system with the coherence examination will be used for fault detection (Table 9).

Test methods for fault detection in analog and mixed signal circuits are given below

1. HHT method
2. Coherence analysis
3. Integral of variable
4. Test strategy

Table 9 classification of various domains (time domain, frequency domain, and time–frequency domain)

S. no	Domains	Scheme	Fault classification	Circuits	Signature
1	Time domain	Minimal effort DC worked in individual test conspire	Catastrophic error	Linear simple circuits	Fault signatures
		Slant voltage increase	Parametric faults	Linear resistive circuits	Fault signatures
		Subband Volterra series	Parametric error	Nonlinear circuits	Fault signatures
2	Frequency domain linear and stationary	FFT Technique (IDD signal)	–	CMOS circuits	Frequency signatures
		Optimal fractional FT	–	–	Fault Signatures
3	Time–frequency domain	Fast WT	–	Analog circuits	Fault signatures
		Subband filter	–	Analog circuits	–

1. **HHT method**

It is explained in part I already.

2. **Coherence analysis**

Coherence analysis will give the estimation of relationship between couple of signals. It is utilized to find faults in amplitude and frequency

$$R_{x,y}(\tau) = \frac{1}{T}\int_0^T x(t)y(\tau - t) \cdot \mathrm{d}t. \qquad (24)$$

$R_{x,y}(\tau)$—cross correlation of variable $x(t)$ and $y(t)$; T—time length of stochastic process; τ—middle variable correlation of $x(t)$, which is

$$R_x(\tau) = \frac{1}{T}\int_0^T x(t)x(\tau - t) \cdot \mathrm{d}t. \qquad (25)$$

$R_x(\tau)$—self-correlation of variable $x(t)$.

FT applying to correlation

$$P_{x,y}(\omega) = \int_{-\infty}^{\infty} R_{x,y}(\tau)e^{-i\omega\tau}\mathrm{d}\tau. \qquad (26)$$

$$P_x(\omega) = \int\limits_{-\infty}^{\infty} R_x(\tau)e^{-i\omega\tau}d\tau. \tag{27}$$

$P_{x,y}(\omega)$—cross-power spectrum of variable $x(t)$ and $y(t)$; $P_x(\omega)$—power spectrum of variable $x(t)$.

Coherence of function is

$$C_{xy}(\omega) = \frac{P_{xy}(\omega)}{P_x(\omega).P_y(\omega)}. \tag{28}$$

$C_{x,y}(\omega)$—coherence function of variable $x(t)$ and $y(t)$.

3. **Integral of variable**

To diminish the information stockpiling of fault references, integral process will be used.

$$I_{c_{xy}} = \int\limits_{0}^{T} C_{xy}d\omega. \tag{29}$$

Ic_{xy}—integral of variable.

4. **Testing strategy**

Using these methods, error detection in analog and blended signal circuits is found [25] (Table 10).

4 Results and Discussions

Wavelet transform (WT), Fourier transform (FT), and HHT are used to decompose the signals. FT uses sine functions particularly for decomposing a signal, whereas WT uses complex and orthogonal waveforms for decomposing of the given signal. But the process involving in the HHT is sifting process for decomposing the signals (nonlinear).

HHT is used effectively in biomedicine, chemistry, financial, ocean engineering, speech processing, queuing theory, hyperspectral imaging, and music signals.

Table 10 Testing strategy

S. no	Fault test	Comparison method
1	Fault detection	Fault detection with fault-free circuit (measured Integral signature)
2	Fault location	Indispensable marks of defective circuit with fault dictionary (associated with predefined fault values)

HHT method does not prioritize on the basis, and it cannot assume the composition of signals. It has two drawbacks: (1) computationally expensive and (2) not having assumptions about the signals.

HHT and its variants have been successfully applied extensively in engineering data analytics. However, there are still some mixing problem, illusive components, non-recursive signals, leakage of energy, etc., which is provided for a challenging research problem.

In this paper, some variants of EMD were discussed. And the assumptions which we could conclude were ensemble EMD that generates better EEG; orthogonal HHT provides an overview analysis about earthquake, and EMD–DMS methods improve the performances of EMD. Finally, it shows that changes in EMD can be very efficiently.

5   Conclusions

In this paper, EEMD method is used for analyzing signals. By using orthogonal EMD method, we can strictly avoid the leakage of energy occurred during the processing of EMD signal. The dyadic masking signal method is used to improve EMD. New approach is named as Improved HHT (IHHT) method. The reviews of variants in EMD show that the EMD method still has a scope for further improvement. Analysis of temperature time series, detection and localization of damage in aluminum plates, environmental time series, and error detection in analog and blended signal circuits are discussed.

References

1. Huang NE, Nii O (2005) Attoh-Okine the Hilbert–Huang Transform in engineering
2. Huang NE, SSP Shen (2005) Hilbert Hunag Transform and its applications
3. Zhu X, Lv S, Fan L, Yu X (2011) The EEG signal process based on EEMD. In: International symposium on intelligence information processing and trusted computing, 978-0-7695-4498-4/11
4. Dacai W, Yingjian W (2011) Pulse signal analysis based on improved ensemble empirical decomposition of Hilbert–Huang transform. Comput Technol Autom 30(1):101–105
5. Garcia JA, Cabria L, Marante R et al (2010) An unbiased dual-mode mixingantenna for wireless transponders. Progr Electro Magn Res PIER 102:1–14
6. Canyang H, Qingyang X (2011) Analysis of non-stationary earthquake ground motion energy changing based on orthogonal HHT theory. In: Fourth international conference on intelligence computation technology and automation, 978-0-7695-4353-6/11
7. Deering R, Kaiser JF (2005) The use of a masking signal to improve empirical mode decomposition. In: Proceedings IEEE international conference acoustics, speech, and signal processing, pp 485–488
8. Yang Y, Deng J, Kang D (2015) An improved empirical mode decomposition by using dyadic masking signals. SiViP 9:1259–1263

9. Solomon S, Qin D, Manning M et al (2007) IPCC: climate change 2007—the physical scientific basis. Cambridge University Press, Cambridge
10. Veltcheva AD, Soare CG (2004) Identification of the components of wave spectra by the Hilbert Huang transform method. Appl Ocean Res 26(1–2):1–12
11. Huang NE, Chern CC, Huang K et al (2001) A new spectral representation of earthquake data: Hilbert spectral analysis of station TCU129, Chi-Chi, Taiwan, 21 September 1999. Bull Seismol Soc Am 91:1310–1338
12. Ma H, Qiu X, Luo J, Gu P, Liu Y (2015) Analysis of temperature time series based on Hilbert–Huang transform. J Hydrodyn 27(4):587–592
13. Watkins SE, Akhavan F, Dua R et al (2007) Impact-induced damage characterization of composite plates using neural networks. Smart Mater Struct 16(2):515–524
14. LeClerc JR, Worden K, Staszewski WJ et al (2007) Impact detection in an aircraft composite panel—a neural network approach. J Sound Vib 299(3):672–682
15. Camarena-Martinez D, Amezquita-Sanchez JP, Valtierra-Rodriguez M, Romero-Troncoso RJ, Osornio-Rios RA, Garcia-Perez A (2014) EEMD-MUSIC-based analysis for natural frequencies identification of structures using artificial and natural excitations. Sci World J
16. Dushyanth ND, Suma MN, Lite MV (2016) Detection and localization of damage using empirical mode decomposition and multilevel support vector machine. Appl Phys A 122:250
17. Ismail DKB, Lazure P, Puillat I (2015) Advanced spectral analysis and cross correlation based on empirical mode decomposition: application to the environmental time series. IEEE Geosci Remote Sens Lett 12(9)
18. Spyronasios AD, Dimopoulos MG, Hatzopoulos AA (2011) Wavelet analysis for the detection of parametric and catastrophic faults in mixed-signal circuits. IEEE Trans Instrum Meas 60(6):2025–2038
19. Deng Y, Shi Y, Zhang W (2012) An approach to locate parametric faults in nonlinear analog circuits. IEEE Trans Instrum Meas 61(2):358–367
20. Vasan ASS, Michael BL (2013) Pecht diagnostics and prognostics method for analog electronic circuits. IEEE Trans Ind Electron 60(11):5277–5291
21. Yuan L, He Y, Huang J, Sun Y (2009) A new neural-network- based fault diagnosis approach for analog circuits by using kurtosis and entropy as a preprocessor. IEEE Trans Instrum Meas 59(3):586–595
22. Roy KMK (2001) Fault detection and location using I_{DD} waveform analysis. IEEE Trans Des Test Comput 18(1):42–49
23. Aminian M, Aminian F (2000) Neural-network based analog-circuit fault diagnosis using wavelet transform as preprocessor. IEEE Trans Circuits Syst 47(2):151–156
24. Roh J, Abraham JA (2004) Subband filtering for time and frequency analysis of mixed-signal circuit testing. IEEE Trans Instrum Meas 53(2):602–611
25. Tang S, Li Z, Chen L (2015) Fault detection in analog and mixed signal circuits by using Hilbert Huang transform and coherence analysis, 0026-2692
26. Dhanalakshmi S, Venkatesh C (2013) Classification of ultrasound carotid artery images using texture features. Int Rev Comput Softw 8(4)

Smoothing and Noise Reduction in Images Using Variable Mode Decomposition

Ajmeera Ravi, P.V. Naganjaneyulu and M.N. Giriprasad

Abstract The error between true colours of a point in a scene relates to real world, and the pixel obtained from an image sensor is known as noise. Various noises were observed, and different filtration and sub-band decomposition techniques were also achieved most success rate in reducing the noise rate. But this leads to harshness on the image pixels they were tuned to more sharp than they need. Now a novel approach discussed in next sections which helps in learning minimization of sharp noise and increases the smooth characteristics of the noisy images. The noise reduction process is carried out by an iterative decomposition method and very much dependent on intrinsic mode function (IMF) and is known as variable mode decomposition (VMD). The database comprised with medical images and synthetic aperture radar images, and the experiments were done on SAR and medical images and shown in results section. This also helps in optimising the direction method of multipliers approach.

Keywords Variable mode decomposition · Clustering · Discretization

1 Introduction

Synthetic aperture associate in measuring system pictures has a transmissible presence of SPECKLE noise, this is often inflicting by victimisation hardware or by radiation of alternative sources severally. The result of speckle noise in pictures is as follows: random displacement of prominently bright or dark from imaginary

A. Ravi (✉)
JNTUH, Hyderabad, AP, India
e-mail: ravi.ajmeera@gmail.com

P.V. Naganjaneyulu
MVR College of Engineering and Technology,
Paritala(v), Kanchikacherla(m), AP, India

M.N. Giriprasad
JNTUA, Anantapur, AP, India

place Scientific opposition [1–6]. For noise reduction, earlier numerous techniques were developed some of them are LEE, FROST, MAP and REDIFINED MAP.

The increasing or speckle noise destructs the image, supported environmental conditions. To introduce noise digitally 'Variance' was used and represented by with respective values are 0.2, 0.5, 0.7 and 1. These values were approximated and solely help in ensure the simulations and values may be modified however the approximation crosses the limit it'll results a lot of noise within the image and there's no likelihood of retrieving the image to traditional condition. The illustration of the speckle noise $b(x, y)$ is given in equation one for input image representing by $a(x, y)$.

2 Scheme of Approach

A spatial domain signal in 2D is used to decompose based on frequency domain by using data fidelity constant, noise slack with temporal measurements and tolerance of convergence criteria typically.

Decomposition is carried out by using the following steps

1. Spectral domain discretization
2. Generalisations
3. Frequency domain conversion
4. Lagrange multiplier
5. Clustering
6. Convergence
7. Gradient ascent

Previously discretization is carried out by using fft & dct & dwt. The date analysis is carried out by time-vary amplitude at instantaneous frequency (W1).

Discretization helps in translating data from continuous time domain to samples form which is discontinuous in time period. For this process, an optimal analytical solution is required for a length of N-expansion. The optimal solution was obtained from orthogonal convergent as $O = 1/N^N$. This helps in minimising the memory and also reduces the error of discretization during change of representation [7, 8]

Here the discretization is called out by the spatial domain parameters.

Hy, $Hx \rightarrow$ resolution of the image

x, $y \rightarrow$ changes of grid values.

This results in the 2D mesh plot and helps in discretization, i.e., converting spatial domain values to frequency domain. As an example, $Hx \rightarrow fx$ (i.e. $fx = 1/Hx$) and vice versa frequency domain conversion based on discretization is carried out by frequency $x = x - 0.5 - fx$ and frequency $y = y - 0.5 - fy$.

Here a discretization sample 0.5 for change in values is considered.

Now converting the discretization data into conventional linear frequency data using Fourier shift process by wiener filtering leads to limit the decompositions

mocks of data. A heterodyne demodulation is caused by [8] shifting the frequency bands and helps in tracing the sub-band data by maintaining data fidelity (alpha). Alpha will be updated for all the le-bands, and vector with direction is given as $\overrightarrow{W_K}$ (Fig. 1).

Results of each are also given as block diagram representation in the below (Fig. 2).

Finally, the process is carried out by Lagrangian gradient normalisation on k clustered data

$$[x, y] = \int\int_0^N (i.\mathrm{d}x, j.\mathrm{d}y)$$
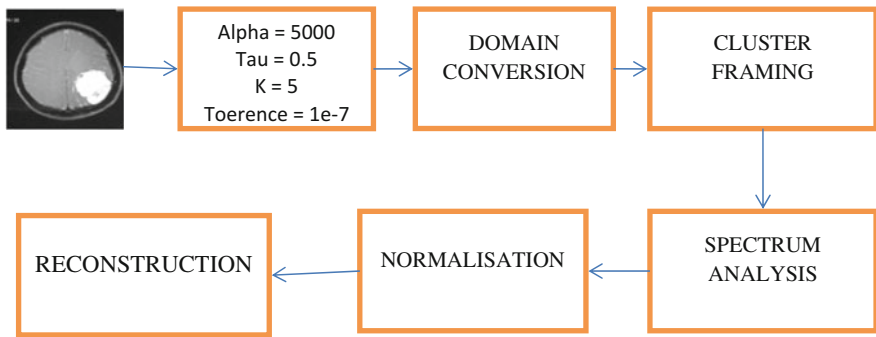
Fig. 1 Block diagram of proposed scheme of noise reduction and smooth of image
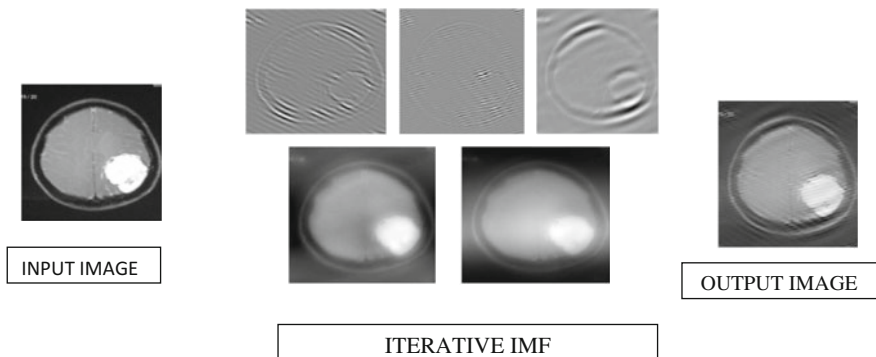
Fig. 2 Results at different sections
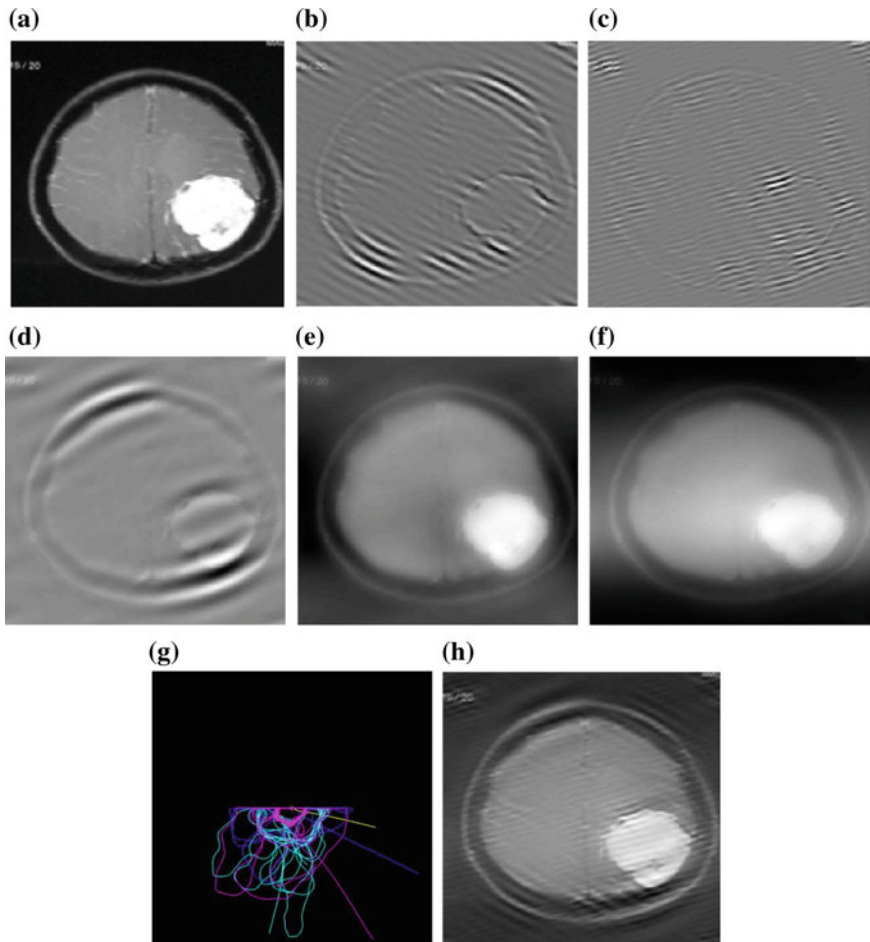
Fig. 3 **a** Input image, **b–f** 5 mode decomposition, **g** spectrum of the input image, **h** reconstructed smoothen output image

3 Results

In the below experimental results, 5 mode decomposition is carried out and spectrum analysis is performed on medical images in Fig. 3 and on SAR images in Fig. 4 all the sharpened values were reduced and results were smooth due to iterative procedure.
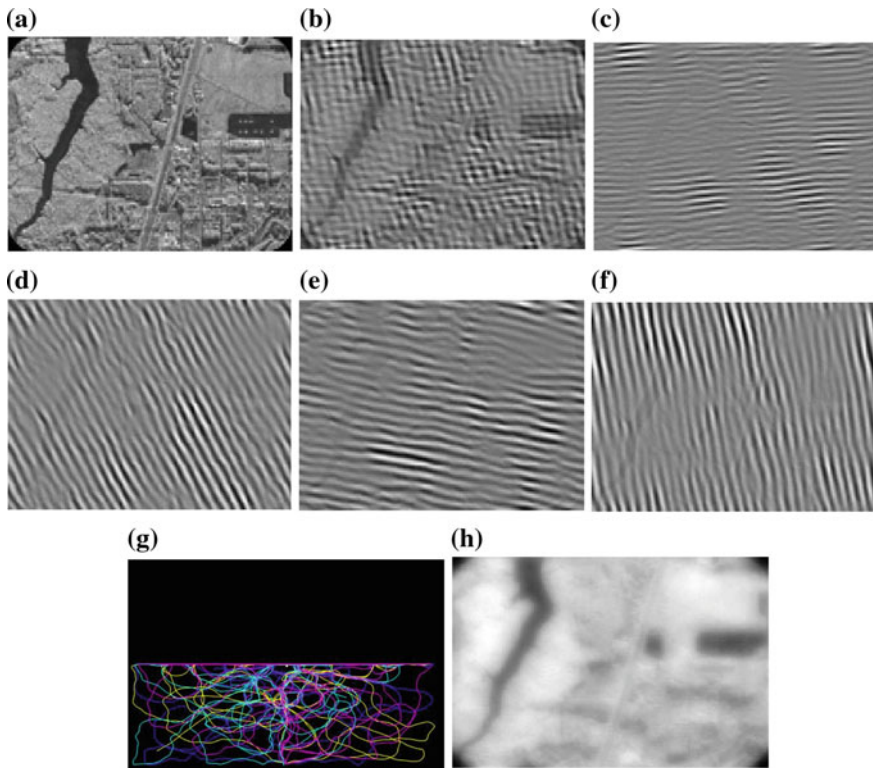
Fig. 4 **a** Input image, **b–f** 5 mode decomposition, **g** spectrum of the input image, **h** reconstructed smoothen output image

4 Conclusion

Under different k-bands of data with their centre of frequencies led to reconstruct the input which is most similar. While each band gets smoothen after demodulation into baseband, this made the usage of directional theorem to reduce. It also reduces the inverse problem since we are opting for an individual mode comprised with a limited band approach and decomposing an image into required modes with least square sense. This process is completely dynamical in acquiring the data and updating the centre frequencies for all the bands. Even here the major limitation boundary effect is also reduced by normalising the data using Lagrangian gradient normalisation on all modes of band frequencies; this is possible with iterative VMD approach.

References

1. Fu X, Wang Y, Chen L, Dai Y (2015) Quantum-inspired hybrid medical ultrasound images despeckling method. Electron Lett 51(4):321–323
2. Bar-Zion AD, Tremblay-Darveau C, Yin M, Adam D, Foster FS (2015) Denoising of contrast-enhanced ultrasound cine sequences based on a multiplicative model. IEEE Trans Biomed Eng 62(8):1969–1980
3. Afsham N, Rasoulian A, Najafi M, Abolmaesumi P, Rohling R (2015) Nonlocal means filter-based speckle tracking. IEEE Trans Ultrason Ferroelectr Freq Control 62(8):1501–1515
4. Maity A, Pattanaik A, Sagnika S, Pani S (2015) A comparative study on approaches to speckle noise reduction in images. In: 2015 International conference on computational intelligence and networks (CINE), Bhubaneshwar, pp 148–155
5. Khusna DA, Nugroho HA, Soesanti I (2015) Performance analysis of edge and detailed preserved speckle noise reduction filters for breast ultrasound images. In: 2015 2nd International conference on Information Technology, Computer, and Electrical Engineering (ICITACEE), Semarang, pp 76–80
6. Rahimi M, Yazdi M (2015) A new hybrid algorithm for speckle noise reduction of SAR images based on mean-median filter and SRAD method. In: 2015 2nd International conference on pattern recognition and image analysis (IPRIA), Rasht, pp 1–6
7. Dragomiretskiy K, Zosso D (2014) Variational mode decomposition. IEEE Trans Signal Process 62(3):531–544
8. Dragomiretskiy K, Zosso D, Two dimensional variational mode decomposition. In: IEEE transactions international conference of image processing

Author Biographies

A. Ravi received his Bachelors degree in ECE from JNT University, and received his M.Tech. from JNTU Kakinada. He is now a research scholar under the guidance of Dr. P.V. Naganjaneyulu and Dr. M.N. Giriprasad. His current research interest includes Satellite and SAR image processing.

P.V. Naganjaneyulu received Ph.D. in Communications from JNTU Kakinada University. And he is currently professor of ECE and Principal in MVR College of Engineering and Technology, Paritala, Krishna dist., he was specialised in communication and signal processing, he published 30 journal papers and 10 conference papers. His current Research area of interest is communications, signal processing and image processing. He is a member of IETE.

M.N. Giriprasad received the B.Tech. degree from JNTUA, Andhra Pradesh, M.Tech. from SVU, Tirupathi and Ph.D. from JNTUH. He is now working as Professor in Dept. of ECE at JNTU College of engineering Ananthapur. His current research interests includes Wireless communications, Bio medical Instrumentation and image processing. He is a member of ISTE, IE & NAFE

A Hybrid Approach for Image Denoising in Ultrasound Carotid Artery Images

Latha Subbiah and Dhanalakshmi Samiappan

Abstract Reliable, effective, and exact estimations of the geometry of the Common Carotid Artery (CCA) are necessary to identify stroke and heart attack prior. The utilization of ultrasound imaging in medicinal diagnosis is well recognized. Denoising speckle in ultrasound is vital. This paper gives thresholding of curvelet for denoising of carotid artery ultrasound images. The hybrid filter preserves radiometric information, edge information, and spatial resolution. Curvelet transform combines speckle reducing and edge preserving properties of the image because it exploits the instantaneous coefficient of variation which is a function of local gradient magnitude and Laplacian operators. Blurred thin edges and low contrast fine features are preserved by thresholding. The filtering methods performance is valuated in terms of eight performance metrics. The proposed filter is compared with existing filters which were used for the carotid artery ultrasound application.

Keywords Speckle noise · Filter · Denoising · Curvelet · Ultrasound

1 Introduction

Medicinal ultrasound imaging has been utilized for successful diagnostics of sicknesses over the previous decades because of its noninvasive, safe, convenient, exact, and financially savvy attributes [1]. When effective human analysis and computer-assisted ultrasound imaging system merge, artifacts complicate the automatic classification of abnormalities. The paper is organized as follows. Section 2 describes the noise model in ultrasound and the available speckle reducing filters with their performance evaluation. Section 3 discusses about the

L. Subbiah (✉) · D. Samiappan
Department of ECE, SRM University,
Kattankulathur, Chennai 603203, Tamil Nadu, India
e-mail: latha.su@ktr.srmuniv.ac.in

curvelet transform and thresholding of the same. Section 4 gives the experimental results, and Section 5 concludes the paper.

2 Denoising Filters

Speckle Noise Model. The random and deterministic speckle noise is the inherent characteristic of ultrasound image. Spot has negative effect on ultrasound imaging. Radical decrease in contrast resolution might be in charge of the poor successful resolution of ultrasound when contrasted with MRI. Noise in ultrasound image is given by

$$X(i,j) = [h(i,j) * Y(i,j)] + n(i,j) \tag{1}$$

where X is the resultant distorted image, Y is the original image corrupted by the multiplicative noise, h is point spread function or impulse response, n is the additive noise component, i and j and axial are lateral indices of the image [2].

2.1 Median and Frost Diffusion Filters

Median filter is a simple nonlinear operator most suitable for removing salt and pepper noise. The filter replaces the middle value of the kennel with the median values of the neighbors. For the same image of different shapes, the performance differs. It cannot differentiate noisy portion and the plaque area. Frost, andLee filters were derived from minimum mean square error (MMSE) criteria. Edge preservation and noise reduction are performed by varying the filter kennel with the local statistical value of the image.

$$\bar{R}(i,j) = \sum_r \sum_s m(i+r,j+s) * I(i+r,j+s); m(i+r,j+s) = K_0 \exp[-KC_I^2(t)\sqrt{(r^2+s^2]} \tag{2}$$

r, s are the indices of the filter window and m is the weighting function. K_0 is the constant of normalization, and K is the damping factor. In diffusion filter, an edge detection function controls the strength and the direction of the diffusion. It evacuates speckles and also upgrades the edges. It expels speckle by adjusting the picture by means of resolving a partial differential equation (PDE).

$$\frac{\partial I}{\partial t} = \partial \text{iv}[c|\nabla I|).\nabla I]; \quad I(t=0) = I_0 \tag{3}$$

where ∇ is the gradient operator, ∂iv is the divergence operator, $c(|\nabla I|)$ is the diffusion coefficient, and I_0 is the original image. Diffusion process is stopped

across the edges when the gradient strength increases and diffusion is stopped [3]. The frost filter hybridized with diffusion filter is performed with 3×3 window size and 1 iteration for all the speckle variances.

2.2 Lee Diffusion, Kuan, and Additive Filter

The Lee filter preserves edges and can denoise the multiplicative speckle noise. It performs well in low-variance areas than high-variance areas where modifications have to be done. Local statistics method is incorporated with it [4]. The general form of Lee and Kuan filter is

$$\hat{R}(t) = \bar{I}(t) + [I(t) - \bar{I}(t)] * W(t). \tag{4}$$

The value of the weighting function W is 1 for areas with high signal activity and 0 for flat areas and the other areas range in between [4]. The Lee and Kuan filter has weighting function

$$W(t) = 1 - \frac{C_u^2}{C_I^2(t)}; \quad W(t) = \frac{1 - (C_u^2/C_I^2(t))}{1 + C_u^2}. \tag{5}$$

where $C_u = \frac{\sigma_u}{\bar{u}}$ and $C_I = \frac{\sigma_I}{\bar{I}}$ are coefficients of variations of the noise u and image I [5]. The Lee filter is combined with the diffusion PDE and is performed for different number of iterations and different kennel size.

2.3 SRAD Filter

Speckle reducing anisotropic diffusion filter filters both additive and the multiplicative speckle noise. Instantaneous coefficient of variation is analyzed by SRAD, which is appeared to be a component of the local magnitude of the gradient and Laplacian operators. The merits of SRAD and DPAD are preservation of mean, edge localization, speckle reduction, detailed preservation and detailed enhancement, and reduction of variance [6]. SRAD also depends on MMSE criteria. The image and noise statistics are estimated better using DPAD filter. First method is more suitable for high-contrast edges, and the second method is suitable for wide regions.

$$g(\nabla I) = e^{\left(-\left(\frac{\|\nabla\|}{K^2}\right)\right)}; \quad g(\nabla I) = \frac{1}{1 + \left(\frac{\|\nabla\|}{K}\right)^2} \tag{6}$$

where K is a constant fixed by noise estimation. The equation of SRAD and DPAD which is an extension of Kuan filter are

$$I_{i,j}^{t+\Delta t} = I_{i,j}^{t} + \frac{\Delta t}{|\eta_{i,j}|}\operatorname{div}[c_{1}(C_{i,j,t})\nabla I_{i,j}^{t}]; \bar{u}^{t}I_{i,j}^{t+\Delta t} = I_{i,j}^{t} + \frac{\Delta t}{|\bar{\eta}_{s}|}\operatorname{div}\left[c_{2}(C_{i,j,t})\nabla I_{i,j}^{t}\right] \quad (7)$$

where $c_{1}(C_{i,j,t}) = 1 - K_{4}^{i,j,t}$ and $|\eta_{i,j}|$ is the number of neighbors required to calculate the factors in the divergence operator [7].

2.4 Gabor Filter

Gabor filter is a wavelet group. By convolving the image with Gabor filter, a group of filtered images can be obtained. Each one represents different orientation and scale.

$$g(x,y) = \frac{1}{2\pi\sigma_{x}\sigma_{y}}\exp[-\frac{1}{2}\left(\frac{x^{2}}{\sigma_{x}^{2}} + \frac{y^{2}}{\sigma_{y}^{2}}\right) + 2\pi jWx]; \quad \sigma_{u} = \frac{1}{2\pi\sigma_{x}}; \quad \sigma_{v} = \frac{1}{2\pi\sigma_{y}}. \quad (8)$$

σ_{x} and σ_{y} are the standard deviations in x-axis and y-axis of the elliptical Gaussian. The fields to be defined are theta, the orientation; lambda, the wavelength of the sinusoidal factor in the given equation; gamma, the spatial aspect ratio; and psi, the phase offset of cosine factor of the Gabor function. The optimum values were found to be $\theta = 90°$, $\lambda = 10$, $\gamma = 0.25$, and $\Psi = 45°$. The Gabor filter performs denoising along with segmenting the edges.

2.5 Performance Evaluation

The performance of the denoising filters is evaluated quantitatively with the following quality metrics: Using SNR and PSNR, the level of the image to the background noise is compared.

$$\text{SNR} = \mu/\sigma; \quad \text{PSNR} = 10\log_{10}(255^{2}/\text{MSE}). \quad (9)$$

The mean square error is the collective squared error between the denoised and the original image, and it measures the quality change between the original and the denoised image. Mean absolute error is the average of the absolute errors.

$$\text{MSE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} (X(i,j) - Y(i,j))^2;$$

$$\text{MAE} = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} |X(i,j) - Y(i,j)|.$$

(10)

RMSE is the square root of the squared error averaged in the window. The structural similarity index between the denoised and the original image is computed as

$$\text{RMSE} = \sqrt{\frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} [X(i,j) - Y(i,j)]^2};$$

$$\text{MSSIM} = \frac{(2\mu_x \mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}.$$

(11)

where $C_1 = (K_1 L)^2$ and $C_2 = (K_2 L)^2$. The dynamic range of the pixel $L = 2^{\text{bits per pixel}} - 1$ K_1, K_2 are constants. SSIM and quality index give the performance based on visual interpretation of the denoised image. IMGQ, correlation between 2 images, is given by

$$\text{IMGQ} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \frac{2\mu_y \mu_x}{\mu_y^2 + \mu_x^2} \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}; \quad r = \sum_{i=-k}^{K} \sum_{j=-k}^{K} X(i,j) Y(x-i, y-j).$$

(12)

The values of SNR, PSNR, MSSIM, IMGQ, and correlation must be high, and MSE, MAE, and RMSE values must be low for good denoising filter.

3 Curvelet Filter with thresholding

Wavelet-based denoising is the most frequently followed method. By thresholding of the undecimated wavelet transform, translation-invariant methods improved denoising. The ridgelet and curvelet transforms became alternatives for wavelet [8]. The asymptotic mean square error of curvelet is smaller than wavelet denoising. Geometrical features are better preserved in curvelet than wavelet and Gabor filters. Multiscale ridgelets and spatial band-pass filtering are combined to separate different scales. Curvelets occur at all scales, locations, and orientations and have variable width, length, and anisotropy. Using curvelet in the ultrasound carotid artery image, optimal sparse representation is achieved in wave propagators and objects with edges. Higher values of hard thresholding are hybridized with curvelet transform to perform better than other methods of denoising carotid artery ultrasound image. Input to curvelet transform is a discrete Cartesian grid $f(n_1, n_2)$, $0 \leq n_1, n_2 < n$, 0. $c^D(j,k,l)$ are a collection of coefficients, and ϕ are digital curvelet waveforms.

Table 1 Filter performance for speckle variance 0.04

Filter	SNR	PSNR	MSSIM	MSE	MAE	RMSE	Quality index IMGQ	Correlation
Median filter	17.4121	27.5418	0.9853	114.5249	3.3130	10.7016	0.9509	0.9568
Lee filter	17.2505	27.1980	0.9810	123.9596	3.5046	11.1337	0.9575	0.9526
Lee diffusion filter	16.9370	**31.9817**	0.9520	**41.2016**	**2.3472**	**6.4188**	0.9765	**0.9847**
Frost filter	17.0364	27.3121	**1.0319**	120.7456	3.5741	10.9884	0.9225	0.9548
Frost diffusion filter	17.3336	27.2931	0.9273	121.2743	3.4688	11.0125	0.9581	0.9538
Kuan filter	**17.4421**	27.1278	1.0055	125.9811	3.6122	11.2241	0.9252	0.9519
Additive filter	17.2161	27.7012	0.9300	110.3984	3.4212	10.5071	0.9594	0.9581
Anisotropic diffusion filter 1	17.3792	29.1779	0.9242	78.5760	3.2139	8.8643	0.9715	0.9703
Anisotropic diffusion filter 2	17.1466	31.2289	0.8999	48.9990	3.3107	6.9999	**0.9807**	0.9815
SRAD	15.3898	19.2235	0.7889	777.5530	6.9770	27.8846	0.8460	0.7747
Curvelet	18.4175	21.3278	0.4199	478.9625	20.5466	21.8852	0.4552	0.9607
Curvelet with thresholding	17.5520	39.4883	0.9817	7.3156	1.4678	2.7047	0.7480	0.9968

Bolded values are the optimum values of the performance metrics

$$C^D(j, k, l) = \sum_{n_1 n_2} f(n_1, n_2) \overline{\varphi_{j,l,k}^D(n_1, n_2)}. \tag{13}$$

4 Experimental Results

An ultrasound carotid artery image is taken, and speckle variances of 0.04, 0.1, 0.8, and 1 with mean zero are added. The denoising methods are applied to the noised image and then compared with the original image using the performance evaluation methods (Table 1).

From Figs. 1 and 2 and the performance evaluation metrics, it is found that SRAD filter gave very poor results for higher speckle noise variances. Lee diffusion filter is better in terms of PSNR, MSE, MAE, RMSE, and correlation. Frost filter gives good structural similarity, and the quality index metric is better in the second method of anisotropic diffusion filter. Kuan filter gives the highest signal-to-noise ratio. Curvelet with thresholding gives the best performance (Table 2; Fig. 3).
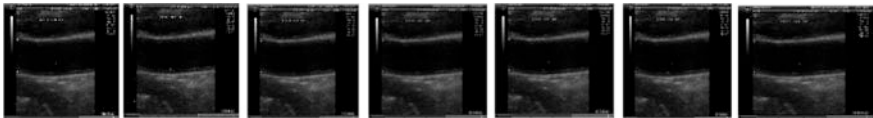
Fig. 1 Original *gray* image, speckle noise with variance 0.04, median, Frost, Frost diffusion, Lee, and Lee diffusion filtered images
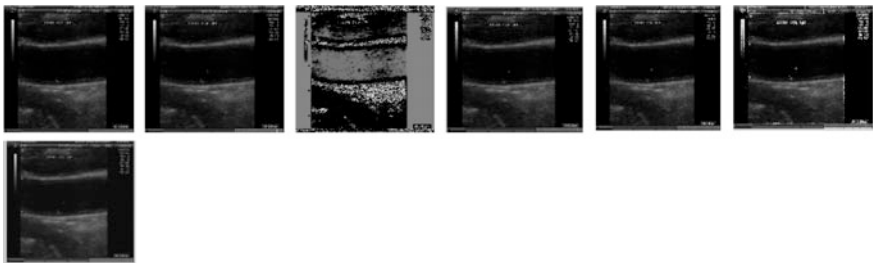
Fig. 2 Kuan, additive, Gabor, anisotropic diffusion 1, 2, SRAD, curvelet filtered images

Table 2 Gabor filter for speckle variance 0.04

Gabor	SNR	PSNR	MSE	RMSE	Correlation	MAE
Psi = 0	9.0900	15.5445	1.8140e+003	42.5908	−0.5930	26.9996
Psi = 45	10.1239	15.5443	1.8141e+003	42.5918	−0.1822	27.1343
Psi = 90	11.1986	15.5533	1.8103e+003	42.5478	0.5506	27.2800
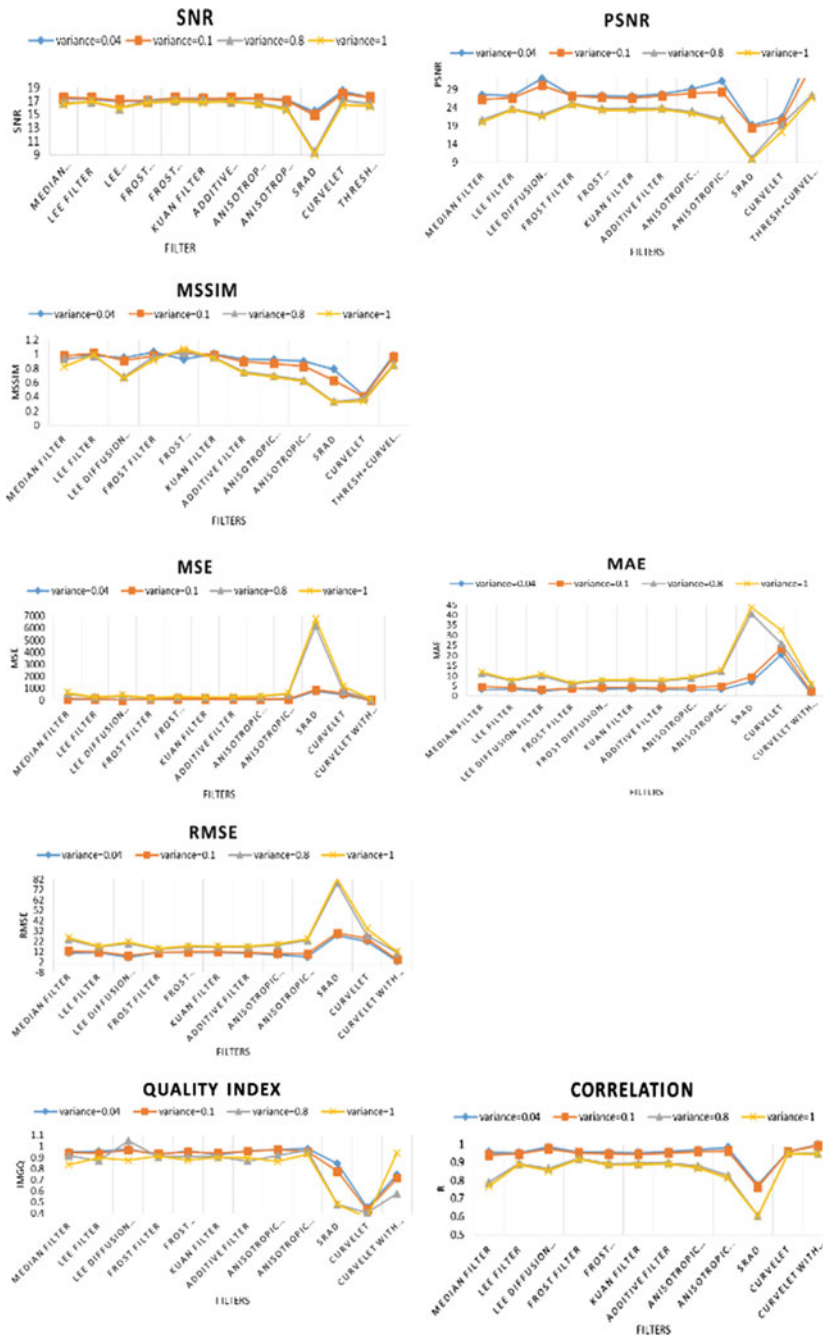
Fig. 3 Performance metrics for different filters for different speckle variances

5 Conclusion

Stroke is caused by atherosclerosis, hardening of the artery walls, and plaque formation. Carotid artery ultrasound images can be used for analysis of intima-media thickness, lumen thickness, and adventitia boundary. Despeckling is an important requirement in enhancement of ultrasonic images. The objective of this paper is to carry out a comparative evaluation of despeckling filters, and a filter combining SRAD and DPAD and curvelet is proposed. Using performance metrics, best filter suitable for carotid artery ultrasound image is to be found out. From literature, it is recognized that in the future, more work is likely to be done to accomplish a fully automated system to identify plaque in carotid artery even in an early stage of plaque formation.

References

1. Ambardar S, Singhal M (2014) A review and comparative study of denoising filters in ultrasound imaging. Int J Eng Technol Adv Eng 4(8)
2. Roomi SMM, Rajee RBJ (2011) Speckle noise removal in ultrasound images using particle swarm optimization technique. In: IEEE international conference on recent trends in information technology (ICRTIT), 926–931
3. Chen G, Liu X, Zhou Z (2007) Modified frost speckle filter based on anisotropic diffusion. In: 2007 IET international conference on radar systems
4. Akl A, Tabbara K, Yaacoub C (2012) Edge-based suboptimal Kuan filtering for speckle noise reduction. In: IEEE international conference on engineering and technology (ICET)
5. Vanithamani R, Umamaheswari G (2010) Performance analysis of filters for speckle reduction in medical ultrasound images. Int J Comput Appl 12(6), Article 4
6. Yu Y, Acton ST (2002) Speckle reducing anisotropic diffusion. IEEE Trans Image Process 11 (11)
7. Aja-Fernandez S, Alberola-Lopez C (2006) On the estimation of the coefficient of variation for anisotropic diffusion speckle filtering. IEEE Trans Image Process 15(9)
8. Starck J-L, Candes EJ, Donoho DL (2002) The curvelet transform for image denoising. IEEE Trans Image Process 11(6)

Data-Dependent Subband Coder
for Image Compression

Jaba Deva Krupa Abel, Dhanalakshmi Samiappan
and Niraimathi Ponnusamy

Abstract Subband coding is a popular technique extensively used in the areas of image and video communications. This technique is widely employed to achieve data compression and efficient data transmission. The general principle of subband coding involves dividing the input image based on its frequency, where the correlation between the subbands still exists. In this paper, we propose a new technique where the subband coder divides the image such that the subbands will have only decorrelated data, thus reducing the redundancy and achieving image compression. To exploit for decorrelation between the subbands, we adopt a polynomial eigen-value decomposition technique which is popularly known as Sequential Best Rotation (SBR2C) algorithm. It is an iterative technique with a set of delay and rotation operations. Using this approach, we design the analysis filter bank whose operation is data dependent and thus splits the input image into subbands that are strongly decorrelated. At the receiver end, we design the synthesis bank by imposing perfect reconstruction condition in the absence of quantization errors. We compare our proposed algorithm with the popular subband coding technique, discrete wavelet transform (DWT). Simulation results show that proposed method can perform better compression because of data-dependent operation of filter banks.

Keywords Subband coding · Polynomial eigen value decomposition
Strong decorrelation · Perfect reconstruction

1 Introduction

Image compression is an important task in the field of image and video processing which allows efficient storage and transmission of signals in multimedia applications [1, 2]. The goal of image compression is to represent the image with less number of bits while the reconstructed image is still similar to original image.

J.D.K. Abel · D. Samiappan (✉) · N. Ponnusamy
Department of ECE, SRM University, Kattankulathur 603203, India
e-mail: sdhanalakshmi2004@gmail.com

Thus, the digital images can be stored using less memory and transmitted using smaller bandwidth.

Image compression can be achieved in several ways. One of the techniques is transform coding where image prior to transmission is subjected to invertible transforms with an aim to remove redundant data. Discrete Cosine Transform (DCT) [3–5] and Vector Quantization (VQ) [6–8] coding are some of popular transform coding approaches. The other popular technique employed to achieve image compression is subband coding. Subband coding is proven to give pleasing reconstructions compared to transform coding. The principle involved in subband coding is to decompose the input image into subbands based on their frequency. The human eyes are less sensitive to high-frequency components and thus compression can be achieved by reducing such components.

Several subband coding techniques are found in the literature. Most popular method used in subband coding is the discrete wavelet transforms [9], where image is projected onto a set of basis functions. Both orthogonal and non-orthogonal wavelets [10] are used in discrete wavelet transforms. In subband coders like DWT, each subband still has redundant data and correlation between the subbands exists. We thus propose a technique where image compression can be achieved by reducing the redundancy or the correlation between the subbands. The novelty of our proposed technique is that the subband coders designed are purely data dependent. For a given image, we design the analysis filters such that the decomposed subbands are strongly decorrelated. This will ensure that each channel will carry only unique information, thus avoiding redundancy between the channels and achieving greater compression. Unlike the existing subband coders, where the filter coefficients are fixed for given order, our proposed approach designs the filter whose order and coefficients change with the input image.

Our algorithm is an iterative approach which involves set of delay and rotation operations and is popularly known as Sequential Best Rotation (SBR2) algorithm. It is based on polynomial eigen value decomposition technique. This algorithm has been successfully applied to broadband extensions of narrowband problems, which traditionally have been addressed by the EVD, including subspace decomposition [11]. Redif et al. [12] proposed modified SBR2 algorithm which is proven to perform computationally better than SBR2 algorithm. We use this approach for our data-dependent subband coder design which can perform Lossy compression.

This paper is organized as follows. In Sect. 2, we present filter design technique. We discuss the polynomial eigen decomposition method and the SBR2 algorithm for designing the filter using this method. In Sect. 3, we discuss the obtained simulation results. Following that we conclude in Sect. 4.

2 Filter Design

In this section, we present the algorithm for designing the data-dependent subband coder. We first give a brief note on two-band subband coder, and following that we discuss our proposed algorithm.

2.1 Two-Band Subband Coder

The structure of two-band subband coder is shown in Fig. 1. In subband coders like DWT, the filters $H_0(Z)$ and $H_1(Z)$ are the wavelet filters whose coefficients are fixed and designed using the standard basis functions. The filters $H_0(Z)$ and $H_1(Z)$ are low-pass and high-pass filters, respectively. Thus, the output at these filters $v_1(n)$ will be a low-pass component which carries smooth information and $v_2(n)$ will be a high-pass component which carries edge information of input image. Edge information can be usually discarded as it gives negligible impact on human perception. These filters cannot guarantee the decorrelation between their outputs.

Thus, we arrive at a new technique where the filters designed split the signal such that their outputs are always decorrelated and also their energies are majorized. That is, each subband contains only unique information and thus resulting better transmission. Spectral majorization facilitates energy compaction, and thus, subbands in noise space can be discarded. These two properties make our proposed subband coder more efficient [13]. Notice that the effect of quantizers is ignored in all derivations.

2.2 Polynomial EVD (Eigen Value Decomposition)

The polyphase representation of the structure shown in Fig. 1 is as shown in Fig. 2. This structure is obtained using noble identities where $\underline{P}_A(Z)$ and $\underline{P}_S(Z)$ are the analysis and synthesis polyphase matrices, respectively. The objective of our proposed algorithm is to achieve decorrelation between the subband signals $\underline{v}[k]$.
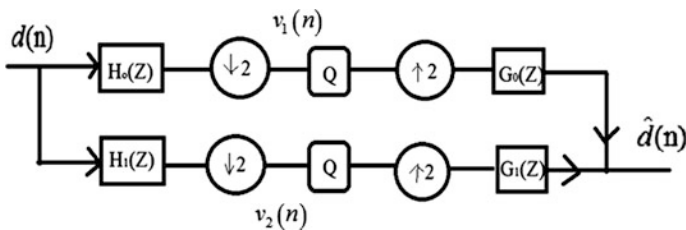
Fig. 1 Two-band subband coder
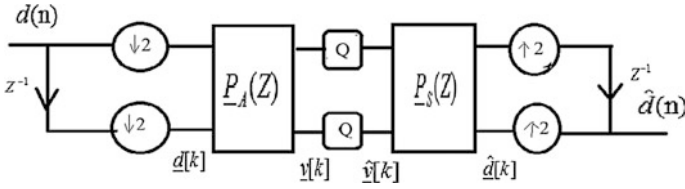
Fig. 2 Polyphase representation of structure shown in Fig. 1

That is, the CSD matrix at the input of analysis filter has to be diagonalized at the filter output. Thus, we propose an eigen value decomposition algorithm which can diagonalize the CSD matrix where $\underline{P}_A(Z)$ acts as modal matrix.

The input to the algorithm is the CSD matrix computed for blocked input vector $\underline{d}[k]$ and let it be denoted as $\underline{R}(Z)$. The procedure for computing this matrix is discussed in Results section. This matrix will be a polynomial matrix. Let $\underline{C}(Z)$ be the CSD matrix at the output of analysis bank, i.e., CSD matrix computed for the signals $\underline{v}[k]$ which should be diagonal to satisfy strong decorrelation between the subbands. The desired operation is given in Eq. (1). We assume that the system is paraunitary.

$$\underline{C}(Z) = \underline{P}_A(Z)\underline{R}(Z)\tilde{\underline{P}}_A(Z) \tag{1}$$

The first step in the algorithm is to find the largest half diagonal element in the matrix $\underline{R}(Z)$. Each element is a polynomial, and the dominant element is the polynomial of largest magnitude. The largest element can be found using Eq. (2).

$$\{p, q, \tau\} = \arg \max_{p,q,p \neq q,\tau} \frac{\left|R_{pq}[\tau]\right|^2}{\left|R_{pp}[0]\right|\left|R_{qq}[0]\right|} \tag{2}$$

In Eq. (2), $R_{pq}[\tau]$ is the element in pth row and qth column of matrix $\underline{R}[\tau] \to \underline{R}(z)$ and τ represents the lag of the elements in the polynomial. Here, the arrow symbol represents that the element on its either side is z-transform pairs. The ultimate goal is to transfer the energy of this element onto the diagonal element. This can be achieved in twofold process as given in Eq. (3).

(i) Operate on delay matrix $\underline{\Lambda}(z)$, so that the lag τ is shifted to zero.
(ii) Operate on rotation matrix $\underline{Q}$, so that the energy is transferred to diagonal.

$$\begin{aligned} \underline{P}'(z) &= \underline{\Lambda}(z)\underline{R}(z)\tilde{\underline{\Lambda}}(z) \\ \underline{P}''(z) &= \underline{Q}\underline{P}'(z)\tilde{\underline{Q}} \end{aligned} \tag{3}$$

The delay and rotation matrix obviously depends on the coordinates p, q, and lag τ of dominant element. The expression for delay matrix is given in Eq. (4).

$$\underline{\Lambda}(z) = diag\left\{\underbrace{1..1}_{q-1} z^{-\tau} \underbrace{1..1}_{M-q}\right\} \tag{4}$$

where $M = 2$ for 2-band. The rotation matrix $\underline{Q}$ is constructed using Givens Rotation matrix. It is basically an identity matrix except at the positions where pth row and pth column and qth row and qth column overlap. The elements at these four positions will be equal to Givens Rotation matrix. The approach can be clearly found in [11].

After the delay and rotation operations, the energy of largest off-diagonal element will be transferred to main diagonal. This operation is repeated till the magnitude of all the largest off-diagonal elements falls below certain threshold. For each new iteration, the input CSD matrix $\underline{R}(z)$ will be the matrix $\underline{P}''(z)$ obtained in previous iteration. The final diagonalized CSD matrix $\underline{P}''(z)$ as obtained in last iteration of the algorithm is given in Eq. (5).

$$\underline{P}''(z) = \underline{M}(z)\underline{R}(z)\underline{\tilde{M}}(z) \tag{5}$$

If K is the number of iteration took for diagonalization of $\underline{R}(z)$, then the matrix $\underline{M}(z)$ is as given in Eq. (6).

$$\underline{M}(z) = \underline{m}_K(z)\underline{m}_{K-1}(z)\ldots\ldots.\underline{m}_1(z) \tag{6}$$

where $\underline{m}(z) = \underline{\Lambda}(z)\underline{Q}$. From Eqs. (1) and (5), we notice that $\underline{P}_A(Z) = \underline{M}(z)$. The synthesis bank $\underline{P}_B(Z)$ is obtained by imposing the perfect reconstruction condition $\underline{P}_B(Z) = kz^\tau\underline{\tilde{P}}_A(Z)$ since the system is assumed to be paraunitary.

In our proposed EVD approach, the order of the filters grows at the end of each iteration. Thus, a trim function [11, 12] can be implemented to keep the order concise based on the overall energy.

3 Simulation Results

The input to the subband coder will be a two-dimensional image. We first convert it into a one-dimensional vector. The pixels values can be read in column-wise or row-wise or zigzag pattern for this conversion. This one-dimensional vector will be the input $d(n)$ for the subband coder. The blocked input signal will be $d_m[n] = d[Mn + m - 1]$, M being polyphase order. We construct the autocorrelation matrix for this blocked vector. There are several methods for finding the autocorrelation matrix. Our approach uses the para-hermitian and pseudo-circulant properties of autocorrelation matrix which makes its construction easier. Different approaches for finding the autocorrelation matrix are discussed in [12]. Any method can be used. The algorithm operates on this autocorrelation matrix. At the receiver

Fig. 3 a Reconstructed image. **b** Error image obtained using proposed technique. **c** Reconstructed image. **d** Error image obtained using DWT technique for camera man image
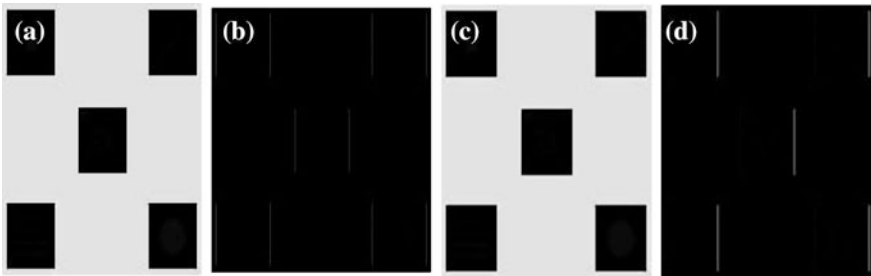
Fig. 4 a Reconstructed image. **b** Error image obtained using proposed technique. **c** Reconstructed image. **d** Error image obtained using DWT technique for camera man image

and, after obtaining the reconstructed vector $\hat{d}(n)$, the output image can be reconstructed back using the same approach followed at transmitter side, while considering the delays. We tested our algorithm on two standard images. One is baboon image, and other is embedded square image. Both the test images are gray scale images of size 256×256. For comparison purpose, we used simple 1-D Haar DWT.

Figure 3a shows reconstructed camera man image and Fig. 3b the error image using proposed approach. Figure 3c, d shows the reconstructed and error images obtained using DWT. Fig. 4a, b shows the reconstructed and error images for embedded square image obtained using proposed technique and Fig. 4c, d using DWT. For better analysis, one can look at Table 1, where we found the quality measuring parameters like PSNR, MSE and also coding gain. For both images PSNR and MSE are better in our approach compared to the standard DWT. Coding gain though similar for both methods for baboon image, it is far better for embedded square image which has very less number of intensity values using our approach.

From Table 1, it can be noticed that our proposed method can guarantee more than 50% coding gain compared to state-of-art technique for images with less

Table 1 Comparison between proposed and DWT technique for two different images

Image approach	Camera man		Embedded square	
	Proposed	DWT	Proposed	DWT
Coding gain	2.4981	2.6982	4.3398	6.3699
MSE	138.5174	384.8827	64.2032	198.1798
PSNR	26.7158	22.2775	30.0552	25.1602

number of intensities. Thus, the proposed technique can perform well for images like embedded square and also for binary images.

4 Conclusion

In this paper, we proposed a new technique for designing the subband coder. The uniqueness of the proposed subband coder is that it is purely data dependent. Unlike the existing compression techniques, it can guarantee for better decorrelation between the subbands and thus reduces the redundancy at greater extent and makes the subband coder more efficient. From the obtained results, we can conclude that the proposed method yields better PSNR and MSE compared to existing DWT technique, which assures better image quality after compression. Though the coding gain obtained for image with good contrast using our approach is comparable to the standard approach, our approach can provide 50% more coding gain compared to standard technique for low-contrast image. Thus, we conclude that our proposed technique can perform in par with standard compression technique and much better than those for low-contrast image in particular.

References

1. Pennebaker WB, Mitchell JL (1993) The JPEG still image data compression standard. Reinhold, New York
2. Taubman DS, Marcellin MW (2002) JPEG2000: image compression fundamentals standards and practice. Kluwer, Norwell
3. Rose K, Heiman A, Dinstein I (1990) DCTiDST alternate-transform image coding. IEEE Trans Commun 38:94–101
4. Ngan KN (1984) Image display techniques using the cosine transform. IEEE Trans Acoust Speech Signal Process ASSP-32:173–177
5. Ngan KN, Leong KS, Singh H (1989) Adaptive cosine transform coding of image in perceptual domain. IEEE Trans Acoust Speech Signal Process 37:1743–1750
6. Hsieh YP, Chang CC, Liu LJ (2008) A two-codebook combination and three-phase block matching based image-hiding scheme with high embedding capacity. Pattern Recognit 41(10):3104–3113
7. Yang CH, Lin YC (2010) Fractal curves to improve the reversible data embedding for VQ-indexes based on locally adaptive coding. J Vis Commun Image Represent 21(4):334–342

8. Wang WJ, Huang CT, Wang SJ (2011) VQ applications in steganographic data hiding upon multimedia images. IEEE Syst J 5(4):528–537
9. Antonini M et al (1992) Image coding using wavelet transform. IEEE Trans Image Process 1(2):205–220
10. Croft LH, Robinson JA (1994) Subband image coding using watershed and watercourse lines of the wavelet transform. IEEE Trans Image Process 3(6):759–772
11. McWhirter JG, Baxter PD, Cooper T, Redif S, Foster J (2007) An EVD algorithm for para-hermitian polynomial matrices. IEEE Trans Signal Process 55(5):2158–2169
12. Redif S, McWhirter JG, Weiss S (2011) Design of FIR paraunitary filter banks for subband coding using a polynomial eigenvalue decomposition. IEEE Trans Signal Process 59(11): 5253–5264
13. Vaidyanathan PP (1998) Theory of optimal orthonormal subband coders. IEEE Trans Signal Process 46(6):1528–1543

Detection of Atrial Arrhythmias
Using Autoregessive Modeling

Ankita Mishra, Chandan Pandey and Sitanshu Sekhar Sahu

Abstract Accurate detection of atrial fibrillation arrhythmia is important for automatic monitoring and treatment of patient. Surface electrocardiography (ECG) is the primary diagnostic tool in detecting and differentiating different type of arrhythmias. In this paper, we have presented an efficient method to detect the atrial fibrillation (AF) and atrial flutter (AFL) using autoregressive (AR) modeling. AR modeling captures the trend present in the signal and thereby fit the model to the signal. The coefficients of AR model act as the features of the ECG signal and then input to support vector machine classifier to predict the AF and AFL from normal sinus rhythm. The efficiency of the method has been assessed with the data generated from bench mark databases in Physionet bank repository. The proposed approach provides 98% accuracy in the fivefold cross-validation test.

Keywords Atrial fibrillation · Atrial flutter · Autoregressive modeling
Electrocardiogram · Support vector machine

1 Introduction

Electrocardiogram (ECG) reflects the underlying electrical activity of the human heart which is nothing but a series of electrical depolarization and repolarization that occurs in the cardiac muscles. Any deviation from the normal heart rhythm is also reflected in the surface electrocardiogram. So analysis of electrocardiogram for

A. Mishra (✉) · C. Pandey · S.S. Sahu
Department of Electronics and Communication Engineering,
Birla Institute of Technology Mesra, Ranchi, India
e-mail: find.ankitamishra@gmail.com

C. Pandey
e-mail: thechandan94@gmail.com

S.S. Sahu
e-mail: sssahu@bitmesra.ac.in

the diseases has been used for a very long period of time. Physicians rely mostly on the visual perception of morphological changes in the heart rhythm. But there are several diseases where the heart rhythm is almost similar in pattern; in such case, detecting the small morphological changes is a challenging task and may lead to false identification. Thus, the diseases should be analyzed quantitatively not qualitatively by means of a computer-aided tool.

Atrial fibrillation (AF) is the most commonly occurring cardiac arrhythmia encountered in clinical practice. It is a supraventricular tachyarrhythmia because the origin of such kind of arrhythmia lies with the upper part of ventricles or in other words atrial region [1]. Hence, it is also called as atrial tachyarrhythmia. In recent years, it has gained a considerable research interest because the underlying mechanisms of initiation, maintenance and termination of AF are not fully understood. Several theories had been developed in this regard but none of them are fully sufficient for the analysis of such an arrhythmia. Wijffes et al. [2] showed that structural remodeling takes place in the atria during AF episodes which in turn tell that AF is a progressive arrhythmia.

The surface electrocardiogram has been proved to be an effective tool in analyzing such kind of arrhythmia. AF is well characterized by rapid and disorganized atrial activity in terms of timing and morphology. Discrete P waves that can be clearly distinguished in a normal sinus rhythm is completely absent and replaced with low-amplitude fibrillatory waves (F waves) resulting in an oscillating baseline. The pattern of F waves varies from patient to patient. In some cases, it may be regular with high amplitude; while in other cases, it may be less regular or completely irregular with low amplitude or both. The fibrillation rate varies from 240 to 540 waves/min [3]. Atrial flutter (AFL) is another type of supraventricular tachyarrhythmia. AFL occurs due to the structural change in the heart that generates a continuous, localized reentrant pathway of electrical signals. Such continuous loop of electrical signals makes the atria to contract very rapidly resulting in an atrial rate of 240–320 waves/min [3]. There exist a substantial overlap between the rates of atrial fibrillation and atrial flutter. In AFL, ECG has the characteristic sawtooth wave shape. The high and rapid atrial activity affects the regular ventricular activity and often leads to an irregular ventricular rhythm. While the ventricular rhythm of AF is irregularly irregular, AFL shows either regular or regularly irregular pattern (maintaining a pattern of irregularity).

In both of these cases, ventricles are not filled to their capacity and hence lead to inadequate blood output. This may affect the normal work of vital organs and tissues and cause many severe diseases. AFL, though less prevalent than that of AF, patients diagnosed with AFL are more likely to develop AF. Hence, in clinical perspective, it is important to distinguish these two since the treatment may be different for them.

The complexity and progressive nature of this arrhythmia has led to develop different signal processing techniques. Several signal processing techniques have been developed in the urge of a detailed and quantitative analysis of this arrhythmia. Stridth et al. [4] have used time–frequency analysis-based techniques to

produce trends of AF frequency where the Wigner–Ville distribution technique and Choi–William distribution techniques are used for the short-term and long-term time–frequency analysis, respectively. Methods based on RR intervals have also been developed, where the standard and test density histograms of RR interval and time difference of successive RR interval are employed for the detection of AF [5]. Methods based on RR interval irregularity have used different techniques such as variance of selected RR intervals [6], hidden Markov model [7], and other statistical methods such as Poincare plot(tells the variability of RR interval series), sample entropy(quantifies the complexity of RR interval series) [8]. P-wave-based method performs a sequential analysis to detect the absence of P waves. All these methods discussed above have some limitations associated. Detecting the absence of P waves is difficult enough because of its small amplitude, for which atrial activity is extracted. Atrial activity and ventricular activity overlaps with each other hence difficult to delineate which in turn make the signal analysis more complicated. There are several episodes of AF have been detected where RR intervals are regular. So methods based on RR interval failed in accurate detection where ventricular activity is regular or the ECG changes quickly between rhythms.

Keeping in view the limitations of the existing methods, in this paper, we have presented an efficient method which uses the ECG signal characteristic to distinguish the different atrial activities such as atrial fibrillation and atrial flutter from normal sinus rhythm. First, the features of the ECG signal are extracted using AR modeling, and then, these features are input to the support vector machine (SVM) to classify the arrhythmias.

2 Materials and Methods

2.1 Database

The datasets used in this study are obtained from three widely used public domain databases such as MIT/BIH Atrial Fibrillation Database [9] and MIT/BIH Normal Sinus Rhythm Database [10] and MIT/BIH Arrhythmia Database [11] available at Physionet bank archive. The Atrial Fibrillation Database comprises of 23 AF ECG recordings and each are sampled at a rate of 250 Hz. The MIT/BIH NSR database consists of 18 normal sinus rhythm recording sampled at a rate of 128 Hz. The MIT/BIH Arrhythmia Database is used to obtain the episodes of atrial flutter which are samples at 360 Hz. It consists of three 3 records of atrial flutter. From each of these databases, we have collected several records of 10,000 samples each (40 s length for fibrillation, 78 s length for normal sinus, and 28 s length for flutter) for the different ECG type. In total, 272 recordings are extracted out of which 170 recordings are of NSR type and 84 recordings are of AF type and the remaining 18 are of AFL type.

2.2 Feature Extraction Using Autoregressive Modeling

Features are the characteristic representation of the signal, which indicates the trend or pattern present in the signal. In this paper, we have considered the statistical properties of the signal and used them to classify the different classes of signals. Typically, ECG signal is a time series or a sequence of data points associated with time. The AR modeling is the most widely used method in time series modeling because it does not require any priori information about the signal to be analyzed and also provides simple linear equations with fewer numbers of coefficients in comparison to other models [12]. It also predicts the data sequence accurately even if in a low signal-to-noise ratio condition. The AR modeling has been used in many applications such as pattern recognition, image processing, radar signal processing, biomedical signal processing, and speech processing. Autoregressive model is a linear model which models the time series data in a linear fashion i.e. the each sample of the signal is modeled as a linear combination of the past values [13].

Let y_1, y_2, y_3,..., y_n be the samples of the time series $Y(n)$. The pth order autoregressive time series ($AR(p)$) is defined as

$$Y(n) = \sum_{k=1}^{p} \alpha_p(k) Y(n-k) + \varepsilon(n) \qquad (1)$$

where p is the order of the model $\varepsilon(n)$ is assumed to be a zero mean white noise sequence with variance σ^2. The AR model parameters α_p can be calculated by different methods. Here, we have computed the parameters using Burg's method and the model order are selected empirically. Burg method is based on the principles of minimization of forward and backward linear prediction errors by selecting appropriate prediction coefficients subject to the condition that they must satisfy the Levinson–Durbin recursive algorithm. The ECG signal is modeled using the AR model and coefficients act as the features to represent the ECG signal.

2.3 Support Vector Machine (SVM)

Support vector machine is a class of learning machines based on optimization principle from statistical learning theory, originally introduced by Cortes and Vapnik [14]. It has been widely applied in many areas of pattern recognition, regression, and classification. The objective of SVM is to separate the training data with a maximum margin while maintaining reasonable computing efficiency. SVM constructs an optimal separating hyperplane in the input space that separates the data into different classes by maximizing the margin between the two classes. The hyperplane is represented as

$$f(x) = \omega^\mathrm{T} x + b = 0 \tag{2}$$

where ω is an adjustable weight vector and b is a bias.

The hyperplane can be found by minimizing the following cost function

$$J(\omega) = \frac{1}{2}\omega^\mathrm{T}\omega = \frac{1}{2}\|\omega\|^2 \tag{3}$$

subject to the constraints

$$d_i[\omega^\mathrm{T} x_i + b] \geq 1 \quad i = 1 \ldots N \tag{4}$$

The constrained optimization problem is solved using the method of Lagrange multiplier.

3   Results and Discussion

To assess the distinguishing capability of the extracted features, we have used the Andrews plot. Andrews plot is a way to visualize the latent structure in high-dimensional data. Andrews curves are based on the principles of Fourier series where the data points are mapped into curves in a higher dimensional space by means of orthogonal basis functions. The Andrews plot of the AR features of fibrillation, normal sinus rhythm, and atrial flutter is shown in Fig. 1. From the
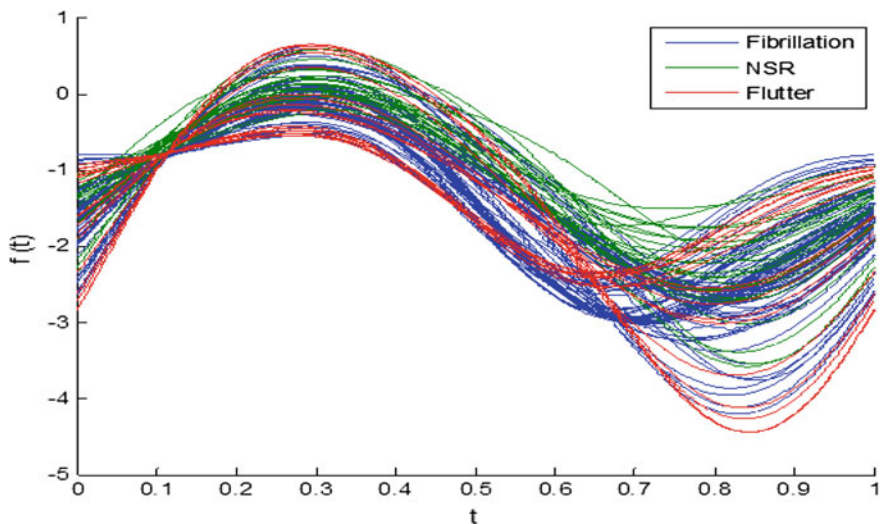
Fig. 1 Andrews plot to visualize the fibrillation, flutter, and normal sinus ECG data

variations in the plot, it is deduced that the features are capable of discriminating between the different classes. Thus, the machine learning approach can be used to classify the different ECG type.

The support vector machine has been used to classify the fibrillation, flutter and normal sinus ECG signals. The performance of the method is evaluated using the performance metrics such as accuracy (Acc) and sensitivity (Sen) and specificity (Spe). All the metrics are based on the parameters, true positive (TP), false positive (FP), true negative (TN), and false negative.

Accuracy of the classifier is defined as the degree of closeness of a measured value to the actual value.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

Sensitivity is the rate of positive instances that are correctly identified as positive.

$$Sen = \frac{TP}{TP + FN} \tag{6}$$

Specificity is the rate of negative instances that are correctly identified as negative.

$$Spe = \frac{TN}{TN + FP} \tag{7}$$

The experimental study has been carried out in MATLAB environment. The performance of the proposed method is computed through fivefold cross-validation test and the results are listed in Table 1. The sensitivity, specificity, and overall accuracy are evaluated in 50 runs and the average of the runs is presented in the Table 1. The optimum value of the σ and C parameter in SVM is found to be 2 and 100. We have studied the classification performance of the method by varying the AR model order. It shows that the model order of 6 provides best result. The atrial fibrillation is predicted with 98.62%, the atrial flutter identified with 99.78%, and normal sinus

Table 1 Performance evaluation by fivefold cross-validation test on the collected dataset by SVM (average over 50 runs)

AR model order	Sensitivity			Specificity			Overall accuracy
	Atrial fabrillation	Normal sinus rhythm	Atrial flutter	Atrial fabrillation	Normal sinus rhythm	Atrial flutter	
3	93.69	99.29	87.78	99.21	94.20	97.46	96.80
4	95.45	99.24	85.89	98.18	93.86	98.20	97.18
6	98.62	97.54	99.78	98.56	99.88	98.89	98.96
8	99.62	95.82	97.00	98.00	98.84	98.09	97.11
10	98.98	95.72	98.44	97.67	99.76	98.18	96.90
12	99.52	96.88	98.44	97.17	99.65	97.76	97.90

Table 2 Confusion matrix of the classifier in model order 6 using SVM

	Atrial fibrillation	Normal sinus rhythm	Atrial flutter
Atrial fibrillation	84	2	0
Normal sinus rhythm	0	167	0
Atrial flutter	0	1	18

rhythm identified with 97.54% accuracy in average. Also the specificity of the method is shown to be around 98% in all the ECG type. The classification accuracy of the method in terms of confusion matrix in a single run for the model order 6 is shown in Table 2. The atrial fibrillation is predicted with 100%, atrial flutter also predicted with 100%, and normal rhythm is predicted with 98.23%. The performance results elucidates that the autoregressive model capture the variation in the signal in different ECG signals in a better way to distinguish them properly.

4   Conclusion

In this paper, we have presented an accurate classification approach to detect the atrial fibrillation and atrial flutter episodes from normal sinus rhythm using autoregressive modeling with support vector machine. It classifies the three classes of ECG signal in 98% overall accuracy. The presented approach is very simple and suitable for tele-monitoring applications.

References

1. Fuster V, Ryden LE, Cannon DS, Crijns HJ, Curtis AB (2006) ACC/AHA/ESC 2006 guidelines for the management of patients with atrial fibrillation: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines and the European Society of Cardiology Committee for Practice Guidelines (Writing Committee to Revise the 2001 Guidelines for the Management of Patients with Atrial Fibrillation): developed in collaboration with the European Heart Rhythm Association and the Heart Rhythm Society. Circulation 114(7):257–354
2. Wijffels MC, Kirchhof CJ, Dorland R, Allessie MA (1995) Atrial fibrillation begets atrial fibrillation. A study in awake chronically instrumented goats. Circulation 92:1954–1968
3. Stridh M, Bollmann A, Olsson SB, Sörrnmo L (2006) Detection and feature extraction of atrial tachyarrhythmias. IEEE Eng Med Biol 25(6):31–39
4. Sornmo SL (2002) Shape characterization of atrial fibrillation using time-frequency analysis. Comput Cardiol 29:17–20
5. Tateno K, Glass L (2001) Automatic detection of atrial fibrillation using the coefficient of variation and density histograms of RR and ΔRR intervals. Med Biol Eng Comput 39:664–671
6. Logan B, Healey J (2005) Robust detection of atrial fibrillation for a long term telemonitoring system. In: Computers in Cardiology, pp 619–622

7. Young B, Brodnick D, Spaulding R (1999) A comparative study of a hidden Markov model detector for atrial fibrillation. In: Proceedings of the 1999 IEEE signal processing society workshop, pp 468–476

8. Mabrouki R, Khaddoumi B, Sayadi M (2014) Nonlinear statistical methods for atrial fibrillation detection on electrocardiogram. In: International conference on electrical sciences and technologies, Maghreb

9. The MIT-BIH Atrial Fibrillation Database (online). http://physionet.org/physiobank/database/mitdb/

10. The MIT-BIH Normal Sinus Rhythm Database (online). http://physionet.org/physiobank/database/mitdb/

11. The MIT-BIH arrhythmia Database (online). http://physionet.org/physiobank/database/mitdb/

12. Arnold M, Miltner WHR, Witte H (1998) Adaptive AR modeling of nonstationary time series by means of Kalman filtering. IEEE Trans Biomed Eng 45:553–562

13. Makhoul J (1975) Linear prediction: a tutorial review. Proc IEEE 63:562–580

14. Vapnik V (1998) Statistical learning theory. Willey, New York

Interpretation of IRS-1C LISS Imagery for Vegetation Extraction: A Spectral Perspective

Rubina Parveen, Subhash Kukarni and V.D. Mytri

Abstract The ecological mapping and monitoring can be done by satellite remote sensing with repetitive, multispectral and synoptic viewing capabilities. This paper reports on the interpretation of vegetation by exploring of some of the existing interpretation methods such as NDVI and K-means clustering, with emphasis on spectral responses of vegetation features. Indian Remote Sensing IC Linear Integrated self-scanning (IRS IC-LISS III) imagery data set was used for the study. The results show that spectral responses play an important role for automatic vegetation feature extraction by setting the predefined thresholds. The interpretation was based on the spectral reflectance of the vegetation pixels. Visual analysis of the resultant image depicts vegetation features extracted from the study image and discriminable classification of the vegetation was also achieved.

Keywords Spectral response · Resolution
NDVI (normalized difference vegetation index) · IRS-1C LISS III imagery

1 Introduction

Remote sensing image analysis faces number of challenges such as satellite sensors, resolution and time of acquisition, atmospheric conditions, qualitative and quantitative interpretations of the image. Land use and land cover maps play a significant role in the ecological management. Due to increased resolution, decreasing cost of satellite imagery and sophisticated automatic image processing techniques, land use mapping by satellite imagery has become more viable [1]. Reflected or emitted

R. Parveen (✉)
VTU, Belagavi, Karnataka, India
e-mail: rubina16det@gmail.com

S. Kukarni
PESIT, Bangalore South Campus, Bangalore, Karnataka, India

V.D. Mytri
AIET, Kalaburagi, Karnataka, India

energy is recorded by the satellite sensors and processed to give the remote sensing satellite image. Identification of features and its discrimination and classification of geographical features depends on the uniqueness of reflectance/emittance properties of those features in different parts of the electromagnetic spectrum [2]. Interpretation of vegetation biophysical properties from remote sensing imagery by spectral vegetation indices is predominant [3]. Spectral signature of any feature in a satellite remote sensing image is the observable characteristics, which leads to the identification of that feature and further analysis can be done to determine its conditions. Principal characteristics of signatures of vegetation can be determined by its spectral, spatial, temporal and polarization variations. Spectral variations occur because of changes in the reflectance or emittance of geospatial features as a function of wavelength. This paper evaluates the capability of several spectral and texture features for vegetation classification at the species level using IRS-1C LISS III multispectral imagery. This method sequentially performs image segmentation and classification to extract vegetation features. Quantitative assessment gives the overall accuracy of vegetation extraction from the study image. Visual comparison gives qualitative assessments. The structure of the paper was arranged as follows: Sect. 1 included introduction and Sect. 2 included the details of study area and the tools used for analysis. The proposed method of vegetation segmentation was explained in Sect. 3. Section 4 shows the results of the proposed work followed by conclusion.

2 Study Area Particulars

The download study image is an image of IRS-1C LISS III sensor of Indian Remote Sensing Satellite Resourcesat-1 LISS III. Ortho-rectified images have green, red, NIR and SWIR four band multispectral data with 24 m spatial resolution and a swath of 141 km, downloadable as image tiles in Geotiff format. Data acquired on 04 January 2013. Simulation was performed by using MATLAB (R2010a) and ERDAS Imagine 9.1 software.

2.1 Earth Observations

The study area is Gulbarga District area (Karnataka State, India), 76.75 E 17.5 N–77.0 E 17.5 N bounding boxes and toposheet no. E43Q15. Gulbarga is situated in Deccan Plateau of the country, black soil. The district has a many of tanks which, in addition to the rivers, this water is not sufficient for cultivation. Gulbarga region is dry area and agriculture is mostly rain fed. Study area is predominantly urban city with very small water bodies and little vegetation. The Fig. 1 shows the area of analysis which is radiometric and geometric preprocessed.
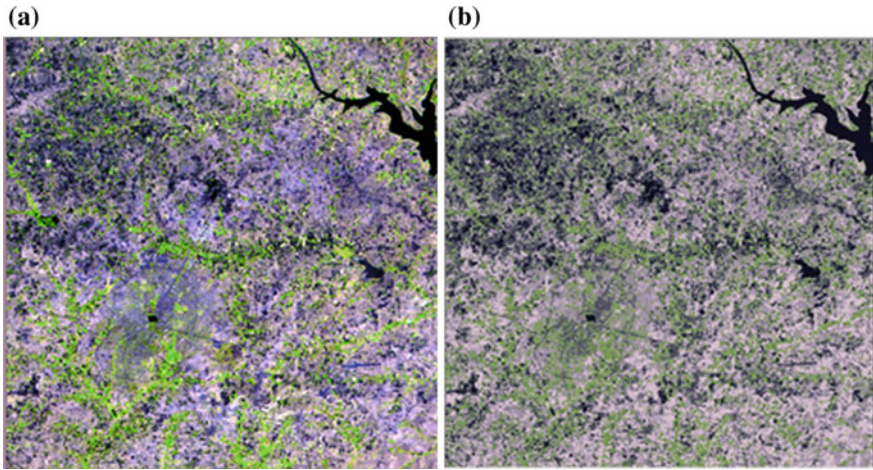
Fig. 1 **a** IRS-1C LISS III Image acquired on 04 January 2013, **b** Classified Image

Table 1 Covariance statistics of IRS-1C LISS III Image acquired on 04 January 2013

Band	Covariance				Univariate statistics			
	Band 1	Band 2	Band 3	Band 4	Minimum	Maximum	Mean	Std. deviation
1	13.6	16.607	15.723	14.756	65	101	78	3.68
2	16.6	67.164	130.519	148.106	43	90	64	8.19
3	15.7	130.519	353.519	383.378	34	128	81	18.8
4	13.8	148.106	383.378	459.234	20	100	66	21.43

Different class pixels show different reflectance at different spectral wavelengths with spectral features such as mean reflectance, maximum reflectance, minimum reflectance and standard deviation. This can be used for spectral classification [4]. As the study image was an urban environment based, statistics in Table 1 shows the less predominance of vegetation. Four radiometric bands of the study image have different covariance and univariate statistics. Band 4 shows maximum covariance, whereas band 2 shows maximum variance with respect to band 1. Maximum standard deviation in band 4 indicates that the intensities of image pixels are spread out over a wider range of values showing large variation.

The multispectral image was divided into spectrally homogeneous but non-contiguous [1] clusters using unsupervised classification [5]. Broad unsupervised classification of Land use of different classes was done in ERDAS Imagine 9.1 shown in Fig. 1b, and its class relative area of occupancy is shown in Table 2. The study image was classified into five classes. Area of occupancy of each class and its relative percentage was calculated. Land cover statistics of classified image was tabulated in Table 2. Green colour pixels in Fig. 1b depict vegetation pixels. These can be interpreted as vegetation, by visual comparison with Fig. 1a.

Table 2 Segmented image is relative classes of land cover

Color	Classes	pixels	Area(Sq mts)	Area in percentage
	Class 1	88545	2080808	6.683645
	Class 2	328728	7725108	24.81339
	Class 3	303742	7137937	22.92737
	Class 4	413085	9707498	31.18091
	Class 5	190701	4481474	14.39469
		1324801	31132824	100

Fig. 2 Comparison of types of vegetation reflectance spectra with respect to the bands
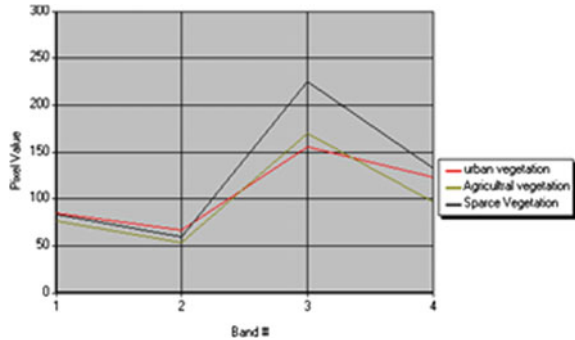
Spectrally homogenous clusters may not be spatially contiguous and may consequently result in salt-and-pepper effect [6]. By visual interpretation, class 3 depicts vegetation covers 22.92% of area with respect to complete image features classes.

2.2 Spectral Analysis of Input Image

Spectral responses of vegetation are unique and types of vegetation shows little difference between each other in that range of vegetation features. The spectral reflectance of vegetation features was quite distinct. Figure 2 demonstrates that vegetation had a very high reflectance in the NIR (near infrared region), and there were three low minima due to absorption [2]. Measurement of wavelengths, measurement of intensities of pixels and (3) interpretation of the image [7, 8] were the important aspects in spectral analysis. Figure 2 shows the spectroscopy of types of vegetation of IRS-1C LISS III Image acquired on 04 January 2013. Modulation of spectral reflectance curve shows there are different spectral signatures for further discrimination of vegetation. Vegetation can be identified by shape of the reflectance spectra.

Table 3 Different spectral band and their applications for vegetation analysis

| Band	Wavelength (mm)	Nominal spectral location	Principal applications
1	0.45–0.52	Blue	Forest mapping
2	0.52–0.60	Green	Peak reflectance of green vegetation
3	0.63–0.69	Red	Plant species differentiation based on chlorophyll absorption
4	0.76–0.90	Short-wave Infrared (SWIR)	Determining vegetation types' discrimination

Total moisture content in the plant, its pigments, structure of leaf and incident solar radiation were important factors, which effects the spectrum in the visible and NIR (near infrared) regions [9]. Absorbed total solar radiations in NIR region are directly proportional to the total water content of plant [10]. Soil reflectance decreases with increase in the organic matter content [2, 11]. Reflectance in the red region (about 0.6–0.7) is low because of absorption by leaf chlorophyll, and the infrared region (about 0.8–0.9) shows high reflectance because of leaf structure. Discriminative vegetation information is based on radiometric pixel property; different land uses features are not entirely spectrally divergent [1]. Table 3 shows wavelengths of four spectral bands and their relative principal applications with respect to vegetation.

Specific pixels based on their attributes are often counted to give estimated area give quantitative analysis. Extracting information based on visual inspection, its success depends upon the analyst exploiting effectively the spatial, spectral and temporal elements present in the image.

3 Vegetation Segmentation

Fundamental mechanism for understanding vegetation features in IRS-1C LISSIII imagery is dependent on reflectivity variations of the land cover of multispectral bands. Blue band with faint reflectance variability and discrimination for vegetation covers is low [1]. First three spectral bands (G, R and NIR) were used to enhance their visible clarity and spatial resolution. Spectrally segmentation on the spectral characteristics of vegetation over different wavelength regions can be done by PCA-based method [12] and spectral mixture analysis (SMA) method [13]. Texture information is dependent on the neighbouring pixels information. To reduce the effect of haziness and simultaneously to preserve the edges, study image was preprocessed with a partial differential-based equations [14]. Figure 3 shows the preprocessed and histogram-equalized image. Vegetation features are visible sharp.

The 3-dimensional feature corresponding to the radiometry in the NIR, R and G bands were used to describe every pixel spectrally [5]. Discrimination in vegetation

Fig. 3 Preprocessed image
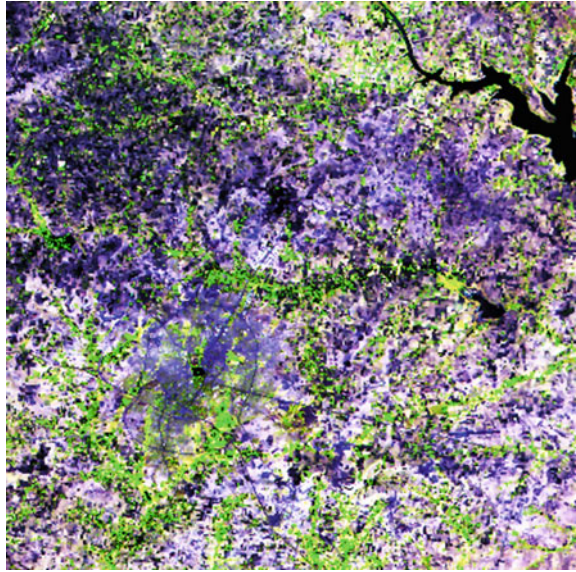
Fig. 4 Segmented image
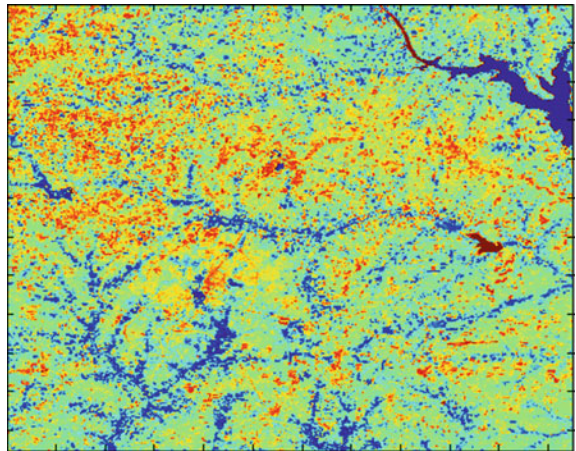
feature was possible by calculating separately the derivatives of the spectral channels. Vegetation appears green from visual spectrum because of lighting and viewing conditions for IRS LISS III images; this affects the classification if proper descriptors were not used. Vegetation indices [3], Gary level co-occurrence matrix, were employed as vegetation feature descriptors for vegetation mapping. The initial segmentation was performed using the spectral values of the study image in conjunction with the NDVI (normalized difference vegetation index) [15]. NDVI was calculated from NIR (vegetation reflectance in NIR band is maximum) and red band

(RED) reflectance values [16] from formula number 1. Figure 4 shows the NDVI image.

$$NDVI = (NIR - RED)/(NIR + RED) \qquad (1)$$

4 Results

K-means clustering was used to group the pixels into a number of clusters using the Euclidean distance. The vegetation was divided into four classes in the study. For hierarchical clustering of vegetation, clusters are either merged into larger clusters or split to smaller clusters into feature space by K-means clustering, into K number of clusters [17, 18]. K-means algorithm randomly assigns the centroid, and then, distance of the object pixel to the centroid is determined followed by grouping of the objects based on minimum distance with centroid. All these steps were iterated until convergence. The initialization input number of clusters in K-means clustering application done in advance, by manual counting [19]. Vegetation and non-vegetation pixels were separated. Figure 5 shows the extraction of vegetation features by masking non-vegetative features. This image shows no contamination if non-vegetation feature spectra. Number of non-black pixels depicts vegetation pixels, as shown in Fig. 5. Total pixels in the resultant image depicting vegetation were 31.42%. As per the general unsupervised classification and Table 1, the

Fig. 5 Vegetation extracted image
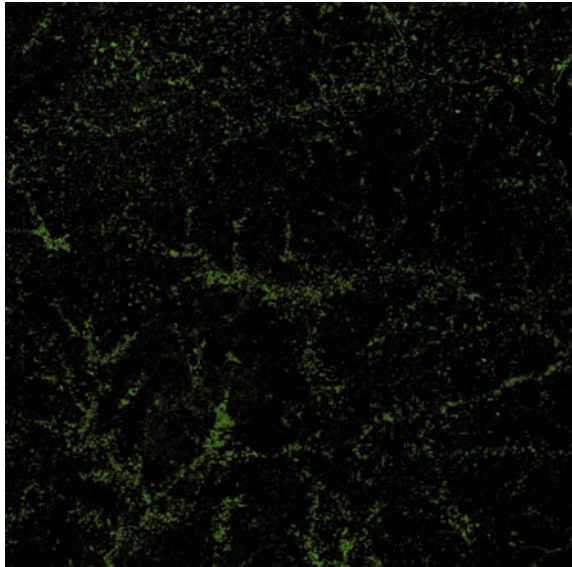
vegetation pixels were 22.92%. Obtained results were about 72% analogous to unsupervised classification, but Fig. 5 was visually clearer.

The spectral modulations of the same features are different in different seasons and different times. Exact matching of these features is difficult with standard spectrum [20]. But the results can be nearer to it. By visual interpretation, pure pixel is chosen to carry out spectral decomposition [21] by spectral mixture analysis or by SVM [4]. Euclidean distance mapping is used to automatic discrimination of the different vegetation features. Euclidean distance represents images as points in a high dimensional image space. If $I = (I_1, I_2, \ldots I_{mn})$ and $J = I+1$, $J = (J_1, J_2 \ldots J_{mn})$ are the two images of size M * N. Ik and Jk are grey levels on Kth pixels. The Euclidean distance [22] $d_E^2(I, J)$ is given by formula 2.

$$d_E^2(I, J) = \sum_{K=1}^{MN} (I^K - J^K)^2 \qquad (2)$$

The pair with more similarity has a larger Euclidean distance. Each pixel was tested and divided into four classes, by calculating the Euclidean distance between given pixel and reference pixel. The classified results are shown in Fig. 5. Colour differencing gives visible discrimination. Four classes of vegetation with significant variations are observed from the resultant image as shown in Fig. 6. Despite the general ambiguity in their spectral response, this helps for vegetation discrimination. Each class can be further discernible as a separate image.

Table 2 shows the vegetation region extracted from the input image. NIR band reflectance changes depending upon the plant leaf structure. Identification of different species and vegetation condition monitoring is possible. Vegetation reflectance is almost inversely proportional to leaf water content. Figure 6 is 72% accurate, as compared with Fig. 1b. Further decimation gives the resultant image shown in Fig. 6. From the Fig. 6, sky blue colour depicts class 3a, brown depicts

Fig. 6 Vegetation discrimination image
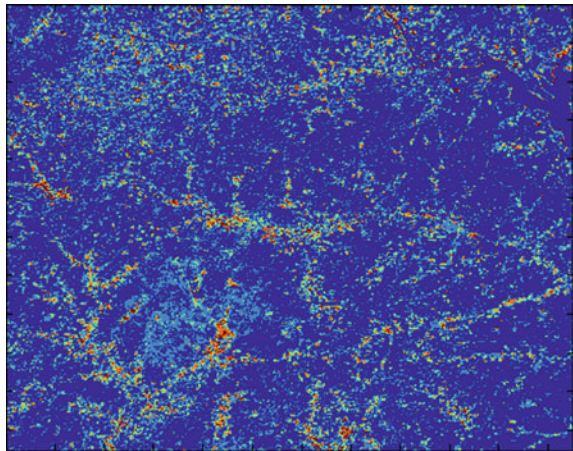
class 3b, green depicts class 3c, orange colour depicts 3d and dark blue depicts masked area or non-vegetation pixels.

5 Conclusion

Vegetation has a strong reflectance in NIR band. Vegetation species shows discriminable spectral variability because of leaf size, shape, water content and geophysical parameters. This property was explored in the study to determine types of vegetation using quantitative models such as NDVI and K-means clustering. Extraction of vegetation was 72% accurate as compared to unsupervised classification. The process was automatic without any biased classification. Further vegetation types were discriminated by the pixel spectral differences analysis by using Euclidean distance mapping. The results were visually clear, and these results can be further used for identification of vegetation species. Vegetation species identification based on spectral signatures can be taken up in the further study.

References

1. Resourcesat-1 data user's Handbook, Balanagar, Hydrabad: National Remote Sensing Agency, Government of India (2003)
2. Reddy MA (2008) Remote sensing and geographical information system. BS Publications, Hyderabad
3. Li Z, Hayward R, Zhang J, Jin H, Walker R (2010) Evaluation of spectral and texture features for object- based vegetation species classification using support vector machines, Vienna, Austria
4. Kavitha K, Arivazhagan S (2010) A novel feature derivation technique for SVM based hyper spectral image classification. Int J Comput Appl 1:25–31
5. Ursani AA, Kpalma K, Lelong CCD, Ronsin J (2012) Fusion of textural and spectral information for tree crop and other agricultural cover mapping with very-high resolution satellite images. IEEE J Sel Top Appl Earth Observ Remote Sens 5:525–535
6. Kartikeyan B, Sarkar A, Majumde L (1998) A segmentation approach to classification of remote sensing imagery. Int J Remote Sens 19:1695–1709
7. Gomez RB (2001) Spectral library issues in hyperspectral issues in hyperspectral imaging applications. In: 5th joint conference on standoff detection for chemical and biological defense, vol 4
8. Burke GA, Burke HK (2003) Spectral imaging for remote sensing. Lincoln Lab J 14:3–28
9. Kalyankar FA, Kalyankar NV Major limitations of satellite images
10. Sanderson DR Introduction to remote sensing
11. Ranganath N, Jayaraman V, Roy PS (2007) Remote sensing applications: an overview. Spec Sect Indian Space Program 93:25
12. Tsai F, Lin EK, Yoshino K (2007) Spectrally segmented principal component analysis of hyperspectral imagery for mapping invasive plant species. Int J Remote Sens 28:1023–1039
13. Zhao H-b, Liu T, Yao-ping Cui (2009) Using multi-spectral remote sensing data to extract and analyze the vegetation information in desert areas-a case in the Western Gurbantunggut Desert, Wuhan, China: IEEE Computer Society

14. Liu P, Huang F, Li G, Liu Z (2011) Remote-sensing image denoising using partial differential equations and auxiliary images as priors. IEEE Geosci Remote Sens Lett 9:358–362
15. Zillmann E, Gonzalez A, Herrero EJM, van Wolvelaer J, Esch T, Keil M, Weichelt H, Garzón AM (2014) Pan-European Grassland mapping using seasonal statistics from multisensor image time series. IEEE J Sel Top Appl Earth Obser Remote Sens 7:12
16. Bannari A, Ozbaldr A, Langlois A (2207) Spatial distribution mapping of vegetation cover in urban environment using TDVI for quality of life monitoring. Barcelona IEEE
17. Senthilnath J, Omkar SN, Mani V, Tejovanth N, Diwakar PG, Shenoy A (2007) Muti-spectral satellite image classification using glowworm swarm optimization, Vancouver, Canada. IEEE
18. Subbiah B, Christopher SC (2012) Image classification through integrated K-means algorithm. IJCSI Int J Comput Sci Issue 9:518–524
19. Koonsanit K, Jaruskulchai C, Eiumnoh A (2012) Parameter-free K-means clustering algorithm for satellite imagery application. IEEE, Suwon
20. Wang L (2011) Study to process of urban sprawl using spectral mixture analysis. IEEE, Hangzhou
21. Powell RL, Roberts DA, Dennison PE, Hess LL (2007) Sub-pixel mapping of urban land cover using multiple endmember spectral mixture analysis: Manaus, Brazi. Remote Sens Environ 106:253–267
22. Wang L, Zhang Y, Feng J (2005) On the Euclidean distance of images. Center for Information Sciences, Beijing

ECG Analysis and Peak Detection Using Zero-Aliasing Correlation Filter

Neenu Jose and Nandakumar Paramparambath

Abstract An electrocardiogram (ECG) is a test that checks how the heart is functioning by measuring the electrical activity of the heart. With each heart beat, an electrical impulse travels through the heart, in turn results in the pumping of blood from the heart. This paper intends to study how correlation filter is used for the analysis and detection of QRS pattern in ECG waveform. Correlation filters are well established and useful for a variety of tasks in signal processing and pattern recognition. The study describes the demonstrated use of zero-aliasing correlation filters which reduces the aliasing effect and offer distortion-tolerant output. By adopting the ZACF design, the study shows that the recognition and localization performance of conventional CF designs can be significantly improved. And hence, QRS detection using this method provides a highly trouble-free path, which in turn helps in the scientific development for the betterment of the society.

Keywords Correlation filter · ECG · QRS complex · Aliasing

1 Introduction

It is the natural capacity of every individual to sense their surroundings with object, numbers, alphabets, voices, faces, scene, vehicle, houses, and roads. Still, developing a computer to make out this type of identification is a demanding task. Pattern recognition plays an important role in diverse fields, and it is an area of artificial intelligence. Correlation filters (CFs) have been successfully applied to a variety of pattern recognition applications. This paper mainly focussed on the human ECG

N. Jose (✉) · N. Paramparambath
Department of ECE, NSS College of Engineering, Palakkad, India
e-mail: neenumol92@gmail.com

N. Paramparambath
e-mail: nandakumarpp@hotmail.com

recognition for cardiologic diagnosis. ECG is the standard tool for monitoring and diagnosing cardiac problems. Processing of ECG signal by using existing ECG library data includes ECG waveforms parameters like P wave, QRS complex, and T wave. In this, QRS complex is most important. The duration and height of QRS gives significant amount of data to physician; thus, he/she can easily understand the condition of heart. Such processed signal undergoes pattern recognition. The diverse forms of noise-like baseline wander, motion artifact, etc., contaminate the ECG signals. Thus, the analysis of ECG signal with naked eye becomes difficult.

Therefore, different algorithms are developed in order to detect the QRS peak. Mainly, those algorithms originated from artificial neural network, genetic algorithm, wavelet transform, and correlation filters. Pam Tompkins algorithm uses the basics of signal derivatives for detection of slop of the QRS complex, which is considered as the first algorithm for QRS detection [1]. Neural network also one of the prominent methods used for QRS detection [2]. R peak detection and single beat extraction ECG signal using cross-wavelet transform was explained in [3]. The structures of these above algorithms are almost similar which includes steps like pre-processing stage, feature extraction stage, and decision stage.

In this paper the analysis and detection QRS complex is developed by using the concept of correlation filter. The newly introduced zero-aliasing correlation filter is used instead of conventional correlation filter. To compute the similarity between a reference pattern and a test pattern, correlation is the natural metric phenomenon that is widely used [4]. In the pattern recognition applications, correlation filter helps in classifying and recognizing patterns in its best possible manner. The template that is specifically designed based on a set of training images is a correlation filter. The training images are from both an authentic class and the non-authentic class. When the test image is correlated by correlation filter, a correlation peak is obtained if the test image comes from an authentic class, while the correlation peak is not sharp if the test image belongs to the non-authentic class [5].
.

2 Zero-Aliasing Correlation Filter

Zero-aliasing correlation filter can be designed by incorporating zero-aliasing constraint into CF design. In the ZACF formulation, minimize localization loss while enforcing both the peak and ZA constraints that force the template's tail to be equal to zero. The resulting aliasing effects caused by circular correlation can be reduced by using this constraint. For example, the ZA constraint for a K channel template $h(n)$ can be expressed as

$$h^k(n) = 0 \quad \text{for} \quad n \geq N \tag{1}$$

Relating the time domain template $h^k(n)$ to the frequency domain $H^k(r)$, by taking the inverse DFT as,

$$h^k(n) = \frac{1}{N_F} \sum_{r=0}^{N_F-1} H^k(r) e^{\frac{j2\pi rn}{N_F}} \tag{2}$$

The zero-aliasing constraint for the template is given by,

$$h^k(n) = \frac{1}{N_F} \sum_{r=0}^{N_F-1} H^k(r) e^{\frac{j2\pi rn}{N_F}} = 0 \quad \text{for} \quad N \le n < N_F \tag{3}$$

This ZA constraint characterizes the ZACF.

Similarly, the peak constraint can express as the correlation filter that are constrained to produce pre-specified value as the correlation peak constraint in response to the training image that is equivalent to the dot product of the training images and the template. In the frequency domain, correlation may be represented as an element-wise multiplication between the DFTs of two signals. However, it is well known that multiplying two DFTs together results in a circular correlation, rather than a linear correlation.

To fix the circular correlation problem, we introduced hard constraints on the CF design that force the tail of the template to zero. We require that training images are padded and a DFT of size $2N - 1$ is used. By zero-padding training images and imposing constraints on the filter design, we obtain a template that is all zeros in the tail. The element-wise multiplication of the DFT of this template with DFTs of padded training images corresponds to a linear correlation, rather than a circular correlation. Thus, the correlation peak is set to 1 for positive class training images and set to 0 for negative class training images. This leads to good localization by exhibiting sharp correlation peak with low side lobes.

3 Designed Methodology

The basic type of processing ECG signal is filtering and normalization. The ECG signal obtained from MIT-BIH database by the Physionet Website. This Website provides the physiological signals for researchers. MIT-BIH database contains sets of normal ECG signal and the sets of ECG signal that affected by abnormalities. This provides an opportunity to test and detect the QRS complex. The QRS pattern is recognized by introducing zero-aliasing correlation filter under the process of testing and training. The methodology includes averaging or normalization of ECG signal along with adaptive thresholding, QRS complex detection, single QRS beat extraction, and correlation of testing and training signals.

3.1 R Peak Detection

QRS pattern is the widely used ECG signal feature in the process of ECG analysis. For better extraction of QRS complex, fixed thresholding is being adopted rather than adaptive thresholding. R peak detection process is channelized by the application of FFT-based correlation which results in the elimination of the aliasing issue that raised due to circular correlation. And then, with the result of the R peak detection, the corresponding RR interval is computed.

Fig. 1 Raw ECG signal
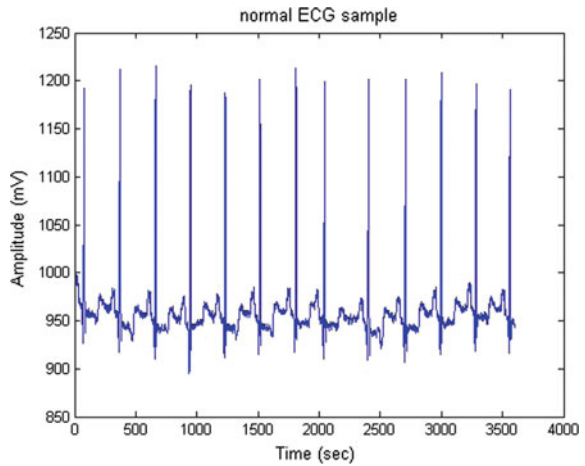
Fig. 2 R peaks detection using ZACF
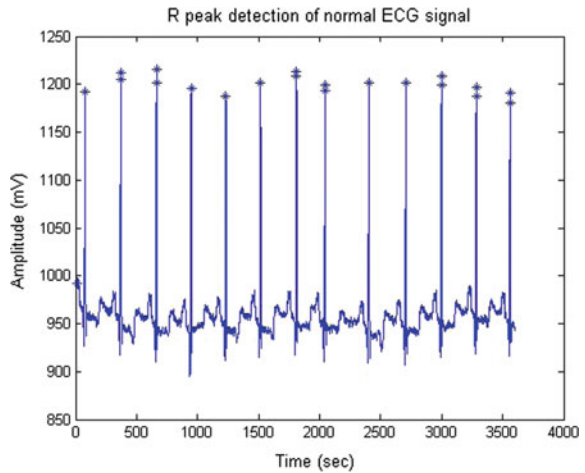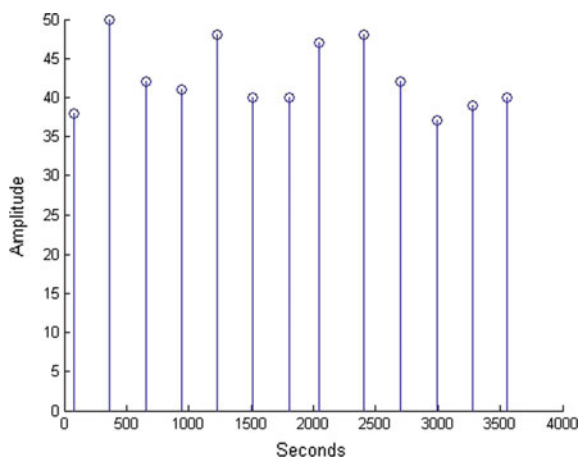
Fig. 3 The detected R peaks after normalization
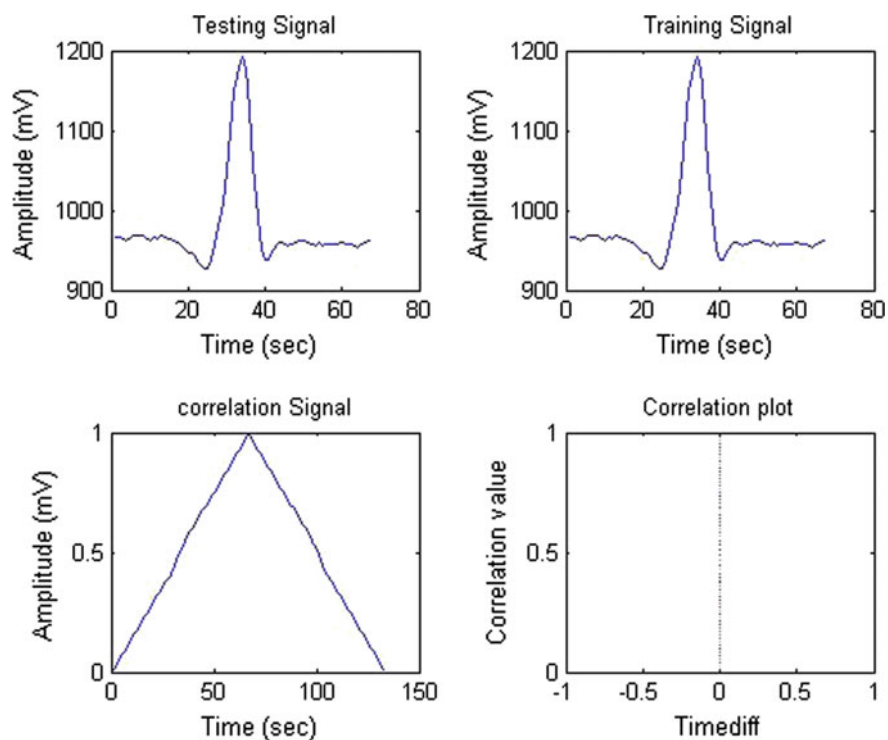
Fig. 4 Result of correlated signal

3.2 ECG Classification

Analysis of single beat QRS complex usually results in the accurate classification of the ECG signals. And hence, separate single beat is being assigned for both training and testing process. Thus, it endures zero padding along with FFT calculation, which resembles the characteristics of linear correlation. In ZACF, the zero-aliasing constraint is adopted mainly in the tail component, which in then accelerates the correlation between the templates. And hence, it positively paves a smooth and easiest way for the ECG classification.

4 Experimental Results

This paper demonstrates the use of the zero-aliasing correlation filter algorithm for the detection of the QRS complex, determination of RR interval and heart rate from the raw ECG data. The data were processed in MATLAB 2013 software. All the data used for this work are taken from the MIT-BIH database by Physionet
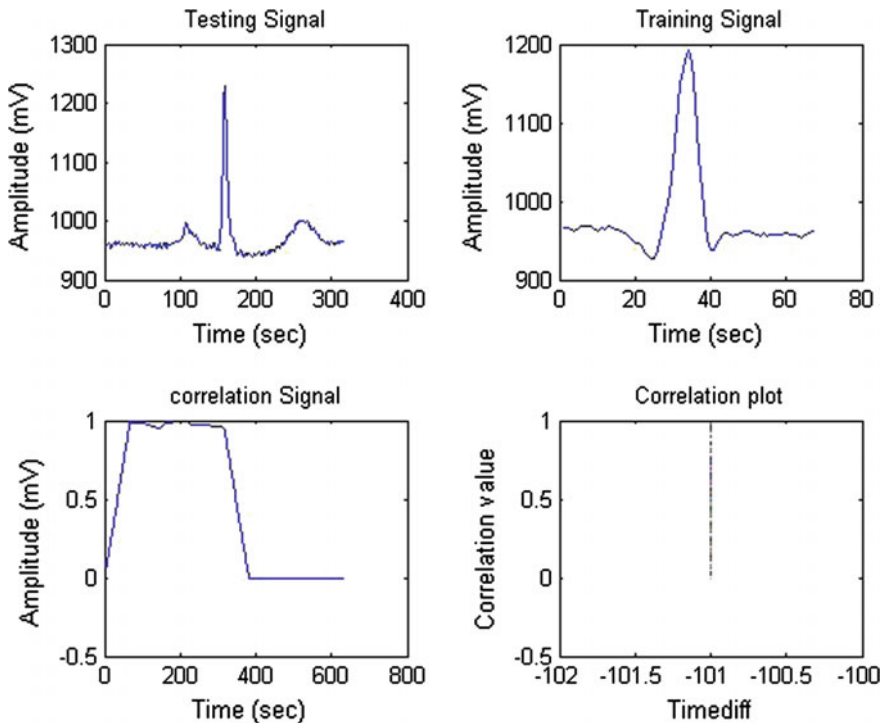
Fig. 5 Result of uncorrelated signal

Website. The sampling frequency of dataset is 360 Hz. The RR interval is the time difference between the R peaks. The heart rate is also calculated from time between the QRS complexes. Normal range is in between 60 and 100 beats per minute (bpm).

The plot in Fig. 1 represents the raw ECG signal. Then the raw ECG signal undergoes zero-aliasing correlation filtering for detection of R peaks (Figs. 2 and 3).

The correct once the R peak detection is completed the single beat is extracted for ECG classification. The simulated output of ECG classification is shown in Figs. 4 and 5. One figure results correlated signal and another figure results uncorrelated signal.

5 Conclusion

Heart diseases are the major concern for many people. So diagnosis of the heart-related issues and prevention of the same has at most importance. The checking of electrocardiogram (ECG) signal is the usual test for identifying disorders of the heart. This paper demonstrated the use of zero-aliasing correlation filter for the analysis and detection of QRS pattern in ECG waveform. Thus, was developed a computerized diagnosis system for doctors, especially cardiologists for easy diagnosis of ECG variations and classification of the ECG waveforms under normal and abnormal classes.

References

1. Banerjee S, Mitra M (2014) Application of cross wavelet transform for ECG pattern analysis and classification. IEEE Trans Instrum Meas 63(2):326–333
2. Vijaya Kumar BVK, Savvides M, Xie C (2006) Correlation pattern recognition for face recognition. Proc IEEE 94(11):1963–1976
3. Fernandez JA, Boddeti VN, Vijaya Kumar BVK (2015) Zero aliasing correlation filters for object recognition. IEEE Trans Pattern Anal Mach Intell 37:1702–1715
4. Peterkova A, Stremy M (2015) The raw ECG signal processing and the detection of QRS complex. In: IEEE European modelling symposium
5. Jaiswal GK, Paul R (2014) ECG classification with the help of neural network. Int J Electr Electron Res 2(2):42–46
6. Pandey S, Ayub S (2014) Wavelet based R peak detection in ECG signals using MATLAB. J Basic Appl Eng Res 1:101–103, print ISSN:2350-0077
7. Sharma A, Toshniwal S, Sharma R (2014) Noise reduction technique for ECG signals using adaptive filters. Int J Res Rev 7(2):187–191
8. Messaoud MB, Khelil B, Kachouri A (2009) Analysis and parameter extraction of P wave using correlation method. Int Arab J Inf Technol 6(1):40–49
9. Akilandeswari K, Sathya R (2014) Feature extraction of ECG signals for early detection of heart arrhythmia. Int J Adv Res Comput Commun Eng 3(12):1–18

10. Akhila N, Tessamma T (2015) Analysis, detection and classification of ECG signal for myocardial infarction using simple cross-correlation function. Int J Emerg Technol Adv Eng 5 (10):81–84

11. Rodriguez A, Vijaya Kumar BVK (2013) Dealing with circular correlation effects. In: Proceedings of SPIE

12. Rafique H, Rafique S (2016) Review of correlation based algorithms in signal and image processing for pattern identification. Int J GEOMATE 11(27):2695–2703

13. Viknesh V, Ram-Prashanth P (2013) Matlab implementation of ECG signal processing. IOSR J VLSI Signal Process 3(1):40–47

An Extensive Study of Visual Search Models on Medical Databases

Padmaja Grandhe, Sreenivasa Reddy Edara and Vasumathi Devara

Abstract Due to the rapid growth of medical images, user-specific ROI and object classification are the significant factors in the region-based segmentation instead of a pixel-based segmentation. Manual image annotation, classification, and filtering are not only a difficult task, but also high memory and time usage. In the visual search system, an unknown query image was given as input, relevant visual images with different diagnoses features are retrieved and then used as clinical decisions. The main goal of the visual search engine is to efficiently retrieve user-specific images that are visually identical to a selected ROI query. In this paper, a survey on traditional visual search methods is analyzed in terms of visual features and accuracy are concerned. Based on the survey performed by different visual search systems, the diagnostic efficiency is increased from 30 to 60% for clinical decision.

Keywords SIFT algorithm · Visual search · Classification · Image retrieval

1 Introduction

User Interest in the potential digital images has increased enormously over the past few years with the rapid growth of image database on the Internet. At the same time, demand for tools which can perform search and retrieval of images also has increased. Image retrieval involves retrieving images based on their visual similarity to images or image features provided by a user. Challenges with conventional

P. Grandhe (✉)
CSE Department, JNTUK, Kakinada, India
e-mail: padmajagrandhe@gmail.com

P. Grandhe
CSE Department, PSCMR College of Engineering & Technology, Vijayawada, India

S.R. Edara
ANU College of Engineering & Technology, Gutur, India

V. Devara
CSE Department, JNTUH College of Engineering, Hyderabad, India

methods of image indexing have lead to the rise of interest in methods for retrieving medical images on the basis of certain features such as texture, color, and shape that are generally referred to as an image retrieval process. Points of interest, which are points on the image that can be uniquely identified, are the most important features due to their robustness. It is usual to associate a scale and a covariant to the point to obtain descriptor invariance. There are two kinds of image retrieval methodologies, namely the text-based methodology and content-based methodology. In the text-based methodology, the images are retrieved through the text that is associated with that image. But this methodology fails in giving the user intent because the text associated with the image may not be relevant. Also, this kind of retrieval requires textual annotations. The subjectivity of the perception and the impreciseness in the annotations may cause mismatches in retrieval processes. Image searching with a specific visual query is one of the advances of the search engines in recent years. Searching for a small region of interest (ROI) in larger images is still a challenging issue. A large number of localization methods have been proposed in the context of detection of objects and recognition of objects in a database; they have not applied in scalable visual search. Traditional methods usually represent an image based on low-level features.

In the Fig. 1, traditional image search query [4, 5, 9, 10] is given input to preprocessing approach. Here, the preprocessing algorithms are used to remove noise or to enhance the brightness of the high resolution spectral images. In the segmentation step, traditional techniques such as K-means, graph-based segmentation, fuzzy-C means are used in [4, 5, 9, 10]. Finally, the number of segmented regions is identified.

CBIR system from a user perspective would find semantically similar images. A set of techniques are used to search the similar images from an image database using image features such as color, shape, and texture. Continuous research on CBIR system expanded its capability to retrieve visual information, images not only

Fig. 1 Traditional visual search using segmentation process
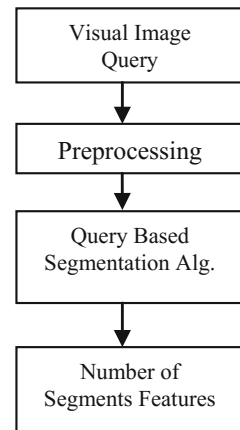
at a conceptual level, but also at the perceptual level using objective measurements of the visual contents.

1.1 Visual CBIR in Medical Domain

As the computer technology is gaining ground in the medical domain, maintenance of digital data that is generated, stored, transmitted, analyzed, and accessed in hospitals has become very complex. In particular, the amount of digital images that are being generated on a daily basis is so huge that it requires some tools to handle data efficiently. The digital images are retrieved using some textual keywords related to the records. Since the current data access techniques are having many limitations, research is being conducted to make medical imaging task easy and efficient. It is proved that visual image search based on image features is more reliable in many cases and visual similarity is easily analyzed than traditional CBIR systems. Also, medical visual characteristics have a strong effect on image diagnosis.

Digital medical image visual search consists of three key steps:

1. Query image features extraction.
2. Analysis of those features for diagnosis.
3. Recommendations for clinical decision making.

2 Related Work

The content-based retrieval systems are improvised in different areas for the better performance.

Johansen Imo, Sebastian Klenk, and Gunther Heidmann worked on **Interactive feature visualization for image retrieval** [4]. The traditional CBIR system is unaware of what features are exactly used and thus fails in analyzing the results. To make it more transparent, image features need to be more visualized for user's selection. They presented a visualization method to extract feature classes such as texture feature and color histogram to evaluate this model.

Manmatha and Ravela [6] implemented similarity of visual appearance. A multiscale vector model is used to filter the query images in the training dataset with linear Gaussian functions at various scales and then computed low-order variations. These differential variations related to feature regions and are compared with those in the training data to find the similarity score for each image.

Gao [3] proposed "**multiple feature-based image retrieval**" model on image color spaces. This system uses color feature by quantifying the color spaces into non-equal intervals using the color histogram. Color histogram is represented by using local patterns. Similarity index is computed between the user query image

and the target image in the training using Gaussian normalization on the distance space and feature space.

3   Structure-Based Image Retrieval Models

One of the most important and fastest growing researches in the field of medical is image retrieval with user's interest in content-based image retrieval. The basic idea of the structure-based image retrieval is to find identical images from the training data with the aid of some key features contained in the images. In structure-based image retrieval process [7], image registrations are performed in offline process so that the ROI of the query image was processed immediately. [7]Implemented image retrieval process in both offline and online process. In the online mechanism, user-specific query and the ROI bounding box are used to find the relevant similar images from the database [4, 5, 8–10].

Following two techniques are applied to the given input image to produce the output.

1. Image classification.
2. Global image registration.

3.1   Image Classification

To divide the X-ray images into 5 different classes such as hand, spine, cranium, chest, and negative. They used visual information to classify the images into classes. The main difficulty comes from the visual appearances of object categories. Multiple kernel learning (MKL) is used to detect a combination of the kernels, which differs either in their functionality, visual features, and parameters. Therefore, it is necessary to apply the multiple similarity measures to the available visual features. For instance, in the X-ray feature vector space has a small rectangular region inside the bounding box, this system computes a similarity value for each box of the object recursively.

Let D_b be the kernel base descriptors, and its related distance procedures are $f_1, f_2 \ldots f_{D_b}$. The descriptors and distance procedures are then kernelised to form base kernels. A novel point can be classified as +1 or −1 by computing Eq. (1).

$$\text{sign}\left(\sum_j \alpha_j y_j k(x, x_j) + b\right) \tag{1}$$

where α_i be the support vector, b is the bounded maximal width. To handle multiclass issues, both one-vs-one and one-vs-all formulations are performed. In one-v-one, region is divided into binary classification and optimal object point is

Fig. 2 SVM-MKL image classification
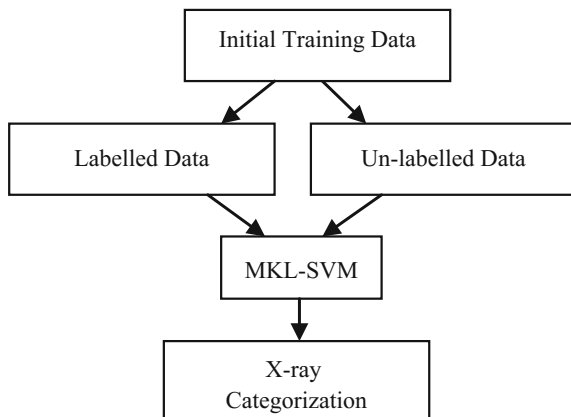
classified by taking majority vote in the classifiers. In one-vs-all, each classifier is learned per class and the optimal point is selected from the hyperplanes. They divided the object's bounding box into a number of rectangular regions and each region used as a feature vector for multiple kernel similarity computation.

As shown in Fig. 2, MKL-SVM based classification is done by using labelled and un-labelled data categories.

3.2 Global Image Registration

Registration is one of the important methods in the area of medical image retrieval for combining multiple image modals, image changes, etc. Image registration is required for fast and efficient image retrieval applications due to the following issues:

- Medical images are heterogeneous and multimodal with temporal features. So, multimodal registration is an integral part of any visual and content-based retrieval application.
- The size of the medical images affects the speed and computational cost for image registration.

Registration is executed using a two-step process. In the first step, points related to current image to the reference image are computed using the block-matching approach. Each block in the current image is compared with the neighborhood of the transformed referenced image. Traditionally, a block matching is computed using normalized cross-correlation and naïve normalization measures.

$$\text{NCC} := (1/N) \sum_{x \in B_r} [B_r(x) - \mu_{B_r}][B_c(x) - \mu_{B_c}]/(\sigma_{B_r} \times \sigma_{B_c})$$

where B_r and B_c are the reference block and current image block, μ and σ are the mean and standard deviation within a block, and N be the size of the block.

Overall Limitations of the Visual and Content-based Search Methods

From this study, it has been depicted that there are many problems which do not have perfect solutions. Some of them are as follows:

- Application of CBIR technique on a large image database containing a wide variety of images result in less accurate results and consume high computation time.
- Traditional CBIR and visual search models heavily depend on segmentation techniques, which are not reliable.
- Problem of accurately representing the user interests in query formulation in ROI (region of interest)-based image retrieval system.
- Problem of considering the relative location of different objects in the query image for increasing the retrieval accuracy while consuming less computation time.
- Traditional local binary pattern (LBP)-based texture descriptors are very useful in analyzing texture of an image. However, there also remain some potential flaws as given below.
- Sensitive to noise magnitude of local difference is not taken and hence unable to represent fine texture details. Global texture properties are not represented.

4 Experimental Results

All experiments are executed with the minimum configurations such as CPU 2.13 GHz, Intel(R) Core(TM)2 4 GB RAM, and Netbeans IDE tool. This framework requires a third-party libraries such as jVisualizer, JAMA, and Apache Math.

Dataset: The dataset contains X-ray images of 5 classes: spine, hand, cranium, chest, and negative taken from the publicly available IRMA dataset [1] (Figs. 3 and 4).

In the results Figs. 5 and 6, it was observed that image orientation and shape are the main problems in the traditional visual search systems (Figs. 7 and 8).

Performance Analysis (Tables 1 and 2).
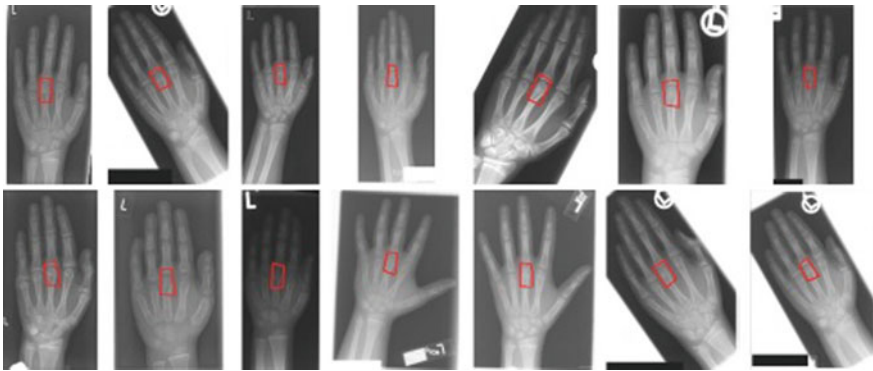
Fig. 3 User query selection

Fig. 4 Top 14 image retrieval results [7]
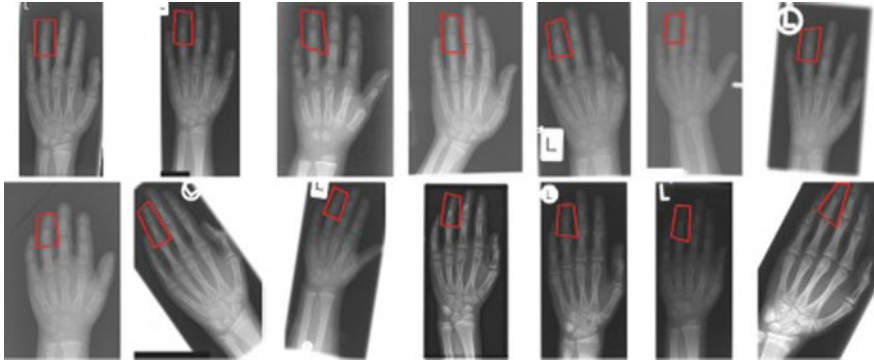

Fig. 5 User query selection

Fig. 6 Top 14 image retrieval results [7]

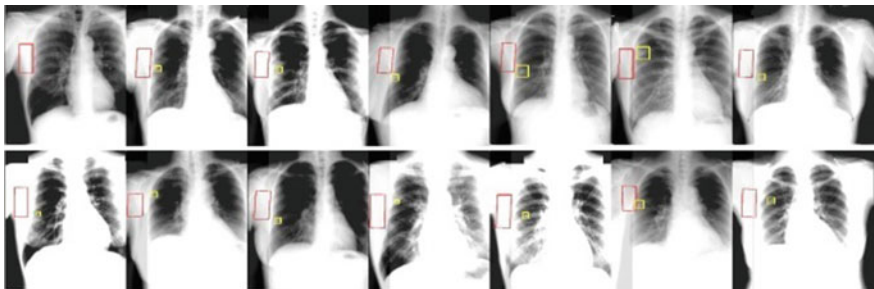**Fig. 7** User selected query

Fig. 8 Lung retrieval results [7]

Table 1 Hand retrieval results using traditional models

Algorithm	Hand1 (Avg. precision)	Hand2 (Avg. precision)	Time (s)
Structure retrieval algorithm	0.85	0.89	10
Local binary pattern	0.814	0.841	16
Region-based algorithm	0.79	0.81	15

Table 2 Lung retrieval results using traditional models

Algorithm	Lung1 (Avg. precision)	Lung2 (Avg. precision)	Time (s)
Structure retrieval algorithm	0.79	0.89	12
Local binary pattern	0.73	0.69	17
Region-based algorithm	0.74	0.77	19

5 Conclusion

The visual captures information in the image database and uses it to improve the performance of user-specific medical image search. Survey of the visual search models on the medical training dataset has proven both advantages and limitations. In the visual search system, an unknown query image was given as input, relevant visual images with different diagnoses features are retrieved and then used as clinical decisions. The main goal of the visual search engine is to efficiently retrieve user-specific images that are visually identical to a selected ROI query. In this paper, a survey on traditional visual search methods is analyzed in terms of visual features and accuracy are concerned. Different image ROI methods along with classification models are explored to highlight the challenges in the visual search process. Based on the survey performed by different visual search systems, the diagnostic efficiency is increased from 30 to 60% for clinical decision.

References

1. Alraqibah HA, Bchir O, Ismail MMB (2014, April) X-ray image retrieval system based on visual feature discrimination. In: Sixth International Conference on Digital Image Processing. International Society for Optics and Photonics, pp 91591S–91591S
2. Foncubierta-Rodríguez A, García Seco de Herrera A, Müller H (2013) Medical image retrieval using a bag of meaningful visual words. In: ACM MM MIIRH
3. Gao Y (2014) Multiple features-based image retrieval. In: 2011 4th IEEE international conference on broadband network and multimedia technology (IC-BNMT)
4. Imo J, Klenk S, Heidmann G (2008) Interactive feature visualization for image retrieval. In: IEEE

5. Philbin J, Chum O, Isard M, Sivic J, Zisserman A (2007) Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of IEEE conference on computer vision and pattern recognition
6. Ravela S, Manmatha R Multi-modal retrieval of trademark images using global similarity. UMass Computer Science Tech Report number TR99-32
7. Simonyan K, Criminisi A, Zisserman A (2011) Immediate structured visual search for medical images
8. Sivic J, Zisserman A (2006) Video Google: Efficient visual search of videos. In: Toward category-level object recognition, Springer Berlin Heidelberg, pp. 127–144
9. Sivic J, Zisserman A (2009) Efficient visual search of videos cast as text retrieval. IEEE Trans Pattern Anal Mach Intell 31(4):591–606
10. Vedaldi A, Varma M, Gulshan V, Zisserman A (2009) Multiple kernels for object detection. In: Proceedings of international conference on computer vision, pp 606–613

Firefly Algorithm-Assisted Segmentation of Brain Regions Using Tsallis Entropy and Markov Random Field

**N. Sri Madhava Raja, P.R. Visali Lakshmi
and Kaavya Pranavi Gunasekaran**

Abstract In recent years, segmentation of medical images attracted the research community because of its significance in medical discipline. In this paper, firefly algorithm and Tsallis entropy-based approach is initially considered to threshold the standard brain MRI dataset. Later, the brain regions, such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CF), are segmented using the Markov random field (MRF) approach. The proposed work is implemented using 256×256 sized benchmark MRI data, of subjects CHIMIC, JANPRZ, and JATKAM. Performance of the proposed approach is validated using a numerical metric that estimates the silhouette index of the estimated clusters. The proposed approach is also tested on other brain MRI dataset available in the literature and obtained better result in the segmentation of WM, GM, and CF. The simulation results in this study confirms that the proposed method offers an average enhancement of cluster classification by 4.44% in terms of silhouette index.

Keywords Firefly algorithm · MRI brain image · Segmentation
Tsallis entropy · Markov random field

1 Introduction

Due to its flexibility and ease of implementation, heuristic algorithm-based approaches are widely employed in various applications. In this paper, heuristic algorithm-assisted thresholding and segmentation is proposed for the brain MR image dataset. A number of procedures to segment the medical images such as multi-thresholding [1–4], clustering [5, 6], and region growing [7, 8] are present. Among these, multi-thresholding seems to be a better preprocessing procedure, which groups the similar pixels based on the chosen threshold values. In the

N.S.M. Raja (✉) · P.R.V. Lakshmi · K.P. Gunasekaran
Department of Electronics and Instrumentation Engineering, St. Joseph's College
of Engineering, Shollinganallur, Chennai 600119, India
e-mail: nsrimadhavaraja@stjosephs.ac.in

processing, a number of thresholding procedures are discussed by the researchers [9–13]. The considered work is aiming in the separation of the brain MRI data into three regions, such as white matter (WM), gray matter (GM), and cerebrospinal fluid (CF). The proposed segmentation process has two states: (i) preprocessing stage and (ii) post-processing stage. Preprocessing stage uses a tri-level thresholding, in which the Tsallis entropy function is used. For minimization of computation cost, firefly algorithm (FA) is adopted. The output is then passed to the post-processing stage, where the thresholded image is segmented into three clusters using the Markov Random Field (MRF) approach. The result is compared with a recent approach by Palani et al. [8] and shows that the cluster classification is improved by 4.44% in terms of silhouette index.

2   Methodology

This section presents the methodology adopted in this work. Figure 1 depicts the stages considered.

2.1   *Image Enhancement*

Image improvement is essential to enhance the image regions. Prior to the preprocessing work, it is necessary to improve the contrast of the brain MRI image frame. Figure 2a, b depicts the Chimic image frame and its histogram with poor contrast. Figure 2c, d depicts the enhanced image and its histogram.

a.   Multi-thresholding

During preprocessing, tri-level thresholding is achieved using the firefly algorithm (FA) and Tsallis entropy. In this work, the FA discussed in the recent paper by Raja et al. [14, 15] is considered.

Generally, in FA the light intensity I at a particular distance d from the light source X_i^t obeys the inverse square law [16]. For I, as the distance d increases in terms of $I \propto 1/d^2$. The movement of the attracted firefly i toward a brighter firefly j can be determined by position update equation;
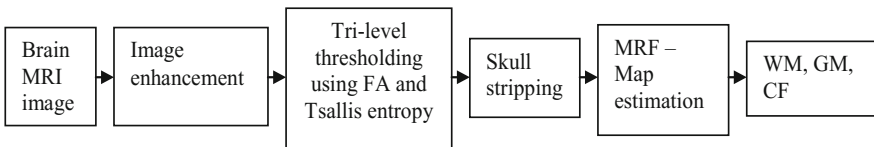
Fig. 1   Block diagram of Tsallis- and MRF-based segmentation procedure
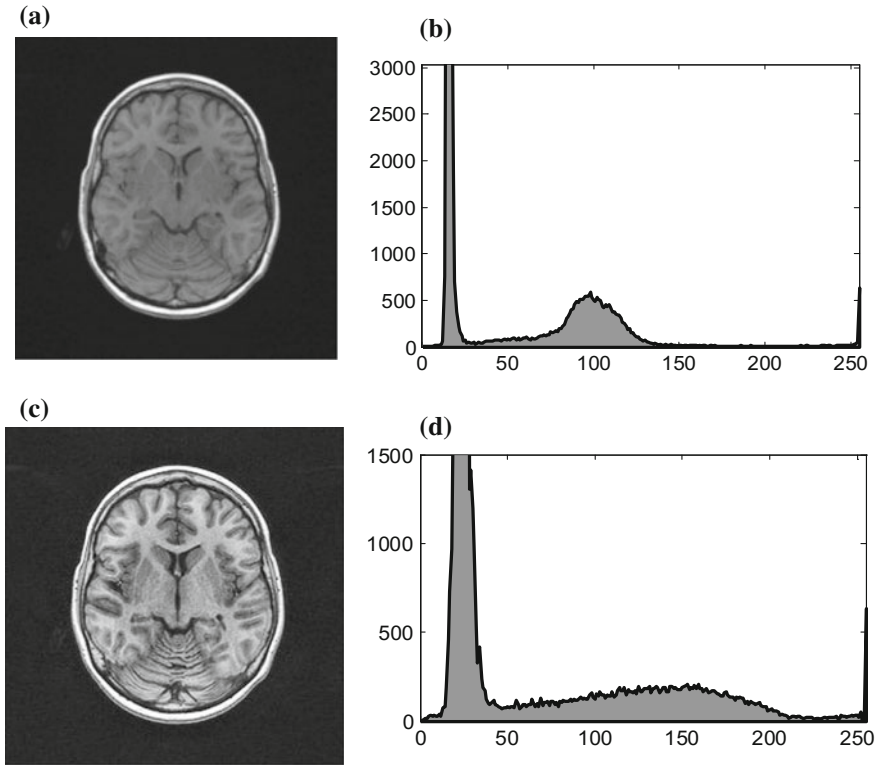
Fig. 2 Image enhancement

$$X_i^{t+1} = X_i^t + \beta_0 e^{-\gamma d_{ij}^2}(X_j^t - X_i^t) + \alpha \cdot \text{sign}(\text{rand}-1/2) \oplus \text{Lévy} \tag{1}$$

where X_i^{t+1} = updated position of firefly, X_i^t = initial position of firefly, and $\int \beta_0 e^{-\gamma d_{ij}^2}(X_j^t - X_i^t)$ = attraction between fireflies.

In this work, the number of fireflies is chosen as 20, the search iteration number is assigned as 2000, and the dimension of search D is assigned as 3.

Non-extensive entropy concept was initially proposed by Tsallis [17] as shown below:

$$S_q = \frac{1 - \sum_{i=1}^{T}(p_i)^q}{q - 1} \tag{2}$$

where T is the system potentials and q is the entropic index. Equation (2) will meet the Shannon's entropy when $q \to 1$. The entropy value can be expressed with a pseudo-additivity rule as:

$$S_q(A+B) = S_q(A) + S_q(B) + (1-q) \cdot S_q(A) \cdot S_q(B) \tag{3}$$

The Tsallis entropy concept is then considered to find optimal thresholds of the image. Let the test image has L gray levels in the range $\{0, 1, 2, 3, \ldots, 256-1\}$ with gray level probability distributions $p_i = p_0, p_1, p_2, \ldots, p_{L-1}$ [11–13]. Tsallis tri-level thresholding can be expressed as:

$$
\begin{aligned}
f(T) &= [T_1, T_2, T_3] \\
&= \arg\max \left[S_q^A(T) + S_q^B(T) + S_q^C(T) + (1-q) \cdot S_q^A(T) \cdot S_q^B(T) \cdot S_q^C(T) \right]
\end{aligned} \tag{4}
$$

Following are the values with their constraints:

$$S_q^A(T) = \frac{1 - \sum_{i=0}^{t_1-1}\left(\frac{Pi}{P^A}\right)^q}{q-1}, \quad P^A = \sum_{i=0}^{t_1-1} Pi$$

$$S_q^B(T) = \frac{1 - \sum_{i=t_1}^{t_2-1}\left(\frac{Pi}{P^B}\right)^q}{q-1}, \quad P^B = \sum_{i=t_1}^{t_2-1} Pi$$

$$S_q^C(T) = \frac{1 - \sum_{i=t_k}^{L-1}\left(\frac{Pi}{P^C}\right)^q}{q-1}, \quad P^C = \sum_{i=t_k}^{L_2-1} Pi$$

$$\left|P^A + P^B\right| - 1 < S < 1 - \left|P^A - P^B\right|$$
$$\left|P^B + P^C\right| - 1 < S < 1 - \left|P^B - P^C\right|$$
$$\left|P^C + P^{L-1}\right| - 1 < S < 1 - \left|P^C - P^{L-1}\right|$$

b. Skull stripping

Skull stripping is essential to remove the skull and the background area from MRI to analyze the brain soft tissue properly by using an image filter. In MR image, generally the skull/bone will have the maximum threshold value (>180) compared to other brain tissues. Hence, the image filter is used to separate the brain regions based on a chosen threshold value [8, 18].

c. Markov random field

Markov Random Field (MRF) is a probabilistic tactic used in gray scale image segmentation problems [8]. The MRF can be expressed as follows: Let $I = \{y(m \cdot n) \,|\, 0 \le y \le L-1; y$—intensity of the image at the pixel location $(m, n);$

L—number of threshold levels. During the segmentation, MRF will approximate the formation of each pixel by mapping into a group of random labels defined as: $X = \{x_1, \ldots x_N)|x_i \in l$. In this paper, the number of labels is assigned as three (since during the multi-thresholding, T is chosen as three as discussed) [8, 19, 20].

3 Results and Discussion

Major aim of this paper is to segment the brain tissue regions automatically with greater accuracy using the heuristic algorithm, as is in the paper by Palani et al. [8] using PSO and Otsu. In their work, the cluster classification of silhouette index is achieved as 4.42%. In the proposed paper, FA and Tsallis approach is implemented to improve the cluster classification >4.42%. Initially, the proposed method is applied on the benchmark MRI of subjects, such as CHIMIC, JANPRZ, and JATKAM [21]. Figure 3 presents the result obtained based on the method discussed in [8], where Fig. 3a shows the skull stripped initial label, Fig. 3b presents the final label by the MRF, Fig. 3c depicts the minimized energy value, and Fig. 3d–f gives the value of GM, WM, and CF, respectively. All the processing works are implemented using Matlab2012a.
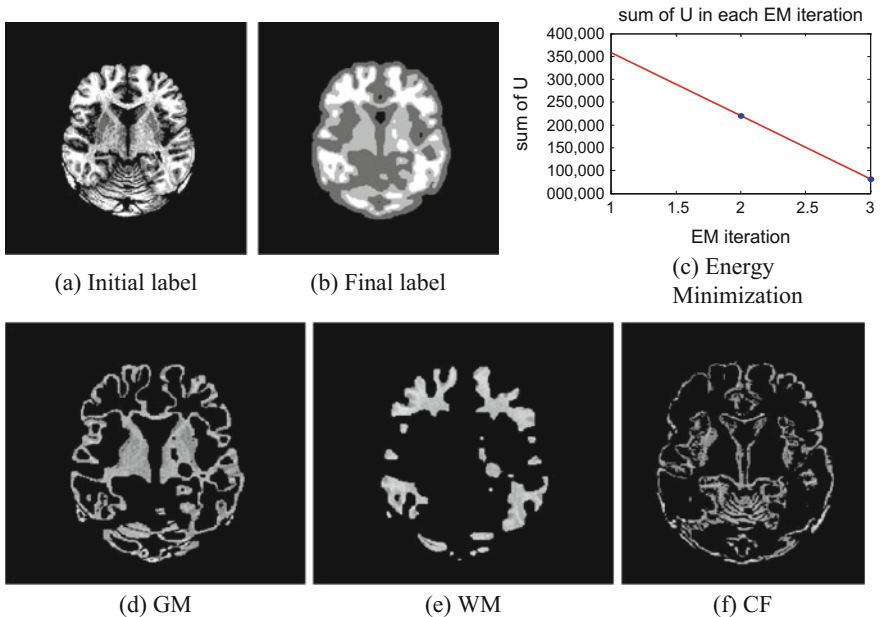

(a) Initial label   (b) Final label   (c) Energy Minimization

(d) GM   (e) WM   (f) CF
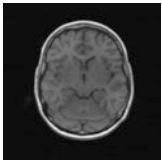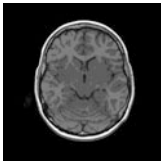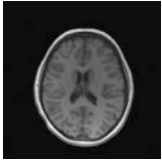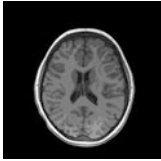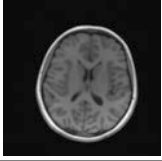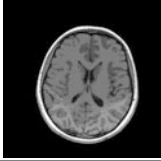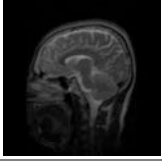
Fig. 3 Results of Chimic with PSO + Otsu + MRF

Table 1 Test image and the preprocessed output

	Test image	Tsallis	Optimal thresholds	Tsallis function
CHIMIC			58, 134, 182	1.5277
JANPRZ			31, 122, 179	1.4924
JATKAM			53, 106, 185	1.4629
Test 1			64, 129, 162	1.8270

Later, similar approach is implemented on CHIMIC, JANPRZ, JATKAM [21], test image [22] brain MRI dataset using the proposed method as discussed in Sect. 2. Table 1 depicts the output obtained during the preprocessing stage. From Tic-Toc function in Matlab, the average time required for this segmentation procedure is around 175.1875 s. This procedure is repeated 10 times and the mean value and the standard deviation (SD) values of the initial and final label are noted in Table 3; evident that, the Silhouette index value is slightly better than the Silhouette index presented in [8] and the average improvement in cluster classification is around 4.44%. Table 2 presents the initial, final label, CF, GM, and WM obtained with the FA- and Tsallis-based approach.

This shows a minimal energy during expectation maximization iteration. From this result, the proposed approach is efficient in separating the brain soft tissue from the brain MRI dataset.

Table 2 Segmented brain regions in the post-processing stage
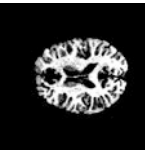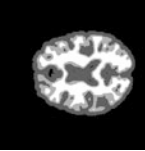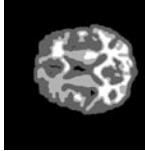
	Initial label	Initial label	CSF	GM	WM
Chimic					
Janprz					
Jatkam					
Test 1					

Table 3 Output parameters in the post-processing stage

Patient	Silhouette index		
	Initial labels Mean ± SD	Final labels	Improvement
Chimic	0.7013 ± 0.2653	0.7851 ± 0.3175	4.568
Janparz	0.7006 ± 0.5167	0.7971 ± 0.3316	5.052
Jatkam	0.6973 ± 0.2864	0.7903 ± 0.2551	4.517
Test image	0.6847 ± 0.3175	0.7826 ± 0.1854	4.049

4 Conclusions

FA- and Tsallis entropy-based preprocessing and MRF-based post-processing is proposed in this paper to segment the brain tissues for MRI. The combination of the tri-level segmentation and the MRF offers better automated segmentation of GM, WM, and CF. The silhouette performance metric of the considered method confirms that it can enhance the segmentation accuracy and offers better clusters. It also proves that the proposed approach improves the spatial clustering performance by an average of 4.44%.

References

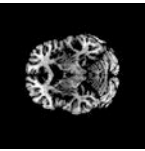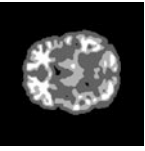1. Rajinikanth V, Couceiro MS (2014) Multilevel segmentation of color image using Lévy driven BFO algorithm. In: Proceedings of the 2014 international conference on interdisciplinary advances in applied computing, ICONIAAC'14, Article No. 19
2. Joyce Preethi B, Rajinikanth V (2014) Improving segmentation accuracy in biopsy cancer cell images using Otsu and Firefly Algorithm. Int J Appl Eng Res 9(24):8502–8506
3. Raja NSM, Sukanya SA, Nikita Y (2015) Improved PSO based multi-level thresholding for cancer infected breast thermal images using Otsu. Procedia Comput Sci 48:524–529
4. Siva Balan N, Sadeesh Kumar A, Raja NSM, Rajinikanth V (2016) Optimal multilevel image thresholding to improve the visibility of *Plasmodium* sp. in blood smear images. Proc Int Conf Soft Comput Syst Adv Intell Syst Comput 397:563–571
5. Kamalanand K, Ramakrishnan S (2015) Effect of gadolinium concentration on segmentation of vasculature in cardiopulmonary magnetic resonance angiograms. J Med Imaging Health Inform 5:147–151
6. Manickavasagam K, Sutha S, Kamalanand K (2014) development of systems for classification of different plasmodium species in thin blood smear microscopic images. J Adv Microsc Res 9:86–92
7. Ali A-R, Couceiro M, Anter A, Hassanien A-E (2016) Particle swarm optimization based fast fuzzy C-means clustering for liver CT segmentation. In: Applications of intelligent optimization in biology and medicine, volume 96 of the series intelligent systems reference library, pp 233–250
8. Krishnan PT, Balasubramanian P, Krishnan C (2016) Segmentation of brain regions by integrating meta heuristic multilevel threshold with markov random field. Curr Med Imaging Rev 12(1):4–12

9. Sezgin M, Sankar B (2004) Survey over image thresholding techniques and quantitative performance evaluation. J Electron Imaging 13:146–165

10. Tuba M (2014) Multilevel image thresholding by nature-inspired algorithms: a short review. Comput Sci J Mold 22:318–38

11. Sathya PD, Kayalvizhi R (2010) Optimum multilevel image thresholding based on Tsallis Entropy method with bacterial foraging algorithm. Int J Comput Sci Issues 7(5):336–343

12. Raja NSM, Kavitha G, Ramakrishnan S (2012) Analysis of vasculature in human retinal images using particle swarmoptimizatio based Tsallis multi-level thresholding and similarity measures. Lect Notes Comput Sci 7677:380–387

13. Akay B (2013) A study on particle swarm optimization and artificial bee colony algorithms for multilevel thresholding. Appl Soft Comput J 13(6):3066–3091

14. Raja NSM, Rajinikanth V, Latha K (2014) Otsu based optimal multilevel image thresholding using firefly algorithm. In: Modelling and simulation in engineering, vol 2014, Article ID 794574, p 17

15. Sri Madhava Raja N, Suresh Manic K, Rajinikanth V (2013) Firefly algorithm with various randomization parameters: an analysis. In: Panigrahi BK, Suganthan PN, Das S, Dash SS (Eds) Proceedings of the 4th international conference on swarm, evolutionary, and memetic computing (SEMCCO'13), lecture notes in computer science, vol 8297, pp 110–121

16. Rajinikanth V, Couceiro MS (2015) RGB histogram based color image segmentation using firefly algorithm. Procedia Comput Sci 46:1449–1457

17. Tsallis C (2002) Entropic non extensivity: a possible measure of complexity. Chaos, Solitons Fractals 13(3):371–391

18. Chaddad A, Tanougast C (2002) Quantitative evaluation of robust skull stripping and tumor detection applied to axial MR image. Brain Inform. doi:10.1007/s40708-016-0033-7

19. Jung HY, Lee KM (2015) Image segmentation by edge partitioning over a nonsubmodular markov random field. Math Probl Eng 2015:9, Article ID 683176

20. Zhang Y, Brady M, Smith S (2001) Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. IEEE Trans Med Imaging 20(1):45–57

21. Zwoliński P, Roszkowski M, Żygierewicz J, Haufe S, Nolte G, Durka PJ (2010) Open database of epileptic EEG with MRI and postoperational assessment of foci—a real world verification for the EEG inverse solutions. Neuroinformatics 8(4):285–299

22. https://sites.google.com/site/hosseinrabbanikhorasgani/datasets-1

Endoscopic Image Enhancement Using Blind Denoising

Amarsinh Deshmukh, Kapil Mundada and Pramod Kanjalkar

Abstract Biomedical images such as endoscopic images, retina, MRI, X-ray play an important role in the analysis and diagnosis of the internal body structure. Endoscopic image is used during pregnancy, plastic surgery, orthopedic surgery, spinal surgery, etc., to examine internal body structure. Endoscopic images are corrupted with various types of noises. The noisy image results into inaccurate diagnosis, and thus, the endoscopic image denoising is essential. In this paper, a method known as blind denoising has been used to improve the visual quality of the images. In the proposed method, we first estimate the noise level in the image obtained. Now having known the noise level, we apply BM3D algorithm to denoise the endoscopic image. By the proposed method, it is found that the PSNR of the test image is improved. The enhanced image will help the doctors for accurate diagnosis.

Keywords Blind denoising · Noise level estimation · BM3D

1 Introduction

Endoscopy is used to observe an internal body organ, structure, or tissue. In this process, long, thin tube is inserted into the body for diagnosis. The main application of endoscopy is imaging, minor surgery, diagnosis, and so on. Few times in

A. Deshmukh (✉) · K. Mundada · P. Kanjalkar
Department of Instrumentation Engineering, Vishwakarma Institute of Technology,
Pune, Maharashtra 411037, India
e-mail: amar9511100045@gmail.com

K. Mundada
e-mail: kapil.mundada@vit.edu

P. Kanjalkar
e-mail: pramod.kanjalkar@vit.edu

endoscopy due to internal bleeding and some other problems we cannot get a clear image. Thus, the noisy image results into inaccurate diagnosis. So we need an engineering solution to this problem.

Every real-time image which is captured from the camera consists of some sort of noise. The noise may be from different types of sources such as photon noise, thermal noise, and quantization noise.

Image denoising is important in many image-processing applications and analyses. The study of image denoising started a few decades ago, i.e., since 1970, but still we are lagging behind the mark of perfection. Image denoising is classified on different bases such as domain-based approach, noise level-based approach. According to the noise level-based approach, denoising is divided into two types non-blind denoising and blind denoising. This classification is based on whether the noise level is known or unknown.

In case of non-blind denoising, the noise level (σ_n) is considered as known parameter, this is conventional way of denoising. On the other hand, in case of blind denoising, the noise level (σ_n) is unknown. We have to estimate the noise level parameter along with the denoising process. The accomplishment of image denoising algorithm predominantly depends upon the noise level (σ_n) estimation.

In most of the commonly used noisy image model, generally the noise is AWGN (additive white Gaussian noise). In the noise level estimation, we mainly estimate the standard deviation (σ_n) for given single noise image. Lots of work is done on this topic, many algorithms [1–7] have been implemented. These algorithms are basically classified into three types of approaches, i.e., filter-based approach, patch-based approach and statistical approach.

In filter-based approach [1, 3, 5], noisy image is passed through the high-pass filter to get the suppressed image structure. Then, the difference between the filtered image and the original image is considered as the noise. The problem with this denoising method is that the difference between the images is considered as noise, but this assumption is not always true especially in case of image with complex structure.

In patch-based approaches [2, 5, 7], the image is divided into number of patches, i.e., rectangular window of size $N \times N$, and select the smooth patch among the separated patches. The smooth patch is selected on the basis of intensity level depending on the standard deviation. Here, the consideration is the smooth patch consists of large amount of noise as compared to the true image contents of the patch. So approximate the true image contents to zero and hence by assuming the smooth patch consisting of only the noise, one can estimate the noise level. But the disadvantage of this method is that if the consideration goes wrong, then overestimation or underestimation of the noise level takes place.

2 Noise Level Estimation

Liu and Tanaka [8] proposed a method for noise level estimation based on the PCA. This method comes under patch-based noise level estimation; here, the input noisy image is divided into number of small patches in raster scan. Then, we slide the window pixel by pixel so the patches are overlapped and the data model of the patches is represented as noisy image patch which is the combination of true image patch and noise. By taking the advantage of properties of the natural image, the data of natural image span only low-dimensional subspace because of redundancy of natural image. If the data patch spans the subspace whose dimension is very less than the patches' dimensions, then that patch is known as low-rank patch. Now here is the assumption that the minimum eigenvalue of the covariance matrix is equal to zero. The variance of the Gaussian noise is equally distributed in all the directions and all eigenvalues are same, so we can estimate the noise level.

The main disadvantage of this method is that our assumption is not always true, especially in case of images with complex structure. When the image with very fine details is given, we can overestimate the noise level. To overcome this disadvantage, we go for proposed method in which we choose the low-rank patches. The low-rank patches may consist of the patches with similar structure which includes the high-frequency components such as edges, corners, or texture.

2.1 Patch Selection

There are many algorithms used for the patch selection depending on their applications. In an image patch, local variance is an important parameter and it is useful to analyze the image structure as well as to select the image patch for noise level estimation. Lee and Popper [9] proposed an algorithm in which homogeneous patches are required to estimate the noise level, but here the homogeneous patches are known as the patches with small local variance. Similarly Pyatykh et al. [7] proposed an algorithm where he discarded the patches with large variances. The advantage of the above two methods is that both the algorithms are simple and fast, but the major disadvantage is that it overestimates the noise level. To overcome the above disadvantage, Shin et al. [5] proposed a method, in which instead of selecting homogeneous patches or discarding the patches with large variance, he suggests to use the adaptive threshold of variance to select the patches. By using this method, the performance is improved but not up to the mark.

To deal with the above problem, Amer et al. [2] proposed an algorithm, in which high-pass operator as well as threshold is used to calculate the homogeneity measures, but the high-pass operator is easily affected by the noise. Hence, in case of high noise level estimation this method fails. So by analyzing above results, we can conclude that noise level estimation using only the variance parameter is not accurate, rather we can say suitable patch selection is the first step for accurate noise

level estimation and it depends not only on the image variance but also on the image structure. Zhu and Milanfar [10] concluded that image structure analysis can be done on the basis of gradient covariance matrix. Liu and Tanaka [8] proposed an algorithm for patch selection which is based on local image gradient matrix and its statistical properties to select low-rank patches. The proposed algorithm for low-rank patch selection is as follows:

Algorithm:

1. Let us take an input patch y_i ($N \times N$)
2. Find the $N^2 \times 2$ gradient matrix G_{yi}; if the gradient matrix is null matrix, go to step 1.
3. Calculate the gradient covariance matrix C_{yi} for the image patch y_i.
4. Find the eigenvalues and eigenvectors of C_{yi} to calculate the dominant direction and its energy.
5. Calculate the texture strength (ξ_n) by using trace operator, i.e., sum of all eigenvalues of covariance matrix.
6. To analyze the statistical properties of texture strength ξ_n, apply the gamma distribution.
7. Apply the threshold to select the weak texture patches which is the function of given significant level δ and noise level σ_n.

Applying the above algorithm to different patches:

CASE 1: Let us take perfectly noise-free flat patch y_f, as the input. Now find the gradient matrix G_{yf},

$$G_{y_f} = [D_h y_f \quad D_v y_f] \tag{1}$$

where D_h and D_v are the horizontal and vertical derivative operators, respectively. As patch y_f is perfectly noise-free flat patch, hence the gradient matrix G_{yf} is,

$$G_{y_f} = [0 \quad 0] \tag{2}$$

CASE 2: Now take noisy flat patch y_f with Gaussian noise,

$$y_f = z_f + n \tag{3}$$

where 'z_f' is actual noise-free image contains and 'n' is the Gaussian noise patch with zero mean and standard deviation σ_n. From case 1, we know that the gradient matrix of flat patch is zero so the gradient matrix of noisy patch is

$$G_{y_f} = [D_h y_f \quad D_v y_f]$$
$$= [D_h(z_f + n) \quad D_v(y_f + n)] \qquad (4)$$
$$G_{y_f} = [D_h n \quad D_v n]$$

Now calculate the eigenvalue and eigenvector of G_{yf}, and by applying the trace operator, we can get the texture strength of the patch as

$$\xi_n = tr(C_{yf})$$
$$= tr(G_{yf}^T G_{yf}) \qquad (5)$$
$$\xi_n = tr\left(\begin{bmatrix} n^T D_h^T D_h n & n^T D_h^T D_v n \\ n^T D_v^T D_h n & n^T D_h^T D_v n \end{bmatrix}\right)$$

To analyze the statistical properties of ξ_n, we have to apply the gamma approximation to the above equation. By simplifying, we obtain

$$\xi_n \sim \text{Gamma}\left(\frac{N^2}{2}, \frac{2}{N^2} \sigma_n^2 tr(D_h^T D_h + D_v^T D_v)\right) \qquad (6)$$

where, in the gamma approximation, first term is shape parameter, second term is scale parameter, and σ_n^2 is the standard deviation of Gaussian noise.

To estimate the unknown noise level, we have to select the weak texture patches, for that we have to set the threshold value. Below that threshold value, the patch is considered as weak texture patch. Liu and Tanaka [8] proposed a formula for the threshold value which is depending upon the gamma approximation of texture strength of patch as follows:

$$I' = \sigma_n^2 F^{-1}\left(\delta, \frac{N^2}{2}, \frac{2}{N^2} tr(D_h^T D_h + D_v^T D_v)\right) \qquad (7)$$

where F^{-1} stands for inverse gamma distribution function, δ is the confidence level (values of texture strength within the threshold range), and σ_n^2 is the standard deviation of Gaussian noise.

2.2 Iterative Structure

The noise level estimation totally depends upon the selection of weak texture patches. We will take the noise level as a variable while threshold for the weak texture patch selection process. As number of iteration increases, the accuracy of estimated noise level increases up to certain point after that we will get a constant value.

The algorithm for iterative structure is as shown below:

1. Estimate initial noise level,
2. Calculate threshold I'_{k+1}
3. Select patch W_{k+1}
4. Estimate new noise level if it is not stable go to stage 2, and
5. Final estimated noise level.

By using the above algorithm, we can get the accurate and stable noise level.

3 Blind Denoising

In case of blind denoising, the noise level (σ_n) is unknown, we have to estimate the noise level parameter along with the denoising process. Till now in this paper, we have estimated the noise level (σ_n) that means now we have to choose or develop the best denoising algorithm for enhancing the image details such edge preserving, structural similarity.

Image denoising is very vast area. Different types of image denoising methods are as shown in Fig. 1. Rather than going for non-blind denoising which is very easy and commonly used method, we have chosen a blind denoising algorithm which is complex but gives more accurate results. Dabov et al. [11] proposed an
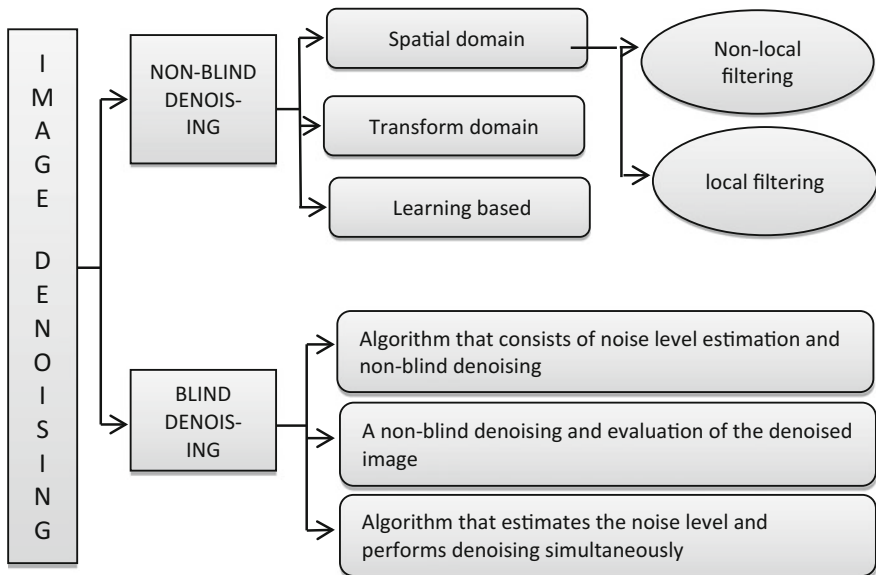
Fig. 1 Different types of image denoising

algorithm block matching and 3D transform which is based on grouping by matching and collaborative filtering which is as follows:

3.1 Grouping and Collaborative Filtering

Grouping of similar patches: Grouping of similar patches or blocks, the name itself indicates the meaning that is collocation of similar patches. For simplicity, divide these stages into two parts, first gets the similar patches or blocks and second stack them together by matching.

Some patches of approximately same intensity level are selected, among them any one patch is selected as reference patch. The similarity of the patch depends on the distance between the patch, as the distance increases similarity decreases. From the reference patch, some fixed distance threshold is taken. The patches within this threshold are selected for grouping. The similar 2D image patches stacked together is known as 'group,' so instead of group, it referred as 3D here one extra D stands for grouping of similar 2D image patches. This group is formed by many block or patch matching so the name is given as BM3D. The forming of groups is as shown in Figs. 2 and 3.

Collaborative Filtering and Reduction in Transform Domain: Given a bunch of n small parts of image which is also known as group of patches. By applying collaborative filtering on that patches, we get an estimate for each individual patch. This estimate preserves the difference between the patches and details of it. Here, we apply filtering to group of patches so the word collaborative filtering occurs.

The collaborative filtering achieves the best results when preforms the shrinkage in transform domain. Now let us consider the 3D groups of similar image patches that are already constructed as discussed above. The collaborative reduction includes following steps:

1. Transform the 3D group,
2. Apply reduction (by wavelet or winner filtering), and
3. Inverse the linear transform.

The collaborative filtering is effective in case of natural images which is characterized by both intrafragment and interfragment correlation. The 3D transform can produce the sparse representation of the signal in a group. Sparsity is defined as

Fig. 2 Formation of 3D group from 2D image patches
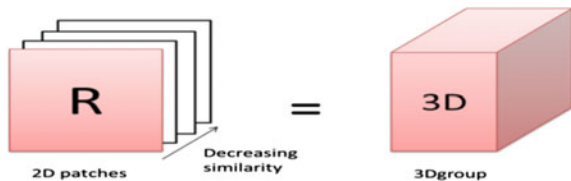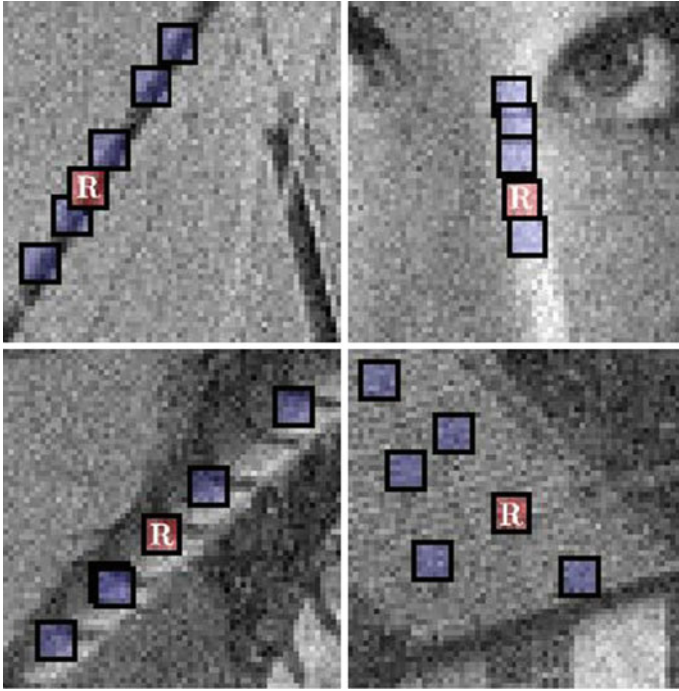
Fig. 3 Example of grouping block or patches from noisy image

number of nonzero elements in a vector or matrix. Sparsity achieves great results while preserving the structural details of an image.

3.2 Algorithm

Dabov et al. [11] proposed an algorithm for image denoising by grouping and collaborative filtering. The image is divided into the number of small patches. Depending upon the structure, intensity, and some more parameters of patch, the similar patches are grouped together. This grouping achieved by block matching and stacking of some patch in a group is referred as 3D. So the name of this algorithm is BM3D.

Algorithm:

1. Form a group of similar patches that is 3D block,
2. Denoise the 3D block by wavelet thresholding, and
3. Aggregate each estimate of the denoised patch to form the image.

   Repeat the same algorithm using Wiener filtering in 2nd stage

4 Experimentation

Liu and Tanaka [8] performed noise level estimation on complex image structure such as *mountain* image, *gravel* image. In this paper, we have taken an endoscopic image (source: http://worldendo.org) as shown in Fig. 4a. Let us consider the noise-free test image in which known white additive Gaussian noise is added, and then, we estimate the noise level (σ_n) as per our proposed algorithm and remove the noise by BM3D. Now, we check the PSNR in both the cases of noise image and denoised image.

Let us take a test image as an input, now add additive white Gaussian noise of zero mean and variance σ_n. We have added an AWGN (additive white Gaussian noise) in an image because it is uniformly distributed over the image. For blind denoising, first we have to estimate the noise level. Then using block matching algorithm, we denoise the image. For noise level estimation, we have to select a weak texture patch from input noisy image. Selection of weak texture patch is an important parameter in the noise level estimation process. Weak texture patch is selected by analyzing the image structure and the strength of the patches (ξ_n). Now in the selected patch, assume that the selected image patch is flat patch. As the selected patch is weak-textured patch in the image, our assumption is true. We can estimate the true noise level by iterative structure.

Apply the grouping and collaborative filtering to the noisy image as discussed in the Sect. 3. In step 1, we denoise the image by smoothing and edge preserving algorithm, in which we use hard (wavelet) thresholding. In step 2, rather than using hard threshold, we have used Wiener filtering keeping rest of the process unchanged.
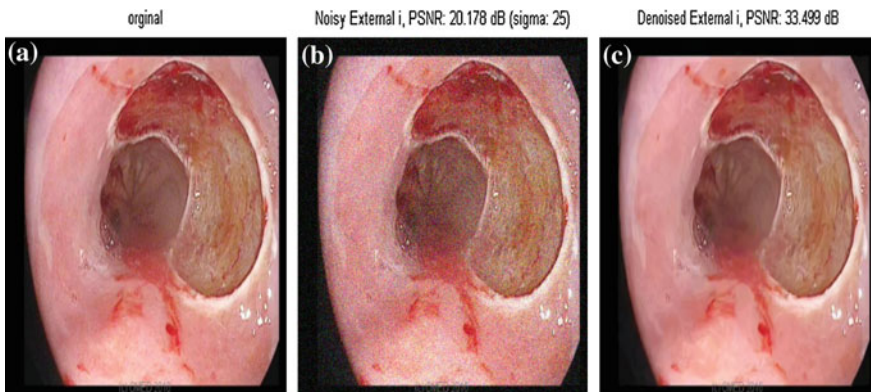
Fig. 4 a Original image (source: http://worldendo.org), **b** noisy image with AWGN, and **c** denoised image

Table 1 The standard deviation versus PSNR value

Noise level (σ_n)	PSNR (dB)
5	43.1819
10	39.8309
15	37.9097
20	36.5403
30	34.5676
40	33.1133

(a) **(b)**
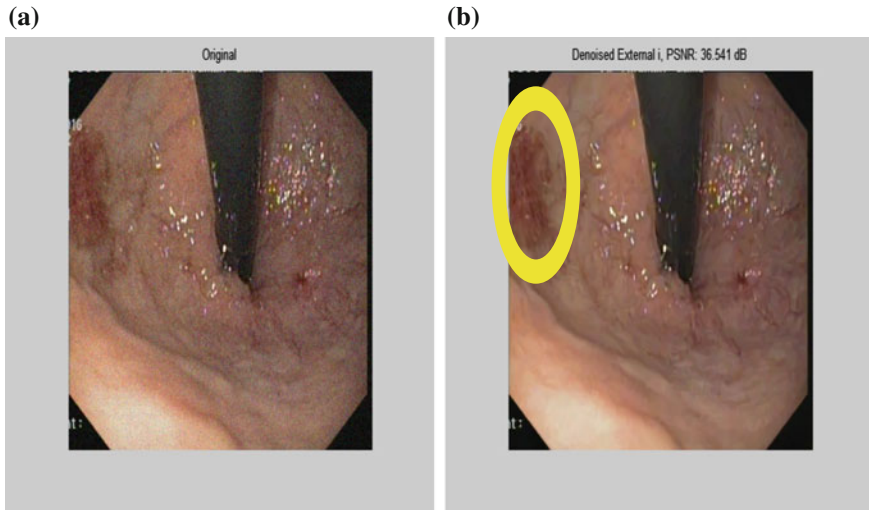
Fig. 5 **a** Original image and **b** denoised image

5 Results and Discussion

In this section, we present and discuss the experimental results obtained by the proposed method. For endoscopic image enhancement, we improve the PSNR value by 60% in case of AWGN. The PSNR of the estimated image ŷ of the true image is calculated by the following:

$$\text{PSNR}(\hat{y}) = 10 \log_{10} \left(\frac{255^2}{(3|X|)^{-1} \sum_{c=R,G,B} \sum_{x \in X} (y_c(x) - \hat{y}_c(x))^2} \right)$$

For the test image 1 as shown in Fig. 3a–c, the PSNR value increases from 20.178 to 33.499 dB, and similarly for test image 2 as shown in Fig. 4a–c, the PSNR value increases from 20.178 to 35.800 dB. The proposed algorithm is tested and sampled with the addition of AWGN and the PSNR is calculated. It was found that the algorithm is working as expected. And thus, the noisy image is taken as

input and PSNR is calculated. Now, we take noise image as a input and then denoise it by using our proposed method (Table 1).

The above table shows the results of PSNR values with different noise levels. From the above table, we can conclude that as the noise level (σ_n) increases, the PSNR value of the test image decreases. Figure 5a is taken as the input from gastroscopy done by Dr. Pankaj Bansode (MS, FIAGES, FICS), Bharati Vidyapeeth Medical College, Pune. Processor used for the gastroscopy is OYLMPUS CV-150. Now, by applying the proposed algorithm to the input image, we have enhanced the image structure as highlighted in Fig. 5b. We have successfully enhanced the image structure which will help the doctors for accurate diagnosis.

6   Conclusion

The proposed work presents the noise level estimation, block matching, and collaborative filtering for blind denoising of endoscopic image. This is a novel approach in case of endoscopic image enhancement. By the observation of PSNR value, it is clear that the blind denoising using BM3D method can enhance the image structure which will help doctors for an accurate diagnosis. The proposed algorithm was applied on real images and found to be successfully working. The enhancement in PSNR is clearly seen.

References

1. Rank K, Lendl M, Unbehauen R (1999) Estimation of image noise variance. IEEE Proc Vis Image Signal Process 146(2):80–84
2. Amer A, Mitiche A, Dubois E (2002) Reliable and fast structure oriented video noise estimation. Proc Int Conf Image Process 51:I-840–I-843
3. Cornera BR, Narayanana RM, Reichenbach SE (2003) Noise estimation in remote sensing imagery using data masking. Int J Remote Sens 24(4):689–702
4. Tai S-C, Yang S-M (2008) A fast method for image noise estimation using Laplacian operator and adaptive edge detection. In: Proceedings of 3rd international symposium communication, control signal processing, pp 1077–1081
5. Shin D-H, Park R-H, Yang S, Jung J-H (2005) Block-based noise estimation using adaptive Gaussian filtering. IEEE Trans Consum Electron 51(1):218–226
6. Liu C, Szeliski R, Kang SB, Zitnick CL, Freeman WT (2008) Automatic estimation and removal of noise from a single image. IEEE Trans Pattern Anal Mach Intell 30(2):299–314
7. Pyatykh S, Hesser J, Zheng L (2013) Image noise level estimation by principal component analysis. IEEE Trans Image Process 22(2):687–699
8. Liu X, Tanaka M (2013) Single-image noise level estimation for blind denoising. IEEE Trans Image Process 22(12):5226–5237
9. Lee J, Hoppel K (1989) Noise modeling and estimation of remotelysensed images. In: Proceedings of 12th canadian symposium IGARSS, pp 1005–1008

10. Zhu X, Milanfar P (2011) Automatic parameter selection for denoising algorithms using a no-reference measure of image content. IEEE Trans Image Process 19(12):3116–3132
11. Dabov K, Foi A, Katkovnik V, Egiazarian K (2007) Image denoising by sparse 3D transform-domain collaborative filtering. IEEE Trans Image Process 16(8):2080–2095
12. Daniel Z, Yair W (2009) Scale invariance and noise in natural images. In: Proceedings of IEEE 12th international conference on computer visible, pp 2209–2216
13. Bishop CM (2006) Pattern recognition and machine learning. Springer, New York

Silhouette Object Recognition Using Edge-Based Method

Praveen Yadav and Manoj S. Nagmode

Abstract Image processing is very vast domain, in which object recognition is toughest challenge in computer vision. As object having different key feathers to describe them we have taken silhouette image of object from recognition proceeding. In this paper edge detection method is used to detect object in frame and then compare their angle with the database angle value is done. To reduce the size of hardware and speed-up the performance we use Raspberry-Pi 2 model.

Keywords Object recognition · Canny · Silhouette object · Raspberry Pi

1 Introduction

Image recognition is an important function in the field of computer vision. Processing on image or video to achieve useful result is very challenging task. Object detection means detecting objects of a certain class (such as books, faces, or cars) in image. Object recognition is a procedure for pointing out an object in an image and tells user about its detail. Object recognition algorithm depends on matching, learning, or pattern recognition algorithms using shape-based or color-based techniques [2, 3]. Recognition algorithm should be accurate and fast enough for real-time applications. Feature selection is one of the most important steps for object recognition. For high accuracy and real-time object recognition, features should be discriminative, robust, and easy to compute. Edge-based algorithm had been widely used features for object detection and object recognition, due to their robustness, simplicity, and speed. Edges are stable object features in the presence of different light intensity and not affected by change in scales [1]. Edge-based algorithm was also used for shape recognition and matching [3, 4].

P. Yadav · M.S. Nagmode (✉)
Department of Electronics & Communication, MITCOE, Pune, India
e-mail: manoj.nagmode@gmail.com

P. Yadav
e-mail: sanupraveen91@gmail.com

This paper presents a recognition system by using image processing algorithm as the core element and the cost can be reduced by using a simple webcam. The image processing will be loaded on a board named Raspberry-Pi 2. It is a single-board credit-card-sized system prepared in the United Kingdom (UK) by the Raspberry Pi Foundation [5]. There are two models and both models are similar except for model B+ have the Ethernet, 512 MB SDRAM and 2 USB ports. Both models run on Linux operating system.

2 Background

All of us have played the jigsaw puzzle. We have a lot of small pieces of images, where we need to assemble them in correct order to form a big image. How it works? This same theory is projecting to a computer program. So answer is, we are looking for specific unique features, which can be easily tracked and compared. But if someone asks us to point out one good feature which can be compared across several images, we can point out one very easily. We try to search for that features in an image and try to match that feature in other images, and we arrange them accordingly. Mathematicians define shape under a group of transformation as an equivalence class. This is incomplete definition in the perspective of graphic finding. This define the same two object can only be same if it has same shape and size. But this definition has to be modified as two objects are same if they are similar to each other. The mathematician's definition, e.g., Bookstein [3] or Kendall [3], talks about the problem that for every viewer same thing look different with different viewing angle. Other geometric tactics for outline comparison do not involve above state statement—e.g., one could compare feature directions containing descriptors such as moments or area—but detailed shape data in this executions often get rejected in the process. Similarity in shape has also been considered in the psychology literature, an early case being Goldmeier [3].

2.1 Feature-Based Methods

An investigation on object matching using the boundaries of solid (fully filled with one color) object in images has been done. Since solid or monochromic image does not have internal markings or holes, the related boundaries are represented by a closed curve in which curve length can be considered as a parameter. Previously method used Fourier descriptors. Blum's medial axis transform [3] has succeeded to capture the part arrangement of the silhouette [6] in the graph structure of the skeleton by Kimia and Collaborators, Zucker, e.g., Sharvit et al. [3]. For real-time screening the nature of silhouette arc leads logically to approaches for matching, which uses the distance between arcs. This algorithm is invariant to several kinds of

transformation and fast [7]. A comparison is been done for comparing solid of different shape descriptors as part of the MPEG-7 standard activity.

2.2 Brightness-Based Methods

Brightness-based methods offer a matching view to feature-/shape-based method. Instead of considering shape of the object or other extracted features, these approaches make use of the grayscale values within the visible part of the object. One can use intensity information in one of two frameworks. In the first type, we have explicitly find alignment using grayscale values. Yuille [3] presents a flexible approach in that in-variance to certain kinds of conversion can be built into the measure of prototypical similarity, but it undergoes from the need for the sensitivity to initialization and human-designed templates when searching via gradient. Lades et al. [3] use flexible graph matching, an approach that involves both photometric and geometry features based on Gaussian-derived jets. Cootes et al. and Vetter et al. compare intensity values but first attempt to change the images into another using a solid correspondence field.

3 Previously Used Methods for Detecting Edge

3.1 Harris Corner Detection

Corners are areas in the image with variation in intensity. First persons to find these corners are Harris and Stephens [8]; therefore, it is called as Harris corner detector. He approaches with simple idea to a mathematical equation. Basically, it finds the intensity difference for a shift of (u, v) in all directions. This is given as Eq. 1:

$$E(u, v) = \sum_{x,y} w(x, y)[I(x + u, y + v) - I(x, y)]^2 \tag{1}$$

Window function is either a gaussian window or rectangular window which gives importance to pixels. For corner detection function $E(u, v)$ should be maximized which is obtain by Eq. 2. That means, maximize the second term. Applying Taylor expansion to term in equation, we get the final Eq. 3 as:

$$E(u, v) \approx [u\, v]M \begin{pmatrix} u \\ v \end{pmatrix} \tag{2}$$

$$M = \sum_{x,y} w(x, y) \begin{bmatrix} I_x I_x & I_x I_y \\ I_x I_y & I_y I_y \end{bmatrix} \tag{3}$$

Here, I_x and I_y are derivatives of image in both x and y directions, respectively. They created an Eq. 4, which will determine whether corner is present or not.

$$R = \det(M) - k(\text{trace}(M))^2 \tag{4}$$

where $\det(M) = e1 * e2$; trace $(M) = e1 + e2$; $e1$, $e2$ are Eigen value of M.

The values of Eigen ($e1$, $e2$) values decide whether a region is flat, corner, or edge.

- When R is small, which happens when $e1$ and $e2$ are small, the region is flat.
- When $R < 0$, which happens when $e1 \gg e2$ or vice versa, the region is edge.
- When R is large, which happens when $e1$ and $e2$ are large, the region is a corner.

3.2 Shi–Tomasi Corner Detector

Shi and Tomasi [9] made a small change which shows better results matched to Harris corner detector. The counting function in Harris corner detector was given by 5:

$$R = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \tag{5}$$

Shi–Tomasi proposed:

$$R = \min(\lambda_1, \lambda_2) \tag{6}$$

If R it is a greater than a threshold, it is considered as a corner in Eq. 6.

3.3 SIFT Method and SURF Method

Harris corner method and Shi–Tomasi method are rotation-invariant, which means, even if the object is rotated, we will not find the same corners. It is understandable that corners remain corners after rotating an object. But scale of that object is changing. A corner may not be a corner if the object is scaled. So, in 2004, Lowe [10] came up with a new algorithm, Scale Invariant Feature Transform, which extract key points and compute its descriptors. SIFT have key point description and detection but it was relatively slow and researcher's required faster method. Tuytelaars et al. [11] in 2006 introduced an algorithm called SURF. In SIFT, descriptors move toward LoG with DoG for finding scale-space. SURF goes a further and gives LoG with Box Filter. Such an approximation is shown in Fig. 1c.
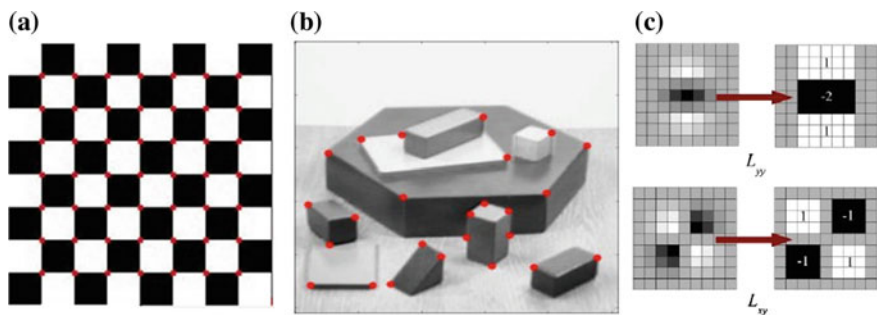
Fig. 1 **a** Harris corner detector output, **b** Shi–Tomasi edge detector output, **c** SUFR working

3.4 ORB (Oriented FAST and Rotated BRIEF)

This algorithm was proposed by, Vincent et al. [12]. ORB is a combination of BRIEF descriptor and FAST key point detector with many modifications to improve the efficiency. It uses FAST algorithm to find key points, and then proceed with Harris corner detector to find top N points for that object. But disadvantage of this method is that it doesn't find the angle orientation. It only computes the intensity centroid of the region and located corner and center. For descriptor purpose, ORB use BRIEF descriptors.

4 Proposed Method

In our approach we selected edge-based detection method and uses canny operator to find the edges of object as well as shape- and color-based characterization. In this original input image firstly resize to 320×380, and then image is converted to gray format from RGB after that canny operator is applied with filter. That gives an edge of every object present in rescaled image. Next step is to find contours of object which gives every value of edge pixel. Using contour we can identify shape and color of object. Final image contain number of objects with label of their color and shape.

Steps includes in our method are:

(1) Frame of input video is taken from simple web cam of 30 fps with resolution 640×480 in RGB24 bit format. All the data of frame has been stored in matrix of dimension 320×280 after resizing of input frame. Let us denote it as $[F_{ij}]$, where value of i is from 1 to 320 and value of j is from 1 to 280. Each pixel contain 24 bits of information about its color. To reduce size and to save processing time step 2 is performed.

(2) Convert the RGB24 bit true color to grayscale of 8 bit value with 0 (white) to 255 (black) so that canny operator can be applied. Matrix $[B_{ij}]$ shows grayscale frame. Apply filter of kernel 10×10 to sharpen the edge and apply canny operator to the grayscale frame $[G_{ij}]$.

(3) After that find contour size and contour point of frame. Contour is a vector function of vector (point). All the edge connected to each other classified as a contour. A frame can contain many contours. Contour is denoted as C_{xy}^i, where x, y are the point on the grayscale frame. Find centroid for contour point using 7.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i A_i}{\sum_{i=1}^{n} A_i} \quad \bar{y} = \frac{\sum_{i=1}^{n} y_i A_i}{\sum_{i=1}^{n} A_i} \tag{7}$$

Here, x, y are the contour point for which centroid is calculated and A_i is area of the contour.

(4) Repeat below step for each closed contour.

(a) Skip small object which have area less that 60 or non-convex objects
(b) Find out the angle at each point in edge of object

Equations 8 and 9 are used for solving angle:

$$b^2 = a^2 + c^2 - 2ac \cos B \tag{8}$$

$$\cos B = a^2 + c^2 - b^2/2ac \tag{9}$$

where a, b, and c are the vector of point AB, BC, and CA, respectively, shown in Fig. 2 and find $\cos^{-1}(B)$ to get angle at that point which is corresponding to other two points.

Fig. 2 Point angle detection
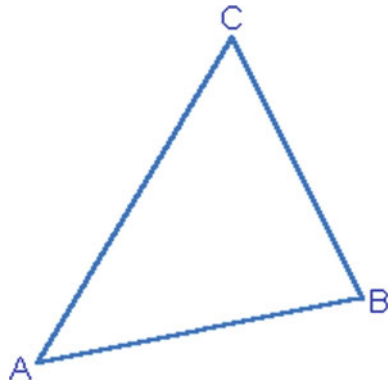
(c) Compare the result with previously saved database result. Database result has been saved with angle information and number of key point.

(d) If object is not present in that database, then it will be classified on the basis of number of vertex and angle it contains.

(e) Find the color of that object at centroid using "color = im.at<Vec3b>(point)".

(f) Name that object according to its classification and color of that object.

5 Experimental Results

To evaluate the efficiency of our proposed method, we have conducted an experiment. Figure 3 shows the steps in recognition of the object. The object contains multi-shaped objects with different colors. The method used edge-based recognition process; we have calculated angles at each key point of the objects. The size of input image is 640×480, but to increase processing speed of Raspberry Pi processor we have resized it to 320×280. Figure 3b, d shows the edge of the object present in that frame, and Fig. 3g shows final output result of our experiment; it contains objects and their classification along with color written at the center of that object.

Accuracy, Speed, and Size on input images saved in database:

- Accuracy on image database:

  1. 100%—When object present in frame and recognize by system.
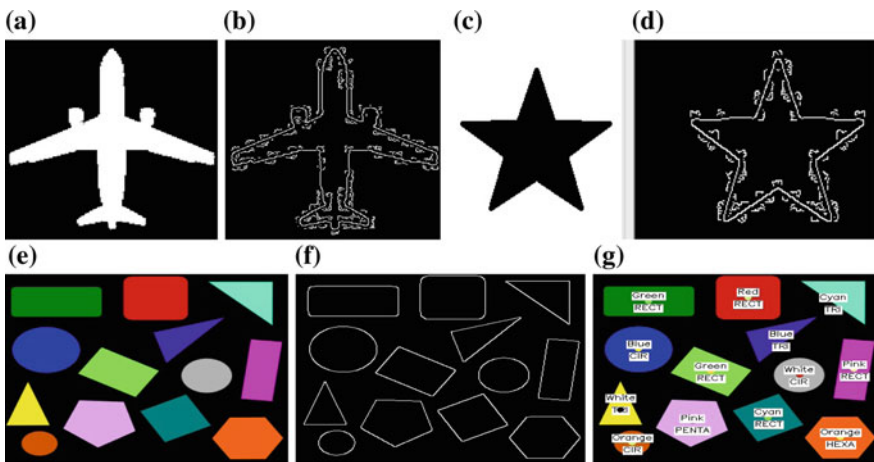
Fig. 3 a Input image of object1 (*airplane*), **b** image after canny operator of airplane, **c** input image of object2 (*star*), **d** detected edge of star, **e** input image of various object3, **f** detected edge of object3, **f** result of various *shapes* and *color*

2. 98.59%—When object present in frame and recognize also some false detection occurs.
3. 92.95%—When object present in frame and recognize also with some true/false detection.

- Processing speed is 1.75951 s, and execution time is 4 s (with image display time and delay time included).
- Size of database:

Seventy-one objects are taken in this experiment. So size of the save database is 1.5 MB, and size of feature save corresponding to object is 7.3 KB.

6 Conclusion

In this paper, a portable small size object recognition system is proposed using embedded system (Raspberry-Pi 2). It can be battery-operated work on 5 V battery with a power consumption of 4 W per hours when connected to keyboard, mouse, LAN, and camera. Further algorithms are proposed to automatically detect and recognize objects present in our database. For object detection edge-based approach is used, and for recognition shaped-/color-based approach is used.

References

1. Ramadevi Y, Sridevi T, Poornima B, Kalyani B (2010) Segmentation and object recognition using edge detection techniques. Int J Comput Sci Inf Technol (IJCSIT) 2(6)153–161
2. Khurana K, Awasthi R (2013) Techniques for object recognition in images and multi-object detection. Int J Comput Sci Inf Technol (IJCSIT) 2(4):1383–1388
3. Belongie S, Malik, J Puzicha J (2002) Shaped matching and object recognition using shape contexts. IEEE Trans Pattern Anal Mach Intell 24(24):509–522
4. Parekh HS, Thakore DG, Jaliya UK (2014) A survey on object detection and tracking methods. Int J Innov Res Comput Commun Eng 2(2)2970–2979
5. Abd Rahim R, Zainudin MNS, Ismail MM, Othman MA (2014) Image-based solar tracker using Raspberry Pi. J Multidiscip Eng Sci Technol (JMEST) 1(5)
6. Zafari S, Eerola T, Sampo J, Kälviäinen H, Haario H (2015) Segmentation of overlapping elliptical objects in silhouette images. IEEE Trans Image Process 24(12)5942–5952
7. Zhan T, Zhang Y, Cai J, Kot AC (2016) Efficient object feature selection for action recognition. In: 2016 IEEE international conference on acoustics, speech and signal processing (ICASSP)
8. Harris C, Stephens M (1988) A combined corner and edge detector. Plessey Research Roke Manor, United Kingdom
9. Shi J, Tomasi C (1994) Good features to track. In: Proceedings CVPR '94 of computer vision and pattern recognition
10. Lowe DG (2004) Distinctive image features from scale-invariant keypoints. Int J Comput Vis 60:91–110

11. Bay H, Tuytelaars T, Van Gool L (2008) SURF: speeded up robust features. Comput Vis Image Underst 110(3):346–359
12. Rublee E, Rabaud V, Konolige K, Bradski G (2011) ORB: an efficient alternative to SIFT or SURF. In: 2011 international conference on computer vision

On the Extraction of Retinal Area from SLO Images Using RBFN Classifier and Its Comparison to the Optimally Trained ANN Classifier

Nimisha and Rana Gill

Abstract Retinal diseases can be detected earlier by scanning laser ophthalmoscope (SLO). Automated detection of retinal area from the SLO images is a crucial task. With the invention of screening technology, the large retinal part must be imaged for better diagnosis of the retinal diseases. During the process of imaging, artifacts (eyelashes and eyelids) are come along with the retinal part. So removal of artifacts is a big challenge. In this chapter, true retinal area is extracted from the SLO image based on machine learning approach for better diagnosis of retinal diseases. To reduce the complexity of an image, the image is divided into group of pixels which is based on its compactness, colour and regional size then that group of pixels are called Superpixel. Then, classifier is used to classify the true retinal area and artifact. The results provide better performance with 96% accuracy. Comparison is performed with respect to accuracy and computational time between ANN and RBFN.

Keywords Retinal area extraction · Feature generation · Superpixel generation
Machine learning approaches

1 Introduction

Nowadays, biometric systems are becoming suitable for conventional methods such as PIN and password. Biometric techniques are depending upon behavioral or physiological trait like hand geometry, blood vessel patterns, gait, signature and iris. Iris patterns consist of unique information such as ridges, crypts, corona, furrows, rings, a zigzag collarette and freckles [1]. The eyelashes hide eyelid boundaries so eyelid detection is difficult. But eyelashes can be detected and eliminated by the wavelet transform [2].

Nimisha (✉) · R. Gill
Department of Electronics and Communication Engineering, Chandigarh University,
Gharuan, India
e-mail: nimisha25.singh@gmail.com

The diseases like age-related muscular degeneration (AMD) and diabetic retinopathy (DR) affect a large number of population, and also it is expected that these diseases must be increased in our coming future. Generally, digital fundus photography is used to screen and identify the nature of retina-related condition which is possible to allow image storage, in a non-invasive examination and for the transmission at different location. So, the digital retinal images are examined by an expert human grader, i.e., Optometrists and Ophthalmologists which perform the whole process, that is, time consuming and difficult [3, 4]. Automated analysis of retinal images has the ability to reduce the time, and also it must detect the problem of the retinal part very easily. Retinal area is obtained from the imaging instruments such as fundus camera and scanning laser ophthalmoscope (SLO) which contain structure of retinal area with artifacts (eyelashes and eyelids) [5].

Removal of artifact is the important step before the detection of retinal diseases. Extraneous objects like eyelashes, dust and eyelids on the optical surfaces may come in focus and also appear bright. Extraction of true retinal area from the SLO images is important for the diagnosis of eye diseases. This is very difficult to differentiate between the artifacts and the true retinal area. With the retinal area, eyelash and eyelid structures are also imaged due to the large field of view (FOV) in the SLO images [6, 7]. If eyelids and eyelashes are removed, then analysis of retinal area along with the diagnosis of diseases in the visible retina can be done [8].

Retinal disease is the most important issue in medical field [9]. So for diagnosis of these types of diseases, it is essential to extract the true retinal part from the eyes. After extraction of true retinal part, it will be easy to diagnose the disease. We can apply different classifiers for the detection of retina [10]. In this chapter, comparison of the accuracy between ANN and RBFN is shown for the detection of retinal part.

2 Proposed Work

2.1 Image Preprocessing

Preprocessing of image can be done by applying Gamma adjustment. Gamma normalization is a nonlinear operation which is used to control the total brightness of the given image [3].

γ can calculated by:

$$\gamma = \frac{\log_{10}(\mu_{target}) - \log_{10}(255)}{\log_{10}(\mu_{original}) - \log_{10}(255)}$$

$\mu_{original}$ is mean intensity of the original image and μ_{target} is mean intensity of the target image. In this chapter, μ_{target} is set to 80.

2.2 Generation of Superpixels

After the gamma normalization, the next step is to generate the superpixel. In this chapter, superpixel is generated by the adaptive SLIC (SLICO) which is similar to the SLIC. SLIC uses some compactness parameter which is chosen by user for every superpixel in the image [7]. The SLIC generates smooth regular-sized superpixels and also highly irregular superpixels. So, it becomes more complicated to select the right parameter for each image. But in the SLICO, users need not to select the compactness parameter. SLICO adaptively selects the compactness parameter for each and every superpixel differently. For both textured and non-textured regions, SLICO used to generate regular-shaped superpixels. And also SLICO is very fast as compared to SLIC [11].

2.3 Feature Generation

The artifacts and the retinal area can be differentiated by textural- and gradient-based features [6].

Textural features: The textures are examined by statistical method that is gray-level co-occurrence matrix (GLCM). GLCM function determines how often the intensity value i takes place with the adjacent intensity value j. In this chapter, the offset value is kept as 1. GLCM calculates various features but for reduction of computational complexity, this chapter has extracted only four features from GLCM, i.e., contrast, correlation, energy and homogeneity which is determined.

Gradient features—Gradient features are necessary to calculate because of the non-uniformity of the artifacts [6]. Mean, standard deviation and variance are calculated.

2.4 Classifier Construction

The next step is the construction of classifier. This chapter used ANN and RBFN classifier. The classifiers are used for the detection of true retinal part. The artificial neural network is the algorithm that is used for classification purpose. Training samples are taken by ANN as input and then decide that which model is suitable to the training samples. Figure 1 shows the diagram of ANN which includes input layer, hidden layer and output layer [3]. More than one hidden layer can be used, but in this chapter only one hidden layer is used.
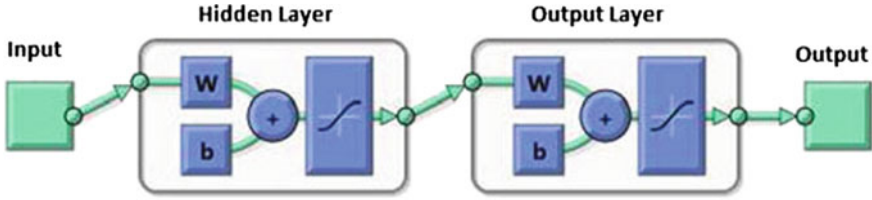
Fig. 1 ANN diagram

2.5 Comparison Study

After the construction of classifiers, the performance using classifiers must be compared in terms of computational time and accuracy. The idea behind using artificial neural network (ANN) is to find the accuracy in small computational time. And artificial neural network with radial basis function (RBFN) is used for high accuracy with large computational time.

3 Results

We performed the experiment on an image which is obtained from Optos [5] that are acquired by the ultrawide field SLO. Field of view (FOV) of each retinal image is up to 200° with the resolution of 14 μm. The retinal image without dilation is captured by the device, over the pupil of 2 mm that is very small. There are two channels in the image, i.e., red and green. The red channel whose wavelength is 633 nm displays the deeper structures of retina against choroid, whereas the green channel whose wavelength is 532 nm gives information about the retinal pigment epithelium to the sensory retina. The dimension of each image is 3900 × 3072, and each pixel is showed by 8 bit on both green and red channels. The dataset of diseased and healthy retina are collected. Many diseased retinal images are collected from diabetic retinopathy patients.

Column (a) from Fig. 2 is the original image that is the dataset. On these images, we can apply different functions for the extraction of true retinal part. Column (b) is the enhanced image by using the gamma normalization techniques. Gamma normalization is a nonlinear operation which is used to control the total brightness of the given image. Column (c) shows the image of superpixel generation. The algorithm which is used to group all the pixels into the different region is called superpixel algorithm, which is used to calculate the feature of the image, and it is also used to reduce the complexity of the image processing task. Column (d) shows the extracted retinal area from the original images of the dataset. Thus, the accuracy of classification is 96.4%.

(a)

(b)

(c)

(d)

| (i) | (ii) | (iii) | (iv) |
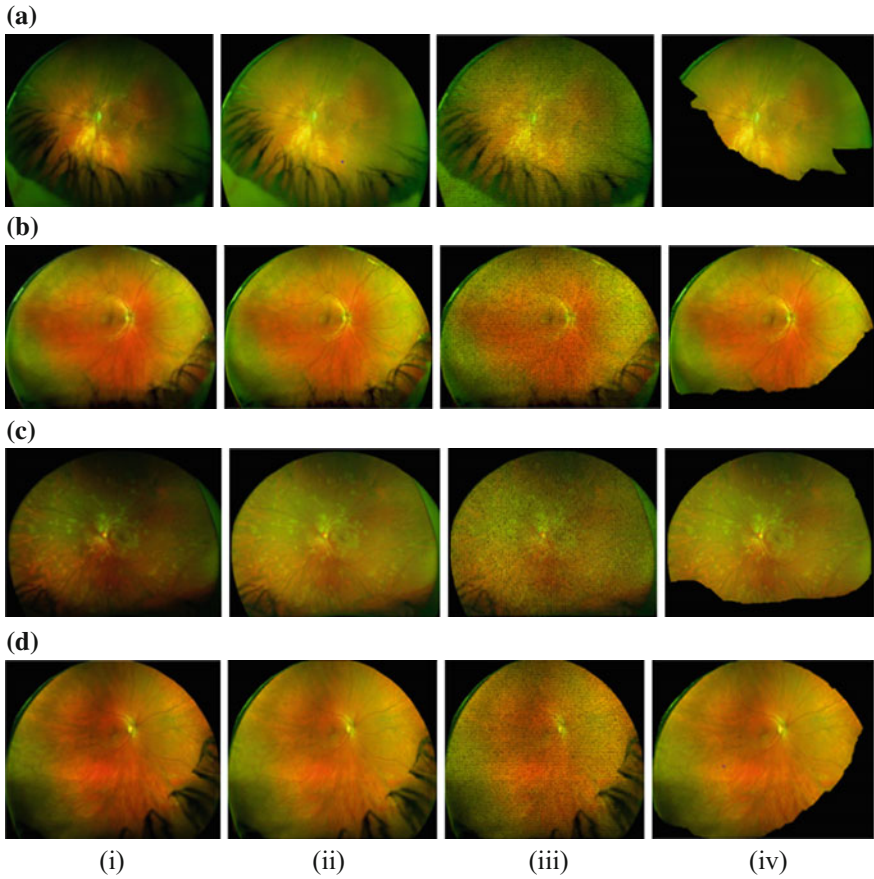
Fig. 2 different results from the four dataset. **a** is the original image1. **b** is the original image2. **c** is the original image3 and **d** is the original image4. Column (i) shows the original images. Column (ii) indicates the image after applying gamma normalization. Column (iii) shows the generation of superpixel of the original image. Column (iv) shows finally extracted retinal part from the original image

Final step is the comparison between ANN and RBFN classifiers. The comparison is performed on the basis of accuracy and computational time. Figure 3 shows the accuracy between ANN and RBFN in which RBFN provides the higher accuracy in comparison with ANN. So from the above figure, it is shown that accuracy is high of RBFN that is 96% in each case, but the computational time of RBFN is less than ANN which is shown in the Table 1.
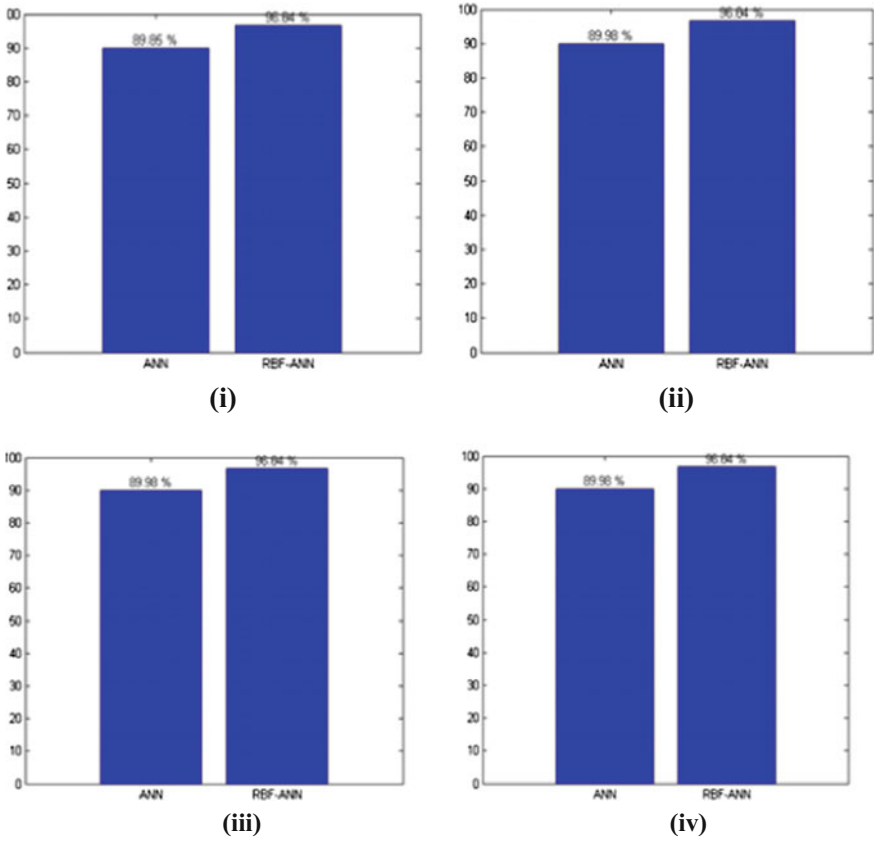
Fig. 3 i Comparison between ANN and RBFN of image (a). **ii** Comparison between ANN and RBFN image (b) **iii** Comparison between ANN and RBFN image (c) **iv** Comparison between ANN and RBFN image (d)

Table 1 Comparison of computational time between ANN and RBFN

Classifiers	Images	Training time (in min)	Testing time (in s)
Artificial neural network (ANN)	Image1	9.24	0.169316
	Image2	12.18	0.070617
	Image3	7.93	0.135918
	Image4	10.82	0.138427
Radial basis function neural network (RBFN)	Image1	32.72	0.164505
	Image2	33.56	0.156286
	Image3	27.83	0.156412
	Image4	33.36	0.199809

4   Conclusion

Extraction of retinal area from the SLO images is very important but difficult task. In this study, we proposed a technique using which true retinal area is detected from the SLO images. We have presented the preprocessing, superpixel generation, feature generation and classifier construction. SLICO is more faster, better performance and more memory efficient for generation of superpixel. Feature generation is very important step to reduce the computational cost. Then classifier construction is used to extract the retinal area. The RBFN classifier has been applied to achieve greater accuracy. Proposed work has been applied on 30 images, and from all these images, it has been concluded that the efficiency achieved is 96% but the computational time is little higher in case of RBFN with compared to ANN. This computational time can be reduced in the future.

References

1. Aligholizadeh MJ, Javadi S, Nadooshan RS, Kangarloo K (2011) Eyelid and eyelash segmentation based on wavelet transform for iris recognition. In: 4th international congress on image signal process, pp 1231–1235
2. Thomas Deserno M (2011) Fundamentals of biomedical image processing. Springer, Berlin
3. Haleem MS, Han L, van Hemert J, Li B, Fleming A (2015) Retinal area detector from scanning laser ophthalmoscope (SLO) images for diagnosing retinal diseases. IEEE J Biomed Health Inf 19:1472–1482
4. Mire V, Dhote BL (2010) Iris recognition system with accurate eyelash segmentation and improved FAR, FRR using textural and topological features. Int J Comput Appl 7: 0975—8887
5. Optos (2014). www.optos.com
6. Abràmoff M, Alward W, Greenlee E, Shuba L, Kim C, Fingert J, Kwon Y (2007) Automated segmentation of the optic disc from stereo color photograph using physiologically plausible features. Invest Ophthalmol Vis Sci 48:1665–1673
7. Li YH, Savvides M, Chen T (2008) Investigating useful and distinguishing features around the eyelash region. In: 37th IEEE workshop applied imagery pattern recognition, pp 1–6
8. Haleem MS, Han L, van Hemert J, Li B (2013) Automatic extraction of retinal features from color retinal images for glaucoma diagnosis: a review. Comput Med Imaging Graph 37: 581–596
9. Gonzalez RC, Woods RE (2006) Digital image processing, 3rd edn. Prentice-Hall, Englewood Cliffs
10. Zhang D, Monro D, Rakshit S (2006) Eyelash removal method for human iris recognition. In: IEEE international conference on image processing, pp 285—288
11. Radhakrishna A, Appu S, Kelvin S, Aurelien L, Pascal F, Su S (2012) SLIC superpixels compared to state-of-the-art superpixel methods. IEEE Trans Pattern Anal Mach Intell 34:2274–2281
12. Kang BJ, Park KR (2007) A robust eyelash detection based on iris focus assessment. Pattern Recognit Lett 28:1630–1639

Compression of Hyperspectral Image Using PCA–DCT Technology

Ramhark J. Yadav and M.S. Nagmode

Abstract Hyperspectral image is a high-resolution image containing large number of details about the presented region. To capture the hyperspectral image, different types of sensor are used in satellite such as AURA, CALIPSO, OCO, and PARASOL, whereas hyperspectral image sensors are used to measure the reflectance of each pixel value in very large amount and stored in 3D cube format. As this type of images is of large data, they possess large amount of storage size and transmission time is also much as compare to other image. Hence, compression of such image is very important to reduce the size and transmission time. Different types of compression methods had been performed on this type of image to compress and to reduce storage size and time. To overcome such problem, a unique compression method should be there to compress the image with less data loss and low transmission period. We introduced a method known as PCA–DCT method for compression which means principal component analysis followed by discrete cosine transform. We are proposing this method in combination pattern to reduce the data dimensionality without loss of data, but as DCT is lossy type of compression, we are going to use DCT to just find the coefficient of the image. The popular known Reed–Xiaoli algorithm is used for the identification of the current pixels with the background image. The quality of image is going to be calculated using the mean square error and peak signal-to-noise ratio value.

Keywords Principal component analysis · Discrete cosine transformation
PSNR · MSE · Compression ratio · Reed–Xiaoli algorithm · Redundancy
HSI · Spectral compression · Spatial compression

R.J. Yadav (✉) · M.S. Nagmode
Department of Electronics and Telecommunication, MIT College of Engineering, Pune, India
e-mail: ramyadav.now@gmail.com

M.S. Nagmode
e-mail: manoj.nagmode@gmail.com

1 Introduction

Hyperspectral imaging is very important part of study in day-to-day life. As this type of image consists collection of large amount of high-resolution spectral detail which is stored in the form of cube were each pixel contain large information. The main difference between the regular image and hyperspectral image is as follows: Regular image is present in two dimension (512×512, 1024×840), whereas hyperspectral image acquires more than three dimensions ($145 \times 145 \times 220$, $640 \times 512 \times 1024$). Hence, reduction in the storage size and transmission period is very useful. Hyperspectral images are widely used on large scale for material identification, target detection, precision agriculture, remote sensing, security and defense, pharmaceuticals, etc. Here, we are using the predetermined database as our input which is present on AVIRIS and HYDICE (e.g., Indian, Salinas, Pavia, Botwana, and KSC). Such type of image is non-contact technique used for thin film manufacturing purpose. On the basis of transformation and dimensionality reduction of the hyperspectral image, correlational and decorrelational techniques have to be considered in mind. Because when dealing with such type of compression losses in image may appear in reconstructed image. Hence, target detection preserves the necessary conditions. As we know, human eyes can only be able to see the visible intensity light in mostly three type (red, green, and blue), other than this color is only seen in combinational manners. Processing such type of image using PCA and DCT, image is first divided into multiple sliding in thin part for obtaining good results.

In this paper, we proposed a simple and time-efficient method to compress the hyperspectral image using principal component analysis further after to discrete cosine transform which mainly works on the spectral domain for reduction of size of storage space (Fig. 1).
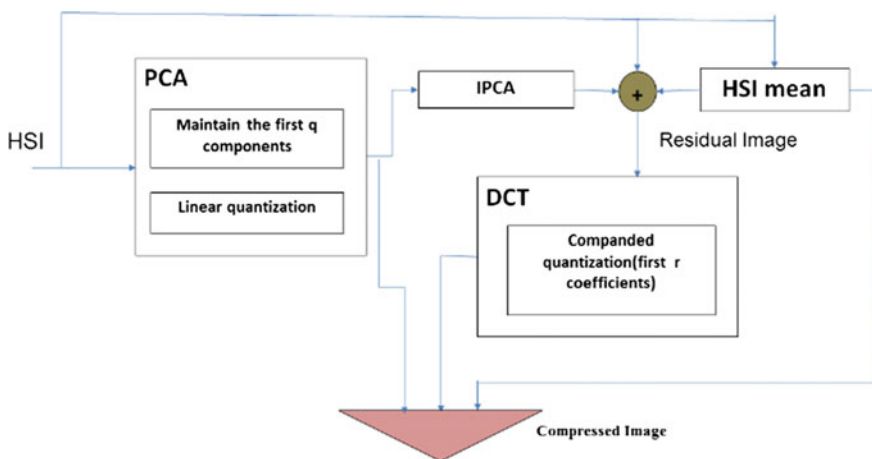
Fig. 1 Block diagram of image compression model

Here, we are using the PCA–DCT method for compression of the hyperspectral image in which principal component analysis is a mathematical formulation for identification of component for reduction of data dimensionality of the image. In principal component analysis (PCA), the standard in the image is matched on the basis of similarities and difference in the image bands in a sequential pattern. The steps which are involved in processing of principal component analysis of the images are first to normalize the original image which is captured by the hyperspectral image sensor to further finding the covariance matrix of the image. Once the covariance matrix is calculated, the eigenvectors and eigenvalues are obtained to find the principal component of the image because principal components are very important to extract the feature vector ($f1$, $f2$,…, fn) and this feature vector is called pattern of the image. Using this pattern, we decide how many feature to select? And which feature value should be selected? When the eigenvector is calculated, we can easily select the principal components (PCs) for finding the similarities and difference in image which further transformed by inverse transformation matrix to reconstruct the image back. Then, this reconstructed image is passed thru the DCT method for finding the coefficient of the obtained hyperspectral image by companding quantization techniques (which is used to identify the first 'r' coefficient). DCT helps to separate the different spectral bands wavelength of the input image. It is very similar to the discrete Fourier transformation (DFT) which consists of many types of transformation methods. The DCT method is mainly bonded by the quantization, encoding and decoding process in such techniques the process is bidirectional in nature to reverse implementation. After applying both methods, we get the compressed image in the same format as we provided in input image. The compression ratio is completely dependent on the selection of the principal component for processing the PCA, and the more the number of (PCs), the higher the quality of the image. PCA–DCT is simplest and fastest method of encoding and decoding the image for compression and maximum levels.

2 Literature Review

Earlier, there are many different types of methods and algorithm that had been used for compressing the hyperspectral image, but most of them are lossy types of compression techniques. Various techniques contain various methods and different pattern.

1. Dimensionality reduction on the basis of spectral and spatial region:

There are basically two type of compression formulation which are either by the 3D compression methods or spatial dimension value. It had been seen that due to higher spectral resolution of the hyperspectral image, if we perform the compression of image in two different regions of operation it will help us reduce storage size, and for that, spectral and spatial reduction is very important. The principal

component analysis and wavelet transformation are few type of compression techniques which are used in the operation of spectral reduction, and it can be done in 2D and 3D spectral bands. The suitable method which is very knowingly used for spectral 2D and 3D compressions are independent component analysis (ICA) and JPEG2000. In independent component analysis, the virtual development concept is used in the processing of image to work on spectral region.

2. Trade-off in data Compression:

The many challenges in compression of data in hyperspectral image are dependent on several factors making the constraints on particular stringent data. First problem which arises is due to the third dimension add-ons the data in image also increases on large scale and it is very necessary to compress the hyperspectral image without loss of contained data in the package which tends to different properties and several challenges in trade-off between reduction and prevention. Second problem which is the difference in spectral and spatial relation between the compression algorithms obsoleted. To resolve this problem, there should be some algorithms to sorting the data package in thin frame pattern works.

3. Redundancy of compression format:

As we know, the capture image by hyperspectral image sensors is of high resolution and used on large scale for target detection by NASA and many satellite in our systems, but in turn it leads to increase a storage space, transmission path, and transmission time without losing the actual data and information. Earlier methods which had been introduced to compress such image format are Minimum/ Maximum Noise Fraction (MNF), principal component analysis (PCA), and independent component analysis (ICA). These are widely used methods in the way of image processing.

Redundancy of image can be identified by calculating the number of bits required to represent the information present in the image or data packets which can be minimized by removing the redundancy from the image. There are mainly three type of redundancies available which are given below:

(a) Spatial redundancy: which occured due to the correlation factor and dependence between the neighboring pixels value.
(b) Spectral redundancy: the difference between the color band or in spectral panels acquire in the image.
(c) Temporal redundancy: temporal redundancy is due to the difference between the different frames in the hyperspectral image.

The compression ratio of the image is calculated on the basis of the input hyperspectral image and the compressed reconstructed image. Consider input image is '$n1$', and the compressed image is '$n2$'. Then, the ratio which is formed from this detailed value is to calculate the compression ratio (CR)

$$CR = \frac{n1}{n2} \tag{1}$$

The compression ratio is readily satisfying the condition for spectral and spatial reduction of the image without loss of the data present in the image. Then, we conclude that the image is redundancy free.

3 Proposed PCA–DCT Compression Method

The proposed compression method mainly combination of both principal component analysis (PCA) and discrete cosine transform (DCT) which are processed in followed thru manner to compress the image to larger scale keeping the mind scale of avoiding the data loss. The PCA process is used to capture the image and process the working on the background extraction by the help of extracting feature function. In this method, the first principal component is maintained 'q' to find the featuring vector of the background image. The residual part of the image is further processed by the DCT technique for separating the bands in different frame pattern using the coefficient of quantization process.

In Markov-I signal, it is shown that the equivalence of PCA and DCT is same in response of compression. As the Markov does not exist, the hyperspectral imaging but then also these two processes can be applicable on it for compressing the image for transmission redundancy. In PCA the characteristics of image taken in consideration whereas if we go toward DCT it is free from characteristics. The steps involved in processing of the hyperspectral image for the reduction of using the PCA–DCT technique have to pass thru the following steps which are given below:

(1) PCA: Considering the hyperspectral image 'X' for processing. The first step for compressing the image of named 'x' which contains row 'j,' column 'k,' and layer 'l.' where each pixel vector is denoted by $\vec{x}$ and $j \times k = N$. Here, the mean of the image factor is obtained as $\vec{m}$ and covariance matrix of the normalized image is given by ϕ. The transformation matrix E and D is used for calculating the eigenvector and eigenvalue in which column is the featuring vector. The eigenvector:

$$\vec{y_n} = \left[\vec{\epsilon\lambda}_1 \ldots \vec{\epsilon\lambda}_g\right]^T . \left[\vec{x_n} - \vec{m}\right], n = 1, Y = \left[\vec{y_1} \ldots \vec{y_n}\right] \tag{2}$$

The quantization matrix of the reconstructed image is denoted by Q_{pca} which extracts the background image, and the remaining part is passed forward for processing.

(2) Construction of residual Image: The qPCs-based extraction image the residual image is further operated by the inverse transformation method for reconstructing the qPCs-based transformation. Inverse transformation is given by

$$X_{qPCs} = (E^T)^{-1}_{[1:l,1:q]} \cdot Y = E_{[1:l,1:q]} \cdot Y \tag{3}$$

Residual image is obtained by subtracting the mean of the image from the original image 'x' where the acquire image is replica of previous image.

(3) DCT and compression output: Discrete cosine transformation is applied on the remaining image from subtraction of the mean of image. DCT is similar to the DFT transformation method in many respect, and it is use for separating the spectral band in different sections for finding the coefficient of the matrix. DCT is performed on each pixel value $\vec{x_n}$ which results in the formation of the vector $\vec{s}:\vec{s} = \text{DCT}(\vec{x_n})$, and the transformed image and quantized image are calculated by using $Y_Q[q \times N]$ from PCA, and the full transformation matrix is denoted in form of $E\ [l \times l]$.

(4) Reconstruction process: The irreversible transformation method is used for reconstruction process which is companded to the quantization of the variance matrix. The quantized coefficients $\vec{S_Q}$ are obtained in S_{rec}.

$$Z_{1DCT} = \text{IDCT} \cdot (S_{\text{rec}}) \tag{4}$$

Dequatizated qPCs-based image is restored using the inverse PCA techniques to restore it into full image form

$$Z = Z_{1DCT} + Z_{1PCA} + m(l \times l) \tag{5}$$

In the final steps, the two resulting matrices are combined to form the mean vector of each pixel values.

4   Flowchat of the Proposed Method

The steps involved in our project are shown in Fig. 2. First step is to read hyperspectral image. After reading the image, image is passed for checking the band value (check band >3). If the band value is greater than 3, then it is considered as valid image for further process or else it is terminated. Once the image is selected, it is processed for background extraction using principal component analysis (PCA). In PCA process, the features are extracted, and on basis of that, the similarities and difference are calculated. And the residual image is further operated
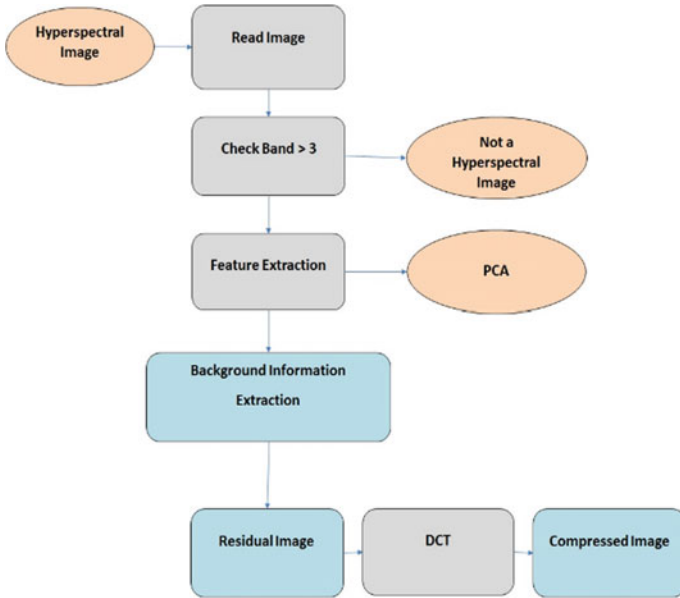
Fig. 2 Flow of the PCA–DCT methods

by DCT to find the coefficients of the residual image. After calculating the point, the compressed image is formed.

5 Experimental and Result

Here, each image goes through a preprocessing phase of automatic noisy and low-variance band removal, before applying compression. The resulting image is considered as the 'original' image to be compressed. The possible compression ratio stems from the value distribution of the two image components of the PCA–DCT transform. The image which is performed in process is taken from the database of HYDICE and AVIRIS (such as Indian, Salinas, KSC, Pavia, and Botwana). The output result of this hyperspectral images is given below (Fig. 3).

RX detection results: Reed–Xiaoli is the method used for identifying the difference between the current pixel and background image. Which is process to detect the region which are processed for compression and at what point does the compression taken part. As shown in the below fig. of RX detection.
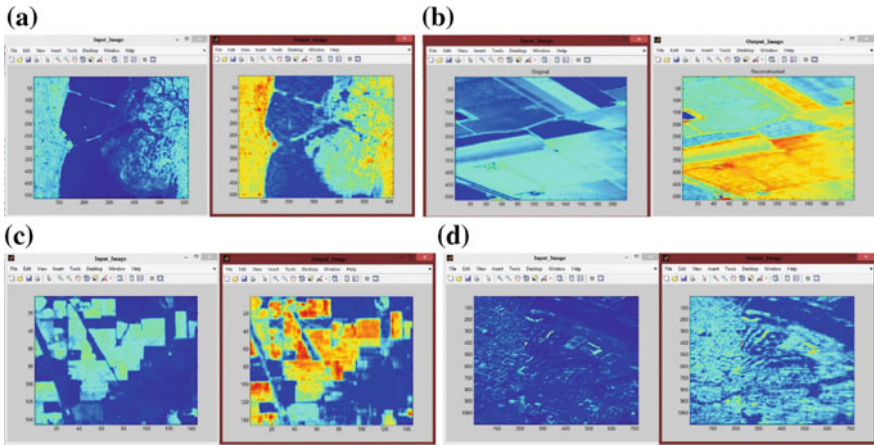
Fig. 3 Resulting image a KSC, b Salinas, c Indian, d Pavia
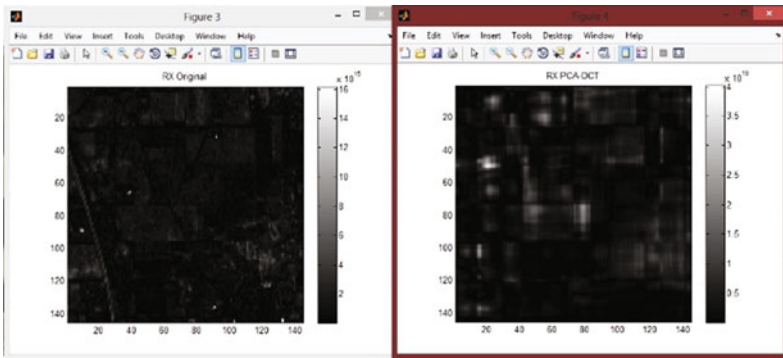
6 Conclusion

The goal of this project is to compress an HSI with the minimal possible degradation of point target detection capabilities. It is obvious that different images are affected differently by the same compression ratio, depending on several characteristics of the image, such as background noise and Gaussianity levels. We proposed a new hyperspectral compression method, which incorporates background extraction and individual compression per pixel. The method does not assume anything about the image and does not require target extraction. It has comparably low computational complexity, with a high compression ratio.

Sr. No.	Original size (MB)	Compressed size (MB)	MSE	PSNR	Compression ratio
Indian (PC = 5)	7.52	1.23	$4.608e{-}07$	22.5781	6.15
Salina (PC = 20)	47.86	4.74	0.1544	33.7543	8.47
KSC	56.85	16.61	0.013268	18.7718	3.43
Data (PC = 20)	140.41	19.28	0	–	7.30

The reconstructed PCA–DCT image achieved the highest similarity to the original image, in MSE terms but with high PSNR value, it maintains the majority of the image information, not only data relevant for target detection, allowing further processing of the decoded image.

References

1. Huber-Lerner M, Hadar O, Rotman SR, Huber-Shalem R (2014) Compression of hyperspectral images containing a subpixel target. IEEE J Select Top Appl Earth Obs Remote Sens 7(6):2246–2255
2. Ramakrishna B et al (2006) Spectral/spatial hyperspectral image compression. In: Motta G, Storer J (eds) Hyperspectral data compression. Springer, New York
3. Huang et al B (2004) Lossless data compression for infrared hyperspectral sounders—an overview. In: Proceedings of the society of photo-optical instrumentation engineers (SPIE), atmospheric and environmental remote sensing data processing and utilization: an end-to-end system perspective, vol 5548
4. Christophe E (2011) Hyperspectral data compression tradeoff. In: Prasad S, Bruce LM, Chanussot J (eds) Augmented vision and reality: optical remote sensing, vol 3. Springer, New York
5. Wang J, Chang CI (2006) Independent component analysis-based dimensionality reduction with applications in hyperspectral image analysis. IEEE Trans Geosci Remote Sens 44 (6):1586–1600
6. Manolakis D (2005) Taxonomy of detection algorithms for hyperspectral imaging applications. Opt Eng 44(6):066403:1–066403:11
7. Caefer CE et al (2008) Improved covariance for point target detection in hyperspectral data. Opt Eng 47(7):076402:1–076402:13
8. Simson Y, Cohen M, Rotman SR, Caefer CE (2004) Point target detection in segmented images. In: Proceedings of the society of photo-optical instrumentation engineers (SPIE), Imaging Spectrom X, vol 5546
9. Du Q, Zhu W, Fowler JE (2007) Implementation of low-complexity principal component analysis for remotely sensed hyperspectral-image compression. In: Proceedings of IEEE workshop signal processing system
10. Indian_pines_corrected' and 'Salinas_corrected' hyperspectral images [Online]

Effective Approach of Finding Missing Children Using Face Age Progressed Prediction Method

Anagha Yogesh Nehete and Krishna K. Warhade

Abstract An improved way of finding missing children is presented in the paper. It is very difficult, time consuming, tedious, and least efficient method to find missing children through traditional methodologies. Using the advanced method, the factor of facial transfiguration of a child is considered with increases the search efficiency. Large numbers of missing children cases are registered in their young age. These lost or kidnapped or missing children undergo huge facial transformation caused by several external, internal, and biological factors. Human face age progressed prediction is the recent upcoming research field which is used to study and decode these facial transfigurations efficiently. The facial image is being fragmented into parts to increase the granularity of face prediction with age progression. The growth curve of feature is used to predict the shape, size, and location of each and every feature at a different age. Hence, this increases the success rate of finding missing child. Various algorithms and image processing techniques are used to carry out this study. This paper aims at finding missing children with the most effective computational method.

Keywords Face age progression prediction · Feature extraction
Aging database · Mahalanobis distance · Basic fitting tool

1 Introduction

Kidnapping is so far the highest reported crime against children. In India, child goes missing in every 8 min, as per survey by the National Crime Records Bureau, India. The most significant application of face prediction is finding a person who went missing for a long time [1]. Face age progression prediction (FAPP) is entirely

A.Y. Nehete (✉) · K.K. Warhade
MIT College of Engineering, Pune, India
e-mail: anagha2605@gmail.com

K.K. Warhade
e-mail: krishna.warhade@mitcoe.edu.in

different concept from age estimation. As time passes, the facial transfiguring based on several factors takes place which makes the search of these missing children more difficult. Hence, computational approaches are introduced in the field of finding missing person from domestic as well as national security point of view. The police in some regions have also begun a database for missing children, like zipnet.in [2]. Detailed study of facial growth curves of child's parent, relatives, and siblings is done to make search more effective and efficient. There exist several hurdles in the path of searching missing child with the approach of

FAPP such as:

1. Growth rate variation: Facial growth rate varies from person to person influenced by several factors such as race, health, socioeconomic stratum, secretion of hormones, lifestyle, and working condition [3].
2. Gender: Growth phase in females is usually early than that of males [3].
3. Aging catalyst: There exist several factors which act as catalysts for premature aging in human face. Some examples of these are working exposure, diet, smoking and liquor addiction, etc. [3].
4. Database: Availability of huge database of images of child along with the parents and relatives at different stages of age is very difficult to get [3].
5. Non-facial factors: Face transfiguration also depends on several global non-facial factors such as hair color, hair style, and baldness of forehead [3].
6. Environmental Factors: Socioeconomic status is one of the major players that can contribute to growth variation [3].
7. Nutrition and Disease throughout the Life Cycle: This factor is further characterized into two i.e., parental growth and infancy and childhood growth [3].

Proper consideration must be done to overcome the mentioned challenges.

Child line [4] is an organization in our country which deals with the child protection and child rights. This organization specially focuses on the children who needs extra care and protection such as children abandoned by parents or guardians, missing children, children who are victims of substance abuse, run away children, children separated from parents due to conflict and disaster, children whose families are in crisis, and victims of child trafficking.

Ministry for Women and Child Development has launched a Web site named khoyapaya.gov.in. [5] This Web site mainly serves the families who have lost their children. This work is done on real time basis i.e., real-time communication among organization, police, and families takes place.

National Center for Missing Children is another organization working for finding missing children across the nation. Strategy used by five zonal areas (north, east, west, south, central). This organization mainly works on two sections: (1) Found children and (2) Separated in childhood [6].

1. Found children: It deals with the tracing of parents of children found from various sources such as orphanages, etc. [6]

2. Separated in childhood: It deals with the information about children who were found missing from their childhood and have grown up now [6].

ZIPNet was introduced in the year of 2004. ZIPNet basically stands for Zonal Integrated Police Network. This is a collaborative effort of Delhi, Chandigarh, Punjab, Haryana, Uttar Pradesh, Uttaranchal, Rajasthan, and Himachal Pradesh Police [2]. ZIPNet project have got several modules in public domain such as missing children, children found, missing person, unidentified dead body, and unidentified person found.

2 Proposed Method

2.1 Feature Extraction

Computer Vision system Toolbox provides MATLAB algorithms for feature extraction, object detection, object tracking, object recognition, and geometric camera calibration and 3D vision (Fig. 1).

2.2 Distance Calculation

In the proposed method, mahalonobis distance computational approach is being implemented. Mahalanobis can also be defined as a measure of dissimilarity between two random vectors x and y of the same distribution with the covariance matrix S:

$$d(x, y) = \sqrt{(x - y)^T S^{-1} (x - y)} \tag{1}$$

Mahalanobis's definition was stimulated by the problem of identifying the similarities of skulls.
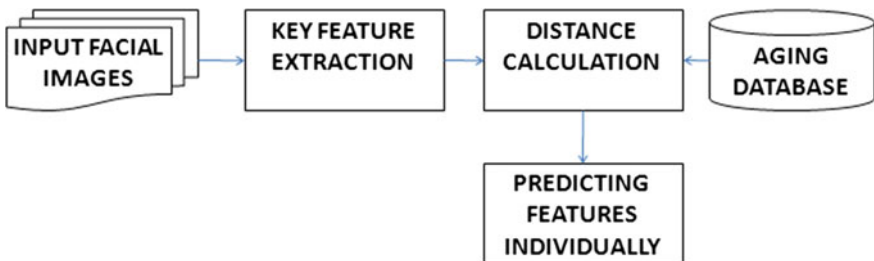
Fig. 1 Flow diagram for FAPP

2.3 Prediction Algorithm

A standard linear regression model with standard estimation techniques makes the prediction of the relationship between the response variables feasible. The vectors obtained from the features extracted acts as response variables. The relation among these variables is computed to find the growth curve of the facial feature of the missing child. Using this growth curves, the future image of facial features is predicted. Polynomial regression is technically a special case of multiple linear regressions. It is often difficult to interpret the individual coefficients in a polynomial regression. The goal of polynomial regression is to model a nonlinear relationship between the independent and dependent variables [7].

2.4 Database

Database is the basic requirement for finding missing children. Size of database increases the search rate of the missing child proportionally. In general, the database provided for finding missing children consist of very few images. These images limit the search efficiency hence the requirement of computation algorithm efficiency increases. To make the age progressed feature sequence study easy, the database is split into five age groups, namely A = A1, A2, A3, A4, A5, where A1, A2, A3, A4, and A5 represent age groups of 0–3, 3–6, 6–11, 11–16, and 16–18 years old, respectively.

3 Results and Analysis

3.1 Feature Extraction

Features from the given facial image are extracted using a computer vision toolbox. The image is initially been divided into three parts, which make the detection of key features (eyes, nose, mouth) easier. Different threshold values are set to detect these features from the image. This threshold is known as merge threshold. The vector of x, y, w, h is computed from the bounding box result of feature extraction of different images from database (Fig. 2).

3.2 Distance Calculation

It basically calculates the pixel distance between the key features of the facial image. The vector of $\{x, y, w, h\}$ is computed for the selected images.
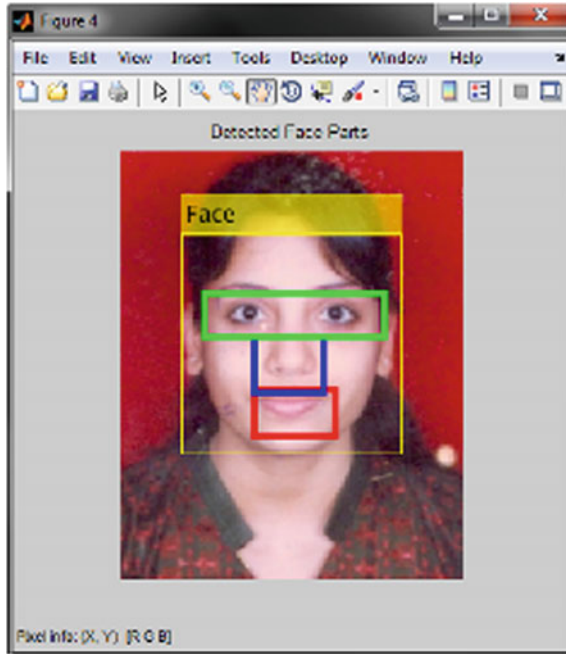
Fig. 2 Snapshot of features extracted from facial image

Table 1 Predicted values of '*x*' coordinate for three distances

Subject no.	Nose-mouth distance	Mouth-eye distance	Nose-eye distance
1	34.468	3.748	0.687
2	65.9	0.856	17.490
3	27.3	0.936	5.43

Table 2 Predicted values of '*y*' coordinate for three distances

Subject no.	Nose-mouth distance	Mouth-eye distance	Nose-eye distance
1	212.836	−7.477	12.958
2	−157.113	0.47	−19.2
3	29.4	−0.199	0.615

For experimentation, 60–70 images are considered of 7 different subjects. This distance calculation gives the rough idea of child's facial growth curve.

3.3 Prediction Algorithm

The method used for predicting future values for distance calculation is polynomial regression. The following table gives the result for distance calculation for different images (Tables 1, 2, 3 and 4).

Following table shows the result for predicted values after 3 years for subject no. 1. Cubic polynomial regression method is used to predict values (Figs. 3, 4).

Table 3 Predicted values of 'w' coordinate for three distances

Subject no.	Nose-mouth distance	Mouth-eye distance	Nose-eye distance
1	9.848	−9.238	7.325
2	4.3	−4.54	3.21
3	−0.0668	−1.14	−1.41

Table 4 Predicted values of 'h' coordinate for three distances

Subject no.	Nose-mouth distance	Mouth-eye distance	Nose-eye distance
1	12.566	−9.125	4.924
2	5.54	0.128	3.49
3	0.775	−0.493	−0.684

X			Y		
NOSE-EYE DISTANCE	MOUTH-EYE DISTANCE	MOUTH-NOSE DISTANCE	NOSE-EYE DISTANCE	MOUTH-EYE DISTANCE	MOUTH-NOSE DISTANCE
0.6868	3.7482	34.468	12.9579	-7.4768	212.836

Fig. 3 Predicted values for x, y for different distances

WIDTH			HEIGHT		
NOSE-EYE DISTANCE	MOUTH-EYE DISTANCE	MOUTH-NOSE DISTANCE	NOSE-EYE DISTANCE	MOUTH-EYE DISTANCE	MOUTH-NOSE DISTANCE
7.325	-9.238	9.848	4.924	-9.125	12.566

Fig. 4 Predicted values for w, h for different distances

4 Conclusion

The precision of result depends on the coefficients of cubic polynomial regression model, which in turn depends on the number of images considered for carrying out the approach. Despite having such a small database, the obtained value for R2 is 1; hence, for huge database, this approach promises excellent results. Currently, accuracy of the proposed method is 63%.

References

1. One lakh children go missing in India every year: Home ministry. http://timeso_ndia.indiatimes.com/india/One-lakh-children-go-missing-in-India-every-year-Home ministry/articleshow/39779841.cms
2. Zipnet. http://blogs.wsj.com/indiarealtime/2012/10/16/indias-missing-children-by-the-numbers/
3. Shen C-T, Lu W-H, Shih S-W, Mark Liao H-Y (2011) Exemplar-based age progression prediction in children faces. In: IEEE international symposium on multimedia
4. International center for missing and exploited children. http://www.icmec.org/
5. The Lost and found website. http://khoyapaya.gov.in/mpp/home
6. National center for missing and exploited children. http://www.missingkids.com/home
7. Linear Regression. http://www.eecs.yorku.ca/coursearchive=2011%20%E2%83%9E%2012%20=%20F=4404%20%E2%83%9E%205327%20=%20lectures%20=%2003%20Linear%20Regression:pdf

Edge Detection Algorithm Using Dynamic Fuzzy Interface System

Vasagiri Venkata Guruteja and Mantosh Biswas

Abstract Edge in an image depends on the viewer's perspective i.e., some viewers may feel it as edge and some may not. Fuzzy logic can be used to solve this partial truth value concept. Many fuzzy-based edge detection methods have been proposed till now, but most of them used the static fuzzy inference system for edge detection, in which we have to change the membership functions for each image in order to get better results. Therefore, to overcome this drawback, we proposed fuzzy logic-based edge detection algorithm with dynamic generation of fuzzy interface system (FIS). The performance of the proposed method is demonstrated through computer simulation results over Sobel, Canny, EFLEDG, and EDFLM edge detection methods in terms of both subjective and fidelity criteria and our proposed method gave good results in terms of F-Measure and visual quality of resultant edge images.

Keywords Digital image · Edge detection · Fuzzy interface system
Membership function · Fuzzification · Defuzzification

1 Introduction

The sudden rise or fall of intensities of gray level values of pixels in the image produces the edges. In segmentation, registration, identification, and recognition, the edge detection plays a vital role.

Many edge detection techniques have been developed in the last few decades. Based on the derivative filter, Sobel [1] and Prewitt [2] edge detectors were proposed. If the edges are blurry or noisy or not flexible, then these gradient-based

V.V. Guruteja (✉) · M. Biswas
Department of Computer Science & Engineering, National Institute
of Technology, Kurukshetra, India
e-mail: vasagirivenkataguruteja@gmail.com

M. Biswas
e-mail: mantoshb@gmail.com

techniques will not generate accurate edges. Marr and Hildreth [3] proposed edge detection algorithm that based on the zero crossings of the Laplacian of an image. Canny [4] proposed a thresholding-based edge detection method which solves the noise problem from gradient operators. Also, the above-mentioned methods use either fixed thickness of edges, or used fixed parameters for good results such as threshold value, σ (width in Gaussian function), and c (Center in Gaussian function). In the last few years, fuzzy logic [5] is emerged as a powerful tool in decision making for the edge detection of the image which can easily overcome the dependency of parameters. Edge Detection methods using fuzzy concept have been increasingly proposed [6–12]. Aijaz et al. [13] proposed rule-based EFLEDG by setting 16 rules to generate a fuzzy interface system (FIS) from the image by 2×2 floating window with the centroid defuzzification method and first-order derivative. This method uses two triangular functions for fuzzification with parameters (0 255 255) for white and (0 0 255) for black, but their uniform fuzzification does not produces effective edges for all images as brighter images, some edges will not be detected and darker images, some false edges will be detected. Suryakant [14] proposed EDFLM-based on 28 inference rule fuzzy interface system which uses 3×3 floating window of the input image as input for the FIS in each pass. This method uses two trapezoidal functions for fuzzification with parameters (0 0 65 90) for black and (80 170 255 255) for white, but their fuzzification process does not produces effective edges for brighter images as the fuzzification mainly focuses on darker and uniform intensity images. However, above fuzzy-based edge detection methods can fuzzify the input image based on some random assumptions, but the fuzzification process should be dependent on the gray values of the given input image; otherwise, it leads to increase in false positive (some non-edge pixels detected as edge pixels) and/or false negative (some edge pixels are not detected) detection of edges. Therefore, to overcome the above shortcoming, we are proposing an edge detection method with automated fuzzification process by finding the parameters for the membership function of fuzzy interface system based on the image information instead of using some constant values that lead to dynamic process of fuzzification. EFLEDG and EDFLM use the static fuzzification process. The static fuzzification process may be easy to implement, but such methods are not well suited for all images. In our work, we use the dynamic fuzzification process to obtain better edge image no matter what the input image will be. The proposed method outperforms the following edge detection methods: Sobel [1], Canny [15], EFLEDG [6], and EDFLM [8] for F-Measure. The rest of the paper is organized as follows. The proposed Fuzzy-based method is explained in second section. We will compute the simulation results and compare the performance of proposed edge detection method with Sobel [1], Canny [4], EFLEDG [13], and EDFLM [14] in third section, and finally conclude in last section.

2 Proposed Fuzzy Interface System

In EFLEDG method, while creating fuzzy interface system, it was assumed that that 100% black for the pixel value '0', 0% black for the pixel value '255', and vice versa for white. Even in EDFLM method, it was assumed that 100% black for pixel value less than or equal to '65' and 0% black for pixel value greater than or equal to '90'. Moreover, 0% white for pixel value less than '80' and 100% white for pixel value greater than '170'. These assumptions for taking constant values of the pixels for black or white are came from the experiments that conducted on their test images with their methods by considering only the best results. In most of these FIS-based edge detection methods, they assume linear functions for categorizing black and white. Therefore, we need to derive an algorithm that will work for any kind of image, and based up on the given input image information, the fuzzy interface system will be generated to produce optimal edge map for the respective input image. So our algorithm fulfills the above assumptions with three stages: Preprocessing Stage, Processing Stage, and Post-Processing Stage to overcome the linear functions of FIS.

2.1 Preprocessing Stage

In this stage, we need to do some preprocessing on the given input image to create the FIS. So, firstly, we find the minimum and maximum gray level value, $g_{\min}$ and $g_{\max}$, respectively, which are present in the input image, $\text{Image}_{\text{Original}}$. Secondly, calculate the number of times, $l(g)$ the gray value 'g' is present in the image where 'g' between the range of $[g_{\min}, g_{\max}]$.

$$l(g) = \text{The number of time the gray value } g \text{ appear in the image.} \qquad (1)$$

Finally, find the first expectation value, second expectation value, and third expectation values: e_1, e_2, and e_3, respectively, for the input image by using the following formulae.

$$e_1 = \frac{\sum_{g=g_{\min}}^{g_{\max}} g \times l(g)}{\sum_{g=g_{\min}}^{g_{\max}} l(g)}, e_2 = \frac{\sum_{g=g_{\min}}^{e_1} g \times l(g)}{\sum_{g=g_{\min}}^{e_1} l(g)} \text{ and } e_3 = \frac{\sum_{g=g_{\min}}^{e_2} g \times l(g)}{\sum_{g=g_{\min}}^{e_2} l(g)}. \qquad (2)$$

2.2 Processing Stage

For creating the FIS and obtaining the edge image from this stage, firstly we scan the each and every pixel of the image with employ the 3×3 scanned window at centered pixel is (i, j), and FIS have 3×3 scanned window with eight inputs and one output pixels: $P_1, P_2, P_3, P_4, P_6, P_7, P_8, P_9$ and P_5 (P_5 is center pixel and other are surrounding pixels), respectively, for each corresponding image pixel of two membership functions: black and white. Here, we used the Gaussian function for fuzzy membership is given below.

$$f(x; \sigma, c) = e^{\frac{-(x-c)^2}{2\sigma^2}}. \tag{3}$$

where, x represents the pixel value. σ (width of the fuzzy set) and c (center of the fuzzy set) are the parameters which we need to calculate.

Secondly, we have to calculate the values of σ and c for both the black and white membership functions. The steps to follow in order to calculate those values are given below.

Consider the black membership function, where we assume 100% black for pixel value '0' and 0.0001% for pixel value Threshold$_{\text{Black}}$ as shown below.

$$f(x; \sigma, c) = 1 \quad \text{at } x = 0. \tag{4}$$

$$f(x; \sigma, c) = 10^{-6} \quad \text{at } x = \text{Threshold}_{\text{Black}}. \tag{5}$$

Therefore, we calculate σ and c values for black membership function from Eqs. (4) and (5) as:

$$C = 0 \quad \text{and} \quad \sigma = \frac{\text{Threshold}_{\text{Black}}}{2\sqrt{3 \log_e^{10}}}. \tag{6}$$

Again, we consider the white membership function, where we assume 100% White for pixel value '255' and 0.0001% for pixel value Threshold$_{\text{White}}$ as shown below.

$$f(x; \sigma, c) = 1 \quad \text{at } x = 0. \tag{7}$$

$$f(x; \sigma, c) = 10^{-6} \quad \text{at } x = \text{Threshold}_{\text{White}}. \tag{8}$$

Therefore, we calculate σ and c values for white membership function from Eqs. (7) and (8) as:

$$C = 255 \quad \text{and} \quad \sigma = \frac{\text{Threshold}_{\text{White}} - 255}{2\sqrt{3 \log_e^{10}}}. \tag{9}$$

Here, the thresholds of black and white pixels, i.e., $\text{Threshold}_{\text{Black}}$ and $\text{Threshold}_{\text{White}}$, respectively, of the 3×3 scanned window are calculated by below function.

$$\text{Threshold}_{\text{Black}} = \begin{cases} e_1 & \text{if } \text{Mean}_{\text{Local}} < e_1 \\ \text{Mean}_{\text{Local}} & \text{otherwise} \end{cases}. \tag{10}$$

$$\text{Threshold}_{\text{White}} = \begin{cases} e_3 & \text{if } \text{Mean}_{\text{Local}} > e_3 \\ \text{Mean}_{\text{Local}} & \text{otherwise} \end{cases}. \tag{11}$$

Here, (x, y) is the center pixel for the scanned window. e_1 and e_3 are obtained from Eq. (2). The local mean intensity, $\text{Mean}_{\text{Local}}$ is given by below function.

$$\text{Mean}_{\text{Local}} = \frac{\sum\limits_{x=-1}^{x=1} \sum\limits_{y=-1}^{y=1} (\text{Image}_{\text{Original}}(x+i, y+j))}{9}. \tag{12}$$

In the proposed work, we used 28 fuzzy inference rules form EDFLM [14]. In the fuzzy interface system, the input gray intensity values are ranged from 0 to 255, and according to the rules, the gray level is converted to the values of the membership functions. The output of the FIS, i.e., after defuzzification, is presented again to the values from 0 to 255. The resultant edge image from this stage will be constructed based on these rules as shown in 28 inference rules table [14]. This edge image may contain some noise i.e., some non-edge pixels may be detected as edge pixels due to our assumptions in Eqs. 7–12. Let us say, the obtained edge image in this stage will be called as $\text{Image}_{\text{edge}}(\text{noised})$.

2.3 Post-processing Stage

In this stage, noise removal is performed on the obtained previous noisy edge image, Image edge(noised) to get Image edge. Some pixels are falsely detected as edge pixels due to the assumptions that we have considered in the above stage. Those pixels can be removed by taking 3×3 scanned window slide over the whole image pixel by pixel rowwise, and the process continues till the whole image is scanned. In every scanned window, if the centered edge pixel is surrounded by non-edge pixels in all eight directions, then that edge pixel changed to non-edge pixel.

2.4 Algorithm

Input: Image$_{\text{Original}}$ $(M \times N)$
Output: Image$_{\text{edge}}$

Step 1: Apply preprocessing on the input image Image$_{\text{Original}}$ and generate the frequency array, the first expectation value (e_1), Second expectation value (e_2), and third expectation value (e_3) using Eqs. (1)–(2).

Step 2: Apply processing to scan the entire image with 3×3 sliding window and use the pixels for the eight inputs ($P1$–$P9$ excluding $P5$) in the FIS system. Calculate C, and σ for black and white Gaussian membership functions by using Eqs. (6) and (9).

Step 3: Fire the fuzzy rules over the eight inputs obtained from the 3×3 scanned window to get edge information of the center pixel and finally obtain the noisy edge image, Image$_{\text{edge}}$(noised).

Step 4: Apply post-processing on the noisy edge image, Image$_{\text{edge}}$(noised), from previous step to remove the noise pixels which have been falsely recognized as edge by the above FIS processing to get final edge image, Image$_{\text{edge}}$.

3 Results and Analysis

We have performed the experiment for edge detection results of considered edge detection methods: Sobel, Canny, EFLEDG, EDFLM, and our proposed on the test images: Plane, Lady, House, and Pepper as shown in Fig. 1a–d. In our experiment, we are considered for objective and subjective qualities of the results in terms of PRECISION [16], F-Measure [16], and visual quality of the restored edge images, respectively. The experimental results are depicted in Tables 1 and 2 contain for PRECISION and F-Measure values, and edge images are shown in Figs. 2a–e, 3, 4, and 5a–e for considered edge detection methods of test images, respectively.

From Table 1, we observe that for Plane, Lady, House, and Pepper images, the PRECISION values are better for the Sobel method when compared with other edge detection methods. This is because of Sobel method detects very less edges which
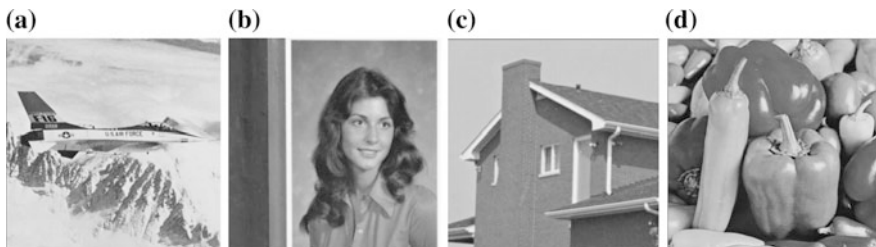
Fig. 1 Test images: **a** Plane, **b** Lady, **c** House, and **d** Pepper

Table 1 Numerical results of PRECISION for Plane, Lady, House, and Pepper images with considered edge detection methods

Image name	Sobel [1]	Canny [4]	EFLEDG [13]	EDFLM [14]	Proposed
PRECISION					
Couple	0.831578	0.795478	0.784157	0.780127	0.801147
Lady	0.854125	0.804789	0.792578	0.791458	0.807895
House	0.867512	0.812478	0.800214	0.799658	0.815789
Pepper	0.841475	0.804415	0.79874	0.791025	0.810214

Table 2 Numerical results of F-Measure for Plane, Lady, House, and Pepper images with considered edge detection methods

Image name	Sobel [1]	Canny [4]	EFLEDG [13]	EDFLM [14]	Proposed
F-Measure					
Couple	0.501458	0.784123	0.794785	0.801414	0.819874
Lady	0.521478	0.791489	0.790145	0.800256	0.813458
House	0.531254	0.799856	0.798547	0.806988	0.820324
Pepper	0.510325	0.781478	0.790142	0.808013	0.818742
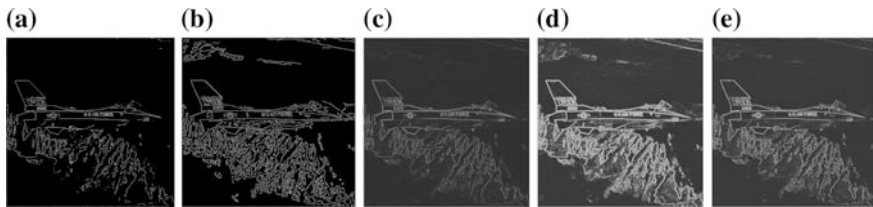
Fig. 2 Plane edge image: **a** Sobel, **b** Canny, **c** EFLEDG, **d** EDFLM, and **e** Proposed
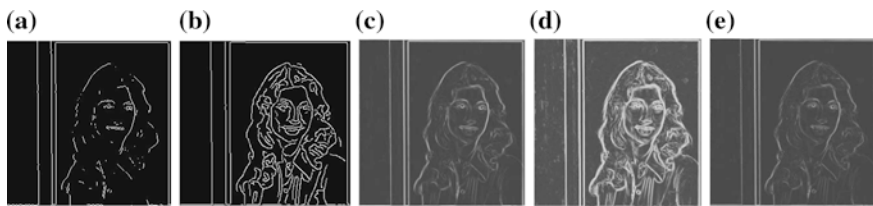
Fig. 3 Lady edge image: **a** Sobel, **b** Canny, **c** EFLEDG, **d** EDFLM, and **e** Proposed

have high probability to occur. So, it may have high PRECISION value, but it got very low recall value. That is why it has a very low F-measure value when compare to other methods as shown in Table 2. Besides Sobel, the proposed method has better PRECISION values for all the considered test images with others considered edge detection methods. This shows that the edges detected by the proposed method actually presents in the original image is very high. The false negative
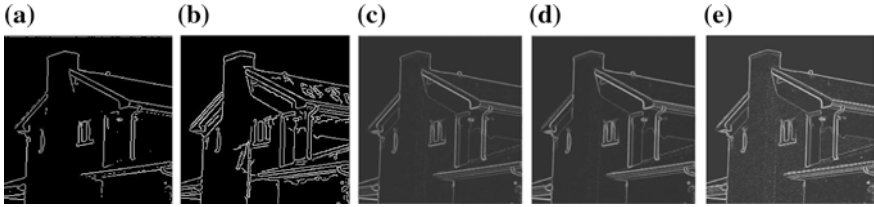
Fig. 4 House edge image: **a** Sobel, **b** Canny, **c** EFLEDG, **d** EDFLM, and **e** Proposed
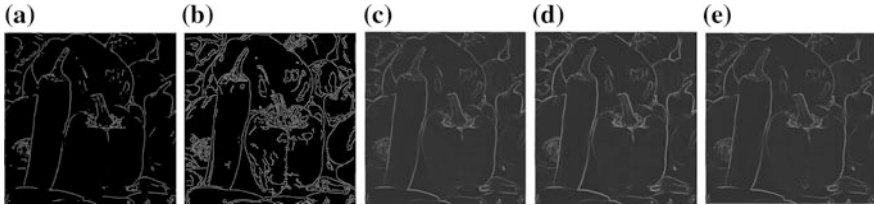
Fig. 5 Pepper edge image: **a** Sobel, **b** Canny, **c** EFLEDG, **d** EDFLM, and **e** Proposed

detection of edges is very low in our proposed method. PRECISION alone cannot be enough to say an edge detection algorithm is effective, because PRECISION cannot consider the missing edges i.e., even if a method does not detect all edges present in the original image still, it may give 100% PRECISION. So, we go for F-measure as well which not only consider the false negative detection, but also the false positive detection of edges.

From Table 2, we observe that for Plane, Lady, House, and Pepper images, the F-Measure value is better for the proposed method when compared with other edge detection methods. This shows that the proposed method not only detects the correct edges, it has a reduced/remove false positive detection of edges.

Form Figs. 2a–e, 3, 4, 5a–e, it was observed that for Plane, Lady, House, and Pepper images, the visual feature is better with visibility and clarity increased for proposed method when compared with Sobel, Canny, EFLEDG, and EDFLM with produced less thick edges and reduced discontinuance.

In our experiment for the edge detection methods, we coded all considered methods are in MATLAB 13.1 running at a Core I5 2.5 GHz processor with 4 GB RAM of computer.

4   Conclusion

In this paper we proposed an edge detection algorithm based on fuzzy with dynamically calculating the membership functions for the fuzzification process based on the image information. It was found that the proposed edge detection

method consistently produces acceptable and better results for all the considered test images based on F-measure and PRECISION with visually edge images over considered traditional methods: Sobel and Canny and fuzzy-based edge detection methods: EFLEDG and EDFLM.

References

1. Sobel I (1978) Neighborhood coding of binary images fast contour following and general array binary processing. Comput Graph Image Process 8(1):127–135
2. Prewitt JMS (1970) Object enhancement and extraction, picture processing and psychopictorics. Academic Press, London
3. Marr D, Hildreth E (1980) Theory of edge detection. In: Proceedings of the royal society of London, pp 187–217
4. Canny J (1986) A computational approach to edge detection. IEEE Trans Pattern Anal Mach Learn Intell 8(6):679–698
5. Zadeh LA (1965) Fuzzy sets. Inf Control 8:338–353
6. Becerikli Y, Karan TM (2005) A new fuzzy approach for edge detection. In: 8th international work-conference on artificial neural networks, pp 943–951
7. Sun S, Liu C, Chen S (2012) Edge detection based on fuzzy logic and expert system. In: Fuzzy inference system—theory and applications, InTech, pp 271–278
8. Bhattacharyya S, Maulik U (2013) Soft computing for image and multimedia data processing. Springer, Berlin
9. Nachtegael M (2003) Fuzzy filters for image processing. Springer, Berlin
10. Ho KHL, Ohnishi N (2005) Fuzzy edge detection by fuzzy categorization and classification of edges. Int Joint Conf Artif Intell 1188:182–196
11. Wu Y, An W, Zhang Q, Chen S (2010) An building edge detection method using fuzzy SVM. Adv Intell Soft Comput 82:819–826
12. Mukherjee S, Majumder BP, Piplai A, Das S (2013) An adaptive differential evolution based fuzzy approach for edge detection in color and grayscale images. In: 5th international conference on swarm, evolutionary and memetic computing, vol 8297, pp 260–273
13. Khan AUR, Thakur K (2012) An efficient fuzzy logic based edge detection algorithm for gray scale image (EFLEDG). Int J Emerg Technol Adv Eng 2(8):245–250
14. Suryakant NK (2012) Edge detection using fuzzy logic in matlab (EDFLM). Int J Adv Res Comput Sci Softw Eng 2(4):38–40
15. Gonzalez W (2008) Digital image processing. Pearson Education, Upper Saddle River
16. Makhoul J, Kubala F, Schwartz R, Weischedel R (1999) Performance measures for information extraction. DARPA Broadcast News Workshop

Super-Pixel Based Segmentation of Urban Area and Its Detection Using Machine Learning Techniques

Inderjeet Kaur and Rana Gill

Abstract In earth observation field, accurate and timely detection of urban area is requisite for military audit, digital mapping, national construction, and change detection which is a major responsibility of urban region planners and government agencies. First, adaptive gamma normalization is used to normalize any given input image without defining target. Adaptive simple linear iterative clustering along with texture feature is proposed for better segmentation of urban area from forest or land. Then, edge and homogeneous region extraction is performed. The comparison of performance results of Sobel operator with two different edge detection operators reveals improvement in detection. Texture feature obtained from gray-level co-occurrence matrix is used to train support vector machine, and further, the edges, which are present in an image, are clustered and labeled by self-organizing map. Then, this output is applied to support vector machine as an input. Finally, morphological hole filling operation makes buildings more perceptible.

Keywords Adaptive gamma normalization · Adaptive simple linear iterative clustering (adaptive SLIC) · Gray-level co-occurrence matrix (GLCM)
Edge extraction · Homogeneous region extraction · Self-organizing map (SOM)
Support vector machine (SVM)

1 Introduction

Detection of urban area in high-resolution satellite images provides valuable results for the analysis of land use and urban region planning [1]. Other applications such as military audit, safe landing of UAV flight in case of emergency, digital mapping, national construction, land exploration and change detection are the reason of detection of urban area [2]. As a prerequisite, urban area monitoring indicates the

I. Kaur (✉) · R. Gill
Department of Electronics and Communication Engineering,
Chandigarh University, Gharuan, India
e-mail: inderjeetkaur60679@gmail.com

man-made objects and the buildings very effectively. To detect and record the changes which may occur in forest, desert, agriculture, and urban area is the major responsibility of urban region planners and government agencies. Government agencies as well as urban region planners are taking useful information from urban area detection for establishing the plans and updating the information of geography system [3].

Rapid growth in development of the urban area leads to change in the land use and the land cover. Thus, utilization of new technique for acquiring the real-time data is becoming essential, to automatically monitor characteristic, growth, and distribution of urban area [4]. Remote sensing images provide important source of information in multiple fields such as analyzing temporal changes [5]. Depiction of urban area in surveillance, geography, cartography, city planning, and remote sensing images is environmentally and socially important [6], so attracted many researchers to work on it with the use of different methods and approaches. Sahin et al. [7] used LISA analysis to detect the urban areas and segmentation of mean-shift based was applied to segment buildings from these areas. Parzen and KNN classifiers were also used to classify buildings. Limin et al. [8] computed two temporal variations, long-term coherence and backscattering intensity, and then applied these extracted features to the unsupervised 2-D thresholding method for accurate change detection. Wei et al. [9] employed histogram peak selection-based unsupervised clustering and edge detection using Canny operator to extract urban building from high-resolution satellite images. Baselice et al. [10] utilized stochastic estimation approach collectively with Markav random field (MRF), where correlation coefficients were estimated between neighboring pixels to detect building edges. Xue et al. [11] used self-organizing mapping network for the detection of urban changes and to avoid the complex requirement of non-changed and changed training area for classification. The results of self-organizing map showed high accuracy rate than that of maximum likelihood classification. Al-Doski et al. [12] employed support vector machine to extract useful information from satellite data, and the postclassification method was employed to monitor changes in the urban area. Bekkari et al. [13] extracted Haralick texture features as space information and spectral descriptor to refine the support vector machine classification results of the urban satellite images. According to literature survey, many researchers have faced different problems while detecting urban area such as roads around buildings were not detected, long rooftop was recognized as non-urban area. The results showed that the proposed method is able to solve these problems (Fig. 1).

The layout of remainder paper is as follows: Sect. 2 illustrates the proposed method, and the experimental results are defined in Sect. 3. Section 4 is a conclusion of proposed work.

Fig. 1 Urban area

2 Proposed Methodology

2.1 Adaptive Gamma Normalization

To enhance the intensity level or visual appearance of image, gamma normalization or correction is becoming the fascinating solution. It is a nonlinear operation which is used commonly to code and decode the radiance or illumination of input image. The advantage of using adaptive gamma normalization is compression over dynamic range is performed by using single adjustable variable factor and is compatible to normalize an image irrespective of its data bits representation and the image resolution [14]. Initial threshold value is computed with Otsu method, to separate two classes of image pixels, i.e., separating foreground pixels from the background pixels on the basis of its gray-level values. Then, the mean for both classes of image pixels is computed as follows:

$$\mu_{\text{new}} = \frac{1}{2}\left[\frac{1}{RC}\sum_{i=1}^{R}\sum_{j=1}^{C}\begin{cases} I(i,j), & |I(i,j)| \geq \text{th} \\ 0, & \text{otherwise} \end{cases} + \frac{1}{RC}\sum_{i=1}^{R}\sum_{j=1}^{C}\begin{cases} I(i,j), & |I(i,j)| < \text{th} \\ 0, & \text{otherwise} \end{cases}\right]$$

This threshold value updates continuously with iterative method till there is no more variations, or absolute difference between initial threshold value and new threshold value becomes less than $\tau = 1e - 10$; it can be formulated as follows:

$$|\mu_{\text{new}} - \mu_{\text{initial}}| = \tau$$

Now the value obtained from above equation is utilized as target mean to normalize an image [15]:

$$\gamma = \frac{\log_{10}(\mu_{\text{target}}) - \log_{10}(255)}{\log_{10}(\mu_{\text{initial}}) - \log_{10}(255)}$$

Gamma normalized image is given as follows:

$$I_{\text{norm}} = \left(\frac{I}{255}\right)^{\gamma}$$

2.2 Adaptive SLIC-Based Super-Pixel Segmentation

Super-pixels are generated by clustering pixels into relevant subregions on the basis of similarity in color features and their proximity. Clustering of pixels is performed in combined 5-dimensional space, i.e., [labxy] where [lab] is pixel color vector in CIELAB color space and [xy] represents the pixel position.

The size of each super-pixel becomes N/k and centroid would be $C = \sqrt{N/k}$, if N is total number of the pixels present in an image and suitable number of super-pixels is k, having equal size. The spatial and color normalization is adapted within each cluster. Maximum color and the spatial distance (m_c, m_s) are observed from previous iteration to normalize the proximity values in following process. Distance measure observed is calculated as follows [16]:

$$D = \sqrt{\left(\frac{d_c}{m_c}\right)^2 + \left(\frac{d_s}{m_s}\right)^2}$$

where

$$d_{\text{lab}} = \sqrt{(l_j - l_i)^2 + (a_j - a_i)^2 + (b_j - b_i)^2}$$

and

$$d_{xy} = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}$$

The main advantage of using adaptive SLIC is more stable compactness of super-pixels (Fig. 2).
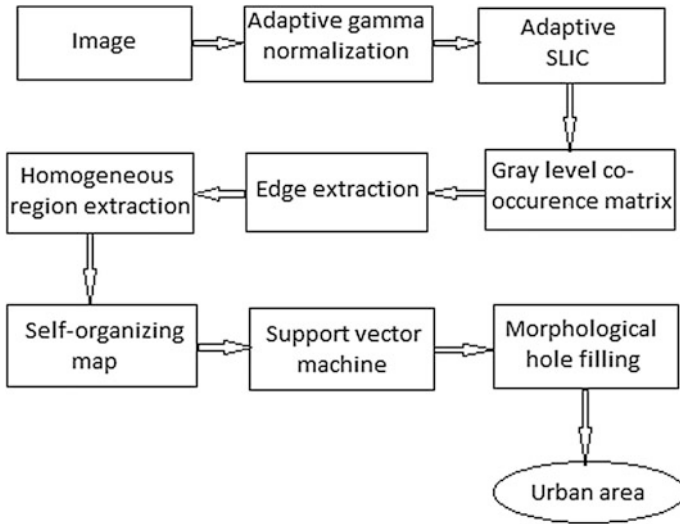
Fig. 2 Simplified diagram of workflow

2.3 Gray-Level Co-occurrence Matrix (GLCM)

It is a well-known method for the analysis of texture where image properties are estimated by measuring second-order statistics. Number of gray-level of image exists in GLCM is same as number of the rows and the columns. Relationship between the pixels is measured to analyze the texture of an image. Frequency of pixel with gray-level *i* occurring at certain distance from the pixel having gray-level *j* is counted in GLCM [17]. Elements in GLCM are computed by following mathematical equation, given as follows:

$$P_{ij} = \frac{P(i,j|d,\theta)}{\sum_{i=1}^{n} \sum_{j=1}^{n} P(i,j|d,\theta)}$$

$0°, 45°, 90°$, and $135°$ are the four directions which are taken into proposed work along with the distance vector to compute four statistics: energy, contrast, homogeneity, and correlation.

Energy is a statistic which is used to represent the existence of homogeneous changes, given as [18]:

$$\text{ENERGY} = \sum_{i=1}^{n} \sum_{j=1}^{n} P_{ij}^{2}$$

The distribution of the matrix values is represented by contrast statistics and shows that with the increase in the value of contrast statistics, clarity of given image increases.

$$\text{CONTRAST} = \sum_{i=1}^{n} \sum_{j=1}^{n} (i-j)^2 P_{ij}$$

The occurrence of the local changes in given image is represented by homogeneity statistic, given as follows:

$$\text{HOMOGENEITY} = \sum_{j=1}^{n} \sum_{j=1}^{n} \frac{1}{1+(i-j)^2} P_{ij}$$

Correlation is used to find the linear dependence of gray among specified and its neighborhood pixels. Joint probability occurrence is computed for specified pixel pairs [19]:

$$\text{CORRELATION} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} (i.j) P_{ij} - \mu_x \mu_y}{\sigma_x \sigma_y}$$

Here,

$$\mu_x = \sum_{i=1}^{n} \sum_{j=1}^{n} i.P_{ij}$$

$$\sigma_x = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (i-\mu_x)^2 P_{ij}}$$

$$\sigma_y = \sqrt{\sum_{i=1}^{n} \sum_{j=1}^{n} (j-\mu_y)^2 P_{ij}}$$

So, four directions for each statistic results in 16 texture features for each super-pixel. Work started with 5000 super-pixels converges in approximately 4900 super-pixels. Thus, GLCM generated the feature vector matrix of size 4900 × 16 from single image.

2.4 Edge Extraction

Edge pixels are those pixels of an image at which change in intensity occurs abruptly and set of those connected pixels is called edge. It transforms an image into the edge image which contains significant image information and meaningful

features. Outlines of roads and buildings in an urban area have so many linear structures. So, edge features can be used to characterize urban area. Sobel operator is used to compute gradient of 'isotropic 3*3' window in horizontal as well as vertical direction to extract edge structure. So, disturbances by non-urban features are effectively avoided by this approach. This is the process of reducing image size very significantly by filtering out the information which is less relevant and preserves the useful structural information of an image.

2.5 Homogeneous Region Extraction

The rooftops of the buildings usually have same texture and color characteristics. Furthermore, the features of neighboring building are also similar. It is found that pixel values of buildings have peak value in one channel at least. On the basis of these characteristics, homogeneous region extraction method is proposed to categorize the object to be detected by some parameter analysis such as size, position, and shape. Maximum value is selected for each channel, and following function is used to normalize the values of pixels [2]:

$$I_N^C = \frac{I^C}{\max(I^C)}, \quad C \in \{r, g, b\}$$

where I^C means the each color channel of an input image I. The pixel value is compared with $\gamma = 0.3$ threshold value using following formula [2]:

$$I_B^C(x_i, y_j) = \begin{cases} 1, & |I_N^C(x_i, y_j) \geq \gamma \\ 0, & |\text{otherwise} \end{cases}, \quad C \in \{r, g, b\}$$

Thus, it will change I_{norm} normalized image into binary image.

2.6 Self-organizing Map

Self-organizing map is a model of neural network, used as unsupervised learning technique named by professor Tuevo Kohenen developed in 1982. It labels the patterns hidden in input images. Topological properties are preserved in an input space by the use of neighborhood function which make it different from the other neural networks [20]. Each node of self-organizing map has weight vector of similar dimension as that of the input data vector and location in space map. Input space of high dimension is mapped into low-dimensional space.

It is a clustering model, which sight to collect the objects which are matching with the statistical distribution. Two approaches which are utilized by self-organizing map to work are training and mapping. In the 'training approach,'

Fig. 3 Architecture of one-dimensional SOM [21] shows that input vectors are mapped to weight vectors in the feature map and preserves the topological structure
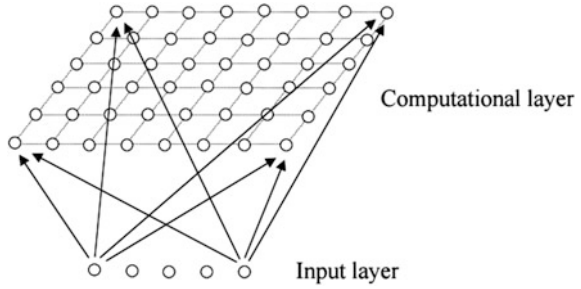
map is built from input examples and classification of new input vector is achieved by 'mapping approach' (Fig. 3).

It clumps all the edges present in an input image and then performs labeling of all the clustered edges. This output of self-organizing map is applied to the following step for the further processing.

2.7 Support Vector Machine

The support vector machine is used as supervised machine learning technique to differentiate roads and buildings of urban area from non-urban area. In a high-dimensional space, it constructs hyperplane for the purpose of classification. Nearest training data points have largest distance in hyperplane, i.e., good separation is obtained [22]. The training data set is applied to train the classifier and then performed the classification over testing data set. The class of testing object is estimated from side of gap it lies (Fig. 4).
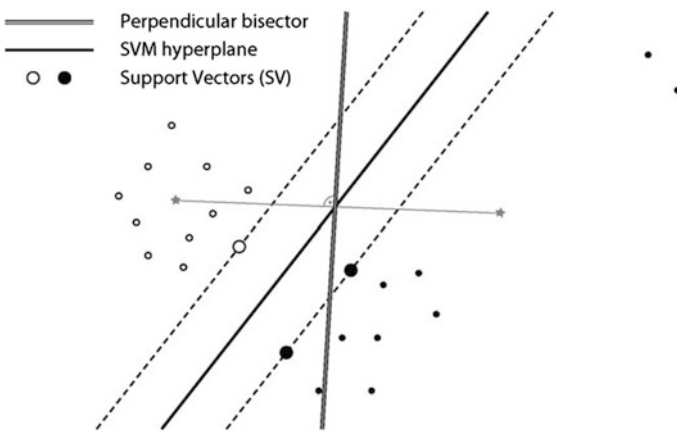
Fig. 4 Example of separating two classes with SVM [23]. SVM constructs a hyperplane (*solid line*) and maximizes the marginal distance between two different classes (support vectors)

Clustered edges obtained from unsupervised self-organizing mapping are applied as an input to the support vector machine. This method is trained over the rough edges also and then, by originating the function from appropriate examples, it becomes able to identify the buildings which are having rough edges in the urban area.

2.8 Morphological Hole Filling Operation

The output of support vector machine has background region which is surrounded by some connected border of the foreground pixels. If direct results are used, some non-urban pixels will be put into an urban category and results into inaccurate output. Thus, morphological hole filling operation is performed and formulated as follows:

$$X_k = (X_{k-1} \oplus B) \cap A^C \quad k = 1, 2, 3. \ldots$$

where A^C is complement of A and B is a symmetric structuring element. This makes the contrast of rooftop stronger than the surrounding area, and building areas become more perceptible.

3   Experimental Results

The performance of this method is evaluated on GeoEye-1 and QuickBird satellite images of different cities with spatial resolution of 0.5- and 0.61-m, respectively. Numerous scenes such as forests, plains, urban areas, mountains, roads, and farmlands are covered in these images. All these satellite images have different sizes.

The results of computation are achieved on personal computer with Intel Core i3-3110 M, processor of 2.40 GHz and RAM of 4 GB. MATLAB R2013b is used for the coding of the proposed method.

The precision and recall quantities are computed to analyze the performance of the proposed method for detection of urban area, given in [2]:

$$P = \frac{TP}{TP + FP} \quad R = \frac{TP}{TP + FN}$$

Here, *TP* denotes number of the true positives, *FP* represents number of the false positives, and *FN* means number of the false negatives.

Some typical results of each steps proposed for the detection of an urban area are depicted in Fig. 5a–f. An original image is taken over which the proposed method is performed. Adaptive gamma normalization benefits the system by controlling the input image's intensity (gamma parameter), thereby making it sure that the input image will easily be operated upon given threshold parameters for a wide range of images. Initial threshold value is computed with Otsu method to separate two classes of image pixels. A new threshold value is computed by computing mean of both classes of image pixels. Iterative method updates the threshold value till the absolute difference between new threshold and initial threshold becomes less than τ and then utilizes as target mean to normalize given input image. Further, the application of adaptive SLIC has enhanced the computation time required for computing texture features, divided to super-pixel level, which have been used in the supervised learning approach. Started 5000 super-pixel converges to 4900 super-pixel. Then, GLCM is used to generate feature set for each super-pixel and it applied to supervised and unsupervised techniques for the accurate detection. Before that, Sobel operator is used to compute gradient of 'isotropic 3*3' window in horizontal as well as vertical directions to extract edge and structure. After the homogenous region extraction, the computed system uses morphological hole filling operation over the image obtained after edge fusion. This helps in detection and marking of the regions like rooftops or planar region with less texture. Finally, the proposed method results into a much better detection of urban area.

Sobel operator detects the edges, where the magnitude of the gradient is high. Due to this, Sobel operator becomes more sensitive to the diagonal edges than the horizontal/vertical edges. The Prewitt operator is also a discrete differentiation operator and uses similar equation of Sobel operator except one constant. The mask of Prewitt operator is easy to implement, but has high sensitivity to the noise. So, there is minor difference in the accuracy achieved by Sobel operator and Prewitt operator. Canny operator is considered as standard and most commonly used edge detection technique. These three different edge detection techniques are applied on more than 20 urban images. Compared to Prewitt operator and Canny operator for the detection of edges, Sobel operator shows an improvement in terms of detection accuracy (Fig. 6).

Table 1 lists the results of precision and recall of using different edge detection method on five different urban area images. As shown in Table 1, performance decreases by using Canny operator, but there is a minor difference in accuracy obtained by using Prewitt operator as compared to Sobel operator.
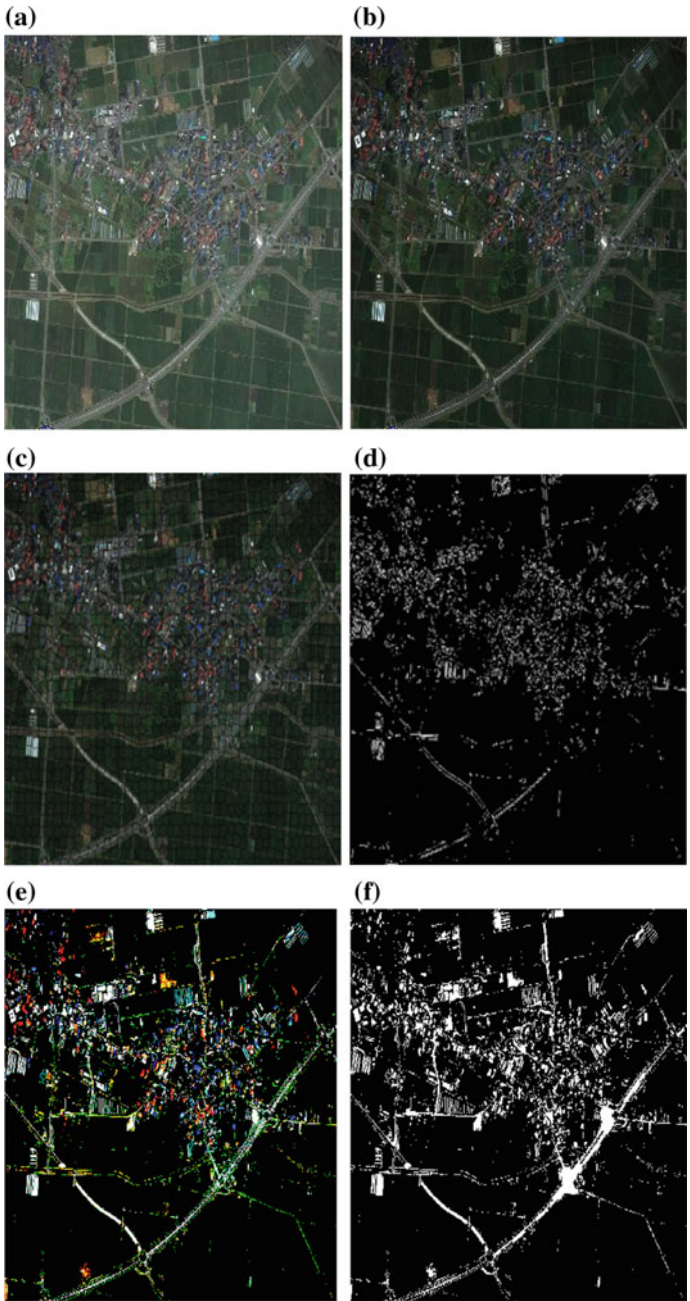
Fig. 5 Proposed method results **a** original image, **b** gamma normalized image, **c** super-pixel generation, **d** edge extraction, **e** homogeneous region extraction, and **f** final result
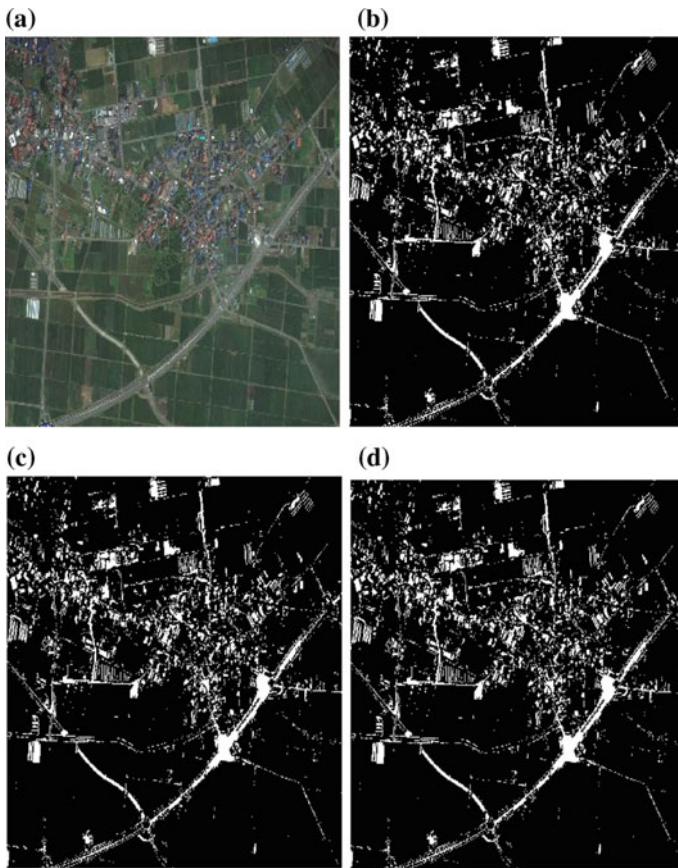
Fig. 6 Different edge extraction techniques' results **a** original image, **b** Sobel operator result, **c** Prewitt operator result, AND **d** Canny operator result

Table 1 Urban area detection performance results obtained by using three different edge detection techniques

Images	Sobel operator		Prewitt operator		Canny operator	
	Precision (%)	Recall (%)	Precision (%)	Recall (%)	Precision (%)	Recall (%)
Image 1	94.53	94.52	94.47	94.54	89.58	88.47
Image 2	95.38	95.48	95.39	95.43	89.69	89.37
Image 3	90.38	90.58	90.69	90.36	86.89	86.89
Image 4	90.50	90.97	90.47	91.01	83.11	83.64
Image 5	87.69	87.81	87.64	87.76	83.14	82.53

4   Conclusion

In this proposed work, detection of urban area over high-resolution satellite images is performed. At first, adaptive gamma normalization is used to enhance the visual appearance of given input image. Then, an adaptive SLIC along with texture feature is proposed for better segmentation of urban area from forest or land. GLCM is used as texture feature extraction method which serves as preprocessing stage for city planning, change detection, and building recognition. Edge extraction and homogenous region extraction are applied for the next step of processing. Both unsupervised and supervised machine learning techniques, self-organizing map and support vector machine, respectively, are applied. SOM clusters all the edges present in an image and labels them. SVM uses the output of SOM as an input to accurately detect the building having rough edges in urban area. Results show that proposed method is also able to detect road around building and buildings having rooftop. At the end, morphological hole filling operation as a postprocessing step is used to make contrast of rooftop stronger and perceptible than surrounding area. Further, there are two main problems according to test results. First, performing the edge learning is a time-consuming process. Second, performance results decrease for the area having low texture variation. Future research will focus on listed problems.

References

1. Sirmaçek B, Unsalan C (2011) Urban area detection from remotely sensed images using combination of local features. In: IEEE international conference on recent advances in space technologies, Istanbul, pp 188–192
2. Shi H, Chen L, Bi F, Chen H, Yu Y (2015) Accurate urban area detection in remote sensing images. IEEE Geosci Remote Sens Lett 12(9):1948–1952
3. Zheng C, Wang L, Zhao H, Chen X (2014) Urban area detection from high-spatial resolution remote sensing imagery using Markov random field–based region growing. J Appl Remote Sensing 8:083566(1)-083566(14)
4. Li S, Zhang B, Gau L, Sun X (2009) Small objects detection of hyperspectral image in urban areas. In: Urban remote sensing joint event, pp 1–5
5. Tao C, Tan Y, Yu J, Tian J (2012) Urban area detection using multiple kernel learning and graph cut. In: IEEE international conference on geoscience and remote sensing symposium, pp 82–86
6. Azmedroub B, Ouarzeddine M (2015) Urban area detection using polarimetric SAR images. In: IEEE international conference on geoscience and remote sensing symposium, pp 3227–3230
7. Sahin Y, Teke M, Erdem A, Duzgun S (2012) Urban area and building detection on high resolution multispectral satellite images using spatial statistics. In: 20th signal processing and communications applications conference (SIU), pp 1–4
8. Liming J, Mingsheng L, Lijun L, Hui L (2005) Unsupervised change detection in urban area using multitemporal ERS-1/2 InSAR Data. In. IEEE international conference on geoscience and remote sensing symposium, vol 6, pp 4002–4005

9. Wei Y, Zhao Z, Song J (2014) Urban building extraction from high-resolution satellite panchromatic image using clustering and edge detection. In: IEEE international conference on geoscience and remote sensing symposium, vol 3, pp 2008–2010

10. Baselice F, Ferraioli G (2012) Statistical edge detection in urban areas exploiting SAR complex data. IEEE Geosci Remote Sens Lett 9(2):185–189

11. Xue C, Xiaowen L, Jianwen M (2004) Urban change detection based on self-organizing feature map neural network. In: IEEE international conference on geoscience and remote sensing symposium, vol 5, pp 3428–3431

12. Al-Doski J, Mansor SB, Shafri H (2013) Support vector machine classification to detect land cover changes in Halabja City, Iraq. In: IEEE business engineering and industrial application colloquium, pp 353–358

13. Bekkari A, Idbraim S, Mammass D, Yassa ME (2011) Exploiting spectral and space information in classification of high resolution urban satellites Images using Haralick features and SVM. In: IEEE international conference on multimedia computing and systems

14. Jang YM, Cho JH, Kim JG, Kim BC, Cho SB (2014) Enhanced tone mapping of high dynamic range to correspond to illumination changes. In: International soc design conference, pp 258–289

15. Haleem MS, Han L, Hemert JV, Li B, Fleming A (2014) Retinal area detector from Scanning Laser Ophthalmoscope (SLO) images for diagnosing retinal diseases. IEEE J Biomed Health Inf 19(4):1472–1482

16. Qin F, Guo J, Lang F (2015) Superpixel segmentation for polarimetric SAR imagery using local iterative clustering. IEEE Geosci Remote Sens Lett 12(1):13–17

17. Chang KI, Bowyer KW, Sivagurunath M (1999) Evaluation of texture segmentation algorithms. In: IEEE computer society conference on computer vision and pattern recognition, vol 1, pp 294–299

18. Bhad VK, Ramteke K (2015) Content based image retrieval a comparative based analysis for feature extraction approach. In: International conference on advances in computer engineering and applications, pp 260–266

19. Yang MD, Su TC, Pan NF, Liu P (2008) Feature extraction of sewer pipe failure by wavelet transform and co-occurrence matrix. In: 2008 international conference on wavelet analysis and pattern recognition, pp 579–584

20. Savita NG, Arora P (2014) Detection and classification of plant leaf disease using image processing techniques. Int J Recent Adv Eng Technol 2(3):1–7

21. Bullinaria JA (2004) Self organizing maps: fundamentals. In: Introduction to neural networks: lecture 16

22. Gavhale KR, Gawande U, Hajari KO (2014) Unhealthy region of citrus leaf detection using image processing techniques. In: IEEE conference for convergence of techniques

23. Rumpf T, Mahlein AK, Steiner U, Oerke EC, Dehne HW, Plumer L (2010) Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. Elsevier J Comput Electron Agric 74:91–99

Curvelet Transform Based Statistical Pattern Recognition System for Condition Monitoring of Power Distribution Line Insulators

P. Surya Prasad and B. Prabhakara Rao

Abstract The power distribution system is considered as the backbone of power system as it is responsible to deliver power to all the consumers. Due to an enormous increase in the power consumption, the damage in insulators at the electric poles is triggering interruption of power, and hence there is substantial loss occurring for the power sector. The power distribution system is protected from heavy transients by the use of insulators. So, monitoring system must be employed which regularly detects the condition of the insulators. Regular monitoring of the overhead power lines along with insulators, sending the images to the processing unit and application of image processing concepts to classify the insulator health condition is the proposed method and hence the determining breakage condition of the insulators. K-means clustering is used for segmenting the acquired image. Then, the insulators are extracted from the acquired image input, and curvelet transform-based features are obtained. These features are given to support vector machine for the determination of health of the insulator. Monitoring of the health of insulators can thus be done consistently, and this method of automatic classification reduces the human efforts too. Hence the efficiency of transmission is improved and continuous supply of power can be delivered.

1 Introduction

The power distribution networks commonly use porcelain insulator to isolate line conductors and supporting structure and so they play a major role in the overhead power lines. As the insulators are mostly mounted in open air and are subjected to diverse atmospheric situations in different geographic areas they are very much prone to contamination. A layer forms on the insulator due to the contaminants or pollutants in severe atmospheric conditions like drizzle or fog. They partially

P. Surya Prasad (✉) · B. Prabhakara Rao
Department of ECE, MVGR College of Engineering,
Vizianagaram 535005, Andhra Pradesh, India
e-mail: suryaprasadp@yahoo.com

dissolve and result in the forming of a conductive layer. This causes a flow of leakage current through the surface of the insulator. Therefore, the insulator's voltage withstanding capability reduces. The insulator's voltage withstanding capability is measured in terms of surface flashover. Finally, breakdown of the insulator occurs which obstructs the reliable operation of power distribution. Therefore, insulator condition monitoring is essential to have smooth and continuous power supply. So, it is very important to monitor the distribution line constantly and identify the signs of failure of the insulators before they actually disrupt the distribution of power [1]. The overhead lines have many other devices than the insulators, which can also be monitored online [2]. The process of health condition monitoring of distribution circuit elements online is generally known as condition monitoring (CM) [3].

The conventional methods in condition monitoring of insulators are pole climbing and aerial surveillance which are dangerous and have many drawbacks. So a video surveillance system with fixed cameras is a viable option for insulator condition monitoring [2] of the overhead power distribution system. Computers and Remote Terminal Units (RTUs) are the tools required to take the images of insulators at regular intervals of time. The RTUs are attached on the poles such that images can be taken properly. With the fast advancement of low-cost video cameras, the cameras can easily be mounted on the poles at low cost, and the images taken from the power distribution lines having the insulators along with them at regular intervals of time are directed to the control room using the RTUs for further processing [2] and determine the insulator's health condition. The objective of this chapter is to determine the condition or well-being of the power distribution system insulators based on image processing and machine learning techniques. The curvelet transform feature vectors are extracted from the images and given to SVM to

Fig. 1 Flowchart of the insulator classification process
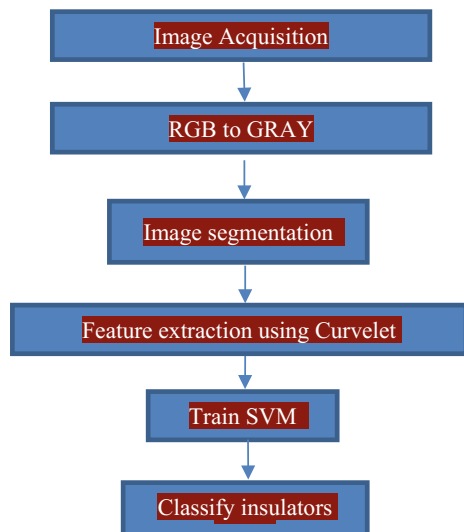
classify the insulators into good, marginal or risky insulators, hence alarming the personnel to repair or replace the risky insulators over the power distribution lines in time and preventing power disruption and the resulting heavy losses. The work proposed is useful in places where human inspection is impossible like hilly terrains and poles located in flowing water. The algorithmic process of the proposed work is depicted in Fig. 1. The acquired images taken at periodic time intervals are converted into gray scale, and K-means clustering procedure is used to separate the insulator portions from that of the pole. The curvelet features of such segmented images are extracted and classified into different categories using Support Vector Machines (SVM).

2 Curvelet Transform

Curvelets are an appropriate basis to represent the images. It is smooth apart from the singularities along smooth curves, where bounded curvature is there for the curves. If such an image is zoomed, the edges contained appear increasingly straight. This property is taken as an advantage by the curvelets. So, they are an effective way to detect the image activity along the curve. In this work, the discrete version of curvelet transform, also called as 'USFFT' (unequally spaced Fast Fourier transform) algorithm, is used. The resulting algorithm, also called as FDCT (Fast Discrete Curvelet Transform) algorithm, is simpler, faster and less redundant.

An image can be decomposed by the curvelet transform into multiple scales and orientations by taking a mother curvelet function, [4] scaling and rotating it. Due to its needle shape, it is suitable for curvelets in describing the curves that are smooth away from the discontinuities. As it detects the curves in an image at different scales and different orientations, it is robust from variations of pixel intensity and also can describe edges along the curves more powerfully compared to the traditional wavelet-based transforms [5]. For image processing applications, it has been widely used in the areas of content-based image retrieval [6], face recognition [7] and SAR image classification [8] among other applications. In this chapter, fast discrete curvelet transform [9] based on the USFFT (unequally spaced fast Fourier transform) is used to calculate the curvelet coefficients. Since it is an extension of the wavelet concept, it also became a widely used tool in the image processing.

A decimated rectangular grid tilted along the main direction of each curvelet is used by the FDCT approach. Using this approach, it is easy to analyze the signal by using basis with large coefficient as the feature of the signal and checking that basis is useful in pattern recognition. Of the various curvelet coefficients, coarse coefficients are used to extract low-frequency component and get general information from the image. The detail and fine coefficients are used to extract the high-frequency component of the image which detects the edges of the image.

3   Classification

In this work, SVM classifier [12] is used for the classification of power distribution system insulators. In the recent past, the SVM machine learning algorithm has become very popular because of its greater performance in pattern recognition tasks compared to the traditional techniques. In general, supervised pattern recognition problems are solved by SVM classifier which is proposed as a very useful method by Vapnik et al. [10, 11]. The SVM classifies each one of the insulator images of the testing dataset into one of the two classes, either good or risky based on the category of the most similar insulator in the training dataset. The samples are separated from different classes with a surface that maximizes the margin between the classes by the application of SVM. And it has been in wide use to separate into defective and good items from a set of observations. Face recognition belongs to the class of nonlinear problem, which increases the difficulty of its recognition, and SVM is a statistical machine learning tool which can be used for such applications in finding the global optimum solutions. When health classification of insulators is done using SVM, a given set of labeled data is separated by the classifier with a hyperplane which is at a distance from them. SVM along with adaptive neuro-fuzzy inference system (ANFIS) was employed in [12, 13] to obtain the insulators' health condition with the help of discrete orthogonal S-transform (DOST)-based features.

4   Results and Discussion

In this work, the dataset is taken from the images of poles having the insulators of the 11 kV overhead power transmission lines as shown in Fig. 2. The dataset contains 80 images of such insulators out of which 50 images are used in the
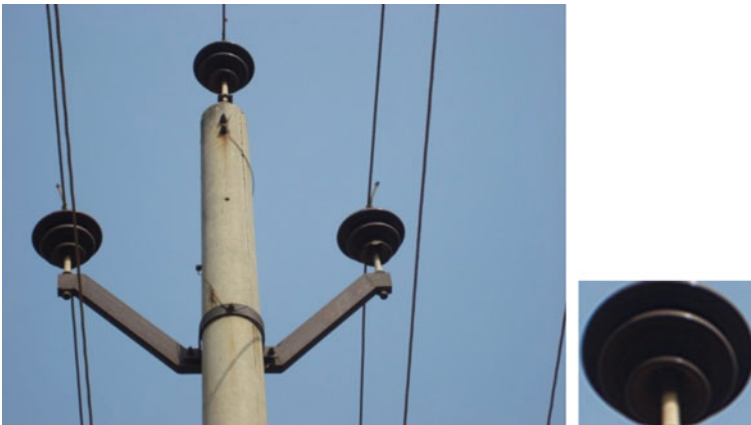

Fig. 2 The source image and the extracted insulator

training phase and 30 images are used for testing purpose. Then classification is done into the corresponding healthy, marginal and risky types of insulators, respectively. The curvelet statistical features extracted for classification purpose are standard deviation, median, mean, mode and entropy, respectively. They are obtained for each image's curvelet coefficients, and feature vectors for the entire dataset are formed. Feature vectors of the testing database are given to SVM for the classification purpose. The classification accuracies reported in this work are average classification accuracy.

In this section, the experiment results are presented. The experiments are conducted, and the performance of the proposed approach is evaluated by determining the algorithm's classification accuracy. From among the 80 image dataset considered for the experiment, training process uses around 60% of the images and the remaining images are used for testing purpose. The SVM classifier is used to classify the testing image dataset into 3 classes. They are Class 2 for good, Class 1 for marginal and Class 0 for risky states, respectively. As per the classification results of the SVM using curvelet features, the directly observed deviation from the actual state of the images is 3 images. Therefore, the classification accuracy or efficiency can be calculated as $((30 - 3)/30) \times 100 = 90\%$.
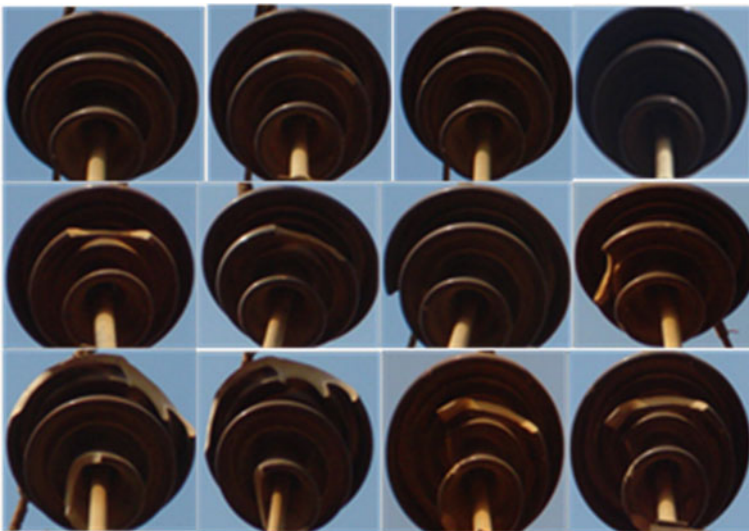

Fig. 3 Insulator images having healthy, marginal and risky states

5 Conclusions

Condition monitoring analysis of the overhead power distribution line insulators is very much essential in improving the efficiency of automation of power distribution system. In this work, the extracted insulators from the power distribution lines are classified using the curvelet transform-based features. These features are extracted from the insulator images for a training set of images, and the test images are classified into the states, good, marginal and risky, respectively, using SVM (Fig. 3). The SVM classifier is used to obtain the predictions from the features obtained for a set of insulator images. The testing data is classified by comparing the features of the testing images with the features of the training data and assigning the corresponding best-matched class. As per the experiment results, it can be established that the power distribution system insulators' health condition can be supervised easily by using the proposed method of classification, and the following necessary actions can be taken.

References

1. Zhang D, Li W, Xiong X (2014) Overhead line preventive maintenance strategy based on condition monitoring and system reliability assessment. IEEE Trans Power Syst 29(4):1839–1846
2. Murthy VS, Tarakanath K, Mohanta DK et al (2010) Insulator condition analysis for overhead distribution lines using combined wavelet support vector machine (SVM). IEEE Trans Dielectr Electr Insul 17(1):89–99
3. Aggarwal RK, Johns AT, Jayasinghe et al (2000) An overview of the condition monitoring of overhead lines. Electr Power Syst Res 53(1):15–22
4. Cand'es EJ, Donoho DL (1999) Curvelets—a surprisingly effective non-adaptive representation for objects with edges. In: Saint-Malo AC, Rabut C, Schumaker LL (eds) Curve and surface fitting. Vanderbilt University Press, Nashville
5. Karoui I, Chauris H, Garreau1 P, Craneguy P (2010) Multi-resolution eddy detection from ocean color and sea surface temperature images. In: OCEANS 2010 IEEE—Sydney, pp 1–6
6. Sumana I, Islam M, Zhang DS, Lu G (2008) Content based image retrieval using curvelet transform. In: IEEE international workshop on multimedia signal processing, Australia, pp 11–16
7. Aroussi ME, Ghouzali S, Hassouni ME, Rziza M, Aboutajdine D (2009) Block based curvelet feature extraction for face recognition. In: International conference on multimedia computing and systems (ICMCS), pp 299–303
8. Uslu E, Albayrak S (2014) Curvelet-based synthetic aperture radar image classification. Geosci Remote Sens Lett 11(6), 1071–1075
9. Cand'es EJ, Demanet L, Donoho DL, Ying L (2006) Fast discrete curvelet transforms. Multiscale Model Simulation 5: 861–899
10. Cortes C, Vapnik V (1995) Support-vector network. Mach Learn 20:273–297
11. Burges C (1998) A tutorial on support vector machines for pattern recognition. In: Fayyad U (ed) Proceedings of data mining and knowledge discovery, pp 1–43

12. Jaya Bharata Reddy M, Karthik Chandra B, Mohanta DK (2013) Condition monitoring of 11 kV distribution system insulators incorporating complex imagery using combined DOST-SVM approach. IEEE Trans Dielectr Electr Insul 20(2):664–674
13. Jaya Bharata Reddy M, Karthik Chandra B, Mohanta DK (2011) A DOST based approach for the condition monitoring of 11 kV distribution line insulators. IEEE Trans Dielectr Electr Insul 18(2):588–595

MRI Image Registration: Data-Driven Approach

C. Hemasundara Rao, P.V. Naganjaneyulu and K. Satyaprasad

Abstract Image registration is the challenging and important step to build computer-based diagnostic systems. One type of image modality is not able to provide all information for better diagnostic, and we combine information from multiple sources/image modalities. In this work, we proposed a canonical correlation analysis (CCA)-based image registration approach. CCA provides us framework to combine information from multiple sources. In this work, we use the information presents in both images for image registration task. We perform multimodal registration on T_1-weighted, T_2-weighted, and FLAIR MRI images. We use public data sets to evaluate our algorithm. Our algorithm performs better with mutual information–based image registration approach.

Keywords Image registration · CCA · MRI

1 Introduction

The goal of registration is to estimate the deformation between the images while taking the domain-specific information into consideration. A closer look at the problem statement intuitively reveals two methods of solving it. The first method operates directly on image intensity values, continuously transforming the entire

C. Hemasundara Rao (✉)
JNTUK, Guru Nanak Institutions Technical Campus, Ibrahimpatnam,
Telegana State, India
e-mail: c_hemasun65@yahoo.com

P.V. Naganjaneyulu
ECE, MVR College of Engineering and Technology,
Paritala, Krishna Dist, AP, India
e-mail: pvnaganjaneyulu@gmail.com

K. Satyaprasad
ECE, JNTUK, University College of Engineering, Kakinada, AP, India
e-mail: Prasad_kodati@yahoo.com

image so as to align it with the other. The image is considered to be registered when
desirable alignment is obtained for the respective transformation. These methods
are called area-based methods [1–3]. The second method relies on a few salient
points which are most prominent in both the images. The goal here is to detect the
corresponding pairs of points/regions across the images from which the deformation
is estimated. These are known as feature-based methods.

Feature-based methods have gained popularity over the area-based methods as
they are more robust to illumination changes, partial overlap between the images,
occlusion, changes in background, and viewpoint. Despite these advantages,
area-based methods are still preferred over feature-based methods in the medical
domain due to two main factors: (1) its ability to handle local deformations, which
is especially the case with human organs, and (2) dealing with information from
different imaging sources.

In general, the process of image registration involves finding the optimal geo-
metric transformation that maximizes the correspondences across the images. This
involves the following components (see Fig. 1).

A transformation model defines a geometric transformation between the images.
There are several classes of nonrigid transformations including parametric and
nonparametric models. Some of these models are well suited for small deforma-
tions, while others can represent large deformations.

A similarity metric measures the degree of alignment between the images. In
cases where features such as landmarks, edges, or surfaces are available, the dis-
tances between corresponding features can be used to measure the alignment. In
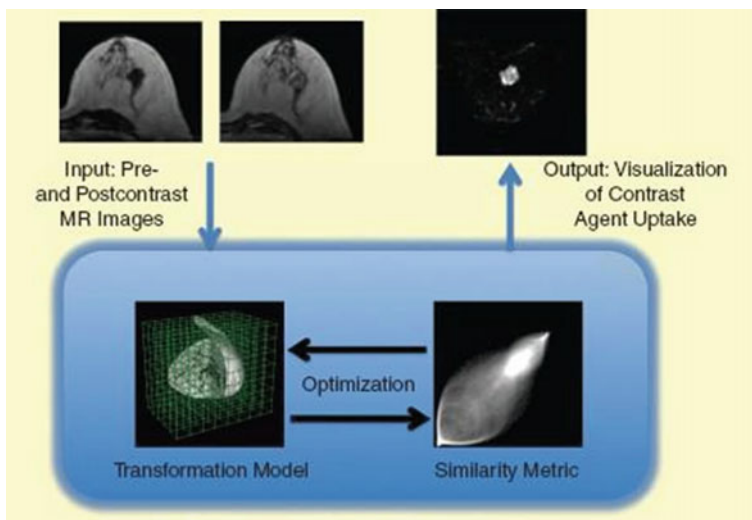other cases, the image intensities can be directly used to measure the alignment.

Fig. 1 Process of image registration

An optimization method maximizes the similarity metric. Nonrigid registration can be formulated as an optimization problem whose goal is to maximize an associated objective function. In addition, a careful validation must be performed to assess measures of performance, such as accuracy and robustness.

Image registration algorithm: An image registration algorithm consists of three parts

1. Transformation model
2. Similarity measure
3. Optimization process

An image registration algorithm defines objective function based on similarity measure and tries to maximize this objective function. In our method, we define a new registration method using canonical correlation analysis (CCA).

Canonical correlation analysis (CCA):

CCA finds the relationship between two multidimensional data sets [4]. The basic formulation of CCA is as follows: For a given two multidimensional data sets X and Y, find the two sets of basis vectors (W_x and W_y) such that the correlation between the projection of variables (X and Y) onto these basis vector is mutually maximum.

$$\rho = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}} = \frac{E\left[\hat{w}_x^T xy^T \hat{w}_y\right]}{\sqrt{E\left[\hat{w}_x^T xx^T \hat{w}_x\right]E\left[\hat{w}_x^T yy^T \hat{w}_y\right]}}$$

$$= \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}}.$$

C_{xx} and C_{yy} are the within-class covariance matrix and C_{xy} is the cross-covariance matrix. Maximum correlation is found as follows

$$\rho = \text{argmax } W_x^{\text{T}} C_{xy} Y^{\text{T}} W_y$$
$$\text{s.t } W_x^{\text{T}} C_{xx} W_x = 1 \quad \text{and} \quad W_y^{\text{T}} C_{yy} W_y = 1$$

This optimization problem is similar to generalized eigenvalue problem which has the standard solution.

Problems in CCA: Basic formulation of CCA has the following disadvantage

1. CCA finds only linear relationship between two data sets.
2. Difficult to extend more than two data sets.

These problems can addressed using the following ways.

1. Nonlinear relationship between the data sets can be addressed using kernel extension of CCA [5]. Kernel CCA defines the nonlinear mapping of two data sets $\varphi : x \rightarrow \varphi(x)$ and $\psi : y \rightarrow \psi(y)$ and performs the traditional CCA on transformed data sets.

2. Neural network-based CCA extracts the nonlinear relationship between data sets.
3. Locality preserving method-based CCA also extracts nonlinear relationship between data sets.

Canonical correlation analysis (CCA) is a method of correlating linear relationships between two multidimensional variables. CCA makes use of two views (or more than two) of the same semantic object to extract the representation of the semantics. CCA is used in image retrieval, image fusion [6], and object recognition problems in computer vision.

2   Proposed Methodology

Image registration methods are trying to find the relationship between two images in intensity domain or feature domain. This relationship is defined in terms of similarity measures. Similarity measures can be classified into two categories: (i) linear similarity measures such as sum of square difference (SSD) or sum of absolute difference (SAD) and (ii) nonlinear similarity measures such as mutual information or divergence measures. In multimodal image registration, the images are captured through different sensors (CT or MRI) or different parameters (T_1, T_2, or FLAIR) so that the intensity relationships between images are highly nonlinear.

In this work, we propose an algorithm which is based on structural representation of images. The input images are replaced by dense set of descriptors which performs the intensity-based registration. The advantage of this method is that after new representation we can use any simple similarity measure such as L_2 norm or SSD for multimodal image registration.

1. Given two images, find projection directions using kernel CCA (Gaussian kernel used for projection)
2. Project original images or features in lower dimension space using projection direction
3. Use L_2 norm as a similarity measure

Gradient descent used as optimization function.

3   Results

We perform two sets of experiments.

1. First set of experiments on T_1 and T_2 MRI images for 3D rigid registration (RIRE data set). Experiments start with 15-mm translation and 10° rotation as deviation from correct position. Run the algorithm 10× with different affine parameter settings. For comparison, we use the mutual information (MI)-based

method for rigid registration. We show the absolute error for translation (in mm), rotation (in degree), and root mean square error (RMS) in Table 1. Consider 1 mm equal to 1 degree for the absolute error computation.

We extend CCA for more than 2 modalities (T_1, T_2, and PD) also. Table 2 shows results for brain Web data set. Comparison purpose uses the MI based on pairwise registration framework.

CCA-based method performs better in terms of accuracy (Table 1) (in translation and rotation) compared to MI-based method. CCA-based method improves overall accuracy to 6.7% in pairwise registration and 13% in groupwise registration compared to MI-based method degree of freedom : 9

2. Second set of experiments on our collected data sets (DWI, FLAIR, T_1, and T_2 images). We divided data sets into 2 parts: (i) data sets which contain large tumor regions ($D_{1, 3}$ volumes) and (ii) data sets which does not contain tumor lesion or very less tumor lesions ($D_{2, 3}$ volumes) (data sets description is given in next section). For the first part of data sets, we use the DWI, T_1, and FLAIR images in groupwise registration model, and we compare our algorithm with mutual information-based method. Results are shown in given table.

For the error calculation, we marked 5 manual points on the MRI image.

In the second set of images, we performed the experiments on similar environment and same method is used for error calculation. Results are shown in following Table 4.

Table 1 Errors for translation (in mm) and rotation (in degree) in T_1 and T_2 MRI

Method	Translation X (mm)	Translation Y (mm)	Rotation (°)	RMS
MI-based method	3.1	2.0	4.2	3.0
CCA-based method	2.9	1.8	4.0	2.8

Table 2 Registration error (translation in mm and rotation in degree) in T_1, T_2, and PD

Data set	Method	RMS
T_1–T_2	MI	3.0
	CCA	2.6
T_1–PD	MI	2.7
	CCA	2.5
T_2–PD	MI	1.9
	CCA	1.5
T_1–T_2–PD	CCA	2.2

Table 3 Registration error in D_1

Method	Error (mm)
MI-based method	14.6
Our method	10.3

Table 4 Registration error in D_2

Method	Error (mm)
MI-based method	10.2
Our method	8.8

4   Conclusion

In this work, we proposed a new algorithm for image registration. This algorithm is based on canonical correlation. We know that in multimodal framework due to different acquisition parameters, the relation between data sets not follows the linear relationship. In this algorithm, we used the kernel version of canonical correlation analysis because the basic formulation of CCA gives only linear relationship between data sets. We show the results (Tables 1, 2, 3, and 4) on two different data sets (i) RIRE data sets (Table 2) and (ii) our collected data sets (Tables 3 and 4).

In the RIRE data sets (T_1, T_2, and PD images), we perform two sets of experiments: (i) pairwise registration and (ii) groupwise registration. From Tables 1 and 2, groupwise registrations perform good compared to pairwise registration because in groupwise registration, we have extra information (due to other modalities) which helps registration. The advantage of using CCA-based framework is that we can easily extend this framework for more than two modalities.

References

1. Pluim J, Maintz JBA, Viergever M (2003) Mutual information based registration of medical images: a survey. IEEE Trans Med Imaging 22(8):986–1004
2. Hill DLG, Batchelor PG, Holden M, Hawkes DJ (2001) Medical image registration. Phys Med Biol 46(3):1–45
3. Maintz JBA, Viergever MA (1998) A survey of medical image registration. Med Image Anal 2 (1):1–36
4. Sun T, Chen S (2007) Locality preserving CCA with applications to data visualization and pose estimation. Image Vis Comput 25:531–543
5. Akaho S (2001) A kernel method for canonical correlation analysis. IMPS2001-International Meeting of Psychometric Society, Osaka, 2001. http://www.neurosci.aist.go.jp/˜akaho/papers/IMPS2001full.pdf
6. Calhoun VD, Adalı T (2009) Feature-based fusion of medical imaging data. IEEE Trans Inf Technol Biomed 13(5)

Analysis of Video Content Through Object Search Using SVM Classifier

Azra Nasreen, H. Vinutha and G. Shobha

Abstract Nowadays the content-based retrieval plays a major role in the area of research in computer vision. Although image or video retrieval is a mature technology, not much work has been done in searching of an object in video sequences. The proposed work proposes a novel method which allows a user to make queries based on visual content properties such as color percentages, layout and texture occurring in frames by using instances of prior matches. Here the author proposes a method that searches representative frames of a digital video sequence containing the required object based on input query provided by the user. The performance measures like color, texture and shape are extracted from the frames of video as well as query image to identify only those relevant frames that are matching. Color correlogram, Gabor filter and morphological operations are used to extract color, texture and shape features, respectively. The proposed work shows a good accuracy of 90% to retrieve the related frames from the video.

Keywords Color correlogram · Gabor · Support vector machine · Object search

1 Introduction

Image retrieval is still one of the major vital research field in the multimedia technology [1]. Video and image retrieval has been a dynamic and testing research zone because of the huge growth of online video data, individual video recordings, computerized photographs and broadcast news videos. Keeping in mind the end

A. Nasreen · H. Vinutha (✉) · G. Shobha
RV College of Engineering, Bengaluru, Karnataka, India
e-mail: vinuthamadhusudhana@gmail.com

A. Nasreen
e-mail: azranasreen@rvce.edu.in

G. Shobha
e-mail: shobhag@rvce.edu.in

goal to viably oversee and use such enormous multimedia resources, end users must be able to use required content at the semantic level. Content-based image retrieval (CBIR) is used to search and retrieve the query image from data set. Various feature extraction methods and algorithms can be used to increase the accuracy of image retrieval [2]. CBIR makes use of visual data which search images from large video databases according to user's requirement, and it's been an energetic and advancing research field in the past couple of decades. However, there are many challenging research problems that proceed to attract researchers from a couple of disciplines. One of the challenges is to search an object from a video sequence. If an object is provided by the user as a query, then the system must be able to search instances of occurrences of that object in a video if at all it is there. An efficient system for searching and retrieving the object based on the content of the query is presented in this chapter. Proposed approach uses efficient techniques like color correlogram, Gabor feature and shape-based feature extraction to efficiently extract the content of interest.

An analysis on low-level visual properties is conducted in [3] that have to be extracted from the video frames. In [4], some of the key contributions in the current decade related to image retrieval and automated image annotation are discussed along with some of the key challenges involved in the adaptation of existing image retrieval techniques to build useful systems that can handle real-world data. System presented [5] is a content-based image retrieval based on texture, color, shape and region. Jun Huang et al. [6] have integrated visual saliency and consistency of image search results. Here they proposed a new re-ranking approach which integrates visual saliency and visual consistency mechanisms. A robust content-based video retrieval (CBVR) system is presented in [6] which retrieves similar video based on a local feature descriptor called SURF speeded-up robust feature (SURF).

Particle swarm optimization (PSO), an evolutionary population-based search algorithm, is employed in [7] to look for frames within the video library. In the system proposed in [8] uses frills-free method of video retrieval based on sample video input. Features like color, shape and texture have been considered for the retrieval of image. In their approach, frames are selected as multiples of a number and then the feature extraction takes place. The time involved is reduced to a great extent, and error rate has been reduced. Color co-occurrence matrix (CCM) is based on an efficient image retrieval algorithm which proposed in [9]. Hue saturation value (HSV) is used to find pixels of an image and then compare with the CCM image in database, and the images are extracted. Proposed system in [8] is an efficient system which use numerical representations called feature vectors to permit content-based looking in substantial image accumulations. Implementation based on color analysis of image is a standout among the most broadly utilized methods since the shading does not rely on image size or orientation. It uses auto color correlogram [10] for recognizing image taking into account the feature [12], and proposed an efficient work for content-based image recovery. This is refined by combining the two components, for example, color and surface [11]. Using so as to shade elements are separated HSV histogram, color correlogram and color moment values. Composition elements are removed by segmentation-based fractal texture

analysis (SFTA). The separated component vector of the inquiry picture is contrasted and removed element vectors of the database pictures to get the comparable pictures. The principle target of this work is arrangement of picture utilizing SVM calculation [13–16] also proposed robust methods on content-based image retrieval. In this proposed work, so for this entire literature surveys highlight in retrieve the related images from videos using one or two performance measures. Here, the author proposes a method in which video is sectioned into number of frames and from these, the related frames are retrieved using SVM classifier for better accuracy. The performance measures like shape, texture and color are taken into consideration to improve extracting content of interest from videos for analysis.

The chapter proceeds as follows: In Sect. 2 discuss Methodology, Sect. 3 discuss the experiment results and Sect. 4 highlights the conclusions and future work.

2 Methodology

In this proposed work, it splits into three sections, (1) preprocessing, (2) feature Extraction, (3) support vector machine (SVM).

The proposed work is as shown in the Fig. 1. The methodology consists of two phases called testing phase and training phase. In the testing phase, the input frame is passed to the preprocessing block. In this step, preprocessing steps like image resizing, color conversion and binarization are carried out.

This step is followed by the feature extraction step. Here features like texture, shape and color features are retrieved from the test image. To get the texture and color, feature methods like Gabor feature extraction and correlogram are followed. Extracted feature is then passed to SVM block for classification.

In the training phase, the video input is passed to frame generation block for frame generation. Every time the frames are picked and passed to feature extraction block. In the feature extraction block, same steps as in testing phase like texture, color and shape-based feature extraction are done. The extracted features from the entire frame are stored in the knowledge base. The test feature is then compared with the features stored in this knowledge base using SVM classifier to retrieve the frame with correct match. All the retrieved frames matching with testing image are then stored in an empty folder.

2.1 Preprocessing

In this step, preprocessing steps like image resizing, color conversion and binarization are carried out. The input for this module is in RGB frame. Then it is converted into resized image, and resized image is converted into grayscale conversion for texture feature extraction using Gabor wavelet algorithm. The resized image is converted into binary format to do morphological operation for shape
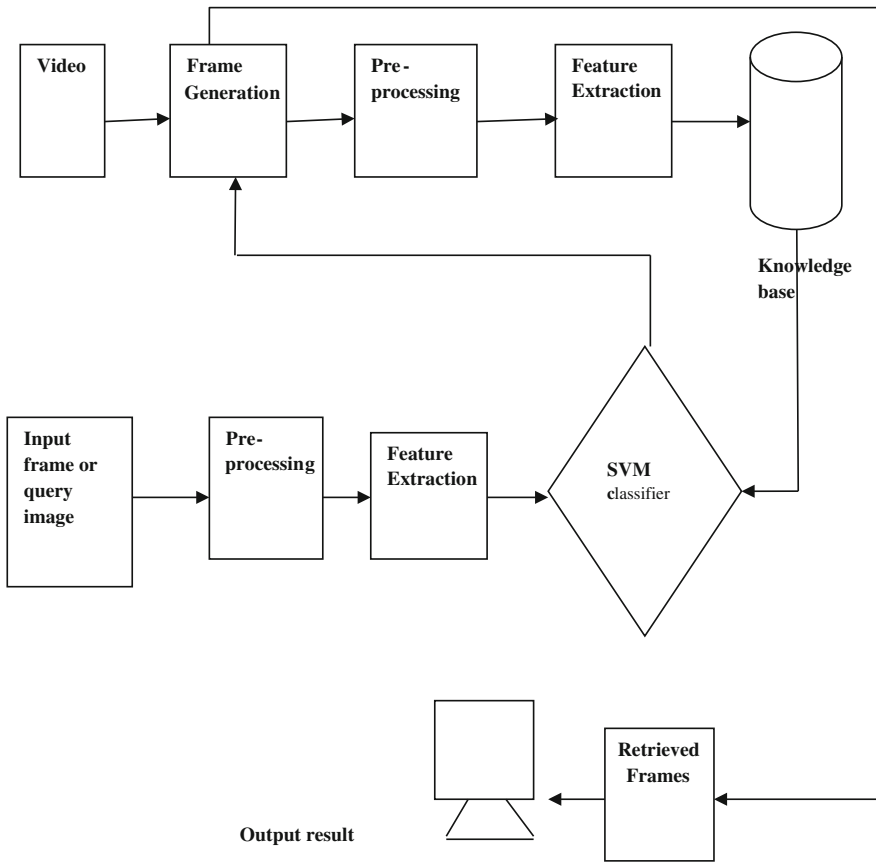
Fig. 1 Block diagram for the proposed work

feature extraction. The same RGB image is send for color feature extraction using color correlogram.

2.2 Feature Extraction

In this step, features like texture, shape and color features are extracted from the test image.

2.2.1 Gabor Filter

The Gabor filter is utilized as a part of the image features. Gaussian coefficient is used to adjust sine wave which is a type of Gabor filter wavelet. The local and

global data are retrieved with the help of Gabor filter. The Gabor filter is tunable band-pass channel, multi-scale and multi-resolution filter [17]. The Gabor filter Eq. (1) utilizes texture segmentation image which prefers ideal resolution in time and space domain, and it gives improved visual representation involved in composition pictures. The current Gabor parameter requires additional time utilization for feature extraction. The Gabor filter works on the frequency, orientation and Gaussian kernel.

$$Gabor\,(x, y, \theta, \varphi) = X.Y \tag{1}$$

$$X = \exp\big(-\big(x^2 + y^2\big) \div \sigma^2\big) \tag{2}$$

$$Y = \exp(2\pi\theta(x\cos\theta + y\sin\theta)) \tag{3}$$

The terms x and y (Eqs. 2 and 3) is the position of the filter relative to the input signal [17]. The angular representation of the filter is represented as 'θ'. The angular orientation of the filter is represented as 'φ'.

2.2.2 Shape Feature

Shape is known as a vital sign for people to distinguish and perceive object like geometrical structures. Example: Radius, Perimeter. Shape feature extraction techniques can be comprehensively characterized into two gatherings viz contour-based and area-based methods. Set of shape properties for each connected component in the binary image is extracted and stored. Properties like area, Euler number, orientation, bounding, box, perimeter, centroid, extrema and other properties are extracted.

2.2.3 Color Features

A color correlogram specifies a spatial correlation which changes with distance aspect between pair of colors. A color distribution will be captured by the color histogram [18]. In an color correlogram image table, there are n entries for row (i, j) which gives the probability for finding the pixel of color j for distance n from pixel color i. This d is chosen from a set of distance values D. Spatial correlation between same colors is captured by an autocorrelogram. This is a subset of the correlogram, and it consists of rows (i, j). A small value of d is sufficient to capture the correlation, because local correlations between colors are more significant than global correlations in an image. Efficient algorithm is available to compute the correlogram when d is small.

2.3 SVM Classifier

SVM is appropriate to distinguish the color, texture and shape effectively [19]. It is a supervised learning process that breaks down and perceives examples like texture, color and shape, and it is used for classification and regression methods. Hyperplane is used for grouping, which has outsized separation to nearest separation training data [20]. In this, perception is bigger than the margin which brings down the generalization of error classifier. The arrangement isolates information into preparing training tests. The objective estimations of test information by giving test information qualities are the goal of SVM.

3 Results

The results obtained at each stage are discussed in this section. The proposed algorithm is evaluated on video sequences, and matched output for one input frame is shown here. Figure 2 (a) depicts the input frame chosen in the set of frames generated from the video sequence. This input frame then undergoes pre-processing and feature extraction steps. These extracted features from the test image are then compared with the trained features already stored in the knowledge base. Matching is done by SVM classifier to give classified result; (b) depicts the set of matched frames for the input frame chosen. The frames with maximum matches are displayed. It is seen that the proposed system gives good result compared to existing method.

Table 1 is the performance analysis of the proposed work taking ground truth value as 20. Ground truth can be defined as the information or data provided by direct observation as opposed to information provided by inference. The accuracy is calculated by the Eq. (4), recall is calculated using Eq. (5) and precision using Eq. (6).

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100 \qquad (4)$$

$$Recall = \frac{TP}{TP + FN} \qquad (5)$$

$$Precision = \frac{TP}{TP + FP} \qquad (6)$$

The computed accuracy for frame retrieval for three different videos is shown in the Table 1.

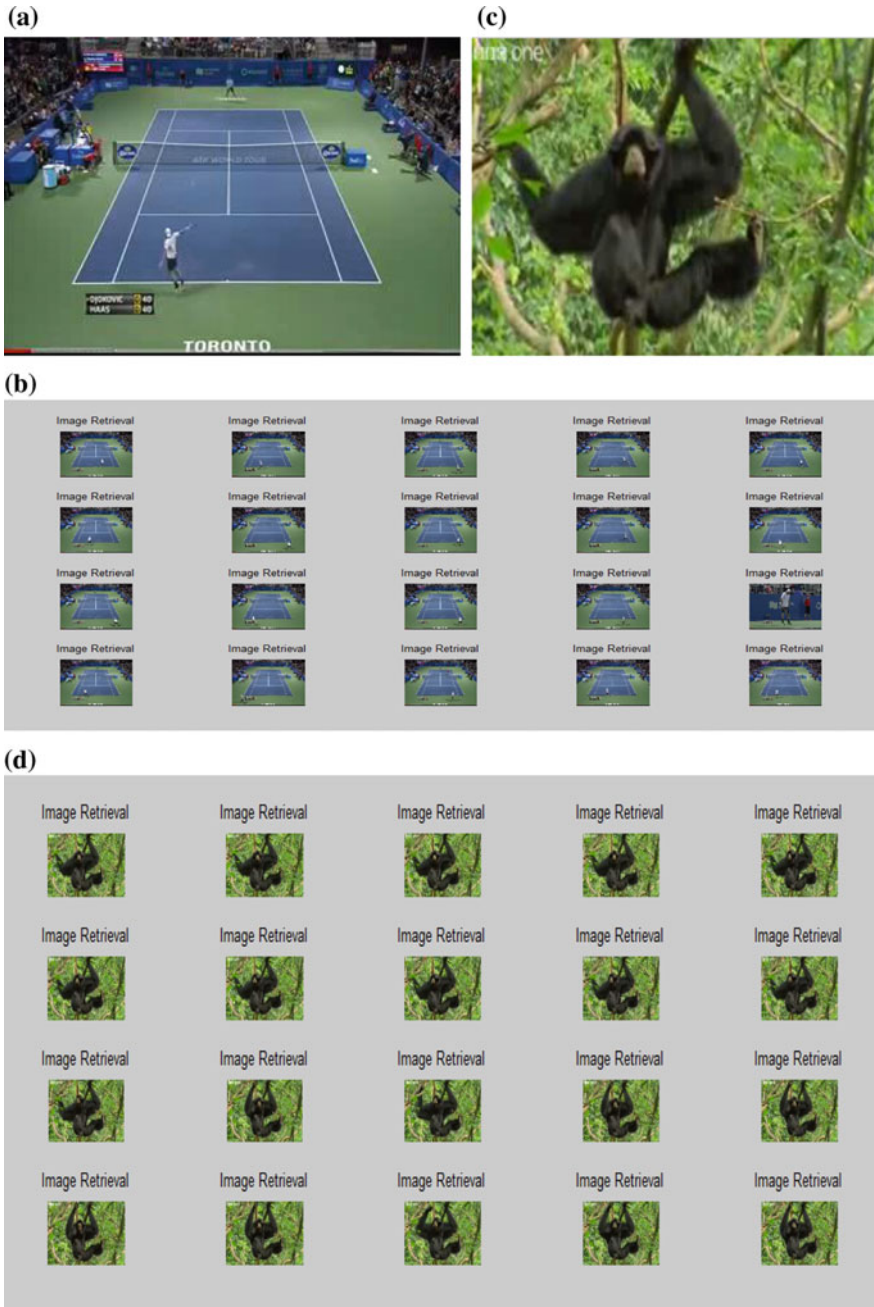**Fig. 2** **a** Input frame, **b** Retrieved matched frames, **c** Input frame, **d** Retrieved matched frames

Table 1 Performance evaluation of the proposed work

	Accuracy	Recall	Precision
RUN 1 Sports video	88.86	0.91	0.94
RUN 2 Jungle video	95.32	0.95	0.97
RUN 3 ATM video	86.36	0.92	0.94

4 Conclusion

This chapter discusses about the searching of instances of particular objects in a video where in the representative frames containing the queried object are picked. Since multiple features like color, texture and shape are extracted from the queried object as well as the video in which search has to be carried out, the accuracy is found to be promising. The system could give accuracy of 90%. The proposed system is a powerful tool for video analysis and saves a lot of time by extracting only content of interest from a long video sequence and is useful in wide variety of video analytic applications such as ATM theft, intrusion detection and in any other form of video surveillance applications.

References

1. Patel BV, Meshram BB (2012) Content based retrieval systems. Int J UbiComp 4(5):737–740
2. Arthi K, Vijayaraghavan J (2013) Content based image retrieval algorithm using color models. Int J Adv Res Comput Commun Eng 2(3):1343–1347
3. Patel BV, Meshram BB (2012) Content based video retrieval systems. Int J UbiComp 3 (2):13–30
4. Datta R, Li J, Wang JZ (2005) Content-based image retrieval: approaches and trends of the new age, 1–10. MIR'05, Singapore
5. Shirazi SH, Noor ul AK (2016) Content-based image retrieval using texture color shape and region. Int J Adv Comput Sci Appl 7(1):418–426
6. Asha S, Sreeraj M (2013) Content based video retrieval using surf descriptor. 3rd international conference on advances in computing and communications, IEEE, 212–215
7. Salahuddin A, Naqvi A, Mujtaba K, Akhtar J (2012) Content based video retrieval using particle swarm optimization. 10th international conference on frontiers of information technology, IEEE, 79–83
8. Balakrishnan S, Thakre KS (2010) Video match analysis: a comprehensive content based video retrieval system. ISSN:0974-0767
9. Arthi K, Vijayaraghavan J (2013) Content based image retrieval algorithm using colour models. Int J Adv Res Comput and Comm Eng 2(3):1343–1347
10. Seetharaman K, Sathiamoorthy S (2013) An improved edge direction histogram and edge orientation auto corrlogram for an efficient color image retrieval. International conference on advanced computing and communication systems, IEEE Coimbatore, INDIA, 1–4

11. Asha S, Sreeraj M (2013) Content based video retrieval using SURF descriptor. In: Third international conference on advances in computing and communications, IEEE, 212–215. doi:10.1109/ICACC.2013.492013
12. Usha R, Perumal K (2014) Content based image retrieval using combined features of color and texture features with SVM classification. Int J Comput Sci Commun Netw 4(5):169–174
13. Anand A, Mala K, Suganya S (2016) Content-based image retrieval system based on semantic information using color, texture and shape features. International science conference on computing technologies and intelligent data engineering, IEEE, 1–8
14. Gill HK, Kaur K (2016) Comparitive study of image features, color models and classifiers for image retrieval. Int J Adv Res Comput Sci Softw Eng 6(6):798–801
15. Gandhani S, Singhal N (2015) Content based image retrieval survey and comparison of CBIR system based on combined features. Int J Signal Process Image Process Pattern Recognit 8 (10):155–162
16. Agarwal S, Verma AK, Dixit N (2015) Content based image retrieval using color edge detection and discrete wavelet transform. International conference on issues and challenges in intelligent computing techniques, IEEE, 368–372
17. ping Tain D (2013) A review on image feature extraction and representation techniques. Int J Multimedia Ubiquitous Eng 8(4):385–396
18. An P, Ajitha T, Priyadharshini M, Vaishali MG (2014) Content based image retrieval (CBIR), using multiple features for textile images by using svm classifier. Int J Comput Sci Inf Technol 2(2):33–42
19. Thiyaneswaran B, Padma S (1989) Analysis of Gabor filter parameter for iris feature extraction. Int J Adv Comput Technol 3(5):45–48
20. Zhao W-L, Tan S, Ngo C-W, Largescale near-duplicate web video search: challenge and opportunity. In: Grants Council of the Hong Kong Special Administrative Region, China (City U 119508)

Noise Destruction and Improvement of Speech Signal Quality Using Group Search Optimization (GSO) Algorithm

V. Anoop and P.V. Rao

Abstract It is often possible that the communication quality of the signal significantly depreciates from existence on account of the noise interruption. In the innovative signal enhancement procedure, the preferred signal quality is considerably increased by initiating a sequence of procedures, including the AMS feature extraction, signal enhancement, GSO optimization, followed eventually by the appraisal of its performance. At first, the input signal troubled with the noise is obtained for which the AMS features are effectively extracted. At the signal enhancement phase, a new signal is achieved as the product of the feature extracted and the original signal. Thereafter, the signal achieved after the enhancement procedure is once again optimized by passing through the GSO optimization algorithm. In the long run, the SNR values are estimated for the entire signal outputs to appraise its performance.

Keywords Speech signal · Group search optimization · SNR
AMS feature · MSE

1 Introduction

Speech enhancement technique refers to the reversal of degradations that have corrupted the speech signal, thereby increasing the quality or intelligibility of the particular speech signal. Speech quality is often measured by considering different aspects, and it is regarded as a subjective measure while speech intelligibility is an objective measure it points to the amount of speech data which are accurately interpreted. Clean signal with best quality can be produced by introducing suitable

V. Anoop (✉)
Department of ECE, VTU Belgaum, Belagavi, Karnataka, India
e-mail: vanoop@gmail.com

P.V. Rao
Rajarajeswari College of Engineering, Bengaluru, India
e-mail: pachararao@rediffmail.com

speech enhancement techniques. This is the prime objective of speech enhancement technique, which is widely used in telephonic applications [1].

Speech enhancement techniques can be classified into single-channel, dual-channel or multi-channel enhancement. Although the performance of multi-channel speech enhancement is better than that of single-channel enhancement, the single-channel speech enhancement is still a significant field of research interest because of its simple implementation and ease of computation [2–4]. Digital signal processing (DSP) techniques for speech enhancement include spectral subtraction, adaptive filtering and suppression of non-harmonic frequencies. Most of these techniques either require a second microphone to provide the noise reference or require the characteristics of noise be relatively stationary [5–7].

Spectral subtraction has been widely used for enhancing speech because of its simplicity and ease of implementation in single-channel systems, but it suffers from the production of musical noise after enhancement and is one of its major drawbacks [9]. Although the MMSE–STSA method gives an estimated speech signal with less musical noise, it requires more complicated computations, for example, the solution required to calculate the modified Bessel function. Moreover, as pointed out by some researchers, real speech histograms do not fit into Rayleigh function employed [8–11].

2 Proposed Methodology

The innovative technique is intended basically for the purpose of augmenting the speech signals. The novel method encompasses the procedure of optimization of the signal input by passing through the group search optimization (GSO) algorithm subsequent to taking the amplitude magnitude spectrum (AMS) features. At first, the preferred signal mixed with noisy signals are taken and extracted with the amplitude magnitude spectrum features.

2.1 Speech Signal Synthesis

The signal input is obtained for which the AMS features are effectively extracted. Let $q(m)$ represents the original signal defined for all m. The original preferred signal habitually consists of certain noise mixed with it. It is highly essential to denoise the signal with a view to achieve the appropriate and accurate communicated signal. Let the noise supplemented with the original signal be represented by the Eqs. 1 and 2.

$$\text{input signal} = \text{desired signal} + \text{noise} \tag{1}$$

$$q(m) = r_m + \sigma n(m) \tag{2}$$

where, $n(m)$—noisy signal.

For the white noise, the noise-contaminated signal $n(m)$ is generally dispersed between 0 and 1.

2.2 AMS Feature Extraction

In the AMS feature extraction phase, the signal input is initially obtained for which the level is adapted, and along with this task the procedure of sampling is carried out simultaneously, where the signal is resized to certain preferred size. Thereafter, the signal has to be segmented into number of frames, known as the signal fragmentation. This is followed by the process of the fast Fourier transformation for the frame segregated signal.

2.3 Signal Enhancement

In the signal enhancement phase, the feature extracted output is multiplied with the original signal, and the product is effectively employed to locate the distance. Let the signal achieved after extracting AMS features be represented by W_i, which has to be multiplied with the noise-tainted signal to achieve a new solution Z_i as illustrated in Eq. 3.

$$Z_i = q(m) \times W_i \tag{3}$$

where, $q(m)$—represents the input signal, W_i—reveals the feature extracted output.

For the purpose of denoising the signal and to improve the preferred signal, the AMS feature extracted signal has to be GSO optimized. In the GSO optimization, the feature extracted output W_i is generated as the initial solution. On additional processing, the original signal is substituted by the product achieved from above equation in the case of minimum fitness value. The minimum fitness value here represents the minimum Euclidean distance between the two solutions.

2.4 Group Search Optimization (GSO)

The innovative method of the group search optimization (GSO) is elegantly employed to improve the speech signal after the extraction of AMS features. The

primary set of solution to the algorithm is labelled as the group, and the members of the group are divided into three distinct types shown below.

Producer

They constitute the members of group which go hunting for the best resources.

Scrounger

These members effectively locate the relation between the resources, which are found out by the producers.

Ranger

These members create the activities in an immaterial way together with the searching in an organized manner, in order that the effective discovery of the resources is duly achieved.

Let us take the initial solution as the final AMS feature extracted signal output. The initial solution to the optimization procedure is effectively in Eq. 4.

$$W_i = \begin{bmatrix} w_{11} & w_{12}.....w_{1m} \\ w_{21} & w_{22}.....w_{2m} \\ w_{n1} & w_{n2}.....w_{nm} \end{bmatrix} \tag{4}$$

2.4.1 Search Direction Vector

The search direction of the kth member at searching space h in the k-dimensional search space is illustrated by means of Eq. 5.

$$V_k^h\left(\xi_k^h\right) = (v_{k1}^h, v_{k2}^h, \ldots, v_{kl}^h) \in R^l \tag{5}$$

The search direction vector may be effectively evaluated with the help of the following Eqs. (6), (7) and (8) by means of the polar and Cartesian coordinate transformations.

$$v_{k1}^h = \prod_{p=1}^{l-1} \cos(\xi_{kp}^r) \tag{6}$$

$$v_{kr}^h = \sin(\xi_{k(r-1)}^h) \prod_{p=r}^{l-1} \cos(\xi_{kp}^h) \text{ for } (r = 2, 3, \ldots, l-1) \tag{7}$$

$$v_{kl}^h = \sin(\xi_{k(l-1)}^h) \tag{8}$$

2.4.2 Head Angle

The head angle is effectively employed in the creation of the best resources, as the producer rotates its head for the purpose of searching for the best resource. The angle to which the producer has to rotate is illustrated as the head angle.

The head angle illustration is furnished as per Eq. 9.

$$\xi_k^h = (\xi_{k1}^h, \xi_{k2}^h, \ldots, \xi_{k(l-1)}^h) \tag{9}$$

2.4.3 Fitness Evaluation

The fitness function is defined as the minimum of Euclidean distance and is illustrated by means of Eqs. 10 and 11.

$$F_f = \min(E_d) \tag{10}$$

$$E_d = \sqrt{\sum_{i=m=1}^{M} (Z_i - r_m)^2} \tag{11}$$

where, Z_i—denotes the new signal created, r_m—corresponds to the preferred signal.

The best resource is created by the producer by arbitrarily producing the rotating angle, which is illustrated as per Eq. 12

$$W^{h+1} = \xi^h + s_2 \gamma_{\max} \tag{12}$$

In Eq. 12, $\gamma_{\max} \in R^1$ and $\gamma_{\max}$ represents the maximum turning angle.

For the minimum turning angle (which usually happens in case of not achieving the best resource), $\gamma_{\max} = 0$ is replaced in the Eq. 12 and is illustrated as per Eq. 13

$$W^{h+x} = \xi^h \tag{13}$$

where, $x \in R^1$ represents a constant.

The relationship between the resources is evaluated by the scrounger as expressed in the following Eq. 14.

$$Wh + 1 = W_k^h + s_3 \circ (W_{pr}^h - W_k^h) \tag{14}$$

where, $s_1 \in R^l$ characterizes the uniform random sequence in the range [0, 1] and "∘"—operator indicating the Hadamard product.

At last, the residual group members representing the rangers are separated from their present locations rooted on the distance measure produced by means of the arbitrarily created head angle. The distance measure is effectively illustrated as per the following Eq. 15.

$$t_s = x \cdot s_1 \cdot t_{\max} \tag{15}$$

The rangers are capable of locating the resources with the help of Eq. 16 shown below.

$$W_k^{h+1} = W_k^h + t_k \, V_k^h \left(\xi^{h+1} \right) \tag{16}$$

In the long run, the procedure of appraisal is initiated, in which the fitness function is once again attained and the solution with superior fitness function gets substituted by the latest solution. The task of ascertaining the best resource goes on till the best one is achieved as the most appropriate resource.

In the ultimate phase, the original signal gets significantly improved, which is achieved by means of the GSO optimization, where the input solution with minimum Euclidean distance is substituted by the preferred signal, which paves the way for the incredible improvement in the attained output.

2.5 Signal Performance Analysis

The improved signal is evaluated by computing the signal-to-noise ratio and minimum mean square values of both the ultimately improved signal and the preferred signal.
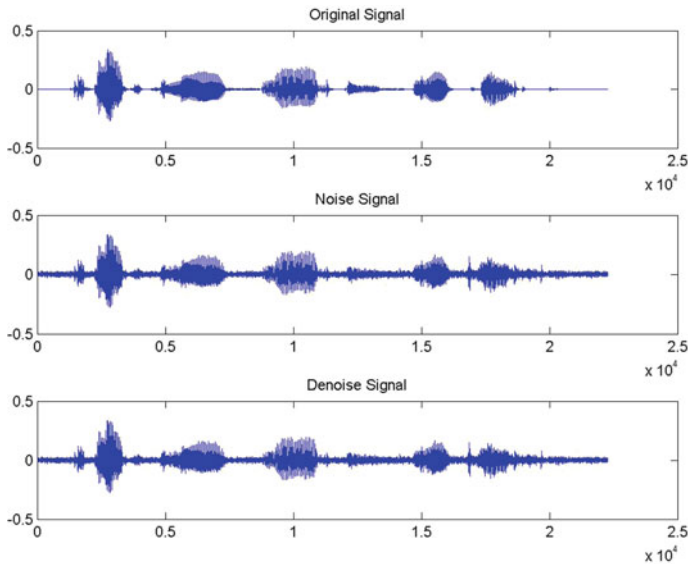
Fig. 1 Input signal, noise signal and denoised signal for 15 db babble noise

3   Result and Discussion

This section puts in a nutshell the upshots realized together with the speech signal enhancement through group search optimization (GSO) algorithm. The experimental result is shown in Fig. 1. The database has been extensively employed for acquiring the productivity from times immemorial (Fig. 2; Table 1).
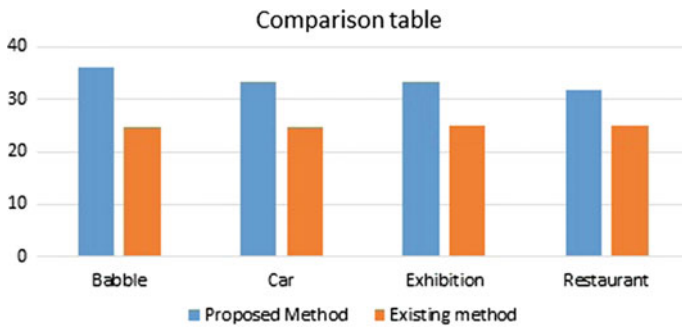

Fig. 2 Comparison of SNR value 15 dB for proposed and existing method

Table 1 SNR and MMSE values for the proposed and existing methods

Amount of noise added (db)	SNR value		MMSE value	
	Proposed method using Group Search Optimization (GSO) algorithm	Existing method using Cuckoo search (CS) algorithm	Proposed method using Group Search Optimization (GSO) algorithm	Existing method using Cuckoo search (CS) algorithm
Babble noise				
5	5.1212383	10.5691	1.15E−05	1.17E−04
10	18.04315683	17.7956	1.62E−05	7.60E−05
15	33.92592455	24.7640	6.49E−06	7.62E−05
Car noise				
5	5.098623322	10.7005	1.16E−05	1.37E−04
10	18.0891978	17.9095	6.46E−06	7.07E−05
15	32.90188183	24.7302	2.88E−05	7.00E−05
Exhibition noise				
5	5.109179851	10.5710	1.80E−05	7.11E−05
10	18.05399375	17.8248	7.22E−05	9.34E−05
15	32.02323124	24.9196	6.13E−06	7.07E−05
Restaurant noise				
5	5.002109971	10.7593	1.16E−05	6.86E−05
10	17.97364836	17.7414	6.46E−06	7.49E−05
15	25.01225084	24.9270	2.88E−05	7.89E−05
10	17.91654387	17.6489	1.43E−05	8.99E−05
15	31.06247205	24.5211	7.69E−06	7.95E−05

4 Conclusion

The denoising related with the communicated signal invariably encompasses the communication to be evidently illustrate and filtered, by eliminating the entire noise and artefacts from the signal, as the entire signal outputs achieved are impacted by the disturbance. The real disturbance decreases the overall precision together with the exactness of appraisal and leads to the reduction in the identification restriction on the strategy. In the innovative technique, the signals achieved and extracted with the AMS features are augmented to perk up its excellence. The signal enhancement is made further suitable with the optimization procedure of the GSO technique. In the long run, the SNR values are effectively estimated for the entire signal outputs and are assessed and contrasted with the original signal to appraise the accomplishment of the signals. The consequential outputs for both the novel and modern techniques are achieved and evaluated to illustrate the effectiveness of the novel method. It produces an average improvement of 8–12% under various noise levels.

References

1. Falk TH, Zheng C, Chan W-Y (2010) A non-intrusive quality and intelligibility measure of reverberant and de-reverberated speech. IEEE Trans Audio Speech Lang Process 18(7): 1232–1237
2. Huang Z, Dong M, Mao Q, Zhan Y (2014) Speech emotion recognition using CNN. IEEE Trans Audio Speech Lang Process pp 23–29
3. Laska BNM, Bolic M, Goubran RA (2010) Particle filter enhancement of speech spectral amplitudes. IEEE Trans Audio Speech Lang Process 18(8):2155–2167
4. Wolfe PJ, Godsill SJ (2003) Efficient alternatives to the Ephraim and Malah suppression rule for audio signal enhancement. EURASIP J Appl Signal Process 2003(10):1043–1051
5. Mao Q, Dong M, Huang Z, Zhan Y (2014) Learning salient features for speech emotion recognition using convolutional neural networks. IEEE Trans Multimed 16(8):73–79
6. Anoop V, Rao PV (2015) Performance analysis of speech enhancement methods using adaptive algorithms and optimization techniques. In: IEEE Digital Library—International Conference on Communication and Signal Processing, April 2–4, 2015, India, ISBN:978-1-4799-8080-2, pp 1322–1326
7. Anoop V, Rao PV (2013) Speech signal quality improvement using Cuckoo search algorithm. Int J Eng Innov Res 2(6). ISSN:2277-5668, 2014
8. Hu Y, Loizou P (2006) Subjective comparison of speech enhancement algorithms. In: Proceedings of IEEE international conference on acoustics, speech and signal processing, vol 1, pp 153–156
9. El Ayadia M, Kamel MS, Karray F (2011) Survey on speech emotion recognition: features, classification schemes, and databases. Pattern Recognit 44:572–587
10. Soon IY, Koh SN, Yeo CK (1998) Noisy speech enhancement using discrete cosine transform. Speech Commun 24(3):249–257
11. Ephraim Y, Malah D (1984) Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Trans Acoust Speech Signal Process ASSP-32 (6):1109–1121
12. Mao X, Chen L, Fu L (2009) Multi-level speech emotion recognition based on HMM and ANN. In: Proceedings of world congress on computer science and information engineering pp 2178–2184

Part III
Communication

SDA Algorithm for Network Path Identification

K. Arun and R. Rejimoan

Abstract Bio-inspired algorithms are applied to computer domain for solving optimization problems. Routing path identification is still a major issue in computer networks. This paper proposes a method inspired from smell detection behaviour of dogs that helps to find optimized path in computer networks. The proposed method reduces the time complexity for finding optimized path in the networks compared to other bio-inspired algorithms.

Keywords SDA · Path · Routing network · Bio-inspired algorithm

1 Introduction

Nature inspires efficient way to solve optimization problems. At present, algorithms are developed from nature behaviour called nature-inspired algorithms [1]. Nature-inspired algorithms are based on the successful characteristics of biological system. Thus, nature-inspired algorithms derived on the basis of our biological system can be termed as bio-inspired algorithms. Bio-inspired algorithms are developed on the basis of behaviour of animals. The best examples for bio-inspired algorithms are Ant Colony Optimization (ACO), Particle Swarm Optimization (PSO), and Artificial Bee Colony Optimization (ABC) [2, 3]. Nature-inspired algorithms along with behaviour of animals help in solving computational optimization problems [4].

Properties of animal behaviour can be imitated to develop new bio-inspired algorithms. Dogs are one of the earliest animals helpful in solving problems such as identification of crimes and hunting down their prey. That is dog locates its prey by

K. Arun (✉) · R. Rejimoan
Department of Computer Science and Engineering, SCT College of Engineering,
Trivandrum, India
e-mail: arun2u@live.com

R. Rejimoan
e-mail: rejimoan@gmail.com

his superior sense of smell. An algorithm was developed that used this dog's behaviour as smell detection agent (SDA)—based algorithm [5].

Each dog has different capacities of finding paths. Here, dogs are referred as intelligent agent. Each agent finds a path from source to destination node based on the smell spot value,. i.e. the agent visits the node with highest smell value such that other agents are not visited the node. This goes on until the destination is reached. Nodes can be either switches or routers in the network.

2   Related Works

Ant Colony Optimization (ACO) was developed by Dorigo et al. during the year 1990 [6]. ACO helps in solving many computing problems like finding optimal path in a graph by the behaviour of ants to find the path between food source and the nest. Omicron Ant Colony Optimization (OA) developed on basis of MAX–MIN ant systems where pheromone levels are within a specific range. OA is used to find paths in dynamic networks with slight variations in ant net algorithm [7]. In this approach, the pheromone value is set to one initially and a solution table at each node contains good solutions to reach the destination. Solution set contains solution matrix to reach all possible destination nodes, and vector solution matrix contains the next node and also time taken to reach the destination. The solution table is updated whenever a new solution is found which ensures dynamic changes in the network. Dual method helps to enhance the performance and route with a random value ranging between 0 and 1. For selecting next best node, value less than 0.5 uses roulette method and above 0.5 uses deterministic method. Thus, algorithm provides a set of better solution to reach destination.

Yanfang Deng and Hengqing Tong have developed a dynamic shortest path algorithm using PSO on fluid neural network [8]. PSO is initialized as a group containing random number of particles. During each iteration, the two values are found, one for best solution and other by particle swarm optimizer. At last, the velocity of each particle is updated. In this method, PSO uses indirect scheme for finding shortest path by not failing in local minimum, which uses energy function to solve problems. Here, the path is selected based on the nodes with highest priority and direct links. A valid path is returned to the terminal node as the destination.

Artificial Bee Colony Optimization (ABC) was introduced by Dervis Karabonga during the year 2005 based on the inspiration of bees on finding their food. Deeban et al. proposed ABC algorithm with Elliptic curve Diffie–Hellman (ECDH) for

finding optimal path and secure data transfer in networks [9]. In this approach, ABC algorithm is more flexible to find shortest path and selection of path is based on greedy process. ECDH algorithm helps in secure and more reliable data transfer.

3 Data Structure for SDA

3.1 Data Structure for Node

In this study, a node is a router or a switch in the network topology. A node has its own identity which is either IP address or MAC address, and its location coordinates denote smell spot value. The data structure for a node is represented as follows

$$D_{\text{Node}} = (\text{nId}, s, V, x, y)$$

where 'nId' denotes the identity of the node which can be a real value or integer, and 's' denotes the smell value of each node which is calculated using the location coordinates 'x' and 'y'. Smell value is a floating point, and 'V' is a Boolean value that gives visiting status of the agent.

3.2 Data Structure of an Agent

Agent is used to identify a path in the network using a set of rules. It finds path by using the smell value of each node in the network.

$$D_{\text{Agent}} = (\text{ID}, \text{current}, \text{Total}, A, \text{Path})$$

where, 'ID' represents the identity of the agent is a string or an integer value, 'current' denotes the current positing of the agent in the network, 'Total' represents the total number of nodes visited by an agent, 'A' is a flag variable which informs the agent to 'Search/Stop' of traversal, and 'Path' represents the path found by the agent.

4   SDA Algorithm for Optimized Path

Smell detection agent (SDA)-based algorithm is a nature-inspired technique, designed to solve optimization problems [5]. SDA uses intelligent agent to find the shortest path from a source to the destination. Each agent has different capacities for finding paths. Thus, for 'N' number of agents, we get 'N' number of paths, and it is required to identify the shortest path. For path identification, two parameters are used, total number of agents and the smell spot value which vary according to the number of nodes in the network.

4.1   SDA Algorithm for Path Identification

Agent is used to identify a path in the network using SDA algorithm that uses linear search to find a path in the network. The agent achieves to find a path through set of smell spot values from a source node to the destination node. If there are 'n' agents, there are 'n' paths returned. From these 'n' paths, shortest path is selected, which reduces solving complex problems.

Shortest path SDA algorithm is shown in Algorithm 1. In this algorithm, first the total number of nodes is retrieved. The initial smell value of each node is updated by the node location coordinates. Next, source and destination nodes are identified to find the best path. With the destination node, smell value of each nodes is calculated by initial smell value, decrement constant which is reciprocal of total nodes and the effective distance. Updating smell value by each total number of SDAs is initialized with their ID, current node, and length. Each agent finds the path based on the smell value of each node. The path is identified based on the highest smell value of each node from the current node. When a highest smell node is found from the current node and the highest smell node is assigned as current node, this will loop until the destination is reached. When the destination is reached, the SDA is assigned with flag 'Stop' which denotes reached the destination. The shortest path is identified from agents who have the less number of nodes traversed. In some cases, more than one SDA returns shortest path and the user can select which path is to be used.

```
Algorithm 1 Smell Detection Agent Algorithm
  N: Number of Nodes in the network
  M: Number of SDA's
  S: ID the source node
  D: ID the destination node
1. Get the total number of Nodes
2. For each node N
   2.1.    D_Node.nId=i
   2.2.    Get the  D_Node.x and D_Node.y
   2.3.    Initialize the smell value of
           D_Node.s=1/sqrt(x²+y²)
   2.4.    D_Node.V= False
3. Get Total Number of agents
4. For i=1 to M
   4.1.    D_Agent.ID = i
   4.2.    D_Agent.current= S
   4.3.    D_Agent.Total=0
   4.4.    D_Agent.A= Search
   4.5.    D_Agent.Path= null
5. Get the source node and destination node
6. For i=1 to N
   6.1.    If current node is less than destination node
           then D_Node.s=1/(a+b*p)
   6.2.    Else if current node is greater than
           destination node then D_Node.s=1/(a-b*p)
   6.3.    Else
      6.3.1. if current is less than destination node
             then D_Node.s=1/(a+b)
      6.3.2. else D_Node.s=1/(a-b)
  where a is the initial smell value of the node, b is
the decrement constant and p denotes the effective
distance
7. For each agent in M
   7.1.       Loop until destination is reached
    7.1.1.   Get the current node of the agent
    7.1.2.   While current node is not the destination
    7.1.3.   Find the node with highest smell spot value
             and the node not visited
    7.1.4.   Update the SDA current node and path length
             and also the Path
   7.2.       Store the path of each SDA's
8. Shortest path is found from agents with less path size
```

4.2   Complexity Analysis of SDA

Assume there are *N* nodes in network topology and *M* agents are used to find best path in networks. The time complexity is represented by big—*O* notation. Assume that for each arithmetic, logic operations, read and write operation take $O(1)$ time. The time complexity of SDA is derived using Algorithm 1. For initializing *N* node

with initial smell value, it takes $O(N)$, and for M agent initialization, it takes $O(M)$. Updating smell value for each nodes takes time $O(N)$. So total time taken is $O(M + N + N)$, which is $O(M + N)$. To find the highest smell value of each node, the time taken is $O(\log E)$, if an agent visits V nodes to reach destination then time taken is $O(V \log E)$. For M agents the time take is $O(M \times V \times \log E)$. $M \times V$ can be termed as B thus time complexity becomes $O(B \log E)$. Thus, total time complexity is $O(M + N + B \log E)$. $M + N$ is constant and is represented as A. So the total time complexity for SDA is $O(A + B \log E)$.

4.3 Working of SDA

Assume a network topology with nine nodes as shown in Fig. 1. The value between nodes represents the hope time.

The SDA algorithm is used to find path from source node (say 1) to destination node (say 9), and total number of agents is assumed as 2. First, the nodes are initialized with initial smell value by the formula $1/\mathrm{sqrt}(x^2 + y^2)$ where x and y denote the location coordinates of the node. After initialization, smell value of each node is updated based on the equation motioned in Algorithm 1(line 6.1 to 6.3). Each agent starts finding the path from node 1 to node 9. From node 1, the agent finds the next node with highest smell value, and the node identified is 2. The current node of the agent is updated as node 2. Next node with highest smell value from 2 is found from the connected nodes and found as 3, like this path identification moves until destination is reached. So the path found by Agent 1 is 1-2-3-6-9. Similarly, Agent 2 finds the path by proceeding this steps. The path found by Agent 2 is 1-4-7-8-9. The optimized path is found by comparing the total number of nodes visited by each agent. In this case, path found by Agent 1 or Agent 2 can be used for data transfer in the network.

Fig. 1 Network topology
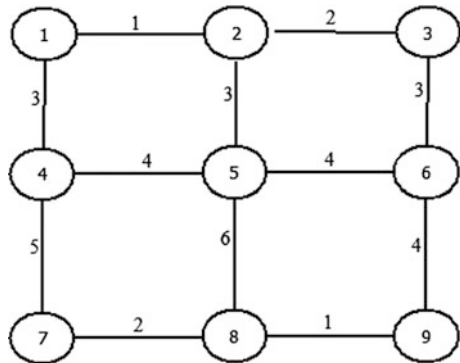
Table 1 Font comparison of time complexity

Algorithm	Time Complexity
Dijkstra's	$O(|M| + |M| \log |N|)$
SDA	$O(|A| + |B| \log |E|)$

4.4 Comparison of SDA with Dijkstra's Algorithm

Dijkstra's algorithm was found by Edsger Dijkstra [10]. Dijkstra's algorithm is used for finding shortest path. Dijkstra's algorithm uses programming methodology for finding shortest path, whereas SDA uses parallel methodology for finding shortest path. That is multiple SDAs find independent shortest paths.

The time complexity analysis of Dijkstra's algorithm and smell detection agent-based algorithm is shown in Table 1. From the time complexity analysis, the SDA is same as Dijkstra's algorithm. But in case of multiple solutions SDA (say two agents) the time complexity becomes $O(2* (|A| + |B| \log |E|))$ which is $O(|A| + |B| \log |E|)$. So the total time complexity is $O(M + N + B \log E)$.

5 Results and Discussion

SDA algorithm reduces the complexity of finding a shortest path in computer networks. This algorithm works on the basis of dog's path tracing behaviour. Each SDA retrieves a path from the paths; the optimized path is identified by the less number of nodes it traversed to the destination.

In ACO shortest path is determined by pheromone concentration along the path, whereas in SDA, the optimized path is determined along the high smell spot values. Pheromone concentration in each node varies according to the ant visits, whereas in SDA, the smell value is updated only once. In ACO, ants are active and used intelligent choice in total process of ACO algorithm, whereas in SDA, agent plots only the path. Thus, SDA is passive in nature and performs effectively.

The SDA algorithm was used to find path from node 3 to node 20 in the network topology as shown in Fig. 2. First nodes are assigned with initial smell value based on their location coordinates. Here, assume 3 agents find the path from node 3 to node 20. The smell values of each node are updated with respect to destination using Algorithm 1. Each agent starts to search path from source node to destination node. Agent 1 returns path as 3-2-6-12-17-18-20, Agent 2 returns path as 3-4-1-5-11-16-19-20, and Agent 3 returns path as 3-9-14-20. From the path returned by agents, Agent 3 returns the optimized path for data transfer between node 3 and node 20.
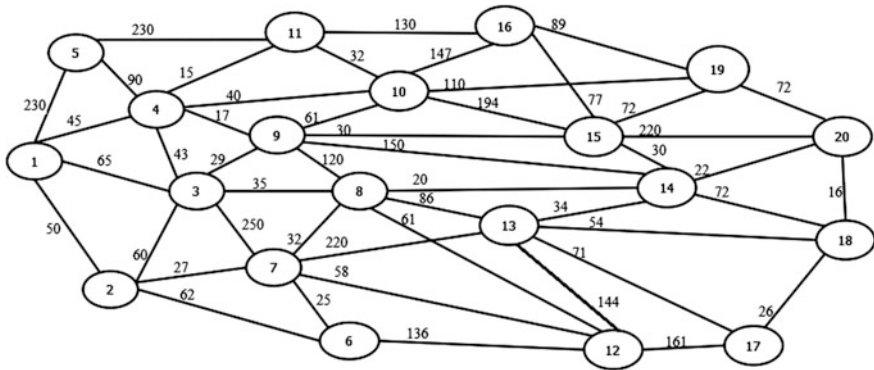
Fig. 2 Network with 20 nodes

The SDA algorithm is used to find the shortest path from node 11 in network topology as shown in Fig. 2. The initial smell value for each node is assigned. Here, 2 agents are initialized to find optimized path in the network. The smell values are then updated by Algorithm 1. Each agent starts finding path from node 11 to node 20 based on smell value. Agent 1 returns path as 11-4-9-14-20, Agent 2 returns path as 11-16-19-20, and the optimized path is obtained by Agent 2.

The SDA algorithm is tested in the network topology (Fig. 2). There are 4 agents marked with source node 1 and destination node 20. The nodes are initialized with initial smell value as mentioned in the Algorithm 1. Each agent is initialized with their ID, current node, and flag values. The smell value is updated for each node with respect to destination using the Algorithm 1. After updating the smell value, each agent starts searching path one after other and reached the destination node. The path found by Agent 1 is 1-2-6-12-17-18-20, path found by Agent 2 is 1-4-11-16-19-20, path found by Agent 3 is 1-3-8-14-20, and path found by Agent 4 is 1-5-4-9-10-15-20. From the search, it is found that agent 3 returns path with less number of nodes, so the optimized path is 1-3-8-14-20. The total time required to find optimized path by SDA is 0.016 s.

The computer network (Fig. 2) is also tested using ACO algorithm with the same source and destination nodes [11]. ACO is iterated 12 times to reach the destination node from same node. The optimized path is obtained during 11th iteration.

We also tested the same network by PSO algorithm [8] from this same source and destination. PSO is iterated 600 times. The time taken to obtain the path is around 0.01 s.

Table 2 shows the path obtained by ACO, PSO, and SDA algorithms with time complexity. By analysing the time complexity of ACO, PSO, and SDA, SDA has time complexity less compared to ACO and PSO. ACO finds the path during 11th iteration, whereas SDA finds the path by 3rd agent. PSO and SDA take almost same

Table 2 Path obtained by algorithms

Algorithm	Path	Time complexity
ACO	1-3-8-14-20	$O((1/\rho) \, n^2 \, m \log n)$
PSO	1-3-8-13-14-20	$O(M \, X)$
SDA	1-3-8-14-20	$O(A + B \log E)$

time to find the path, but the total number of nodes that need to reach destination by PSO is 4, whereas SDA needs to traverse only 3 nodes to reach the destination. Thus, the SDA is more suitable for computer networks, exclusively path identification.

6 Conclusion

SDA algorithm reduces time complexity to find optimized path in computer networks. When comparing the performance of SDA with other bio-inspired algorithms, it is found that SDA gives optimized routing path in less amount of time. This algorithm can also be implemented in software- defined network (SDN) for finding optimized path. In future, SDA algorithm can be enhanced to find path from a single source to multiple destinations.

References

1. Fister I Jr, Yang XS, Fister I, Brest J, Fister D (2013) A brief review of nature-inspired algorithms for optimization. Electrotech Rev 80(3):116–122
2. Saritha R, Vinod Chandra S S (2016) An approach using particle swarm optimization and rational kernel for variable length data sequence optimization. In: Proceedings of Springer ICSI, pp 25–30
3. Vinod Chandra SS, Anand HS (2014) Artificial intelligence and machine learning. PHI Publishers, New Delhi
4. Saritha R, Vinod Chandra SS (2014) A novel algorithm based on honey bee foraging principle for transportation problems. ACCIS Proc Elsevier 1:203–210
5. Vinod Chandra SS (2015) Smell detection agent based optimization algorithm. J Inst Eng (India): Ser B 97:1–6. https://link.springer.com/article/10.1007/s40031-014-0182-0
6. Keisam Thoiba Meetei (2014) A survey: swarm intelligence vs. genetic algorithm. Int J Sci Res 3:231–235
7. Verma OP, Sharma M, Gupta N, Nanda P, Chawla S (2011) A new approach to dynamic network routing using omicron ant colony algorithm. IEEE 5:177–181
8. Deng Y, Tong H (2011) Dynamic shortest path algorithm in stochastic traffic networks using PSO based on fluid neural network. J Intell Learn Syst Appl 3:11–16

9. Deeban Chakravarthy V, Sivarajan S, Gayathiri N (2014) Artifical Bee Colony (ABC) Algorithm with ECDH Algorithm for Finding Optimal Path and Secure Data Transfer. Int J Sci Eng Res 5(3):436–439
10. Sanan S, Jain L, Kappor B (2013) Shortest path algorithm. IJAIEM 2(7):316–320
11. Alshaheen HS (2013) Finding shortest path in routing problem by using ant colony optimization. J Univ Thi-Qar 8(3):125–131

Selective Encryption Algorithm Based on Natural Language Processing for Text Data in Mobile Ad hoc Network

Ajay Kushwaha, Hari Ram Sharma and Asha Ambhaikar

Abstract In today's era, security is extremely recommended for data which is send over network. The research problem states that although various approaches toward security are proposed day by day, the security loop holes are also propagating and thus constant innovation is essential toward protection of data. The primary objective of this paper was to introduce a selective significant data encryption method (SSDE). The SSDE provides sufficient uncertainty to the data encryption process as it selects only significant data out of the whole message using natural language processing (NLP). This in turn reduces the encryption time and enhances the performance. The research work shows that proposed work is superior to the existing work.

Keywords Cryptography · Tokenization · Selective encryption
Natural language processing · Stop words · Natural Language Toolkit

1   Introduction

Nowadays, world is drifting toward wireless network, and thus, ad hoc networks are also acquiring importance. An ad hoc network is defined as a wireless network in which all the nodes are able to communicate with each other directly without the need of a central access point. The performance of ad hoc network is good when less number of nodes are involved, but when the number of nodes increases, the performance gets affected and becomes difficult to manage. As mobile ad hoc

A. Kushwaha (✉) · H.R. Sharma · A. Ambhaikar
Rungta College of Engineering and Technology, Bhilai, Chhattisgarh, India
e-mail: ajay.kushwaha@rungta.ac.in

H.R. Sharma
e-mail: hrsharma44@gmail.com

A. Ambhaikar
e-mail: dr.asha.ambhaikar@rungta.ac.in

networks are widely used nowadays, so the security requirements for the network are also increasing which can be provided by means of cryptography.

The encryption and decryption could be performed by the use of keys. There are two types of key-based encryption: symmetric and asymmetric algorithms. In case of symmetric algorithms, the keys are same, while asymmetric algorithms possess dissimilar keys. Symmetric algorithms may have stream ciphers and block ciphers. The key here is called secret key used in both sender's and receiver's end. One of the examples of symmetric algorithm is DES.

In asymmetric algorithm, the public key is available at both ends, while private key is available at only one side. When data is encrypted by public key, it can be decrypted by only private key and vice versa. The algorithm, also called as public key cryptography, provides the fit of authenticating the source as a means of digital signature. An example for asymmetric algorithms is RSA.

Natural language processing (NLP) is related to the area of interaction between humans and computers. The biggest challenges in natural language processing are natural language understanding, word processing, information management, and enabling computers to obtain meaning from humans.

The paper is organized in the following way: Related work is provided in Sect. 2. In Sect. 3, proposed methodology SSDE is introduced, which is followed by the result analysis in Sect. 4, and 5 contains the conclusion part.

2 Related Work

Yonglin et al. [1] exhibit a probabilistic methodology that increases the uncertainty in the process of message encryption. Matin et al. [2] propose a method which emphasizes on the security provided at the application level. The key size makes the task difficult for the intruder to break. Zhou and Tang [3] study the issues related to implementation of a RSA public key algorithm. Uma parvathi et al. [4] had a deep study in symmetric encryption algorithms such as AES, Blowfish, and 3DES with respect to power consumption. Roy et al. [5] introduced a selective video encryption algorithm which deals with real-time security in which higher intense bits were used to achieve higher visual results.

Shen et al. [6] presented a two-layer selection scheme for selective video encryption method which was also compared with SEH264 algorithm. The proposed method reduces the complexity by 50% over existing methods. Xu et al. [7] exhibit a selective image encryption method based on hyper-chaos. Under this algorithm, the image is divided into blocks, and coefficient blocks are obtained by DCT transformation and quantization on the blocks. Massoudi et al. [8] provide various evaluation standards for JPEG 2000 compressed image transfers such as cryptographic security and format compliance.

3 Proposed Methodology: Selective Significant Data Encryption

3.1 Concept of Selective Encryption

Selective encryption algorithms reduce the overhead spent on data making it more popular nowadays. The purpose of selective encryption algorithms is to provide trustworthy safety so as to secure the transmitted message confidentiality. It also improves the scalability of data transmission. NLP is used for selective encryption of messages. Natural Language Toolkit 3.0 with python 3.5 versions is used for analyzing messages under this proposed method [9].

3.2 Selective Significant Data Encryption

In SSDE, the significant data in the message is encrypted before sending over the network, while the remaining data is kept as it is [10].

3.3 Algorithm of the Proposed Methodology

Step 1　Input messages.
Step 2　Remove special characters from the messages like *$&? etc.
Step 3　Process of tokenization in which it extracts all the words present in the messages.
Step 4　Dropping stop words (common words) and collecting significant data (key words) from the messages.
Step 5　Encrypt significant data (key words) using Blowfish method and send it to the network.
Step 6　Stop words are sent without encryption on to the network.

The above Fig. 1 shows the flowchart representation of SSDE showing the flow of control from one step to another.
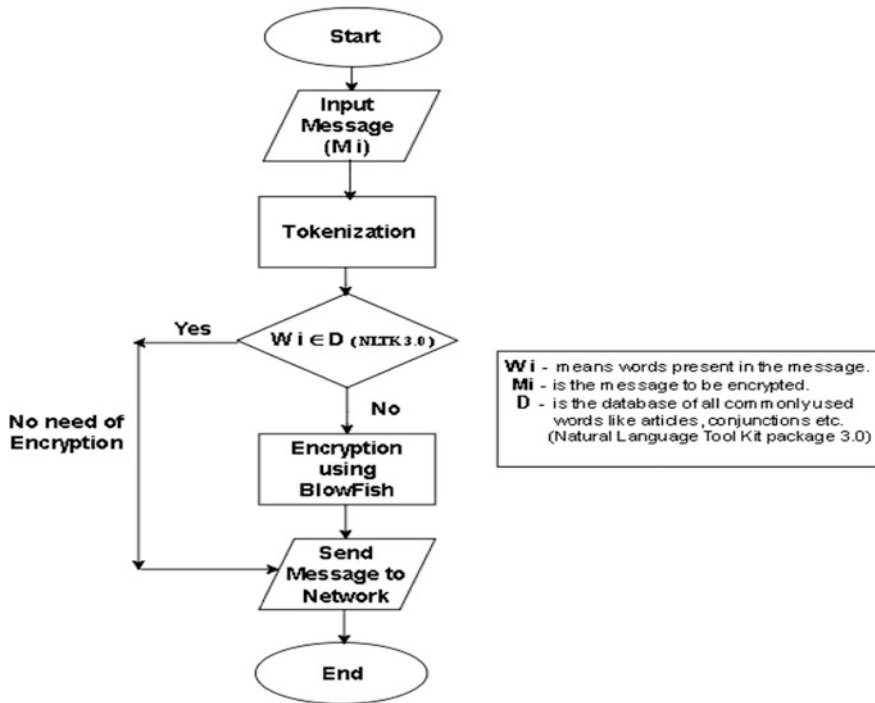
Fig. 1 Flowchart for proposed methodology

4 Result Analysis

In order to observe the characteristics of the proposed method SSDE, set of experiments were carried out in a wireless environment. In this work, the proposed method SSDE is compared with full encryption method and Toss-A-Coin method. Each test is executed for 50 ns under identical conditions. The various parameter used for evaluating the performance of SSDE are encryption time percentage, decryption time percentage, and encryption proportion.

The SSDE method is compared with 2 approaches as stated above. The first method (i.e., full encryption) encrypts all messages without leaving any text unencrypted and thus termed as full encryption. In the second approach, half of the data is encrypted and is termed as toss-a-coin method.

Figures 2 and 3 represent the comparison of encryption with simulation time and decryption with simulation time, respectively, for the above three approaches.

It is shown from Fig. 2 that toss-a-coin and SSDE methods are efficient than full encryption method, whereas Fig. 3 given below shows that full encryption method takes more time than the other two methods (i.e., toss-a-coin and SSDE). SSDE is encrypting the keywords (i.e., significant words) making the intruder's task difficult, whereas toss-a-coin is less time consuming, but it does not focus on significant
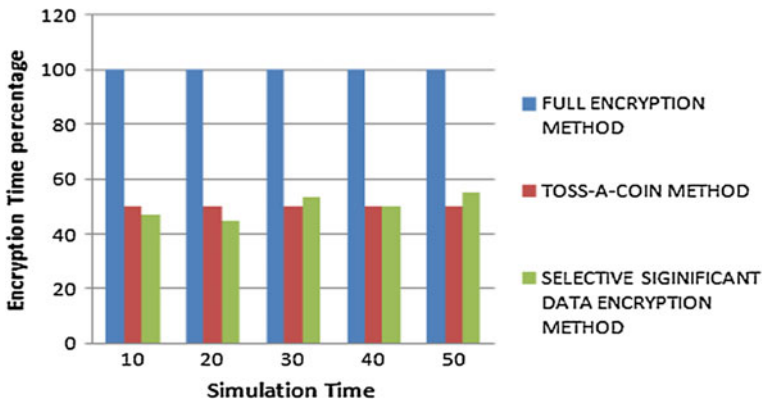
Fig. 2 Encryption time percentage versus simulation time
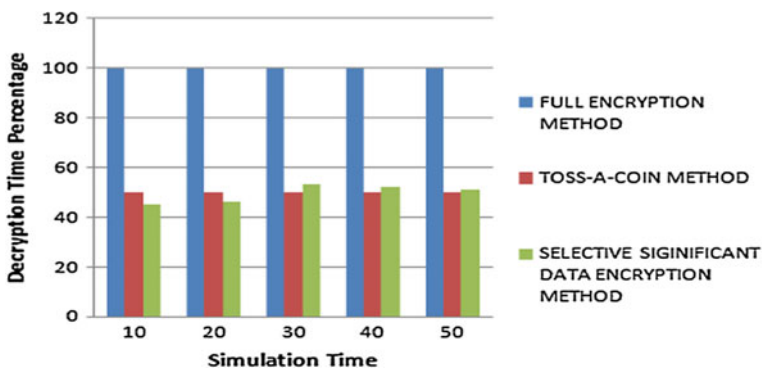
Fig. 3 Decryption time percentage versus simulation time

words. It encrypts random words making the intruder's task easier. From the security point of view, SSDE is more secure and efficient than toss-a-coin and full encryption method.

5 Conclusion

This paper introduces a better solution for data encryption in wireless networks.

The performance of the method is evaluated by carrying various experiments. The outcome shows the edge of SSDE over other methods in wireless network. It can be used in social chatting apps, military security, corporate world communication, and government activities involving text data encryption. This method can be used for text data only. In future, this method can be extended for other file formats (i.e., audio, video).

References

1. Boukerche A, Mokdad L, Ren Y (2011) Performance analysis of a selective encryption algorithm for wireless ad hoc networks. In: IEEE, pp 1038–1043
2. Matin MA, Hossain MM, Islam MF, Islam MN, Hossain MM (2009) Performance evaluation of symmetric encryption algorithm in MANET and WLAN. In: International conference for technical postgraduates (TECHPOS), Kuala Lumpur, 14–15 Dec
3. Zhou X, Tang X (2011) Research and implementation of RSA algorithm for encryption and decryption. In: 6th International forum on strategic technology
4. Umaparvathi M, Varughese DK (2010) Evaluation of symmetric encryption algorithms for MANETs. In: International conference on computational intelligence and computing research (ICCIC), pp 1–3, IEEE
5. Roy M, Pradhan C (2011) Secured selective encryption algorithm for MPEG-2 video. Int Conf Electron Comput Technol (ICECT) 2:420–423
6. Shen H, Zhuo L, Zhao Y (2014) An efficient motion reference structure based selective encryption algorithm for H264 videos. Inf Secur, IET 8(3):199–206
7. Xu P, Zhao J, Wang D (2011) A selective image encryption algorithm based on hyper-chaos. In: 3rd international conference on communication software and networks (ICCSN), pp 376–379, IEEE
8. Massoudi A, Lefebvre F, De Vleeschouwer C (2008) Overview on selective encryption of image and video: challenges and perspectives. In: 10th IEEE international symposium on multimedia, pp 31–38
9. Steven B, Loper E, Klein E (2009) Natural language processing with python. O'Reilly Media Inc, Sebastopol
10. Ajay K, Sharma HR (2014) Designing an enhanced selective encryption method for securing mobile adhoc network. In: International conference on computational intelligence and communication networks (CICN), IEEE

All Optical Health Monitoring System: An Experimental Study on Visible Light Communication in Biomedical Signal Transmission

C.R. Uma Kumari and S. Dhanalakshmi

Abstract In the biomedical signal analysis and transmission, procreation of radio frequency communication technology is recurrently perplexed by electromagnetic interference. This effect can be eliminated by using visible light as a medium of transmission. In this work, Light Fidelity (Li-Fi) transmitter and receiver sections have been designed using white light emitting diode (LED) and photodiode, respectively. For demonstration, we transmitted electrocardiogram signal from Li-Fi transmitter to receiver and the performance has been analysed for various transmission distances. Further enhancement of this work has been proposed as an all-optical health monitoring system that is suitable for hospital environment. In hospitals, VLC is the efficient wireless data communication technology in health monitoring. VLC ensures secured, faster way of communication and use of unlicensed visible spectrum. This enhanced scheme uses biomedical sensors that detect specific biological, chemical, or physical changes that occur in a patient's body. Biological data are transmitted by an array of LEDs. Biomedical signal processing in both transmitter and receiver sections is performed by Arduino unit. The proposed research work focuses on developing a high-speed eco-friendly local area communication network with low-cost, off-the-shelf components for interconnecting electronic devices within hospital environment.

C.R. Uma Kumari (✉) · S. Dhanalakshmi
Department of Electronics and Communication Engineering, SRM University, Chennai, India
e-mail: umakumari.cr@gmail.com

1   Introduction

Visible light communication (VLC) uses light in the visible region as a medium for transmission. VLC has some unique advantages such as security against hacking, as transmitted light is confined within the systems' coverage area and also it is less hazardous to human health. Health monitoring system has drawn considerable attention since past few decades. With advancement in wireless technology, health monitoring system can be widely deployed to monitor health condition of a patient inside and outside the hospitals. VLC along with Free Space Optics (FSO) technology would play a vital role in future wireless communication domain. The FSO market has achieved a growth of 51.8% during the forecast period, and VLC market has faced 91.8% growth between 2015 and 2020. Sensors that form Internet of Things require fast data rates that could be achieved by either VLC or FSO. FSO makes use of ultraviolet (UV) rays, infrared (IR) rays, or visible light as a carrier of data, whereas VLC uses light under visible spectrum (380–780 nm).

Visible light communication using white LEDs provide illumination along with data transmission in the field of communication. VLC plays a major role in RF-restricted areas such as hospitals, mines, and aircraft. All kinds of indoor environment can be replaced by white LEDs as they have longer life time, high data rate, and low energy consumption with no impact on health. Li-Fi can also be called as a form of green technology. Underwater communication will be possible between deep sea divers and explorers. Current research has been focused on inter-satellite communication using VLC. Standard for VLC is IEEE 802.15.7 that supports high-data-rate visible light communication [5] up to 96 Mb/s. It involves ultraspeed modulation of light sources.

The following passages summarize some of the recent research works that were performed in the area of VLC, FSO, and Li-Fi.

Allen et al. [1] make use of light emitting diodes that are operated in the range of 420–620 nm that generates photoacoustic signals that are employed in vascular phantoms [1]. They observed changes in velocity with 200-ns pulses that operates at less duty cycle. It enables energy of pulses in the range of 10 µJ. Dhatchayeny et al. [2] demonstrated the EEG (electroencephalogram) signal transmission through visible light. In their work, red–green–blue (RGB) LEDs and photodiodes with colour filters were used for transmission and reception, respectively. This system achieves a bit error rate (BER) greater than 1.5×10^{-5} at a signal-to-noise ratio (SNR) value of 7 dB using 10 channels of raw EEG signals.

Han et al. [3] proposed a colour-clustered optical Red-Green-Blue (RGB) Multiple-Input-Multiple-Output (MIMO) system. On-Off Keying (OOK) modulation was adopted to modulate 30 RGB LEDs. Later, selection combining algorithm was performed at receiver end that produces a diversity effect within one particular colour cluster.

Josline Priya et al. [4] presented VLC technology-based downlink wireless system that makes use of pulse width modulation (PWM). They demonstrated a transmission at a distance of 30 cm with 9600 baud rate (peer to peer

communication) and a baud rate of 9600 at a distance of about 10 cm (broadcast communication).

Lawrentschuk et al. [6] stated that clinically relevant EMI was secondary to mobile phones that potentially endanger patients. This was concluded by experiments conducted in 45 out of 479 mobile phones that are operated at 900 MHz and, similarly, 14 devices of 457 that are examined at 1800 MHz. In further studies, it was predicted that clinically significant EMI occurred when mobile phones were used within 1 m of any medical equipment.

2 Proposed Model

2.1 Li-Fi Transmitter Section

The Li-Fi system consists of two main parts: transmitter and receiver. Transmitter part modulates the input signal with required time period and transmits the data in the form of 1s and 0s using LED bulb. These 1s and 0s are flashes of the bulb. The receiver part catches the flashes using a photodiode and amplifies the signal.

The proposed work uses PIC16F877A microcontroller for signal processing. ECG sensor used in this work has the following specifications: current consumption: 4 mA, gain: 1100, range: 1.5 mV (with VCC = 3.3 V) and ECG sensor senses heart signal of human. A simple VLC system has two qualifying components: (1) a light source (LED) equipped with a signal processing unit (PIC microcontroller) and (2) at least one device that is capable of receiving light signals (PIN photodiode). LED bulbs are excellent to realize Li-Fi transmitter. LED cells can modulate thousands of signals without the human eye ever noticing. Changes in light intensity from the LED light source are interpreted by photodiode. It converts light energy to electrical signal. Electronic signal is further demodulated to obtain sets of binary data that can be used by any Internet-enabled devices.

Block diagram of transmitter is shown in Fig. (1). The signal from ECG sensor is given to PIC16F877A. It converts analog signal to digital for processing. The input voltage for microcontroller is 5 V. Further, MAX232 acts as a voltage regulator that converts 5–12 V, and the signal is passed to LM555 timer circuit. LM555 timer generates carrier wave at 4.9 kHz. This carrier wave is modulated by the processed ECG signal. On-off keying (OOK) modulation is implemented in this
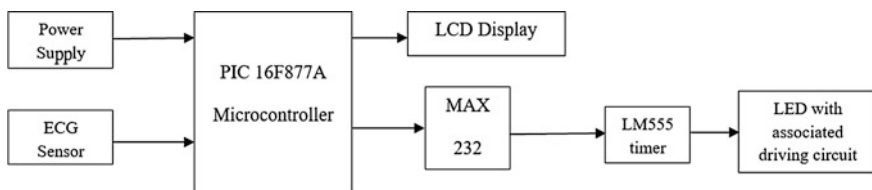
Fig. 1 Block diagram of Li-Fi transmitter
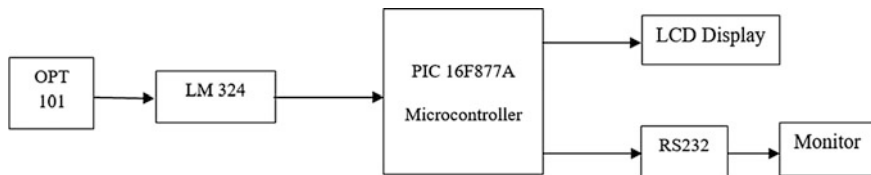
Fig. 2 Block diagram of Li-Fi receiver

experiment. It is the simplest modulation technique where the presence of the digital data is represented as binary one and absence of the digital data with zero. The advantage of OOK modulation is the reduced power consumption compared to other modulation techniques such as phase-shift keying (PSK), binary phase-shift keying (BPSK) and quadrature phase-shift keying (QPSK). When compared to RGB LEDs, the use of white LED is cost effective and less complicated. The resistor in LED driver prevents enormous flow of current.

2.2 Li-Fi Receiver Section

Li-Fi receiver includes OPT101 which is a monolithic photodiode with on-chip Trans-Impedance Amplifier (TIA), voltage amplifier (LM324), and microcontroller unit followed by monitoring unit. Connection of all the components is shown in Fig. 2.

The monolithic photodiode and single-supply Trans-Impedance Amplifier (TIA) are driven in photoconductive mode to achieve perfect linear operation. TIA that presents on the same chip converts electrical current into equivalent voltage. The signal is then passed through High Pass Filter (HPF) to amend external noise and light from other sources. The signal from photodetector will be in the range of millivolts, and hence, it should be amplified. This amplification is performed by LM324 (voltage amplifier circuit).

PIC microcontroller in this section is programmed to perform as a voltage comparator to convert the data signal into digital format. Thus, the transmitted signal is recovered back at the receiver side. Finally, the data streams are collected from microcontroller with the help of RS232 interface. The installation cost of this scheme is less compared with other RF communication systems.

3 Result and Observation

Whenever the ECG signal is transmitted, for a normal signal (high output value of logic 1) 'NOR' will be displayed in the screen. If it is an abnormal signal, it will display (low output value of logic 0) 'ABN' in the screen as shown in Fig. 3. There is no delay in getting the output that leads to faster response.
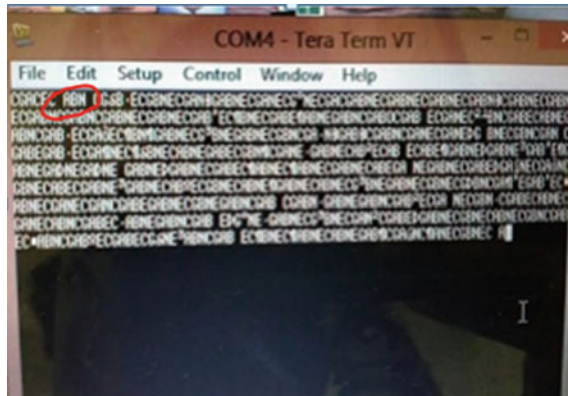
Fig. 3 Display of monitoring unit
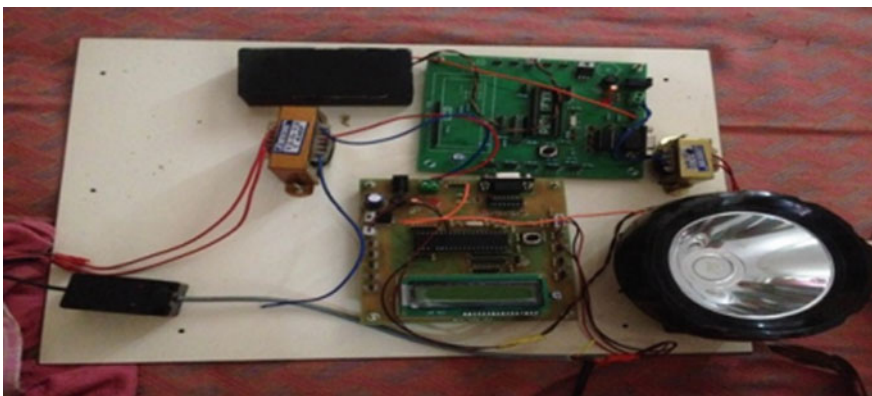
Fig. 4 Final setup after fabrication

Table 1 Distance versus amplitude

Distance (cm)	Received signal strength (mV)
0	50
5	30
10	30
20	10
30	8
40	4
50	2

The setup in Fig. 4 is compatible till the distance of 20 cm, after that amplitude of the signal decreases. This is evident from Table 1.

The corresponding graph in Fig. 5 shows the variation in received signal strength with respect to distance between transmitter and receiver.

Fig. 5 Plot of distance versus received signal strength
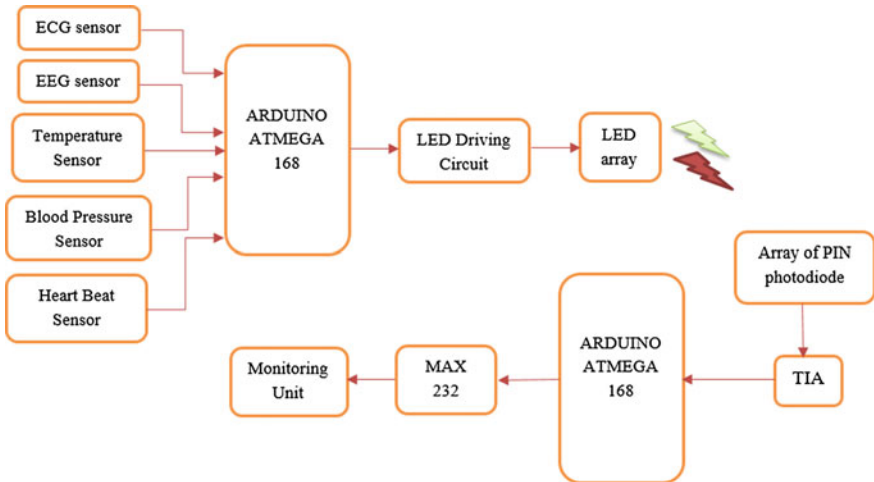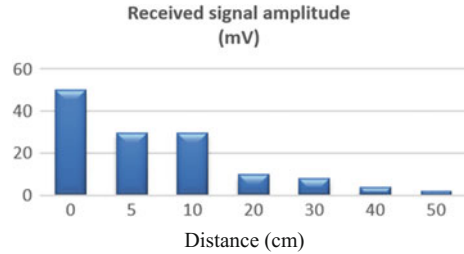
Fig. 6 Optical health monitoring system using VLC

4 Enhancement of Proposed Scheme

4.1 All Optical Health Monitoring System

The previously mentioned experiment can be enhanced to a health monitoring system by using an array of five light sources (either LEDs or LASER diode) for transmitting and monitoring five different sensor outputs. This enhanced scheme can make use of pulse width modulation (PWM). Similarly, in the receiver section, an array of five photodetectors [either PIN or avalanche photodiode (APDs)] receives the light. Signal processing in transmitter and receiver sections is performed using Atmega 168 microcontroller that is built in Arduino module. Personal computer (PC) is used as monitoring unit. The fact that optical signals do not create electromagnetic interference (EMI) will make this health monitoring system well suitable for use inside operation theatres and in intensive care unit (ICU).

Block diagram of all optical health monitoring system is shown in Fig. 6.

Table 2 Sensor unit

Biomedical sensor	Parameter to be accessed	Output form of sensor to be interfaced with Arduino
ADS1292R Electrocardiogram (ECG) sensor	Electrical and muscular functions of the heart	Analog
Electroencephalogram (EEG) sensor	Electrical activity of human brain	Analog
LM35 temperature sensor	Body temperature	Analog
BME441 Blood pressure sensor	Blood pressure level	Analog
Polar heart rate receiver	Rate of heart beat	Digital

4.2 Sensor Unit

Any biomedical sensor can be used based on patients' health monitoring system. The selection of sensor should be done carefully such that the output from sensor is compatible with the input ports of Atmega 168 microcontroller. Some of the commercially available biomedical sensors that are used in this work are listed in Table 2.

4.3 Interfacing with Arduino Module

In this project, Arduino unit with Atmega 168 microcontroller has been used for processing analog and digital outputs from biomedical sensor unit.

Atmega 168 has 28 pins out of which 6 pins support analog inputs as shown in Fig. 7. All the signal processing is done in digital domain. To support digital inputs and outputs, 14 pins are available. Because of all these features, Atmega 168 is well suitable to use more number of biomedical sensors simultaneously.

As depicted in the pin diagram, four analog input pins (0, 1, 2 and 3) and one digital input pin (2) are used in this proposed model to get output of sensors. Pulse width-modulated (PWM) digital output can be taken from pins 5, 6, 9, 10, and 11 in a serial manner.

4.4 Programming Arduino

User-defined program for biomedical signal processing is coded using Arduino software (Arduino Duemilanove or Nano w/ATmega168). The ATmega168 on the Arduino Nano comes preburned with a bootloader that can be used to upload new code to it without the use of an external hardware programmer. It communicates using the original STK500 protocol.

Atmega168 Pin Mapping

Arduino function				Arduino function
reset	(PCINT14/RESET) PC6	1	28 PC5 (ADC5/SCL/PCINT13)	analog input 5
digital pin 0 (RX)	(PCINT16/RXD) PD0	2	27 PC4 (ADC4/SDA/PCINT12)	analog input 4
digital pin 1 (TX)	(PCINT17/TXD) PD1	3	26 PC3 (ADC3/PCINT11)	analog input 3
digital pin 2	(PCINT18/INT0) PD2	4	25 PC2 (ADC2/PCINT10)	analog input 2
digital pin 3 (PWM)	(PCINT19/OC2B/INT1) PD3	5	24 PC1 (ADC1/PCINT9)	analog input 1
digital pin 4	(PCINT20/XCK/T0) PD4	6	23 PC0 (ADC0/PCINT8)	analog input 0
VCC	VCC	7	22 GND	GND
GND	GND	8	21 AREF	analog reference
crystal	(PCINT6/XTAL1/TOSC1) PB6	9	20 AVCC	VCC
crystal	(PCINT7/XTAL2/TOSC2) PB7	10	19 PB5 (SCK/PCINT5)	digital pin 13
digital pin 5 (PWM)	(PCINT21/OC0B/T1) PD5	11	18 PB4 (MISO/PCINT4)	digital pin 12
digital pin 6 (PWM)	(PCINT22/OC0A/AIN0) PD6	12	17 PB3 (MOSI/OC2A/PCINT3)	digital pin 11(PWM)
digital pin 7	(PCINT23/AIN1) PD7	13	16 PB2 (SS/OC1B/PCINT2)	digital pin 10 (PWM)
digital pin 8	(PCINT0/CLKO/ICP1) PB0	14	15 PB1 (OC1A/PCINT1)	digital pin 9 (PWM)

Fig. 7 Pin mapping of Atmega168

Table 3 Light sources of different colours

S.No	Colour of LED	Wavelength range
1	Red (ECG sensor)	610–760 nm
2	Yellow (EEG sensor)	570–590 nm
3	Green (temperature sensor)	500–570 nm
4	Blue (blood pressure sensor)	450–500 nm
5	Violet (heart beat sensor)	400–450 nm

Bootloader can also be bypassed, and programming the microcontroller can be done through the ICSP (in-circuit serial programming) header using Arduino ISP (in-system programmer).

4.5 Interfacing Arduino Output with Light Source Array

Serial digital outputs from Arduino module are used to drive light emitting diodes (LEDs) that produce modulated output. Spectral response curve of optical output should be measured for each light source used in the array.

4.6 Selection of Light Source

Table 3 shows the list of LEDs that are used to transmit outputs of sensors, and the corresponding wavelength range is also mentioned. Since there are differences in wavelength range, wavelength division multiplexing (WDM) can also be employed to enhance efficiency of this network.

4.7 Photodiode Array

Modulated visible light that reaches receiver section is detected by an array of five PIN photodiodes. Each PIN photodiode is equipped with an individual colour filter corresponding to the colour of light sources used in the transmitter section. Output current from all the detectors should be measured.

4.8 Trans-Impedance Amplifier (TIA)

TIA is implemented for current-to-voltage conversion. In this research, an array of five LM358AN Trans-Impedance Amplifiers is used to amplify the signal that is received from photodetector array. TIA section should be interfaced with Arduino unit that is inbuilt with an Atmega 168 microcontroller in the receiver section.

4.9 Arduino Module and Monitoring Unit in Receiver

Processing of received signal is performed in digital domain in this microcontroller unit. Noise cancellation should be done in this part of work. Digital-to-analog conversion can be performed on specific data by suitable coding. Received data can be compared against reference values that are coded in the Arduino program.

4.10 Transmission of Arduino Output to Monitoring Unit Through MAX 225

This is the final step in which received data should be displayed and stored in a personal computer (PC) or in a laptop. It can be accomplished by using MAX 225 IC that converts TTL output of Arduino to RS485 form that is compatible with the monitoring system under consideration. MAX 225 is selected since it has five I/O pins corresponding to five biomedical sensors. Output of MAX 225 is connected to RS485 bus in monitoring unit.

5 Conclusion and Potential Applications of Proposed Work

The proposed system can be deployed in homes and hospitals for health monitoring. LED bulbs were used to produce high data rates, which is faster than average broadband connection. This system is ideally suited for healthcare

environments. Like conventional broadband and Wi-Fi, Li-Fi also functions as a bidirectional communication system. In the enhanced work, Arduino unit has been used instead of PIC microcontroller for interfacing more number of sensors.

In terms of application, FSO market has got its place in the domains of Mobile Backhaul, Enterprise Connectivity, Disaster Recovery, Defence, Satellite Communication, Healthcare, Security and in Engineering and Design. In the same essence, scope of VLC can be viewed in smart stores, consumer electronics, defence and security, vehicle and transportation, aviation, hospital, underwater communication and in hazardous environment (RF-restricted area).

6 Future Research

The proposed system makes use of LEDs and PIN photodiodes for transmission and reception of visible light signals, respectively. This can be replaced using an array of LASER diodes for transmission and avalanche photodiodes (APDs) for signal reception, and its performance can be compared with that of LED and PIN photodiode, respectively. If this technology can be put into practical use, every bulb (either LED or LASER) can be used like a Li-Fi hot spot to transmit wireless data and we will proceed towards a cleaner, greener, safer and brighter future.

References

1. Allen TJ, Beard PC (2016) High power visible light emitting diodes as pulsed excitation sources for biomedical photoacoustic. Biomed Opt Exp 7(4):1260–1270
2. Dhatchayeny DR, Sewaiwar A, Tiwari SV, Chung YH (2015) EEG biomedical signal transmission using visible light communication. In: International conference on industrial instrumentation and control (ICIC) on IEEE 2015
3. Han PP, Sewaiwar A, Tiwari SV, Chung Y-H (2015) Color clustered multiple input multiple-output visible light communication. J Opt Soc Korea 19(1):74–79
4. Josline Priya D, Daya N (2015) Visible light communication using white LEDs for indoor wireless data transmission. Int J innovative res Electr Electron Instrum Control Eng 3(1), April 2015
5. Rajagopal S, Roberts RD, Lim S-K (2012) IEEE 802.15.7 visible light communication: modulation schemes and dimming support. IEEE Commun Mag 50(3):72–82
6. Lawrentschuk N, Bolton DM (2004) Mobile phone interference with medical equipment and its clinical relevance. Med J Aust 181(3):145–149
7. Nezhad MH, Subari KS, Yahyavi M (2013) Improvement of wireless transmission system performance for EEG signals based on development of scalar quantization. J Elect Bioimpedance 4:62–72

Optimum Error Control Code for Underwater Acoustic Communication

Y. Venkata Ratnam, V. Malleswara Rao and B. Prabhakar Rao

Abstract Maintenance of quality of data is one of the challenges facing in ocean observing communication systems. Underwater acoustic channels are characterized by variable channel (by multipath reflections, Doppler shifts in frequency and acoustic noise in the channel) conditions and variable distances between sensor nodes due to water currents. Reliability is usually achieved by errorless communication. Errors can be reduced by using error-resistant modulation schemes or by implementing error control coding (ECC) schemes. The cheapest of these, in terms of cost and power, is ECC. Hence this project attempts to build such a mechanism for the UWA system of interest. In this chapter, one type of channel coding is described. Algorithms are developed for encoding and decoding of text data. Performance of this code is analyzed in terms of bit error rate for different parameters such as Doppler frequency shift, multipath delays compared with convolution-coded data. A software package is developed in MATLAB over AWGN channel, Rayleigh channel and Rician channel using M-ary FSK modulation scheme.

Keywords Error control coding (ECC) · M-ary FSK modulation
Bit error ate · Convolution codec · AWGN channel · Rayleigh channel
Rician channel · MATLAB

1 Introduction

In underwater due to attenuation of electromagnetic waves, it is not feasible to travel over long distances, optical waves suffered by scattering hence needs accurate fine laser beams [6], so acoustic waves become the better solution [1] for

Y. Venkata Ratnam (✉)
Department of ECE, SBTETAP, Hyderabad, India
e-mail: yvr10061976@gmail.com

V. Malleswara Rao
Department of ECE, GITAM University, Visakhapatnam, India

B. Prabhakar Rao
Department of ECE, JNT University, Kakinada, India

underwater communications. Due to multi-path delays, doppler frequency shifts, Inter-Carrier Interference (ICI) and Inter-Symbol Interference (ISI), it is difficult to transmit the acoustic signal at high data rate with adequate reliability in the ocean acoustic channel [2]. Due to attenuation of high frequency wave in seawater, carrier frequencies preferred for underwater are in the range of 10–20 kHz [3]. In the ocean environment, the reliability of reception, bandwidth efficiency and energy efficiency of acoustic signals are more challenging tasks than its propagation [6]. In such cases, optimum Error Control Coding (ECC) schemes in connection with efficient error-resistant modulation schemes are needed [8]. The cheapest of these, in terms of cost and power, is ECC. Hence we develop an optimum code that provides low bit error rate (BER) along with bandwidth efficiency.

Automatic repeat request (ARQ) stresses retransmission and hence provides high reliability data delivery ratio; however, the throughput decreases quickly in high bit error rate cases. In forward error correction codes (FEC), overheads are added for error reduction leads to reduction in data rates. Block codes or convolution codes are commonly used FECs used for low bit error rate of acoustic signal transmission [5]. Error control techniques designed for terrestrial communication require some sort of changes to suit ocean acoustic channels. In this chapter, an optimum code is developed that provides low bit error rate (BER) over convolution codec.

2 Method of Implementation

2.1 Optimum Encoder

Message to be transmitted will be entered through a keyboard. This text data is encrypted by using cryptography model into integer symbols [12] for safeguard security. In this model, the text data is modeled as a random set of integer symbols, which are combined to represent an "alphabet." Each alphabet is expressed in terms of any two integer symbols in the range of 0, 1, 2 … 7. Each integer symbol is again coded into any of the three symbols in the range of 0, 1, 2, 3, … 14, 15, becomes integer symbol in an optimum error control code (Fig. 1).

Fig. 1 Communication system design model
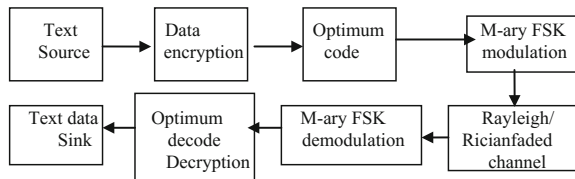
2.2 M-ary FSK Modulation

The software interfaced to the processor generates signal(s) (different frequencies of M-ary FSK) corresponding to each integer symbol transmitted through either in Rayleigh or Rician-faded sinusoidal waveforms in AWGN. The frequency bins in the MFSK [11] signaling are orthogonal to each other, leads to no leakage of the symbol energy from one channel to the other [10]. But in practice leakage takes place because of time variant nature of the channel. In order to reduce the effect of Doppler frequency shift, frequency interval must be much larger than the uncertainty in the Doppler shift estimation [13]. Frequency range used for M-ary FSK modulation is 9–13 kHz, 16 frequencies with interval of 0.25 kHz. The bit error rate performance can be judged not only from the noise but also from the ISI and inter-channel interference (ICI). Hence bit error rate modeling/prediction [14] in Rayleigh channel and Rician channels is considered.

2.3 Optimum Decoder

The output of stage is set to 16 channels of filters. Each filter is a narrow band amplifier (which is a notch filter) tuned to one of the frequencies of M-ray [7]. The output of each notch filter is led to envelop detector then to priority encoder. The output of envelop detector is almost DC voltage and is maximum at the frequency for which notch filter is tuned. Through priority encoder, all the 16 channel outputs are converted to a 3 integer symbols. Sets of 3 integer symbols in a sequence are converted as message.

3 Simulations

This section covers results/simulations done on MATLAB. Analysis of performance of the optimum code is done in terms of BER. This section covers the BER plots for optimum coded and convolution coded of bits. Then a comparison is made between them for Rayleigh-faded sinusoidal waveforms in AWGN channel shown in Tables 1 and 2 with different Doppler frequency shifts (fd), noise, with and without multipath delays, respectively, are considered. Rician-faded sinusoidal waveforms in AWGN channel shown in Table 3 when different Doppler frequency shifts (fd) in Hz, multipath delays and noise are considered.

From the above results, it is declared that for low Doppler frequency shifts (fd), i.e., below 25 Hz, optimum error control code and convolution code perform well. But for Doppler shifts (fd) below 125 Hz, optimum error control code provides low BER over convolution code in both Rayleigh-faded sinusoidal waveforms in

Table 1 Comparison between optimal-coded and convolution-coded data while not considering multipath effect in Rayleigh channel

Optimum code			Convolution code	
fd	Number of errors	BER	Number errors	BER
25	0	0	0	0
75	0	0	45	0.12
125	1	0.009	283	0.5
175	3	0.0026	297	0.5228
225	21	0.0179	291	0.5028

Table 2 Comparison between optimal-coded and convolution-coded data while considering multipath effect in Rayleigh channel

Optimum code			Convolution code	
fd	Number of errors	BER	Number of errors	BER
25	0	0	0	0
75	2	0.0017	122	0.2075
125	3	0.0026	277	0.4711
175	14	0.0119	293	0.4983
225	61	0.0519	327	0.5561

Table 3 Comparison between optimal-coded and convolution-coded data (13) while considering multipath effect in Rician channel

Optimum code			Convolution code	
fd	Number of errors	BER	Number of errors	BER
25	0	0	0	0
75	0	0	17	0.289
125	1	0.0009	238	0.4048
175	3	0.0026	256	0.4354
225	16	0.0136	282	0.4796

AWGN channel and Rician-faded sinusoidal waveforms in AWGN channel with SNR of 5 dB.

In Fig. 2, fd versus BER plot for optimal-coded and convolution-coded data is shown for Rayleigh-faded sinusoidal waveforms in AWGN channel with M-ary FSK modulation. It is observed that as Doppler frequency shifts is increasing from 25 to 225 Hz bit error rate is also increasing from 0 to 1.7×10^{-2} in optimum error control code whereas in convolution code bit error rate is increasing from 0 to 5.5×10^{-1}. It is quite clear from the figure that optimum error control code provides low BER for Doppler shifts (fd) below 125 Hz over convolution code with SNR of 5 dB without considering multipath effect.

In Fig. 3, fd versus BER plot for optimal-coded and convolution-coded data is shown for Rayleigh-faded sinusoidal waveforms in AWGN channel with M-ary FSK modulation, three path channel delay vector is [0, 1, 2 ms], and power gain vector is [0, −1, −9 dB]. It is quite clear from the figure that optimum error control
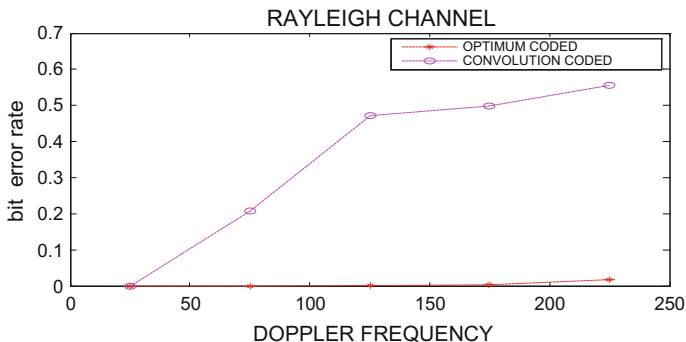
Fig. 2 fd versus BER for while not considering multipath effect in Rayleigh channel
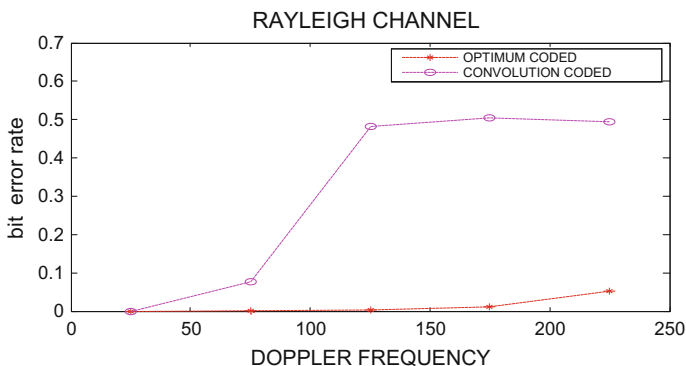
Fig. 3 fd versus BER for while considering multipath effect in Rayleigh channel

code provides low BER even considering both different Doppler shifts (fd) and multipath effect with SNR of 5 dB. It is observed that as Doppler frequency shifts (fd) are increasing from 25 to 225 Hz, bit error rate is also increasing from 0 to 5×10^{-2} in optimum error control code, whereas in convolution code bit error rate is increasing from 0 to 4×10^{-1}.

In Fig. 4, fd versus BER plot for optimal-coded and convolution-coded data is shown for Rician-faded sinusoidal waveforms in AWGN channel with M-ary FSK modulation, three multipath channel delay vector is [0, 1, 2 ms], K-factor is 3. Typically, Doppler frequency shift (fd) is increasing from 25 to 225 Hz, while bit error rate is also increasing from 0 to 1.3×10^{-2} in optimum error control code whereas in convolution code bit error rate is increasing from 0 to 4.7×10^{-1}. It is quite clear from the figure optimum error control code provides low BER even considering both Doppler frequency shift (fd) and multipath effect with SNR of 5 dB over convolution code especially fd is varying between 75 and 175 Hz.
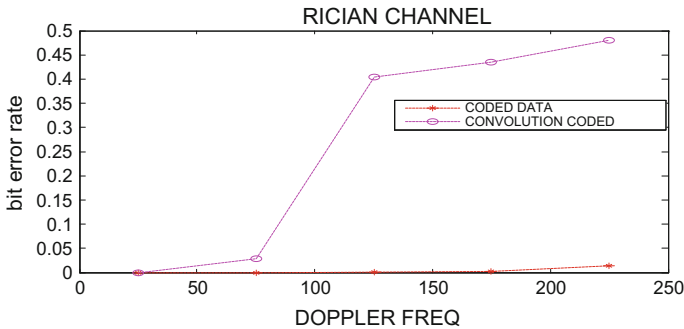
Fig. 4 fd versus BER for while considering multipath effect in Rician channel

4 Conclusions

This chapter covers the performance of error control code for reliable communication in adverse channels by using text data. It is shown that optimum error control code provides much better error performance as compared with the data in which convolution code is used. This developed code is sufficiently strong to overcome both channel effects and ISI. It is shown that how the bit error rate is effected by varying with Doppler frequency shift and multipath delays. It is also observed that for low Doppler frequency shift when a text data is encoded before transmission and decoded after reception, we get almost the same data using optimum error control code which was transmitted through faded channels such as Rayleigh and Rician. Hence it is concluded that optimum error control code performs well over convolution code for underwater acoustic channel.

References

1. Baggeroer A (1984) Acoustic telemetry—an overview. IEEE J Ocean Eng 9:229–235
2. Catipovic JA (1990) Performance limitations in underwater acoustic telemetry. IEEE J Ocean Eng 15:205–216
3. Nabeel A, Abdul B (2012) Implementation and analysis of convolution codes using MATLAB. Int J Multidiscip Sci Eng 3(8):9–12
4. Kilfoyle DB, Baggeroer AB (2000) The state of the art in underwater acoustic telemetry. IEEE J Ocean Eng 25:4–27
5. Chitre M, Shahabudeen S, Stojanovic M (2008) Underwater acoustic communications & networking: recent advances and future challenges. Mar Technol Soc J 42(1):103–116
6. Hanson F, Radic S (2008) High bandwidth underwater optical communication. Appl Opt 47:277–283
7. Freitag L, Stojanovic M, Singh S, Johnson M (2001) Analysis of channel effects on direct-sequence and frequency-hopped spread spectrum acoustic communications. IEEE J Ocean Eng 26(4):586–593

8. Bingham JAC (1990) Multicarrier modulation for data transmission: an idea whose time has come. IEEE Commun Mag 28(5):5–14
9. Li B, Zhou S, Stojanovic M, Freitag L, Willett P (2008) Multicarrier communication over underwater acoustic channels with non-uniform Doppler shifts. IEEE J Ocean Eng 33(2): 198–209
10. Chitre M, Motani M (2007) On the use of rate-less codes in underwater acoustic file transfers. In: OCEANS 2007—Europe, pp 1–6
11. Green MD, Rice JA (1997) Error correction coding for communication in adverse underwater channels. The Oceans'97 Halifax, Nova Scotia, Canada
12. Trubuil J, Goalic A, Beuzelin N (2012) An overview of channel coding for underwater acoustic communications" In: MILCOM 2012: IEEE Military Communications conference, Orlando FL, United States
13. Labrador Y, Karimi M, Pan D, Miller J (2009) 'Modulation and error correction in the underwater acoustic communication channel. Int J Comput Sci Netw Secur 9(7):123–130
14. Wu L, Trezzo J, Mirza D, Roberts P, Jaffe J, Wang Y, Kastner R (2012) Designing an adaptive acoustic modem for underwater sensor networks. IEEE Embed Syst Lett 4(1):1–4

Secure Energy-Efficient Clustering Protocol for Wireless Sensor Network

N.A. Sangeetha and G.S. Binu

Abstract Every operation in wireless sensor network requires energy. So energy conservation is a major problem in WSN. As a solution for this problem, a joint optimization between medium access control and network layers can be done using received signal strength as a cross-layer parameter. A special kind of clustering is used for cluster formation, wherein each sensor node calculates its received signal strength from the base station. This parameter also considered for new cluster head selection. Moreover, security is also provided for data protection. Using an NS2 simulator, it is demonstrated that this cross-layer optimization reduces energy consumption and delay and increases network throughput and packet delivery ratio as compared with the existing method.

Keywords Wireless sensor network · Energy conservation
Cross-layer · Clustering · Encryption

1 Introduction

Wireless sensor networks (WSN) are highly dispersed networks of sensor nodes which mainly used to monitor events in a particular environment. The main drawback of wireless sensor network is that sensors are extremely sensitive to energy consumption. So efficient energy consumption makes wireless sensor network more powerful in terms of network lifetime. Mainly, sensors have three major functions such as collecting data, process the collected data, and communicate the processed data. Communication phase leads more amount of energy wastage. So designing an energy-efficient communication protocols is a demanding area in wireless sensor network for diminishing energy utilization.

N.A. Sangeetha (✉) · G.S. Binu
Department of ECE, NSS College of Engineering, Palakkad, India
e-mail: na.sangeetha477@gmail.com

G.S. Binu
e-mail: binu_g_s@reddiffmail.com

Wireless sensor network applications have great importance in nowadays because of its ability to operate in harsh environment or unattended areas [1]. In WSN, clustering can be used to lower the energy consumption by conveniently applying every node's energy and split management duties to every node for supporting the scalability of the management system in heavily deployed sensor networks. Distribution of management tasks in this way utilizes node resources effectively in a large scale WSN [2]. When clustering is applied, whole sensors are not required to enroll in communication, only cluster head in each cluster requires to take parts in transmission–reception process, so grouping of sensor nodes into clusters leads to significant energy savings and prolonged network lifetime. Security is an important concern in wireless sensor network because every communication requires data to be secure. So security criterion such as confidentiality, integrity, availability, and authenticity must be considered for developing a network environment. Due to limited energy of sensor nodes, conventional security methods with large overhead of computation and communication are not possible in WSNs. The cryptographic algorithm called RSA [3] is presently the most used among asymmetric algorithms. Complication of factoring massive prime numbers is the principle used in this algorithm, when number increases, that much of difficulty increases to calculates its factorization, so overall algorithm become very complex, attacker cannot break the algorithm and there by increases its security. RSA is a famous precaution in Internet transactions.

Wireless sensor network system performances were determined by multiple factors, optimization of individual layer often leads to inefficient solution, so joining properties of more than one layer leads to higher performance than their layered counterparts. This is called as cross-layer design. Here, cross-layer interaction between medium access control layer and Network layers are considered. Received signal strength is taken as a cross-layer parameter. Here, encryption and decryption are provided for data protection. Since clustering is applied here, encryption of complete data causes an excess delay in the network and also leads to packet loss. So data transferred between clusters only encrypted here. An RSA algorithm is used for encryption and decryption.

1.1 Related Works

Many algorithms have been designed to choose cluster heads in wireless sensor networks. In [4], a clustering-based data collection protocol called LEACH (low-energy adaptive clustering hierarchy) is presented. This protocol is designed in such a way so as to lessen the total energy expenditure by distributing traffic load among different nodes at different times. This can be achieved by selecting cluster heads and randomly rotating these cluster heads so as to uniformly share traffic load among all sensors. In LEACH protocol, cluster heads were selected in a random manner and do not require any contact with the base station or knowledge of the network for its operation. ALEACH (Advanced LEACH) is a modified type of

LEACH protocol in which nodes create independent decisions without any central innovation [5]. An algorithm for rotating cluster heads for balancing the traffic load is also proposed here. A single-level clustering algorithm is described in [6], where the sensors join individual cluster head on gathering advertisement from them. A hierarchical clustering algorithm is also proposed here where more than one level of clustering is done. There are two types of cluster heads are here called level 1 cluster head and level 2 cluster head. The sensors first sense the data and send it to their corresponding first level cluster heads, and these cluster heads will combined this data and forward it to second level cluster heads and other. This distributed algorithm helps to lower nodes energy decay and then improves network lifespan more than LEACH. To alleviate hot spot problem [7], an unequal cluster-based routing is proposed in [8]. Here, both rotation and selection of cluster heads with higher residual energy are accepted. In this routing, networks are divided into clusters of uneven numbers. By lowering higher relay nodes which are adjacent to base station, reliable energy utilization can be maintained in the network. Akin to LEACH, the process of unequal cluster-based routing also splits into different rounds. The chance of being a master node is rotated among sensors in each round to donate energy usage over the network. Energy-efficient unequal clustering [9] is a disbursed cluster head competitive algorithm, where the cluster head election mainly depends on residual energy of unsettled cluster heads. Moreover, energy-efficient unequal clustering produces clusters of different sizes to weaken the hot spot problem. Nearby clusters of base station have smaller sizes, thus will exhaust low energy throughout the intracluster data processing, and can sustain some more energy for the intercluster relay traffic. In [10], hybrid energy-efficient distributed clustering protocol (HEED) is presented. This is a cluster-based protocol in which cluster head election depends on the residual energy of sensor nodes and nodes join the cluster heads in such a way so as to decrease the whole communication cost of networks. Cluster head selection is done with more number of iteration, and it terminates with uniform cluster head distribution throughout the network. HEED also considers node density, intracluster, and intercluster transmission ranges. Proper limits for each of these criteria are also calculated, and with these bounds, HEED provides better connectivity for the network. In [11], time-controlled clustering algorithm (TCCA) is presented which uses the rotating cluster head election for lowering the network wide energy consumption. Operation of TCCA is done in two phases called cluster setup and steady-state phase. Data collection, aggregation and data transfer occurs in second phase called steady-state phase. Residual energy of nodes is also considered for cluster head election. TCCA differs from other protocols because here the cluster formation is controlled by TTL (Time to Live) messages.

From the literature, it is clear that there are so many algorithms are in wireless sensor networks which reduce energy expenditure of nodes in an efficient way. Residual energy is considered as an important parameter for cluster head selection. Since optimization of single layer is inefficient for wireless sensor networks because its performance depends on multiple factors. In [12], a literature review of various cross-layer protocol is presented. Cross-layer protocols are classified based on the

interaction among different layers such as physical layer, data link layer, network layer, and application layer. The authors showed that wireless sensor network mostly uses the cross-layering protocol which mainly takes the interactions among medium access control layer and network layer. Here also proposed a cross-layering protocol between MAC and Network layer and received signal strength is selected as a cross-layer parameter.

2 System Model

In this section, a wireless sensor network with secure energy-efficient clustering protocol is discussed.

2.1 Clustering-based Protocol

Consider a wireless sensor network subsist huge number of sensor nodes. After deployment of nodes, some clusters are developed in network. The main assumptions taken are

(a) Static nodes are considered in the simulation
(b) Each node will aware of the location of all other nodes
(c) Each node has fixed transmission range

A wireless sensor network consisting of fifty nodes are considered. Clusters are formed by using thesed fifty nodes in which each cluster will contain eight nodes. At first, the sensor nodes divided into clusters. Then, cluster head is elected randomly. Each cluster head in each cluster sends a status signal to all other nodes in the cluster. Transmission of data is carried through these cluster head. Here used a special kind of cluster head selection such as each node will calculate distance from the base station. For calculating distance of each node use

$$D\left(m,n\right) = \sqrt{\left(x_2 - x_1\right)^2 + \left(y_2 - y_1\right)^2} \tag{1}$$

where m and n are two nodes in the network. By using this distance, calculate received signal strength (RSSI) [13] of each node in network by using the formula.

$$\text{RSSI} = A - 10N \log d\left(m,n\right) \tag{2}$$

where RSSI is the received signal strength, A is the received signal power in the distance between two nodes, and N is the path loss index and relates to the background. In practical operation model parameters A and N are depends on some obstacles in environment. So received signal strength of each node is obtained.

During the next cluster head selection phase, this energy is also takes into account in addition to residual energy. Hence, significant energy can be saved in the network.

2.2 Security

Security is an important concern in wireless sensor network. Cluster security is provided for data protection. Once the malicious node become cluster head, an attacker can attack the whole network easily. So making the cluster head secure leads to whole network protection [14]. Here, each cluster head collects the private key of all other nodes and thereby certificate authority before transmitting the data. Any node from the same cluster or intermediate node outside the cluster can share information with cluster head only after the approval of certificate. In this way, any untrusted node tries to take part in communication can be avoided. An RSA algorithm is used here for data encryption and decryption.

3 Simulation Results

For the simulation, 50 sensor nodes are set up in a sensor field of size 1000×1000 m is considered. Simulation is done using network simulator NS2. Sensors are deployed randomly in the network area. After the deployment of nodes, six clusters are formed in which each cluster will contain eight nodes and one node is set as a base station. Each node will produce data packets of length 30 bytes. Basic energy of a node is set to 100 J. Energy consumed in transmitter and receiver circuit for transmitting and receiving one packet is taken as 50 and 29 mw, respectively (Table 1).

Figure 1 is a sensor network with 50 nodes are deployed. Then, six clusters formed in which each cluster contains eight nodes. One node is set as base station. Then, the cluster formation completed. The cluster head for each cluster is selected randomly. Then, data transmission is only permissible through these cluster heads. Then, received signal strength of each sensor node is calculated, and during the second phase of cluster head selection, this parameter is also considered for the

Table 1 Simulation parameters

Parameter	Specification
Number of nodes	50
Packet length	30 bytes
Node initial energy	100 J
Simulation time	30 s
Field	1000×1000 m

cluster head selection. Since node with highest residual energy and received signal strength, more energy will be saved when compared with other protocol.

Figure 2a shows the average remaining energy of the nodes in network. Sensor nodes are provided with battery power of 100 J. Average remaining energy for both cross-layering protocol and cross-layering protocol with encryption are plotted. Both protocol almost used same energy during the simulation. Because after 25 s the cross-layering protocol uses about 94 J and that with encryption uses 96 J energy.
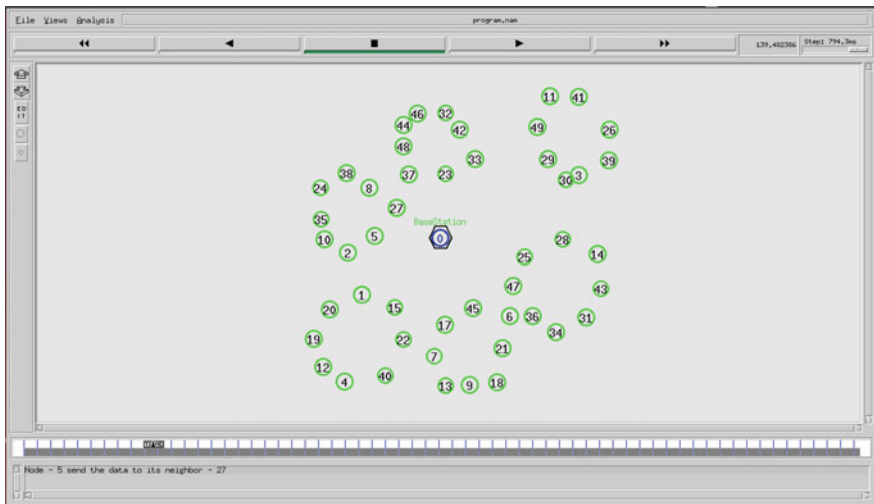
Fig. 1 Network setup
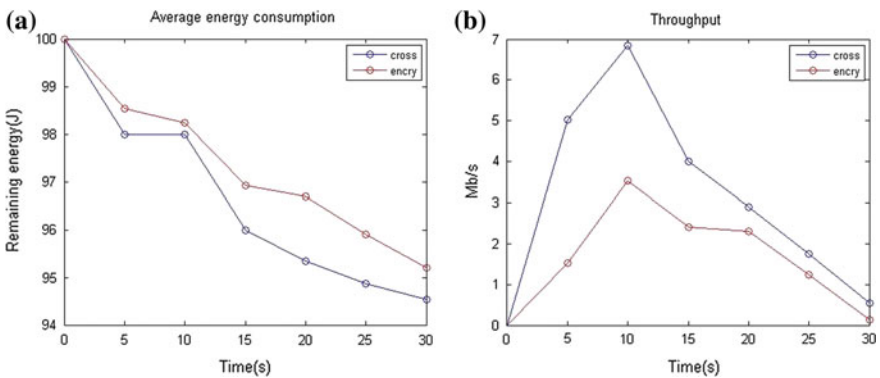
Fig. 2 **a** Average energy consumption, **b** throughput

Figure 2b depicts the throughput of the sensor network for both cases. Total number of bits successfully received at receiver side in each time interval is called as throughput. Initially, throughput got increased due to the efficiency of protocol used. After 10 s, throughput gets decreased due to an increased delay in the network. When security is added to the simulation, more time will consume for encrypting and decrypting the data, thereby an increased delay was experienced in the whole network. Hence, throughput of the network is less when compared with the other protocol. Figure 3a shows the packet delivery ratio obtained from network. It can be defined as the ratio of the number of forwarded data packet to the destination. This shows the level of transported data to the destination.

$$PDR = \frac{\text{Total Number of packet received}}{\text{Total Number of packet send}} \quad (3)$$

When packet delivery ratio increases, that much of better performance have seen in the protocol. Here, initially higher packet delivery ratio obtained due to efficiency of cross-layer protocol and then decreased due to increased delay. But when encryption and decryption are added to simulation, an increased delay is experienced in the whole network, and hence, the PDR is less when compared to the first case.

Figure 3b represents the average delay in network. The average time used by a data packet to reach the receiver side is called as delay. Delay caused by route discovery process and the queue in data packet transmission is also included. Only the data packets that are successfully forwarded to destinations are counted.

$$Delay = \frac{\text{Total send time}}{\text{Total received time}} \quad (4)$$

Here when encryption and decryption are added to the simulation, an increased delay is experienced in the network. So delay will be more when compared with existing protocol.
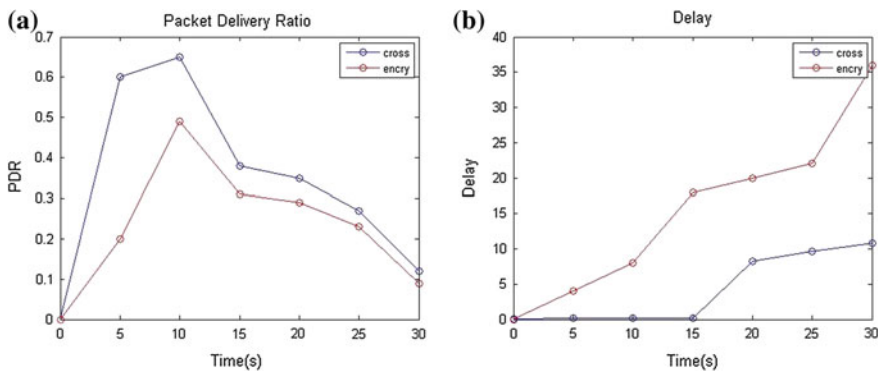
Fig. 3 a Packet delivery ratio, b delay

4 Conclusions

In this work, secure energy-efficient clustering protocol for wireless sensor network is implemented. Here, clustering was performed in an efficient way in which an RSSI-based cluster head election is performed. For electing cluster head, RSSI is also considered as an additional energy parameter in addition to residual energy. The algorithm provides less energy consumption when compared with the existing method. An RSA algorithm is used for data protection. Data will be more secure when encryption and decryption are added to the simulation, but it will cause a decrease in packet delivery ratio and network throughput due to greater delay than the other protocol.

References

1. Abbasi AA, Younis M (2007) A survey on clustering algorithms for wireless sensor networks. Comput Commun 30:2826–2841
2. Sadouq ZA, El Mabrouk M, Essaaidi M (2014) Conserving energy in WSN through clustering and power control. IEEE. doi:10.1109/CIST.2014.7016654
3. Rivets RL, Shamir A, Adleman L (1983) A method for obtaining digital signatures and public-key cryptosystems. Commun ACM 26(1):96–99
4. Sridharan A, Bhaskar K (2004) HEED: a hybrid, energy efficient, distributed clustering approach for ad hoc sensor networks. IEEE Trans Mob Comput 3(4):3–4
5. Ali MS, Dey T, Biswas R (2008) ALEACH: advanced LEACH routing protocol for wireless microsensor networks. In: IEEE conference, pp 909–914, Dec 2008
6. Heinzelman WR, Chandrakasan A, Balakrishnan H (2000) Energy-efficient communication protocol for wireless micro-sensor networks. In: 33rd Hawaii international conference on system sciences
7. Ahn S, Kim D (2006) Proactive context-aware sensor networks. In: European workshop on wireless sensor networks. Zurich, Switzerland, Feb 2006
8. Bandyopadhyay S, Coyle EJ (2003) An energy efficient hierarchical clustering algorithm for wireless sensor networks. In: IEEE Transactions on wireless sensor network, INFOCOM 2003
9. Li CF, Ye M, Chen G, Wu J (2005) An energy-efficient unequal clustering mechanism for wireless sensor networks. In: IEEE international conference mobile adhoc and sensor systems, pp 8, Nov 2005
10. Chen MG, Li C, Ye M, Wu J (2007) An unequal cluster-based routing protocol in wireless sensor networks. IEEE Trans Wirel Netw 15(2):193–207
11. Sangeetha NA, Binu GS (2016) Energy management in wireless sensor networks: cross layer approach brief review. IJARECE 5(4):870–875
12. Selvadurai S, Sukunesan S (2006) An energy-effcient clustering algorithm for multi hop data gathering in wireless sensor networks. IEEE Trans J Comput 1(1):40–47
13. Zheng J, Wu C, Chu H, Xu Y (2011) An improved RSSI measurement in wireless sensor networks. Int J Proc Eng 5:876–880
14. Zhang K, Wang C, Wang C (2008) A secure routing protocol for cluster-based wireless sensor networks using group key management

Simulink-Based Estimation of Spectrum Sensing in Cognitive Radio

Kavita Bani and Vaishali Kulkarni

Abstract Cognitive radio is an upcoming new technology for future wireless generations for its reconfigurability and adapting in nature under environment change. Spectrum sensing (SS) is one of the great important functionalities of cognitive radio to analyze and learn its surrounding radio environment. The main objective of the spectrum sensing is to achieve maximum amount of probability of detection and minimize complexity of the system. Here, there are three spectrum sensing techniques analyzed which are classified based on its transmitter detection such as energy detector (ED), cyclostationary feature detection (CFD), and matched detector (MD). Spectrum sensing techniques are implemented in software MATLAB and Simulink. Results of all these techniques are discussed here with individual FFT spectrum graphs.

1 Introduction

In Mitola's dissertation [1], he has described the term 'cognitive radio' which can adapt to changing conditions and cleverly change its transceiver parameters when needed. Today, however, cognitive radio has become an all-round with broad varieties of techniques which can enable radios for its self-configuration, wireless access, and dynamically accessing the spectrum for a future device midmost interference.

K. Bani (✉) · V. Kulkarni
Mpstme, Mumbai, India
e-mail: kavitabhatu@gmail.com

V. Kulkarni
e-mail: vaishali.kulkarni@nmims.edu

The definition of cognitive radio (CR) according to US Federal Communications Commission (FCC) [2] is a radio which can make change in various transmitter parameters by interacting the environment surrounding. The maximum number of cognitive radio will almost certainly be software-defined radio (SDR), and CR is not required to have software and fixed hardware.

Figure 1 shows spectrum measurement across 900 kHz–1 GHz band of USA. Vacant spectrum is utilized by secondary users. CR is investigated by regulation committee to enable the technology for opportunistic access which is lead to access TV white spaces.

Amount of large parts of TV bands are available by geographically because nowadays analog communication is switched over digitalization. The UK and the USA have adapted the CR model and decided 802.22 standards for utilization of these TV white spaces. So in future worldwide, CR is expecting to become main area in advanced technology [2].

Spectrum utilization is one of the primary mottoes of the CR. By using cognitive radio with MIMO radio, it gives flexibility in terms of carrier frequency, transmit power, channel bandwidth, and multiplexing gain [3]. Spectrum opportunities in opportunistic secondary spectrum allocation (SSA) can be done in spectrum holes or white spaces, where a portion of the band is left completely idle by the primary users such as radar systems or TV channel.

The wireless applications are growing more rapidly and without having space on spectrum, there are not possible wireless telecommunications and wireless Internet services. So available spectrum need to share very efficiently among the users. Primary users (PUs) are having license for using available spectrum while some times it is found that PUs are not using the spectrum which is allocated to them. This case secondary user (cognitive user) can use the PU spectrum without causing interference to PU signal level for particular time interval when PUs are not present.

Fig. 1 Spectrum measurement across the 900 kHz–1 GHz band (Lawrence, KS, USA) [2]
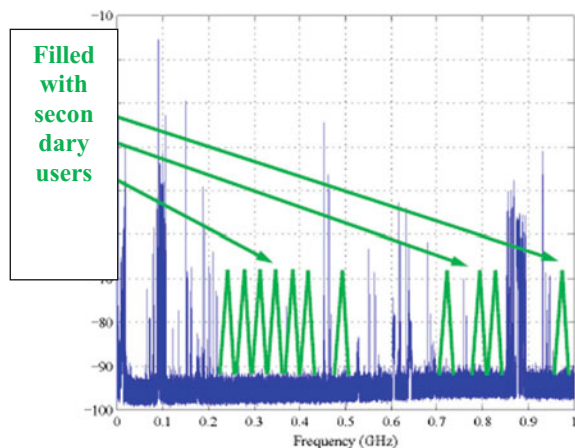
In village areas, where optical fiber is economically not possible to install and where there is nearly no spectrum use. In many village areas, if we find the source of broadcasting TV might be at far away and because available TV service will be not or less. Here in this scenario, we can find out good amount of opportunities for Internet facilities and telecommunication devices by using this vacant or less used spectrum.

To sense the available channel or free channel from its spectrum, three spectrum sensing methods are discussed here such as energy detector (ED), cyclostationary feature detection (CFD), and matched filter or detection (MD).

2 Literature Survey

Geete et al. [4] have explained four primary role of cognitive radio are explained in this paper, spectrum mobility, spectrum sensing, spectrum sharing, and spectrum management. Spectrum sensing techniques such as matched filter, energy detector, and cyclostationary feature detection (CFD) are examined under AWGN, Rician fading, and Rayleigh fading channel environment.

Mehta et al. [5] have discussed that cognitive radio network (CRN) requires the many advanced techniques such as interference management, CR reconfiguration management and cooperative communications, and distributed spectrum sensing. A method which is going to use for spectrum sensing, it should be dependable and primary user should not have problem from secondary user (SU) disturbance. In this paper, advantages and disadvantages of various SS methods are listed in terms of operation, accuracy, complexity, and implementations. Among these all SS methods, energy detector (ED) is very common detection technique because of low complexity and implementation.

Lavanya et al. [6] have implemented based on transmitter detection, Eigen value detection, energy detector (ED), and matched detector (MD) using MATLAB, and results are compared. ED suffers from SNR wall problem; after 4 dB, ED performance is high. While eigenvalue detection is good under low SNR values. For SS approach, this adaptive SS technique reduces the complexity.

Singh et al. [7] have done Simulink models which are used for modeling ED and CFD spectrum sensing techniques. Compared to ED, CFD is better under heavy noise environment, which is shown through the results.

Bagwari et al. [8] have explained spectrum sensing techniques ED, CFD, and MD, in terms of probability of detection (PD), probability of miss detection, and probability of false alarm. ED is no longer good at low SNR values. ED and CFD do not require the prior information about primary users (PUs) while information of PUs is needed in MD. CFD is better than ED and MD, but computation time is large and implementation is also complex.

3 Spectrum Sensing Techniques

Spectrum sensing (SS) is one of the main functions of cognitive radio to acquire knowledge about surrounding environment. The main aim of SS is to acquire maximum amount of probability of detection and reduce the complexity. SS is classified in three different techniques such as energy detector (ED), matched filter or detection (MD), and cyclostationary feature detection (CFD) which is shown in Fig. 2 based on to serve different purpose with advantages and disadvantages.

3.1 Energy Detector (ED)

Energy detection is a basic way of detection of primary user because ED has less calculation and reducing complexities. In energy detection technique, receiver does not require any previous knowledge of primary users' (PUs) signals while in other techniques such as matched filters and other approaches, is needed. This is given by (1).

The purpose of the spectrum sensing is to find correct hypotheses among two, which are described the following,

$$
\begin{aligned}
X(t) &= W(t), && H0(\text{Null Hypotheses when Primary user is absent}) \\
X(t) &= h\,S(t) + W(t), && H1(\text{One Hypotheses when Primary user is present})
\end{aligned}
\tag{1}
$$

where $X(t)$ is received signal by the CR user, $S(t)$ is the primary user's signal, $W(t)$ is the AWGN channel noise, h represents the amplitude gain of the band. H0 is a null hypothesis, which indicates only noise is present, means channel is vacant and secondary user can transmit through the same band for particular time interval. ED is finding primary user based on fast Fourier transform (FFT) [9] which can convert a signal from time domain to a frequency domain and calculates the energy in each frequency of the signal showing as the PSD (power spectral density).

Figure 3 represents the general flow of energy detector [10]. Initially, signal is generated and applied to band-pass filter (BPF) to choice band of frequencies than

Fig. 2 Transmitter based of spectrum sensing techniques
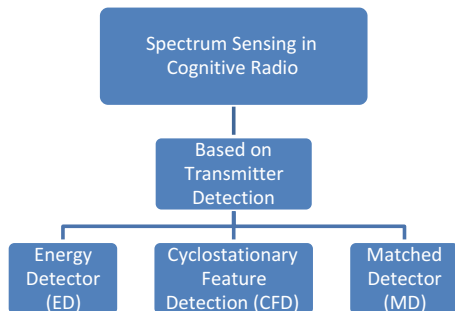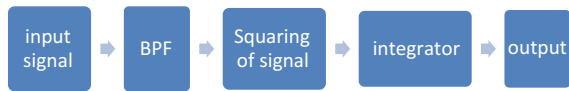
Fig. 3 Energy detector block diagram
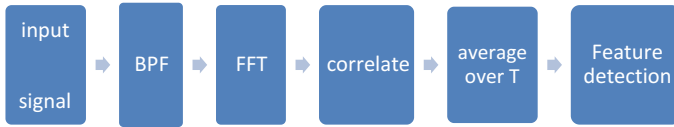
Fig. 4 Cyclostationary feature detection (CFD) block diagram [11]

further done by squaring of the signal and after squaring, signal is integrated through specified time interval to calculate the final received energy. At the end, output of the integrator is compared with a throughput for determining that licensed user is absent or not. The throughput can be a fixed or different depending on the channel conditions.

3.2 Cyclostationary Feature Detection (CFD)

CFD working principle: Based on statistical analysis of licensed user's graph, Cognitive user can find the signal with or without interference of noise. Autocorrelation function is used to extract the information about primary users.

Figure 4 shows block diagram of CFD technique. Input signal is given to band-pass filter to select the specified signal and remove the other not required signals from the band. Selected signal is then converted from time domain to frequency domain by using FFT. Output of FFT is given to the correlation where autocorrelation function is used to correlate signal with itself, by taking the average value of signal compared with fixed value of threshold level for channel.

3.3 Matched Detector (MD)

The matched detector is very efficient technique for spectrum sensing under heavy noise interference. Matched detector required the primary knowledge about the licensed user's signal for its linear filtering operation. Convolution is been applied between received and detector signal which is nothing but a time-shifted property of the reference signal.

Figure 5 outlines the principle of matched detector. The input signal is passed through AWGN channel and further passed through band-pass filter to select the channel. The output $b(n)$ is given as in Eq. (2)

Fig. 5 Matched detector block diagram

Input signal → AWGN → BPF → Matched filter → output
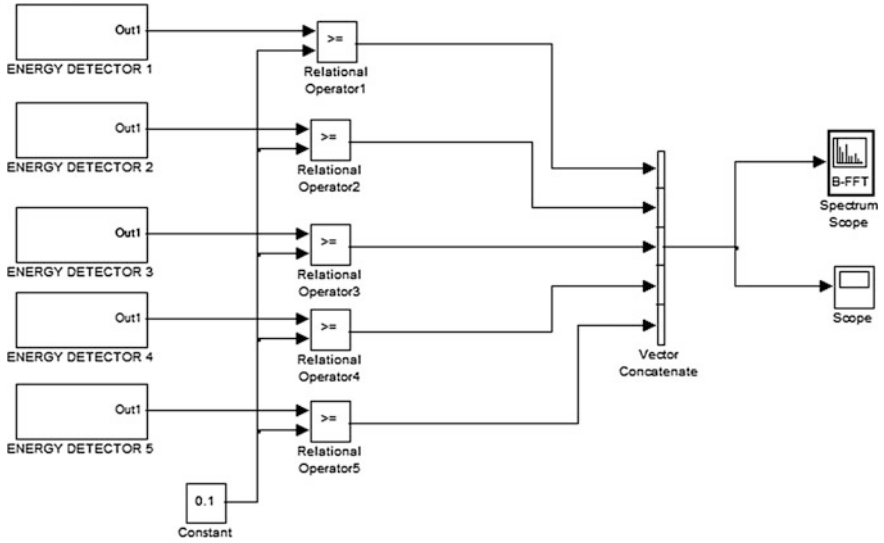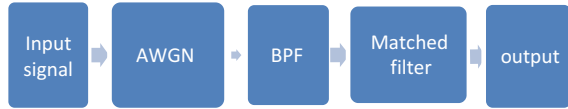
Fig. 6 Energy detector Simulink model for multi (five)-users

$$b(n) = \sum_{k=-\infty}^{\infty} s[k]h[n-k] \qquad (2)$$

where $s[k]$ is the received signal and $h[n-k]$ is the matched detector signal with time shifted.

4 Implementation of SS Techniques

SS techniques are implemented with the help of MATLAB and Simulink software.

4.1 Energy Detector (ED)

Figure 6 shows Simulink model of energy detector. Input signal is generated through random number which is further passed into band-pass filter (BPF) to choose a specific band of frequencies and block the other frequencies. After the BPF, magnitude of the received input signal is squared using absolute (Abs) math function. Then, window integrator is used to integrate the received signal. Integrated signal undergoes rising edge detection, and for that, edge detector is used which can detect rising edge, falling edge, or either edge. Then to generate output signal, relational operator is used to compare the input signal and constant threshold signal (0.1). Output of energy detector is plotted on time scope as well as on FFT spectrum.

4.2 Cyclostationary Feature Detection (CFD)

Figure 7 shows Simulink model of cyclostationary feature detection (CFD). Applied input sine wave signal to AWGN channel. Output of AWGN channel is connected to peak notch filter which can reject very few ranges of frequencies and select majority of all other range of frequencies. Now for quantization of input signal, ADC quantizer is used. Uniform encoder converts quantizer output into integers. FFT is then converted the signal from time domain to frequency domain. At the end, signal is compared with specified threshold level and output is drawn on scope.

4.3 Matched Detector (MD)

In Fig. 8, transmission side, square-root-raised cosine filter is used to shape the pulse while at receiver side, the same matching receiver is used with square-root-raised cosine filter. Rectangular 16-QAM modulation is used
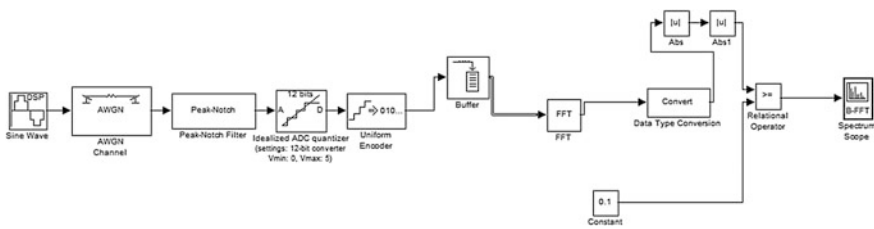
Fig. 7 Cyclostationary feature detection Simulink model
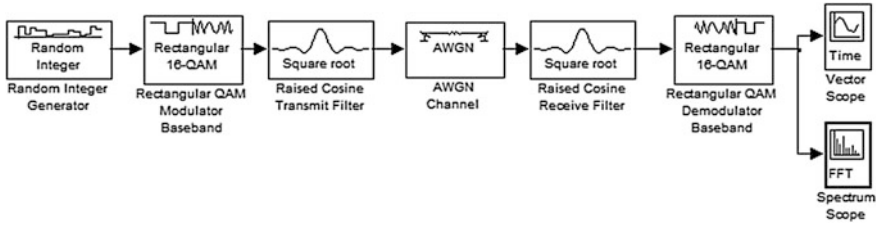
Fig. 8 Matched detector (MD) Simulink model
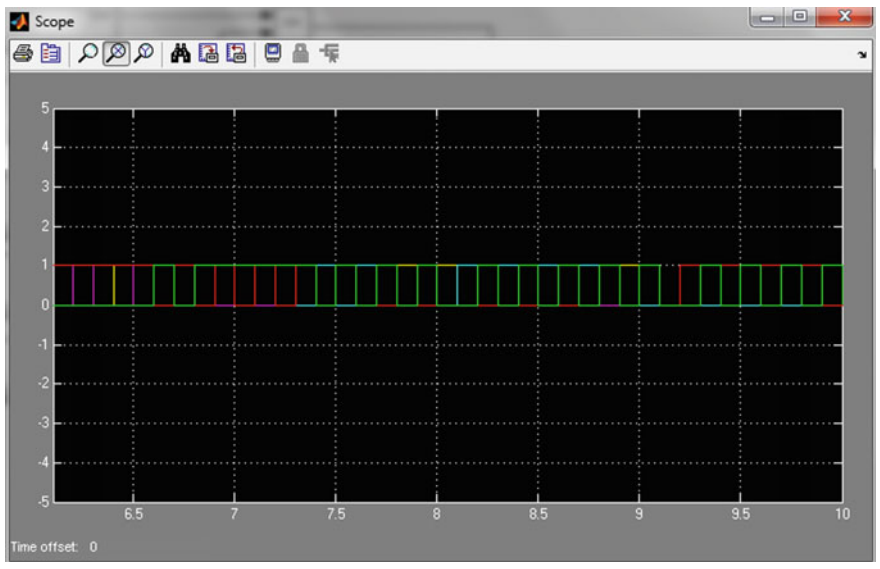
Fig. 9 Scope output of energy detector for multi (five)-primary users

for amplitude and phase both modulation and demodulation. Filter roll-off factor
and group delay can be varied. FFT and vector output are shown in Results section.

5 Results

Figure 9 shows scope output of energy detector for multiuser. Five random users
have been simulated by varying seed of random numbers. It can be seen through the
scope that there are some spectrum holes at time interval, between 9.1 and 9.2,
where all the five primary users are absent.
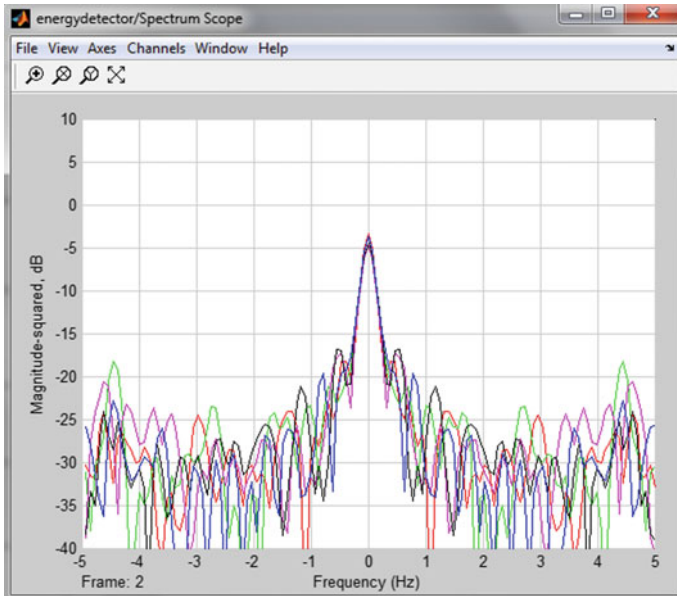
Fig. 10 FFT spectrum output of energy detector for multi (five)-primary users
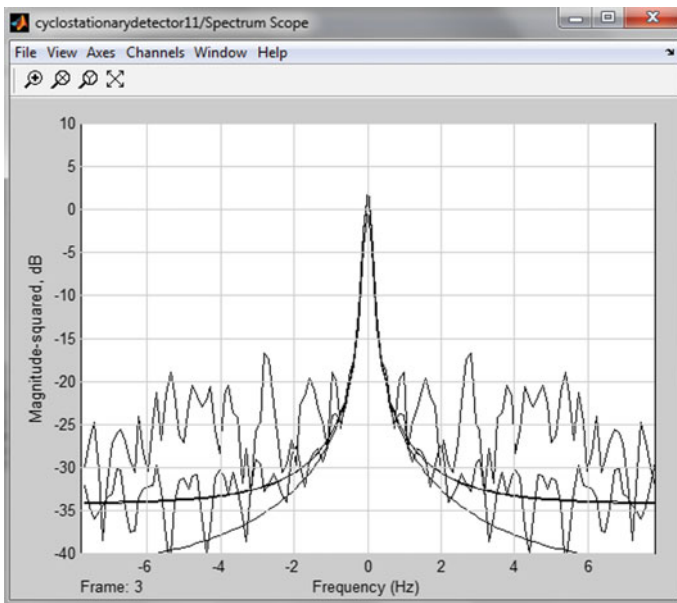
Fig. 11 FFT spectrum output of cyclostationary feature detection (CFD)
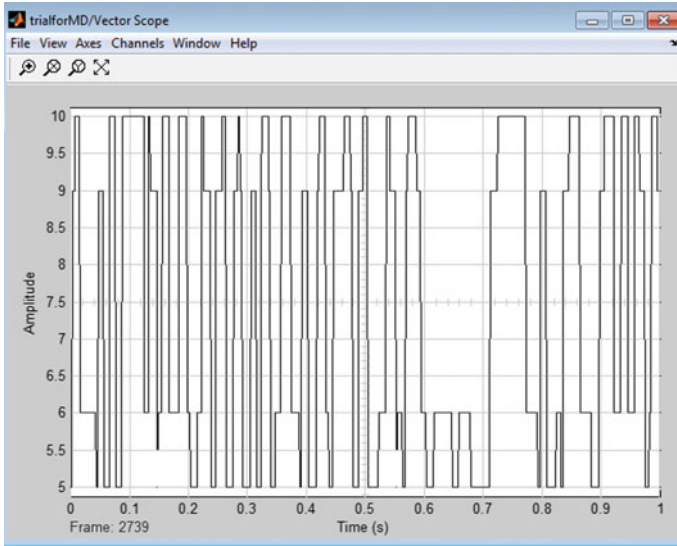
Fig. 12 Vector scope output matched detector (MD)
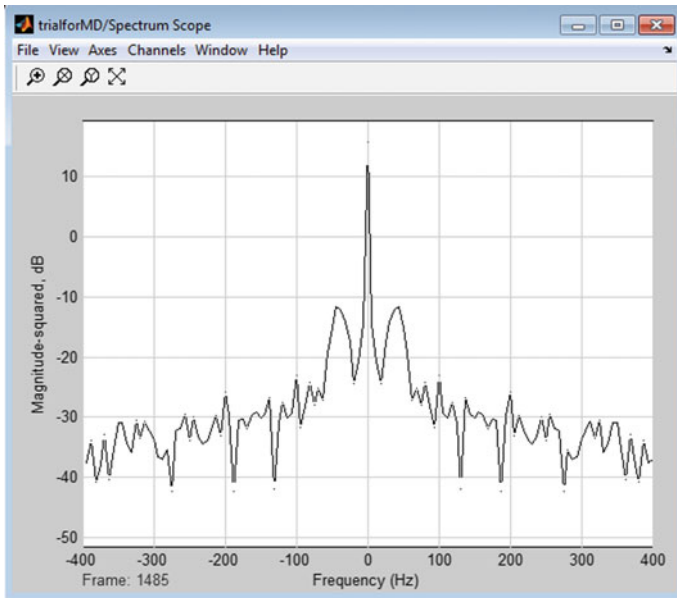
Fig. 13 FFT spectrum output of matched detector (MD)

Figure 10 shows the FFT spectrum output of energy detector for multi (five)-primary users with different five colors, all PU's peaks at −5 dB and frequency is 10 Hz, rest all peaks are noise.

Figure 11 shows the FFT spectrum output of cyclostationary feature detection (CFD). Signal which is shown at peak is 2 dB while rest other peaks are noise. And Fig. 12 shows vector scope output of CFD.

Figure 13 shows the spectrum output of matched detector. At 12-dB highest peak, clear indication of primary user is present.

6 Conclusion and Future Scope

Estimation of different spectrum sensing in cognitive radio has been done in MATLAB and Simulink. Detection of the PU signal is very important in the cognitive radio to decide whether licensed user (primary user) are present or absent in their allocated spectrum depends on the received signal detection.

ED is very common and easy to implement but suffers low at low SNR values, while CFD can use at low SNR values, but processing time of CFD is more compared to others. By having spectrum sensing techniques, energy detector (ED), matched filter or detection (MD) and cyclostationary feature detection (CFD), FFT spectrum scope output of matched detector or filter is very sharped and clear at 12 dB, higher in magnitude compare to other two techniques without having noise disturbances.

In future, spectrum sensing techniques can be implemented by combining adaptively two SS techniques under the environment it is needed.

References

1. Mitola J III (2000) Cognitive radio: an integrated agent architecture for software defined radio. PhD thesis, Royal Institute of Technology (KTH), Stockholm, Sweden
2. Wyglinski AM, Nekovee M, Hou YT (2010) Cognitive radio communications and networks: principles and practice. Elsevier, Amsterdam
3. Haykin S (2005) Cognitive radio: brain-empowered wireless communications. IEEE J Sel Areas Commun 23:201–220
4. Geete P, Motta M (2015) Analysis of different spectrum sensing techniques in cognitive radio network. IRJET 573–577
5. Mehta T, Kumar N, Saini SS (2013) Comparison of spectrum sensing techniques in cognitive radio networks. IJECT Apr-June 33–37
6. Lavanya S, Sindhuja B, Bhagyaveni MA (2015) Implementation of an adaptive spectrum sensing technique in cognitive radio networks. In: 2015 International conference on computing and communications technologies (ICCCT'15). IEEE
7. Bagwari A, Singh B (2012) Comparative performance evaluation of spectrum sensing techniques for cognitive radio networks. In: 2012 Fourth international conference on computational intelligence and communication networks. IEEE

8. Fette B (2006) Cognitive radio technology. Elsevier, Boston
9. Trivedi PM, Aarthy ST, Sharm SK (2014) A comparison of different local spectrum sensing techniques in cognitive radios. IJEDR 2(1). ISSN: 2321-9939
10. Shahzad A et al (2010) Comparative analysis of primary transmitter detection based spectrum sensing techniques in cognitive radio systems. Aust J Basic Appl Sci 4(9):4522–4531 INSInet Publication
11. Singh K, Mahajan R (2014) Comparative analysis of energy detection and cyclostationary feature detection using Simulink. Int J Sci Eng Res 5(5). ISSN 2229-5518

Efficient Pre-distortion Power Amplifiers for OFDM Transmitters

P.R. Bipin and P.V. Rao

Abstract This paper presents a new linearization technique for power amplifiers. Here, ABC, PSO, and modified ABC–PSO algorithms are used to solve the existing problems in pre-distortion of power amplifiers. Algorithms are implemented in MATLAB2014. The performance of the ABC, PSO, and modified ABC–PSO algorithms are compared for the purpose of linearization. From the results, it can be observed that the modified ABC–PSO algorithm has obtained better results by acquiring better linearity and power gain.

Keywords Pre-distorter · Particle swarm optimization · Artificial bee colony High-power amplifier · Wiener model

1 Introduction

Power amplifiers are used in almost all wireless communication systems to improve the strength of the signal need to be communicated so that large distance communication is possible. It also enables the effective reproduction of the transmitted signals. However, the HPAs are habitually deformed and show a tendency to generate the nonlinear outputs, thereby miserably failing to attain the saturation level [1]. In order to rectify the issues due to the nonlinear effects in power amplifier operation, a new method based is proposed here.

P.R. Bipin (✉)
Department of ECE, Ilahia College of Engineering & Technology,
Muvattupuzha, Kerala, India
e-mail: bipinpr@gmail.com

P.V. Rao
RajaRajeswari College of Engineering, Bengaluru, India
e-mail: pachararao@rediffmail.com

From the large variety of linearization methods available in the literature, pre-distortion technique [1, 2] is considered here in this paper to rectify the problems due to deformations. In the new method proposed, the Wiener HPA model is employed to devise the high-power amplifier [3, 4]. The optimization process is carried out by means of different optimization techniques [5–11]. The innovative technique here is simulated with the help of MATLAB2012 and its effectiveness is assessed.

Section 2 discusses the proposed methodology used, Sect. 3 illustrates the results obtained, and Sect. 4 concludes this paper by illustrating the merits of the proposed technique compared to the results for methods for pre-distortion based on other optimization algorithms.

2 Methodology

2.1 Power Amplifier Modeling

Among the various models available to model the power amplifier characteristics, the Wiener model is the most commonly used model. Wiener model is illustrated in Fig. 1, where NL indicates the nonlinearity part and LTI indicates linear time-invariant system.

2.2 Proposed Methodology

In the proposed technique, the modified ABC–PSO [3] is used for improving the power gain of the amplifier employed in OFDM transmitters with less nonlinearities. For validating the performance of developed pre-distortion algorithm, a test setup has been developed as shown in Fig. 2.

The OFDM signal has been generated for transmission, and it has fed to developed pre-distortion amplifier. In the proposed algorithm, generated OFDM signal has been send to Weiner model-based power amplifier.

Fig. 1 Wiener model
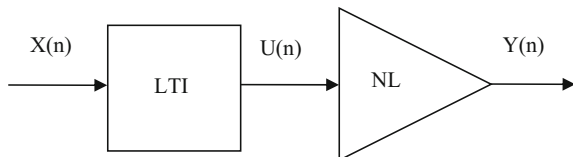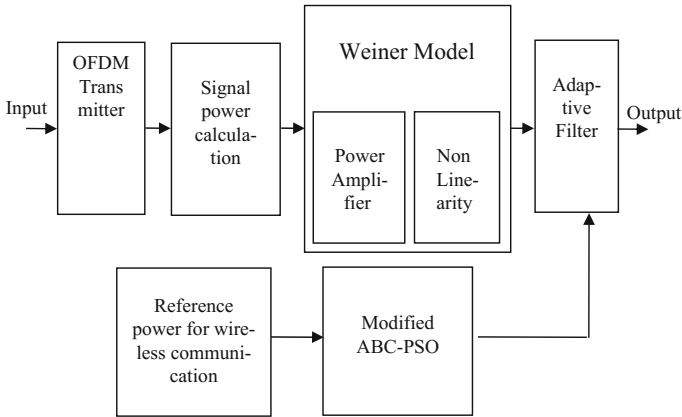
Fig. 2 Proposed methodology

3 Results and Discussion

Developed methodology has been modeled and validated in MATLAB2012, and simulation results has been plotted. In Figs. 3 and 4, generated OFDM signal and transmitted OFDM signal in channel are shown, where x-axis indicates time and y-axis indicates amplitude. From Figs. 3 and 4, it is observed that the generated OFDM signal's power is very low.
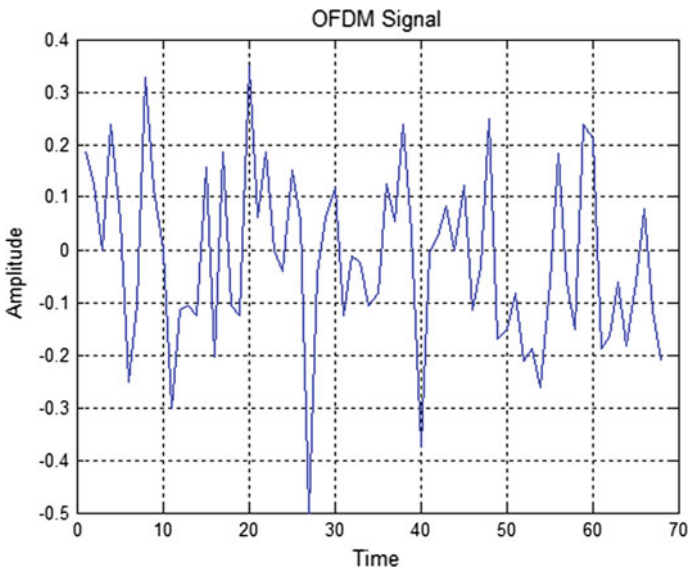
Fig. 3 Generated OFDM signal
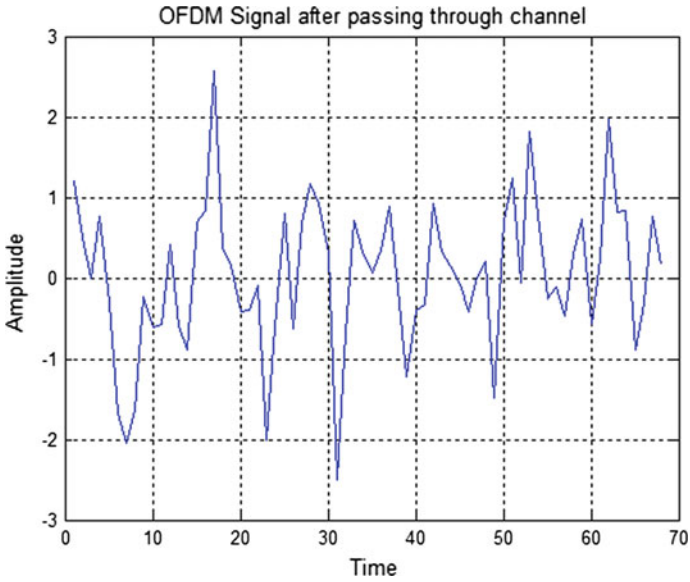
Fig. 4 OFDM signal after passing through channel
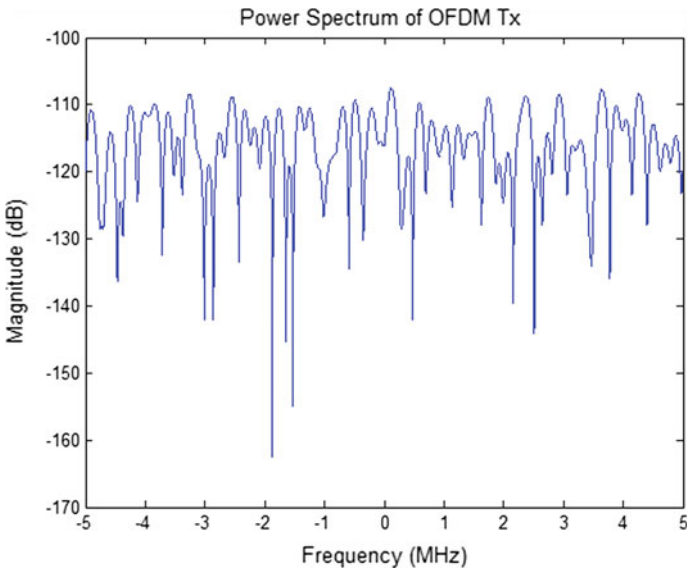
Fig. 5 Power spectrum of generated OFDM signal

For evaluating the power gain of the power amplifier for the OFDM signal, the power spectrum of the OFDM signal is plotted in Fig. 5, and from this, it is observed that the power gain is 110 dB.
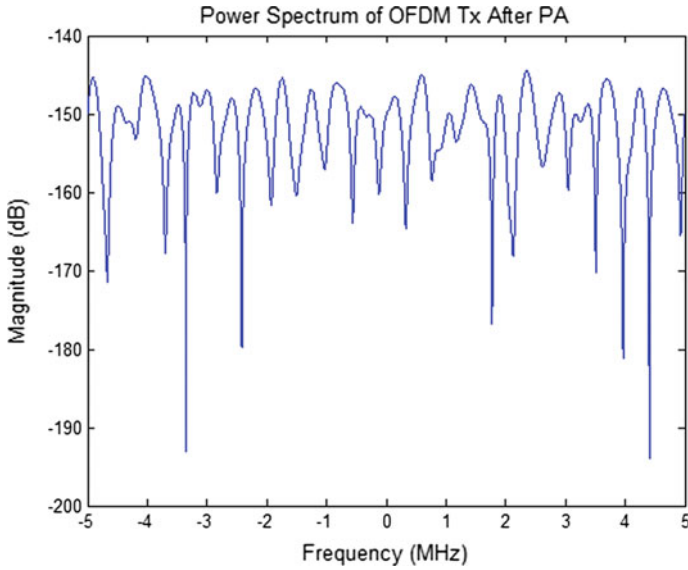
Fig. 6 Power spectrum of amplified signal

The OFDM signal has been amplified with Wiener model-based power amplifier, where nonlinearities are being added. Here, as shown in Fig. 6, the power gain is reduced to −144.5437 dB after amplification.

In Fig. 7, the power spectrum plot of PSO algorithm is shown. With the help of PSO algorithm-based pre-distortion, the power gain is improved to −98.0827 dB. The gain has also been improved with the use of multi-balanced ABC–PSO
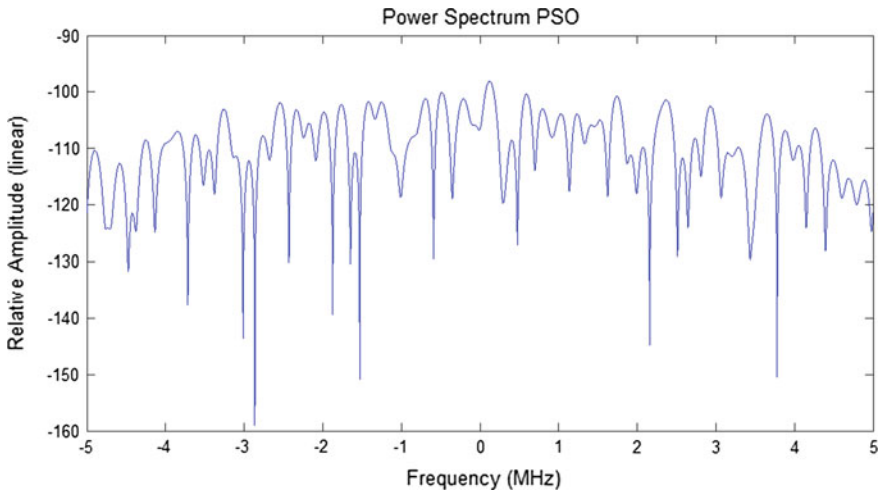
Fig. 7 Power spectrum using PSO algorithm
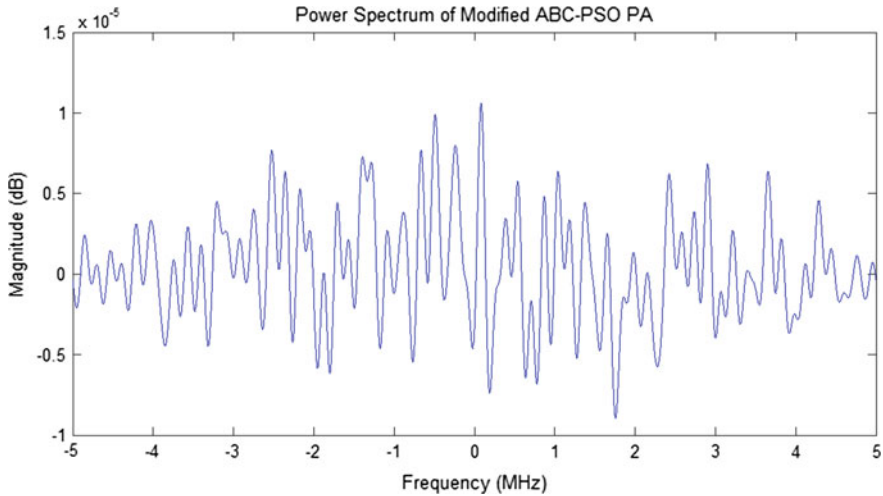
Fig. 8 Power amplifier with pre-distorter

Table 1 Power gain for different pre-distorters

Pre-distortion algorithm	Power gain (dB)
Wiener model PA	−144.5437
ABC algorithm	−101.0499
PSO algorithm	−98.0827
ABC–PSO algorithm	−98.0827
Modified ABC–PSO algorithm	$4.8143E{-}06$

algorithm. The power spectrum for output of developed power amplifier with pre-distortion using modified ABC–PSO is plotted in Fig. 8. From Fig. 8, it is observed that the power gain is improved to $4.8143E{-}06$ dB due to the modified ABC–PSO algorithm.

Table 1 illustrates the power gains produced by the power amplifier for pre-distorters using different algorithms. The proposed pre-distortion using modified ABC–PSO algorithm gives better gain to the input signal and provides less distortion and less nonlinearity.

4 Conclusion

For the development of efficient pre-distortion-based power amplifier, for OFDM transmitters, several algorithms have been used such as ABC, PSO, and modified ABC–PSO algorithms with Wiener model power amplifier. From the results, it can be concluded that the pre-distortion using modified ABC–PSO algorithm has provided the highest power gain by reducing the problems due to nonlinearity.

References

1. Chen S (2011) An efficient pre-distorter design for compensating nonlinear memory high power amplifiers. IEEE Trans Broadcast 57(4):856–865
2. Bipin PR, Rao PV (2016) A new pre-distorter for linearizing power amplifiers using adaptive genetic algorithm. Springer Advances in Intelligent Systems and Computing Series. ISSN 2194-5357
3. Bipin PR, Rao PV (2016) Linearization of high power amplifier using modified artificial bee colony and particle swarm optimization algorithm. Proc Technol J 25:28–35
4. Lin B, Wang Y, Feuvrie B, Dai Q (2010) A novel power amplifier linearization technique based on pre-distortion principle and wiener model. IJAST 22:59–70
5. Bipin PR, Rao PV (2015) Pre-distortion of non-linear power amplifiers using ABC and PSO algorithms. In: IEEE 4th international conference on communication and signal processing. Adiparasakthi Engineering college, Tamilnadu
6. Saleh AA (1981) Frequency-independent and frequency-dependent nonlinear models of TWT amplifiers. IEEE Trans Commun 29(11):1715–1720
7. Bipin PR, Rao PV (2013) Artificial bee colony for pre-distortion of nonlinear power amplifiers. IJEIR 2(6):514
8. Bipin PR, Rao PV (2016) A novel predistorter based on MPSO for power amplifier linearization. In: International conference on emerging trends in Engineering technology and science (IEEE), Kings College of Engineering, Thanjavur, February 24–26, 2016
9. Yang X-S (2009) Firefly algorithms for multimodal optimization. In: SAGA, LNCS 5792, pp 169–178
10. Bipin PR, Rao PV (2013) Artificial bee colony for pre-distortion of non-linear power amplifiers. Int J Eng Innov Res 2(6):514
11. Ding L, Zhou GT, Morgan DR, Ma Z, Kenney JS, Kim J, Giardina CR (2004) A robust digital baseband predistorter constructed using memory polynomials. IEEE Trans Commun 52 (1):159–165

Link Statistical Pattern of Undersea Communication Networks

Sudhir K. Routtay, Anita Mahto, Divya R. Kumar, Karishma Gunesegar and Rumela Choudhury

Abstract In this chapter, we show the statistical patterns of undersea networks found around the world. Though the continental networks have been studied to a large extent, the undersea networks have not been studied for the statistical patterns. In this work, we study 333 different undersea cable networks and analyze their common characteristics. We present the obtained statistical patterns in terms of best-fitting distribution.

1 Introduction

Undersea cable networks are instrumental in the intercontinental information transfer. These networks are deployed along the seabed through the submarines. These networks are extremely important in the modern communication. They interconnect the continents through the high-speed optical communication networks. The Global Internet relies on these networks. Without these networks, the Internet would have been limited only to the continents. For the planning and

S.K. Routtay (✉) · A. Mahto · D.R. Kumar · K. Gunesegar · R. Choudhury
Department of Telecommunication Engineering, CMR Institute of Technology,
32, AECS Layout, Kundalahalli, Bangalore 560037, India
e-mail: sudhirkumar.r@cmrit.ac.in

A. Mahto
e-mail: armo12tc@cmrit.ac.in

D.R. Kumar
e-mail: dydb12tc@cmrit.ac.in

K. Gunesegar
e-mail: karg12tc@cmrit.ac.in

R. Choudhury
e-mail: ruch11te@cmrit.ac.in

dimensioning of these networks, we need proper statistical models which are not available at the moment. Here, we address the main issues of the statistical models for these undersea communication networks.

Statistical models of networks are needed for several purposes such as cost estimation and dimensioning. These aspects are important for the early-stage planning of the communication networks. In [1], dimensioning for optical networks is presented which is helpful in the estimation of the CAPEX and OPEX. In this approach, the networks are planned and dimensioned without the complete information of the networks. In [2], another similar approach is provided for the fast estimation of optical network parameters. Authors in [2] use the methods developed for fast estimation of the network needs and their associated parameters. Both [1 and 2] are effective in the network-related estimations. Network classification and characterization is important for the understanding and engineering of the network dynamics. In [3], optical network classification has been presented on the broad categories such as random, scale-free, small-world, and complex. This classification is based on the measurements of the networks. In [3], the links too have been characterized. In [4], a survey has been presented for the statistical modeling of networks. Authors have presented and analyzed both the past and present trends of network statistics and their importance. In [5], the historical developments of network modeling and their use have been addressed. The author in [5] addresses the very beginning of the network statistical modeling which was started during the time of Euler. A link length statistical model has been developed in [6]. This model does not need the complete information of the network and this model can be estimated from just the node locations (i.e., number of nodes and convex area). In [7], a method based on the circumferential ellipses of the networks has been proposed and expression for convex area has been developed. In [8], the statistical model developed in [6] has been used for the estimation of the link-related parameters of optical networks. In this work, it has been shown that the estimations based on the link statistical models provide much better accuracy than the estimations based on just the average link length [8]. Statistical model for the shortest path lengths of networks is also needed for the estimation of several network parameters such as types of modulation techniques to be used in transparent networks and estimation of the optimal routing paths. In [9], a statistical model for the shortest path lengths of core networks has been developed which does not need the complete information of the networks for its estimation. Topologies used for optical network statistical analysis in [6–9] can be found in [10]. These topologies are mainly the continental networks used for the communication within the continents (only a few cases cover intercontinental communication). In [11], we show the undersea cable networks which are the main intercontinental traffic carriers.

In this work, we measure the link lengths of the undersea communication networks. We collect the data from several publicly available sources and also measure them using realistic tools. We show the statistical patterns of undersea communication networks. The presented patterns are based on the real measurements.

The reminder of this chapter is organized in 4 different sections. In Sect. 2, we present the basics of the undersea communication networks. In Sect. 3, we present the measurement of the network parameters used for this study. In Sect. 4, we show the statistical patterns of the networks obtained from the analysis and modeling. In Sect. 5, we conclude the work with main points of the analysis.

2   Undersea Communication Networks

Undersea communication networks are quite old when compared to the wireless communications. The first transatlantic undersea cable was deployed in 1856 which was used for telegraphic communication between Europe and America. After the invention of the telephone, many such transatlantic and transpacific cables got deployed. Since the 1980s, almost all the undersea cables deployed are optical fiber cables and they provide huge data rates.

These days, undersea communication networks are found across all the oceans. The transatlantic optical communication networks carry highest amount of data when compared with the undersea networks of other oceans. This volume of data carrying capacity is followed by the optical networks of the Pacific and then the Indian Ocean. In this work, we have analyzed 333 different optical networks deployed across different parts of the oceans. We collect these network topologies from TeleGeography and several other sources [11].

In Fig. 1, we show the overview of the main undersea cables deployed along the ocean floors. Most of the cables are found in the Atlantic Ocean followed by the Pacific Ocean and the Indian Ocean. The global bandwidth for the Internet and other forms of communications are mainly contributed by these cables for the intercontinental data transfers.
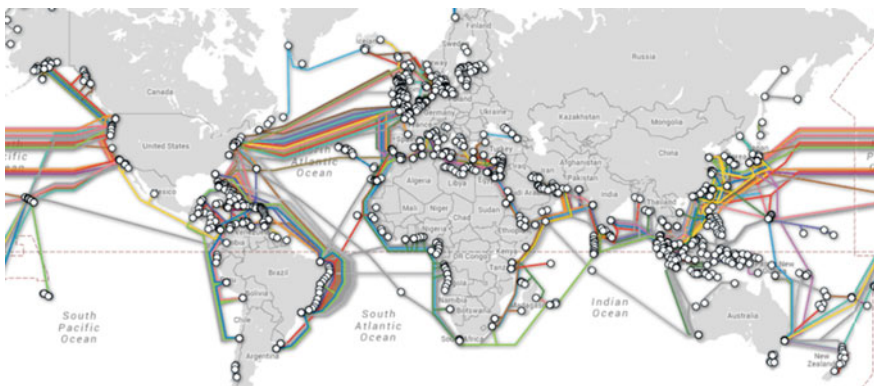
Fig. 1 Overview of the main cable networks found in the world. The small circles represent the cable landing points, and the colored lines are the approximate positions of the optical fiber cables deployed along the seabed

3 Measurement of Network Parameters

We collected 333 different topologies of undersea communication networks and measured their nodal degrees and link lengths. Out of these 333 networks, we found 175 have just two links to connect the points (or gateways or hubs as they are commonly known in networking terminology) across either shores of the oceans. However, there are several networks for which there are several nodes along its path from one end of the cable to the other end. We collected several such topologies to check their statistical patterns.

In [11], some of the link information is provided along with the total length of the fiber used. We measured the link lengths of the fibers using the distance measurements tool Google Earth Professional (version 6.5). The sum of the entire individual link measurements are the same as the total length of the fiber as provided in [11]. In case of some small mismatches, we calibrate all the individual link lengths according to the following expression shown in (1) where l_o and l_m are the rectified and measured link lengths; S_o and S_m are the exact total sum of the link lengths (as given in [11]) and the measured total sum of the link lengths in km (Fig. 2).

$$l_o = \frac{S_o}{S_m} l_m \qquad (1)$$

4 Statistical Pattern Analysis

From the above measurements, we found link lengths of the networks. All the available undersea networks were studied. Majority of the networks (175 out of total 333 we have used for this work [11]) have just two nodes (one each on either side of the coasts). This information is used for the final statistical pattern matching.

Fig. 2 Topology of Flag Europe-Asia undersea cable network. It has 16 nodes or cable landing points in Europe and Asia
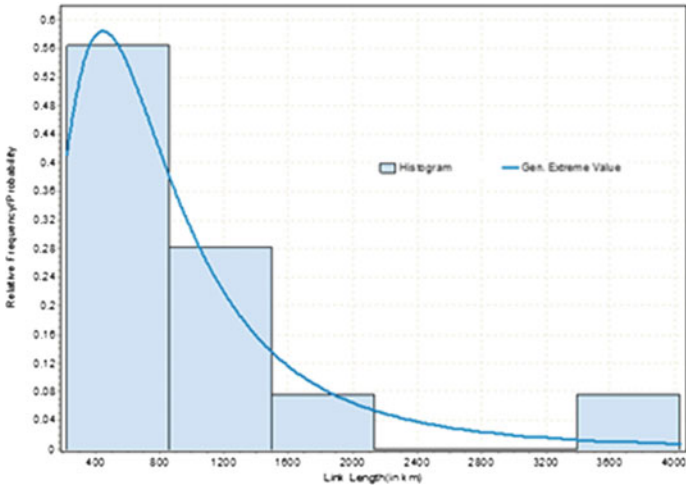
Fig. 3 Link length statistical pattern for ARCOS [11] network. GEV distribution is found to be the best-fitting distribution for this case

We used both MATLAB and EasyFit to analyze the statistical patterns of the link lengths of the undersea communication networks.

We statistically analyzed the link lengths using EasyFit and found that generalized extreme value (GEV) is the best-fitting distribution for individual networks. One of the obtained patterns is shown in Fig. 3. GEV distribution has three parameters: the location parameter (it shows where the samples of the distribution are located), the scale parameter (it is proportional to the standard deviation of the samples), and the shape parameter (it determines the shape of the PDF of the distribution). This distribution is a generalized form of three extreme value distributions. The parameters of this distribution can be estimated from the average value of the samples or the first moment [6].

The statistical validity of the obtained patterns has been assessed using the goodness-of-fit tests. In this case, we choose Kolmogorov–Smirnov statistic (KSS) to asses these statistical patterns. GEV distribution provides smallest average KSS for the networks assessed in this study.

5 Conclusions

In this work, we collected the topological data of 333 different undersea optical networks. We measured their link lengths. Based on the measured data, we did the statistical analysis of the network parameters. For the link lengths of the undersea cable networks, we found a generalized extreme value pattern which can be estimated from the average link lengths. Overall, the obtained patterns are statistically valid and significant for the characterization of the networks.

References

1. Korotky SK (2004) Network global expectation model: a statistical formalism for quickly quantifying network needs and costs. IEEE/OSA J Lightwave Technol 22(3):703–722
2. Labourdette JF, Bouillet E, Ramamurthy R, Akyamac AA (2005) Fast approximate dimensioning and performance analysis of mesh optical networks. IEEE/ACM J Trans Netw 3(4):906–917
3. Routray SK, Pavan C, Sahin G, Rocha JRF, Pinto AN (2015) Statistical analysis and modeling in optical transport networks: characterization and application. In: Proceedings of conference on telecommunications (ConfTele), Aveiro
4. Goldenberg A, Zheng AX, Fienberg SE, Airoldi EM (2010) A survey of statistical network models. Found Trends Mach Learn 2(2):129–233
5. Fienberg SE (2012) A brief history of statistical models for network analysis and open challenges. J Comput Graph Stat 21(4):825–839
6. Routray SK, Morais RM, da Rocha JRF, Pinto AN (2013) Statistical model for link lengths in optical transport networks. IEEE/OSA J Opt Commun Netw 5(7):762–773
7. Routray SK, da Rocha JRF, Pinto AN (2014) Estimating the parameters of optical transport networks from their circumferential ellipses. In: Proceedings of IEEE International Conference on Telecommunication (ICT), Lisbon, Portugal, pp 119–123
8. Routray SK, Sahin G, da Rocha JRF, Pinto AN (2014) Estimation of link-dependent parameters in optical transport networks from statistical models. IEEE/OSA J Opt Commun Netw 6(6):601–609
9. Routray SK, Sahin G, da Rocha JRF, Pinto AN (2015) Statistical analysis and modeling of shortest path lengths in optical transport networks. IEEE/OSA J Lightwave Technol 33 (13):2791–2801
10. Reference Optical Networks (2015) http://www.av.it.pt/anp/on/refnet2.html
11. TeleGeography Undersea Communication Networks. http://www.submarinecablemap.com/#/

Cooperative Spectrum Sensing in CR-VANET with Small-Sized Data Sets

S. Lakshmi Nandan and T. Sudha

Abstract The advent of cognitive radio (CR) technology in intelligent transportation system (ITS) paved the way for new advancements in vehicular ad hoc network (VANET) communication. The prompt sensing of spectrum is an important issue in CR-VANET minimizing the interference to primary user (PU). In real time, less time means small data. The employment of cooperative sensing along with clustering in CR-VANET mitigates this problem. In this chapter, a clustering scheme using limited data sets is used. Due to the dynamic topology of VANET, a clustering algorithm is used considering node degree, speed and location. The simulation results show that the proposed clustering detects the PU signal in a relatively low SNR condition compared to single-user spectrum sensing.

Keywords Cognitive radio (CR) · Vehicular ad hoc network (VANET)
Cognitive radio vehicular ad hoc network (CR-VANET) · Spectrum sensing
Clustering · Cooperative spectrum sensing

1 Introduction

Nowadays the increasing vehicles on road led to the need for improving road safety and in-vehicle entertainment in vehicular communication. Many applications are developed in tune with this rising demand. These include safety and traffic monitoring, collision avoidance, vehicle to vehicle communication [1]. These applications in vehicular communication are effectively met by vehicular ad hoc network (VANET). VANET is a network of vehicles formed either between vehicles or between a vehicle and road side unit for increasing safety. The dedicated short

S. Lakshmi Nandan (✉) · T. Sudha
Department of Electronics and Communication,
NSS College of Engineering, Palakkad, India
e-mail: slakshminandan@gmail.com

T. Sudha
e-mail: sudhat@nssce.ac.in

range communication (DSRC) is used for point-to-point communication between vehicle and road side unit [2–4]. The channels in 5.9 GHz are reserved for this communication. The total available 70 MHz bandwidth is split into one control channel and six service channels allotting 10 MHz bandwidth for each. The IEEE 1609 Dedicated short range communication group is developing wireless access in vehicular environment (WAVE) supporting communication in dynamic environment based on IEEE 802.11p. The rising number of vehicles results in overcrowding mainly in urban area for communication causing spectrum congestion. Thus, the available bandwidth is not enough to meet the consumer needs. This led to the advent of cognitive radio where the licensed spectrum band is monitored.

When the absence of licensed user (primary user) in a band is detected, the vacant band is opportunistically utilized by secondary user (cognitive user). The secondary users need to vacate the band when a PU is detected. The vacant spectrum in the PU band is known as spectrum holes. The application of cognitive radio (CR) is well utilized in vehicular ad hoc network (VANET) [5]. The high mobility of vehicles results in dynamically varying topology in VANET. The spectrum holes are identified by spectrum sensing techniques, also known as signal detection. So the spectrum sensing should be fast for PU detection. The performance of single-user spectrum sensing is affected by effects of multipath fading, shadowing etc. To mitigate this, cooperative spectrum sensing is used. The sensing results of individual nodes are forwarded to data fusion centers (FC). For efficient organization of CR network, the vehicles are grouped into small groups called clusters. This helps in formation of dynamic groups supporting QoS and for efficient routing of nodes [6].

The CR is a fast emerging area of interest. In [7, 8], the functionality and behavior are explored. In [8], the flexibility of software radio in CR is enhanced through a Radio Representation Language. CR application is exploited effectively in VANET. According to the works in [9, 10] the vehicles nowadays are a platform for computing, communications (vehicle to vehicle, vehicle to infrastructure) with sensing abilities. Thus, CR concepts are incorporated with VANET for improving vehicular communication performance. In [11], the author studies the requirements and several challenges in this field to understand how spectrum sensing is fundamental in CR network. The wireless and mobile environment poses limitation to the spectrum sensing, and these aspects together with the cooperative spectrum sensing are studied in [12]. It compares the stand-alone sensing with cooperative sensing to conclude the former is less accurate as latter considers multipath fading, shadowing and hidden terminal problem. Another work [13] considers the CogV2V framework giving three contributions. First research explains a lightweight cooperative sensing to exchange sensing information in vehicles and detect spectrum holes along the path. The latter research leverages the vehicle mobility and how each vehicle uses received signal information to decide in advance the channel to use in future location. Third research proposes a framework integrated on existing IEEE 802.11p. The PU presence is detected using the spatial and temporal correlation of received signal. Apart from this, signal correlation is used in [14] across different bands for channel occupancy estimation. Single-user spectrum sensing is not only

inaccurate but also unreliable. This is mitigated by employing a collaborative spectrum sensing in [15] which uses belief propagation method. Here, each vehicle combines its own belief of existence of PU along with beliefs of other vehicles to generate a new belief. In [16], the limitation of spectrum sensing using limited samples is solved using quickest detection where the change of two different random variables with shortest delay is found. In [17], a cluster-based scheme uses cluster heads to collect sensing data from vehicles in different cluster, but fails to use the shadowing effect which differs from vehicle to vehicle due to dynamic environment. The spectrum sensing method in [18] considers the single-user sensing with limited samples from primary user, to reduce sensing time. This research extends the aspects of sensing to VANET application.

This chapter aims to understand the performance of CR-VANET when cooperative spectrum sensing is used by clustering the vehicles over a Rayleigh fading channel. A density-based clustering algorithm is used by vehicles performing spectrum sensing individually using limited data sets. Limited data set helps to perform sensing in limited time.

The remainder of this chapter is organized as follows: In Sect. 2, the system model is described where the vehicles in a network are clustered to perform spectrum sensing. The proposed algorithm for clustering and spectrum sensing algorithms are explained in Sect. 3. Section 4 describes the cooperative spectrum sensing process in CR-VANET. Finally, results are explained in Sect. 5.

2 System Model

This section describes a cognitive radio vehicular ad hoc network (CR-VANET) system model. The vehicles in the network are clustered into groups. Spectrum sensing is performed among the groups, and decision regarding the spectrum availability is taken at the fusion center, i.e., the road side unit (RSU). There are two communication links; one between cluster head and cluster member, and the other between cluster head and RSU.

The CR-VANET configuration consists of vehicles equipped with CR devices. The secondary users (CR users) monitor the licensed spectrum band continuously to find vacant spectrum band. These available spectrum holes are allotted to the secondary users. The cognitive users use this PU spectrum to transmit data among other cognitive users. As shown in Fig. 1, the vehicles moving on the road communicate with neighbors and with the RSU (fusion center) in that area. The RSU controls the vehicles within the transmission range. To improve the sensing performance, the vehicles in an area are grouped into clusters according to a clustering algorithm. There is a cluster head assigned to each cluster. Each vehicle in a cluster communicates with the corresponding cluster head. The CRs (vehicles) individually sense the spectrum. The spectrum availability information is then forwarded to the corresponding cluster head. This information is passed on to corresponding cluster heads via cluster head–cluster member link (ch-cm link). The cluster heads in an
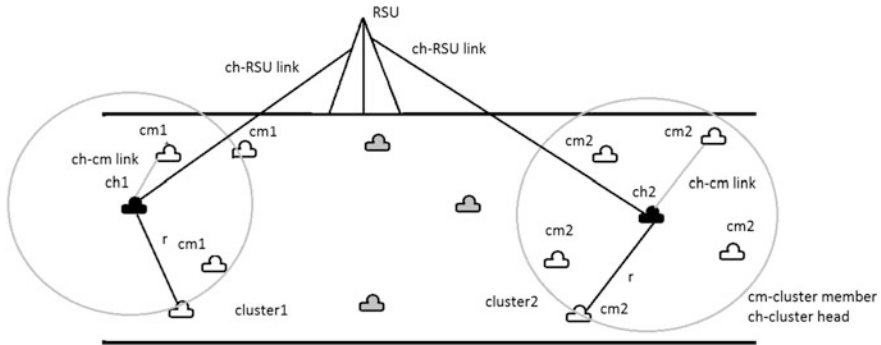
Fig. 1 System model for CR-VANET

area forward the sensed information in each cluster to the fusion center (FC), i.e., RSU in that area via cluster head-RSU link (ch-RSU link) [19]. The sensed information contains the vacant bands of PUs. According to logical AND operation, the collected data from the cluster heads in the area is processed at RSU to obtain final spectrum availability information. The RSU allocates these vacant bands to the needed CR users.

2.1 Clustering in VANET

Clustering is done in VANET for efficient routing and supporting QoS. The vehicular nodes are grouped into clusters each with a cluster head that coordinates and manages the cluster.

A density-based clustering algorithm is used considering the degree, speed and location of each node [20]. Two parameters are considered for clustering:

- The maximum radius of neighborhood.
- Minimum number of points in neighborhood to form a cluster.

Periodic messages containing position data of vehicles are transferred among vehicles. The cognitive receivers set a transmission range for each considering the maximum and minimum location data of vehicles at an instant. The nodes (secondary/cognitive radio users) within this transmission range forms a cluster. For coordinating, a cluster head is elected. As shown in Fig. 2, transmission range is represented by r and $r1$, $r2$, $r3$ the distances of cluster head from the members. For a vehicle from a distance r' to cluster head is greater than r, hence not a member of this cluster.

Stability of a cluster poses a challenge in a dynamic environment. For a stable cluster, the number of clusters formed in a network should be minimal. The goal of clustering is to minimize the effect of dynamically changing topology, considering the node location, speed and degree.

Fig. 2 Clustering of vehicles according to density-based clustering algorithm
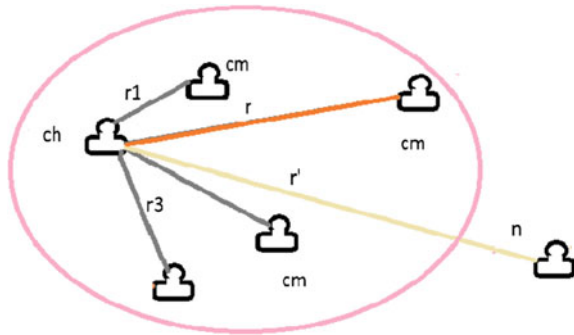
2.2 Spectrum Sensing

Here a spectrum sensing algorithm is used for abrupt detection of PU as explained in [24]. This is similar to collecting of small data sets of signal. A cumulative spectrum sensing process considering small data sets is used. Each vehicle is incorporated with a receive antenna to sense the PU presence. The continuous received signal with unknown channel is $x(t)$

$$x[n] = x(nT_s) \tag{1}$$

represents the received signal sample with T_s the sampling interval.

The presence and absence of PU is represented by the hypothesis given below. H_0 representing the absence of PU and H_1 the presence of PU

$$H_0 : x[n] = w[n] \tag{2}$$

$$H_1 : x[n] = s[n] + w[n] \tag{3}$$

where $w[n]$ is additive white Gaussian noise, independent identically distributed with zero mean and variance σ_n^2. $s[n]$ is the sample receive signal with signal distribution unknown and noise considered white. To evaluate the performance of the CR system, two probabilities are considered: probability of detection of primary user, P_d defined by hypothesis H_1 and probability of false alarm, P_{fa}.

The samples collected from the PU is converted to sample sensing segment, $\Gamma_{x,i}$ with N_{tot} data size.

$$\Gamma_{x,i} = \left\{ \mathbf{x}_{(i-1)N+1}, \mathbf{x}_{(i-1)N+2}, \ldots, \mathbf{x}_{(i-1)N+N} \right\} \tag{4}$$

$$\mathbf{x_i} = [x[i], x[i+1], \ldots, x[i+L-1]]. \tag{5}$$

N represents sensing vectors (sample size), and L (smoothing factor) represents consecutive output samples in each sensing vector.

$$N_{\text{tot}} = N + L - 1. \tag{6}$$

Sample covariance matrix is calculated from the observed sampled signals.

$$\mathbf{R}_x = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x}). \tag{7}$$

Assuming sample mean to be zero, sample covariance matrix is simplified as

$$\mathbf{R}_x = \frac{1}{N} \sum_{i=1}^{N} x_i \bar{x}_i. \tag{8}$$

$\frac{N}{L}$ represents number of observations per variable is small or less than one. In such a condition, estimation of covariance matrix is considered [16].

2.3 Rayleigh Fading Channel

The fading model for the primary signal received is given by,

$$f(r) = \frac{2r}{\Omega \Gamma} \exp\left\{ \frac{-r^2}{\Omega} \right\}, \quad r \geq 0 \tag{9}$$

where Ω is the spread of distribution showing the average power gain for channel and Γ is the gamma function [21].

3 Clustering and Spectrum Sensing Algorithm

3.1 Clustering Algorithm

The clustering algorithm in this chapter considers the node speed, degree and location to form stable clusters according to Algorithm 2. The vehicles in each cluster perform spectrum sensing individually, collected by FC to make the decision about spectrum availability.

The vehicles in a cluster are managed by a cluster head. The slowest vehicle in the cluster is elected as the cluster head [18]. For a vehicle to be stable neighbor of the cluster, the distance of vehicles from cluster head should be at a threshold distance denoted as epsilon,

$$\begin{aligned} \varepsilon = &((\text{prod}(\max(x) - \min(x)) * k \\ &* \text{gamma}(.5 * n + 1)) / (m * \text{sqrt}(\text{pi}.^{\wedge} n))).^{\wedge}(1/n) \end{aligned} \tag{10}$$

here x denotes the node location array containing positions of all vehicles, and n denotes the number of vehicular nodes.

The Eq. (10) represents probability distribution function of a circular uniform distribution. Also the relative speed between the vehicles should be less than or equal to a threshold velocity [22]. The vehicles outside the cluster are considered in the formation of next cluster.

Algorithm 1: Density-based clustering algorithm

1: Epsilon = min distance for vehicles to be considered
2: Clusterset = empty
3: for each vehicle in group/road s
4: if vehicle not in any cluster
5: clusterHead = vehicle
6: related_neig = neighbors with distance < epsilon
7: depth_related_neighbors = neighbors of vehicles in cluster set with distance < epsilon
8: Clusterset(i) = clusterHead + related_neighbors + depth_related_neighbors
9: $i = i + 1$
10: continue

3.2 Spectrum Sensing Algorithm

Since the data set considered is small, the sample covariance matrix estimator is a poor estimator to find statistical relationship between the samples. Hence oracle-approximating shrinkage (OAS) estimator is considered which also minimizes the mean square error.

$$\hat{\varepsilon} = (1 - \rho)\widehat{R_x} + \rho\hat{F} \tag{11}$$

$\hat{F}$ is the shrinkage target and ρ is the shrinkage coefficient.

$$\hat{F} = \frac{\text{Tr}\left(\widehat{R_x}\right)}{L}I \tag{12}$$

where I is an L dimensional unitary matrix.

Assuming samples to be independent and identically distributed Gaussian random variable with covariance matrix R_x. But since R_x is difficult to obtain hence ρ is found by repetitive iterations.

$$\rho_{\text{OAS}} = \min\left(\frac{\left(1 - \frac{2}{L}\right)\text{Tr}\left(\widehat{R_x}\right)^2 + \text{Tr}^2\left(\widehat{R_x}\right)}{\left(N + 1 - \frac{2}{L}\right)\left[\text{Tr}\left(\widehat{R_x}\right)^2 - \frac{\text{Tr}^2\left(\widehat{R_x}\right)}{L}\right]}\right). \tag{13}$$

Finally estimated covariance matrix,

$$\hat{\varepsilon}_{\text{OAS}} = (1 - \rho_{\text{OAS}})\widehat{R_x} + \rho_{\text{OAS}}\hat{F}_{\text{OAS}} \tag{14}$$

For detection, eigen values of estimated covariance matrix are considered. The ratio of max–min eigen values (MME) is calculated from the values $\lambda_1 \geq \lambda_2 \geq , ..., \lambda_L$

$$T = \frac{\lambda 1}{\lambda L}. \tag{15}$$

This ratio is further used for calculating the detection metric, Q_N which is the average of T values for all samples collected

$$Q_N = \frac{1}{N}\sum_{k=1}^{N} T_k. \tag{16}$$

As discussed in [23], the solution of CUSUM test is used that minimizes the delay and algorithm for non-Bayesian quickest detection. It calculates the total stopping data samples detecting PU signal. In case of harsh environment, total stopping data samples are equal to the sample size.

The detection metric is compared with a threshold value. Since the information of signal is unknown, the threshold is not set based on probability of detection (P_d), but based on probability of missed detection (P_{fa}). The estimated covariance matrix is equivalent to sample covariance matrix with a large data set, N_{eq}. The distributions are similar. So the final detection metric, Q_N follows a Gaussian distribution $N(\mu, \sigma)$. The multiple T values are generated from the advantage of sample and estimated covariance matrix. Since sensing is a cumulative process, the variance of random variable is reduced by a factor N to σ. The pdf of H_0 follows a Gaussian distribution. The whole detection process is according to Algorithm 2 [24].

Algorithm 2: Spectrum sensing with limited data samples

1: γ is set (decision threshold)
2: $\Gamma_{x,i}$ the sensing segment
3: while $k \leq N$ do
4: calculate sample covariance matrix, $\widehat{R_x}$
5: find the shrinkage target, $\hat{F}$ and shrinkage coefficient, ρ
6: from above find the estimated covariance matrix, $\hat{\varepsilon}$
7: find the maximum minimum eigen values of $\hat{\varepsilon}$ (MME) to find ratio, T
8: find detection metric, Q_N and compare with γ
9: if $Q_N > \gamma \rightarrow$ PU present
10: else if $Q_N < \gamma \rightarrow$ PU absent
11: end
12: increment k, till $k > N$
13: end while

The false alarm probability transformed to standard Gaussian distribution,

$$P_{\text{fa}} = P\left(\frac{Q_N - \mu}{\sigma} > \frac{\Upsilon - \mu}{\sigma}\right) = Q\left(\frac{\Upsilon - \mu}{\sigma}\right). \tag{17}$$

Threshold is obtained as,

$$\Upsilon = \mu + Q^{(-1)}(P_{\text{fa}})\sigma. \tag{18}$$

From (16) it is clear that the threshold is not affected by noise power or environmental changes.

4 Cooperative Spectrum Sensing with Small Data Sets

The spectrum sensing in VANET is performed individually by vehicles in each cluster. The sensing result is forwarded to the corresponding cluster heads. At the cluster head, the maximum of the sensing results of vehicles in each cluster is found, i.e., aggregation of results. The results are forwarded to the RSU by the cluster heads. RSU is the FC where the final spectrum sensing result is calculated. A logical AND rule is used that finds the spectrum availability information. Whenever a channel is needed, the vehicle sends a request to the RSU. The RSU sends spectrum availability information to the cluster heads in its transmission range. The cluster heads forward the information to the vehicles within each cluster. This information is passed by RSU to the vehicles. This information can be used as a history for the CR users in future [25].

5 Results and Analysis

Simulation results are obtained in MATLAB for single-user spectrum sensing and cooperative spectrum sensing in CR-VANET with clustering are presented. The PU signal considered is a TV (DTV) signal captured at Washington, DC, USA. An Additive white Gaussian noise is used. The smoothing factor, $L = 32$ and probability of false alarm, $P_{\text{fa}} = 1\%$ is considered. A total of 100 sample size is used, i.e., 131 total data. Table 1 presents the parameters used for simulation.

In CR-VANET, a number of vehicular nodes are considered. Each with a receive antenna to capture PU signal and a transmit antenna to send the spectrum-sensed results to RSU.

Figure 3 shows the performance of single-user spectrum sensing, i.e., non-cooperative spectrum sensing. A probability of detection against SNR in dB is plotted. Hundred percentage probability of detection is achieved for an SNR of −5 dB. In case of VANET, the effect of channel fading and the dynamic environment highly deteriorate the sensing performance, which can be improved by cooperative sensing. The clustering further improves the sensing performance.

Table 1 Simulation Parameters

Parameter	Value
SNR range	−20:20 (in dB)
Network area for clustering	20 × 20 (in m) (two dimensional)
Node degree	100 vehicles
Node speed	40:60 (in km/h)
Node location	20 × 20 (in m) area
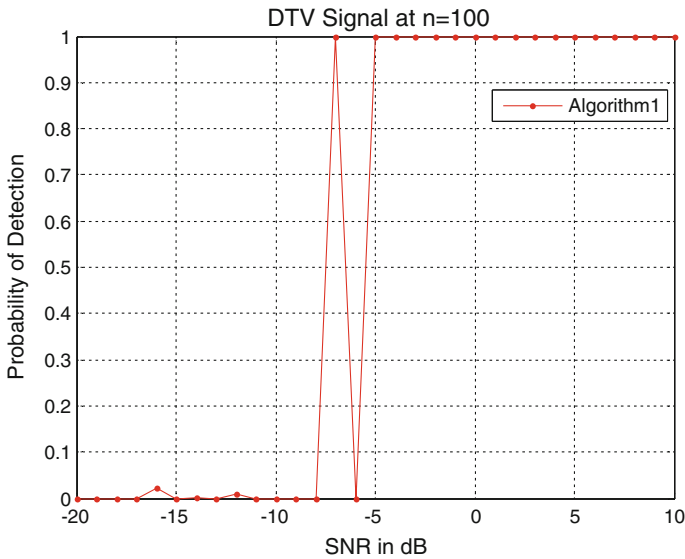
Fig. 3 Performance of spectrum sensing with small data sets

Figure 4 shows the clustering of vehicular nodes, the slowest vehicle in each cluster as cluster head, denoted as ch, the cluster members with corresponding cluster number and vehicle not in clusters with letter N. The increased vehicles in cluster result in improved sensing.

Figure 5 shows the cooperative spectrum sensing in CR-enabled VANET network over Rayleigh fading channels. The vehicles are clustered according to the density-based clustering algorithm. With increase in number of vehicles, a better sensing result is obtained. The results show that with cooperative spectrum sensing, 100% detection is possible in a relatively lower SNR compared to Fig. 3.
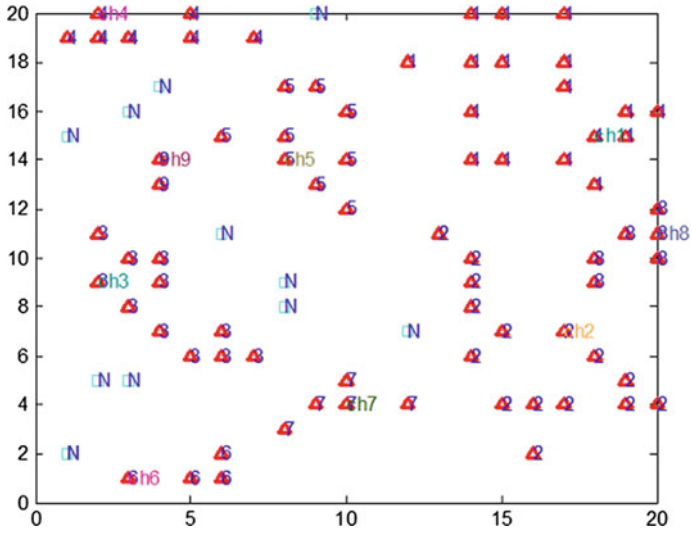
Fig. 4 Clustering of 100 vehicles with 5 cluster heads
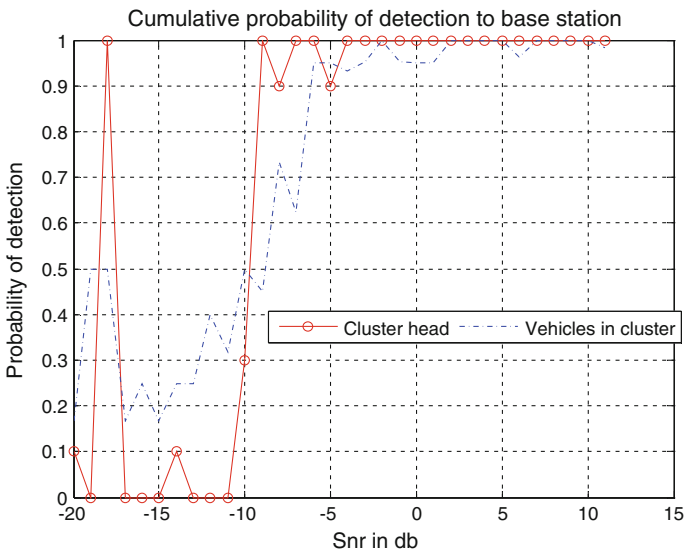
Fig. 5 Performance of cooperative spectrum sensing with small data sets

6   Conclusions and Future Work

This study has verified that with the employment of cooperative spectrum sensing with limited data sets in CR-VANET, the PU detection is possible in a relatively low SNR compared to single-user spectrum sensing. The clustering mechanism employed in this chapter further reduced the overhead problem in network, making the sensing process less complex, resulting in accurate decision for primary user detection.

References

1. Rawat P, Deep Singh K, Chaouchi H, Bonnin J-M (2013) Wireless sensor networks: a survey on recent developments and potential synergies. J Supercomput. doi:10.1007/s11227-013-1021-9
2. Cseh C (1998) Architecture of the dedicated short-range communications (DSRC) protocol. In: 48th IEEE vehicular technology conference, 1998. VTC 98, vol 3. IEEE, 1998, Ottawa, 18–21 May 1998, pp 2095–2099
3. Zhu J, Roy S (2003) MAC for dedicated short range communications in intelligent transport system. IEEE Commun Mag 41(12):60–67
4. Cheng L, Henty BE, Stancil DD, Bai F, Mudalige P (2007) Mobile vehicle-to-vehicle narrow-band channel measurement and characterization of the 5.9 GHz dedicated short range communication (DSRC) frequency band. IEEE J Select Areas Commun 25(8):1501–1516
5. Di Felice M, Doost-Mohammady R, Chowdhury KR, Bononi L (2012) Smart radios for smart vehicles: cognitive vehicular networks. IEEE Veh Technol Mag 7(2):26–33
6. Rawashdeh and Mahmud (2012) A novel algorithm to form stable clusters in vehicular ad hoc networks on highways. EURASIP J Wirel Commun Netw 2012:15
7. Haykin S (2005) Cognitive radio: brain empowered wireless communications. IEEE J Select Areas Commun 23(2)
8. Papadimitratos P, de La Fortelle A, Evenssen K, Brignolo R, Cosenza S (2009) Vehicular communication systems: enabling technologies, application, and future outlook on intelligent transportation. IEEE Commun Mag, November 2009
9. Venkatesha Prasad R, Pawelczak P, Hoffmeyer JA, Steven Berger H (2008) Cognitive functionality in next generation wireless networks: standardization effects. IEEE Commun Mag, April 2008
10. Di Felice M, Doost-Mohammady R, Chowdhury KR, Bononi L (2011) Smart radio for smart vehicles. IEEE Commun Mag, January 2011
11. Deep Singh K, Rawat P, Bonnin J-M (2014) Cognitive radio for vehicular ad hoc networks (CR-VANETs): approaches and challenges. EURASIP J Wirel Commun Netw 2014:49
12. Abeywardana RC, Sowerby KW, Berber SM (2014) Spectrum sensing in cognitive radio enabled vehicular ad hoc networks. In: 7th International conference on ICIAFS, December 2014
13. Di Felice 7th International Conference on, Chowdhury KR, Bononi L (2010) Analyzing the potential of cooperative cognitive radio technology on intervehicle communication. 978-1-4244-9229-9/10 IEEE
14. Sun Y, Chowdhury KR (2014) Enabling emergency communication through cognitive radio vehicular network. IEEE Commun Mag, October 2014
15. Li H, Irick DK (2010) Collaborative spectrum sensing in cognitive radio vehicular adhoc network. In: Networks: belief propagation on highway. IEEE

16. Lin F, Student Member, IEEE, Qiu RC, Browning JP (2015) Spectrum sensing with small-sized data sets in cognitive radio: algorithms and analysis. Trans Veh Netw 64(1)
17. Brahmi IH, Djahel S, Doudane YG (2012) A hidden Markov model based scheme for efficient and fast dissemination of safety messages in VANETs. In: Proceedings of the 2012 IEEE global communication conference, pp 177–182
18. Chai R, Yang B, Li L, Sun X (2013) Clustering-based data transmission algorithms for VANET. In: IEEE 2013 international conference on wireless communications and signal processing (WCSP), pp 1–6
19. Backlund H, Hedblom A, Neijman N, DBSCAN A density-based spatial clustering of application with noise. Linkopings Universitet—ITN
20. Qiu RC, Zhang C, Hu Z, Wicks MC (2012) Towards a large scale cognitive radio network testbed: spectrum sensing, system architecture, distributed sensing. J Commun 7(7):552–566
21. Souid I, Ben Chikha H, Attia R (2014) Blind spectrum sensing in cognitive vehicular ad hoc networks over Nakagami-m fading channels. In: IEEE 2014 international conference on, pp 1–5
22. Rawash deh ZY, Masud Mahmud S (2012) A novel algorithm to form stable clusters in vehicular ad hoc networks on highways. EURASIP J Wirel Commun Netw 2012:15
23. Page E (1954) Continuous inspection schemes. Biometrika 41(1/2):100–115
24. Lin F, Qiu RC, Browning JP (2015) Spectrum sensing with small-sized data sets in cognitive radio: algorithms and analysis. IEEE Trans Veh Technol 64(1):77–87
25. Abbassi SH, Qureshi IM, Abbasi H, Abbasi H (2016) History-based spectrum sensing in CR-VANETs. EURASIP J Wirel Commun Netw, October 2016

Performance Enhancement of Eight-Channel WDM-RoF-PON System at 80 Gbps Data Rate Using Raman Amplifier

Jayesh C. Prajapati and K.G. Maradia

Abstract In today's time due to the spread of smart phones, tablets, various machine-to-machine (M2M) communication-based applications and the rapid arrival of the internet of things (IoT)-based applications have raised request of huge bandwidth. Services for applications like transfer of various types of images, high-quality video streaming, and upcoming concepts of clouds in real time demand wireless broadband access. To cater such demands wireless access schemes using radio over fiber (RoF) technology can be used faithfully. RoF is becoming matured technology in terms of security, reliability, and coverage. Use of Raman amplifier can be done to mitigate limitations such electrical power attenuation, multipath fading, and chromatic dispersion for comparatively long distances in many applications. In this paper we have proposed the design of WDM-RoF-PON link operated at 80 Gbps data rate which comprises 8 multiplexed channels with 100-GHz channel spacing with PSK modulation technique. Due to their abundant bandwidth RoF networks are expert of supporting multiple RF subcarrier signals at a same time. Performance of proposed system is evaluated in terms of Q factor and BER parameters which provides better insight into quality of received signals in many applications.

J.C. Prajapati (✉)
Department of Electronics and Communication Engineering,
Ganpat University, Kherva, Gujarat, India
e-mail: jcp.nitk@gmail.com

K.G. Maradia
Government Engineering College, Sector 28, Gandhinagar, Gujarat, India
e-mail: kgmaradia@yahoo.com

1   Introduction

Radio over fiber (RoF) denotes the technique by which radio frequency signals are transmitted over optical fiber to provide wireless communication services up to user ends. It is fundamentally an analog communication scheme. Such RoF architecture is also called sometimes fiber wireless (Fi-Wi) architecture which is shown in Fig. 1 in which RF signals coming from a central base station (CBS) are travelling via an optical fiber to a remote site (remote antenna) and then reaching to various users via wireless channel.

It is known that optical fiber links have enough bandwidth up to tens of GHz to transmit radio waves with little distortion and moreover they offer very low attenuation. Optical fiber, moreover, offers very low attenuation (0.2 dB/km as per theoretical limit), which would allow multi-GHz radio signals to be conveyed over several kilometers with very low loss in contradiction of electrical wires. Loss in the optical fiber is a function of the optical wavelength which does not be influenced by frequency of the radio signal being transported. By reason of plentiful bandwidth and frequency-independent low-loss properties, more than one RF carrier signals can readily be frequency division (or subcarrier) multiplexed and transmitted through a single optical fiber [1, 2].
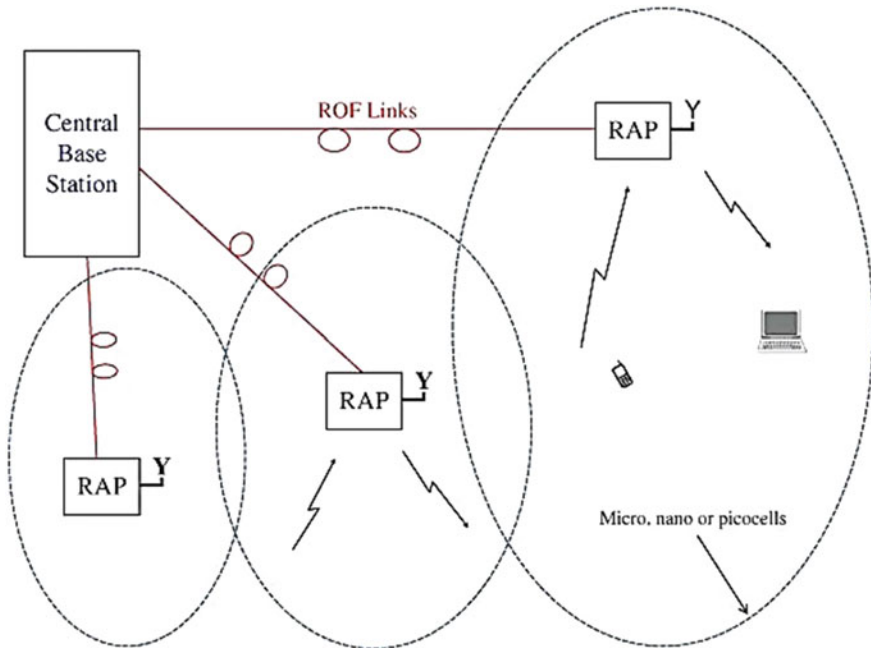
Fig. 1   Basic block diagram of Fi-Wi access scheme with point-to-point fiber links [1]

Millimeter-wave wireless systems can also be installed through RoF concept which is a versatile and prominent also. While working in millimeter-wave bands such signals are highly affected by attenuation effects but only optical fiber cables can serve such signals with very small amount of loss with controlled and tolerable amount of distortions [1].

Use of passive optical network (PON) with RoF network structures is one of the solutions to device point-to-multipoint topology at receiver side [3]. Using such concept received one channel signal can be delivered to many users means 1–2, 1–4, etc. We must have to pay attention to split loss for each such operation which is also of major concern. Such arrangements avoid costly optic-electronic conversions and require minimal number of optical transceivers and feeder fibers because here cost is divided among connected N users [4]. PON can be installed anywhere, and they don't need power for its operation [5].

2 Working of RoF System

Figure 2 depicts an analog RoF system which consists of a central site (CS) and remote site (RS) which is connected by an optical fiber feeder network. RF signal processing functions such as frequency up-conversion, carrier modulation, and multiplexing are performed at the Central Station which makes design of RS very simple with reduced installation and maintenance cost [6].

Due to reduced complexity a single high-capacity CS can handle multiple BS to avoid the problem such as cell-edge problem or dead-zone problem. RoF system is combination of wired and wireless systems in which the signal from CS to RS is guided through fiber and from RS to BS wirelessly. RoF is a capable technology for millimeter-wave wireless systems also.
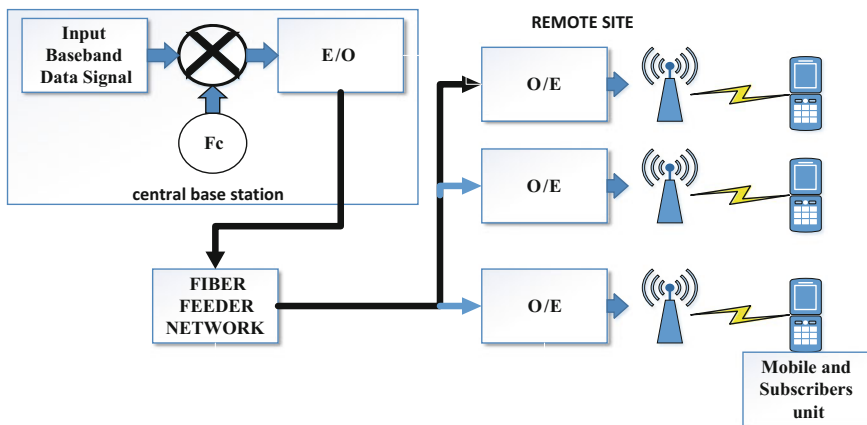
Fig. 2 Block diagram of RoF system

3   System Simulation Arrangement

As shown in Fig. 3a, eight different channel signals were obtained from various
transmitters, and then they were modulated using PSK modulators for baseband
modulation, and then they were used to optically modulate signal from CW lasers
using MZM modulators. Then such modulated signals were multiplexed using one
8 × 1 WDM multiplexer. For simulation work, initially 2.5 Gbps per channel
transmission data rate was used which is the latest one used in GPON
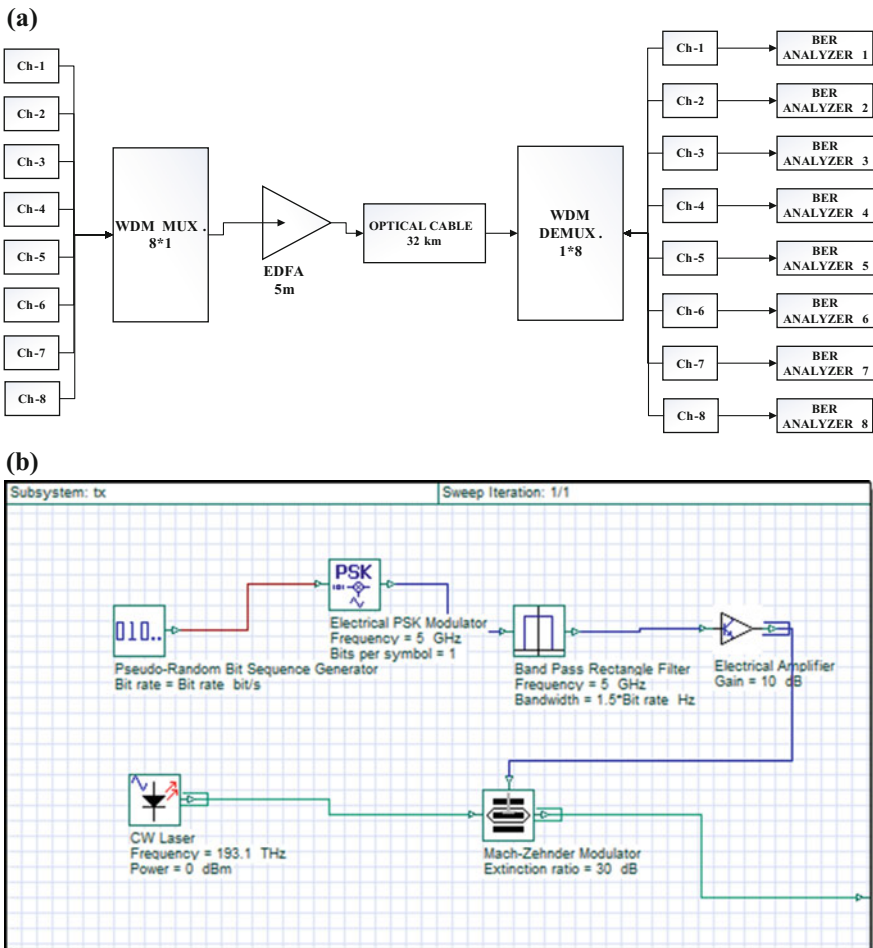
Fig. 3 a Simulation setup for eight-channel WDM-RoF-PON system. **b** View of subsystem at
transmitter side. **c** Simulation setup for eight-channel WDM-RoF-PON with split ratio 2 system.
d Simulation setup eight-channel WDM-RoF-PON system with Raman amplifier

(c)
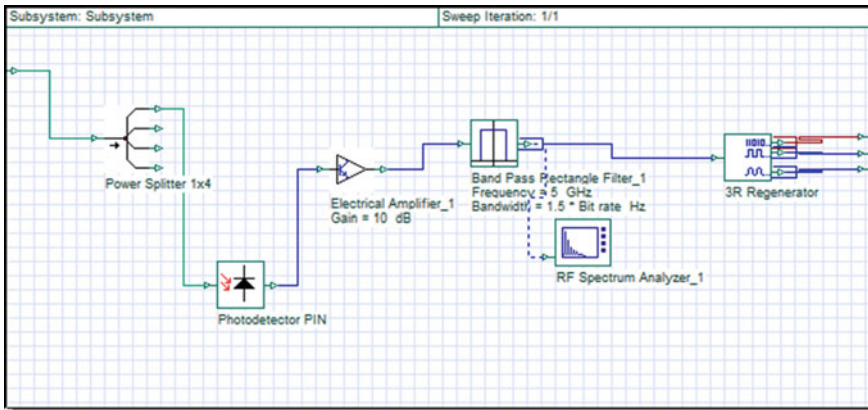
(d)
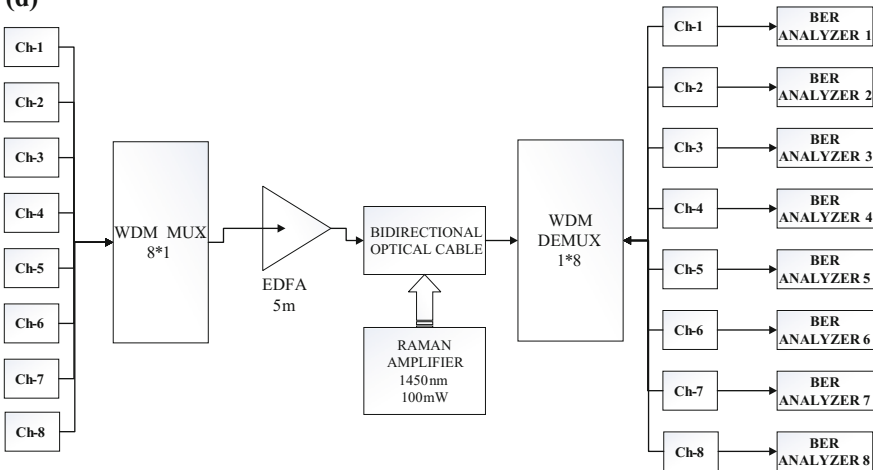
Fig. 3 (continued)

configurations and then same configuration was simulated using data rate 10 Gbps per channel. Such multiplexed signal transmitted over fiber maximum up to of 32 km. EDFA with 5 m length used to compensate optical loss occurring with signal for this much span of channel as shown in Fig. 3a. At receiver side, we have demultiplexed signal using 1:8 demultiplexer, and then each received signal splitted is through power splitter of 1:4; at ONU side, using PIN photodetectors optical signal is detected and then passed through band-pass rectangle filters. As shown in Fig. 3c from a received single using 1:4 PON concept at each channel total 32 end users can be provided service signals form 8 channels WDM RoF system. Results obtained using BER analyzer for intrinsic link and link with 5 m EDFA are as shown in Fig. 4a.

(a)

	1	2	3	4	5	6	7	8
■ Q FACOR WITHOUT EDFA	3.801312	3.38506	3.44795	3.42346	3.4101	3.35673	3.33322	3.63211
■ Q FACTOR WITH EDFA	4.01412	3.4818	3.4341	3.44099	3.62698	3.40896	3.87714	6.80039

(b)

	1	2	3	4	5	6	7	8
■ Q WITHOUT RAMAN AMPLIFIER	3.801312	3.38506	3.44795	3.42346	3.4101	3.35673	3.33322	3.63211
■ Q FACTOR WITH RAMAN AMPLIFIER	6.84245	6.60429	6.68718	6.54376	6.47666	6.43649	6.37987	6.95576

(c)

	1	2	3	4	5	6	7	8
■ DISTANCE	5	10	15	20	25	30	31	32
■ OUTPUT OSNR	50.396374	50.580081	50.284592	50.628873	50.358574	50.432961	50.319289	50.435795

Fig. 4 a *Q* factor improvement using 5-m EDFA. **b** *Q* factor improvement using Raman amplifier (1450 nm, 100 mW power level). **c** Values of output OSNR with increased distance. **d** Values of output OSNR with increased distance. **e** Values of BER with and without use of Raman amplifier

(d)

	1	2	3	4	5	6	7	8
■ INPUT OSNR	96.654051	96.848341	96.537666	98.864923	96.59103	96.662025	96.53558	96.664607
■ OUTPUT OSNR	50.396374	50.580081	50.284592	50.628873	50.358574	50.432961	50.319289	50.435795

(e)
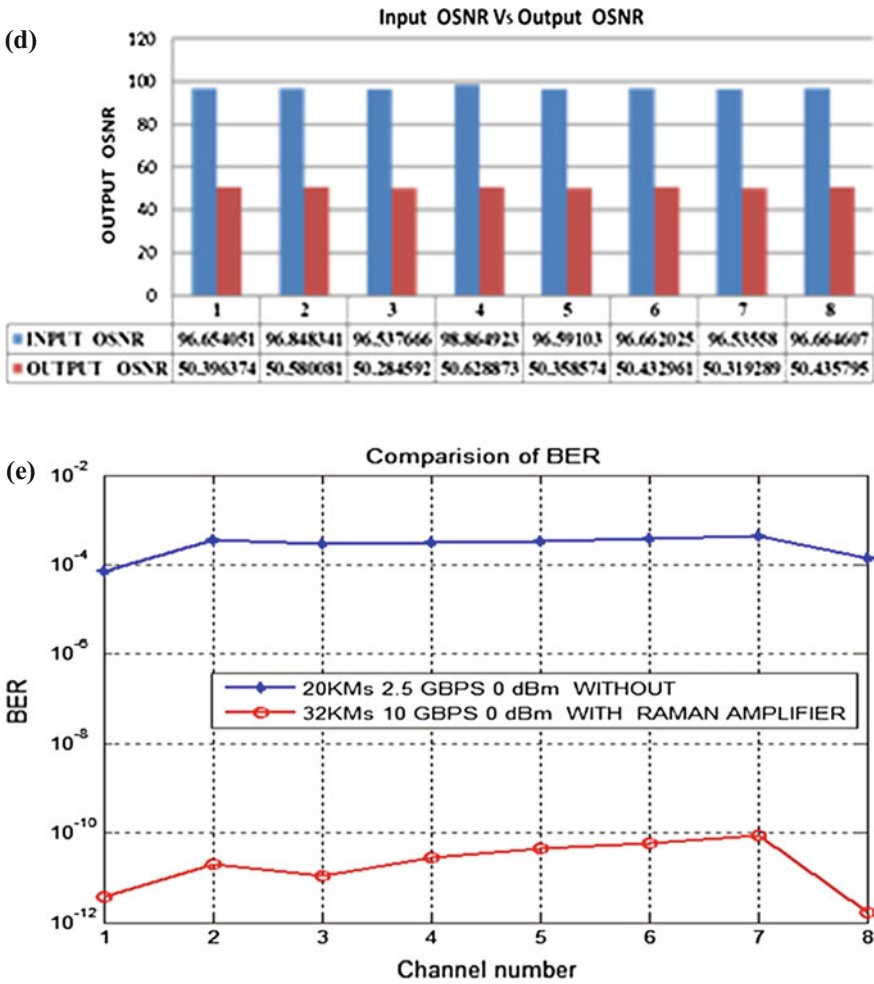
Fig. 4 (continued)

4 Simulation Setup

During simulation of our proposed work, sequence length was 256 bits and samples per bit used were 64 [7] which in turn given to PSK modulator after passing through NRZ pulse generator. Then optical signals from CW lasers with frequency 193.1–193.8 THz were modulated using MZM modulators with power level of each channel used 0 dBm. 100 GHz frequency spaced 8 multiplexed signals launched to optical fiber. Using BER analyzer values of Q factors and BER parameters were measured. Simulations were carried out for different distances up to 35 km. We have also simulated same configuration of system with Raman

Table 1 Simulation parameters

S. No.	Parameters	Values
1	Data rate (per channel)	2.5 and 10 Gbps
2	Reference wavelength	1550 (nm)
3	Max PON splits	32 (Simulated)
4	Max distance/fiber length	32 (Developed) (km)
5	Fiber optic attenuation	0.2 dB/km
6	Fiber optic dispersion	16.75 ps/nm/pm
7	Laser diode power	0 dBm
8	PSK carrier frequency	5 GHz

amplifier (1450 nm and 100 mW power), and its arrangement is shown in Fig. 2d. We have seen that PSK modulation technique is best for 2.5 GPON architecture, and it has given better result at 10 Gpbs per channel [8] (Table 1).

5 Results and Discussion

From simulation results we concluded that eight-channel WDM-RoF-PON system with PSK modulation scheme along with use of PON system having split ratio 4 gives Q factor value 3.4101 and BER value 0.000323625 for 20 km channel span without use of EDFA (Fig. 4a), (Fig. 4b) for channel number 5. Same system configuration with use of 5 m EDFA gives Q factor value of 3.62698 and 0.000274287 BER value for channel 5 which shows improvement. Same system configuration with use of Raman amplifier (1450 nm and 100 mW) gives best possible Q factor value 6.47666 and BER value $4.61E-11$ obtained for 32 km channel span as shown in Fig. 4b, e, which clearly shows that use of suggested Raman amplifier offers significantly improved results. Figure 4e shows results comparison of values of BER for eight-channel WDM-RoF-PON system having PSK modulation scheme, with and without use of Raman amplifier. Nearly constant maintained values of output OSNR nearer to 50.35 for all eight channels show assured reasonable quality of all splitted signals for distance up to 32 km, normally OSNR value decrease as distance increases. This is shown in Fig. 4c. For such WDM RoF systems normally average values of measured parameters for all channels can be taken for showing system performance improvement. As shown in Fig. 4d constant maintained values for input OSNR and output OSNR value show nearly equal quality of all signals at receiver side.

6 Conclusion

From simulation results we concluded with data rate 10 Gbps per channel, eight-channel WDM-RoF-PON system using PSK modulation scheme with the use of Raman amplifier gives significantly improved results in terms of Q factor and

BER parameters which are well above acceptable levels which actually assure the quality of received signals. Use of Raman amplifier (1450 nm and 100 mW) gives Q factor value 6.47666 and BER value $4.61E-11$ up to 32 km channel span which shows clear improvement in the performance of proposed system. Obtained results show that after using split ratio of 4 quality of all received 32 signals is good enough which are nearer to acceptable levels. Maximum BER value $8.77586E-11$ and minimum Q factor value 6.37987 were observed. Dispersive nature and non-linear behavior of transmission channel are becoming limiting factors for the long haul and high-speed optical communication applications. Influence of attenuation loss and dispersion on traveling signal in optical channel can be compensated using appropriately chosen EDFA and dispersion compensation techniques such as DCF. Further enhanced performance of such system can be obtained with use of higher order QAM modulation schemes like 8 QAM, 16 QAM, and 32 QAM along with dispersion compensation techniques. Higher order modulation schemes may offer higher data rate due to their better spectral efficiency. Use of appropriate optical amplifiers can even give longer channel span. Use of pre-coding techniques and filtering schemes at transmitter side, and appropriate DSP-based algorithms at receiver end can give further enhanced results.

References

1. Fernando XN (2014) Radio over fiber for wireless communications from fundamentals to advanced topics, 1st edn. Wiley, London
2. Lee S, Cho S-H, Lee JH (2014) Future-proof optical mobile converged access network based on integration of PON with RoF technologies. In: IEEE 2014
3. Alateeq A, Alatawi K, Almasoudi F, Matin MA (2012) Design of broadband RoF PON for the last mile. Commun Netw 4: 269–277
4. Chang G-K, Chowdhury A, Jia Z, Chien H-C, Huang M-F, Yu J, EllinasJ G (2009) Key technologies of WDM-PON for future converged optical broadband access networks [Invited]. J Opt Commun Netw 1(4): C35–C50
5. Zin AM, Idrus SM, Zulkifli N (2011) The characterization of radio-over-fiber employed GPON architecture for wireless distribution network. Int J Mach Learn Comput 1(5): 522
6. Kassim NM (2008) Recent trends in radio over fiber, 1st edn. Universiti Teknologi Malaysia, Malaysia
7. Aldhaibani AO, Yaakob S, Shaddad RQ, Idrus SM, Abdul Kadir MZ & Mohammad AB (2013) 2.5 Gb/s hybrid WDM/TDM PON using radio over fiber technique. Optik 124 (18):3678–3681. doi:10.1016/j.ijleo.2012.11.013
8. Dane P, Kaushal H (2013) Characterization of RoF GPON performance for different modulation schemes. IEEE

Analysis of Statistical Pattern in Multi-hop Cellular Networks

Alibha Sahu, Henna Kapur, Prathiba Anand and Sudhir K. Routray

Abstract Statistical models of communication networks are important to understand the network dynamics and characteristics. In the early stage of network planning, they play key roles in the network resource estimation. Network statistics plays pivotal role in the network performance analysis. In this work, we analyze the statistical pattern of link lengths of multi-hop cellular networks. For this work, we use the real multi-hop wireless communication network (i.e., cellular network) topologies. The link lengths were measured for various multi-hop wireless networks. The measured data were analyzed statistically, and we find appropriate patterns for them.

Keywords Statistical pattern · Statistical analysis · Multi-hop network statistics Characterization of multi-hop networks

A. Sahu (✉) · H. Kapur · P. Anand
Department of Electronics and Communication Engineering,
CMR Institute of Technology, Bangalore 560037, India
e-mail: alibhasahu13@gmail.com

H. Kapur
e-mail: hennakapur93@gmail.com

P. Anand
e-mail: prathibanand1995@gmail.com

S.K. Routray
Department of Telecommunication Engineering,
CMR Institute of Technology, Bangalore 560037, India
e-mail: skr@ieee.org

1 Introduction

Statistical analysis and modeling are important for understanding the complex characteristics of communication networks. Planning and design of communication networks is quite complex when complete information of the networks is not available. These tasks can be simplified through the statistical models. In the characterization and application of the networks too, statistical models can be utilized. We do not have any statistical model for the multi-hop wireless networks. In this work, we find such a model for multi-hop wireless networks.

From the recent works, it has been found that tools of statistical mechanics offer suitable framework to describe complex interwoven systems [1]. The discovery of small-world and scale-free properties of many natural and artificial complex networks has stimulated this interest even further. The ubiquity of complex networks in science, technology, social science, and mathematics has naturally led to a set of common and important research problems concerning how the network structure facilitates and constrains the network dynamic behaviors, which have largely been neglected in the studies of traditional disciplines [2–5]. Statistical models of communication networks are important to understand the network dynamics and characteristics. In the early stage of network planning, they play a key role in the network budget estimation [5–9]. They are also used for performance evaluation, parameter optimization, and testing of communication systems. In communication traffic engineering, network statistics plays a pivotal role in network performance analysis [3]. For complex systems, engineers do not have proper guidelines based on which they can utilize them efficiently. There is need of a model that can predict the outcomes in order to utilize them effectively and to avoid difficulties and catastrophes [1, 5]. In [5–9], several model for the optical networks have been developed which can help in the estimation and prediction of the network parameters. For multi-hop cellular networks, openly available information can be used for real analysis [10–12].

In this work, we analyze the statistical pattern obtained from the measurement of link lengths in multi-hop wireless network. We collect the network topologies from [12]. Then, we measure the link lengths of each of these networks. We do the analysis of these link lengths using appropriate statistical tool. Finally, we extract the best fit probability density function (PDF)-based model for multi-hop wireless networks.

The remainder of the paper is organized as follows. In Sect. 2, we present the common properties of multi-hop communication networks in brief. In Sect. 3, we present the measurement of the link lengths of multi-hop wireless communication networks. In Sect. 4, we show the statistical analysis of the link lengths. In Sect. 5, we present the results obtained from the analysis. Finally, in Sect. 6, we conclude the paper with the main points.

Fig. 1 Different routing paths for intra-cell and inter-cell traffic
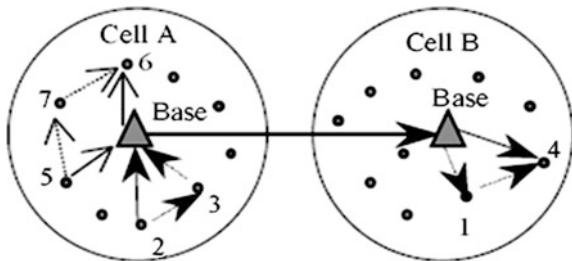
2 Multi-hop Cellular Networks

Multi-hop cellular networks (MCNs) are the networks that cannot always be reached through the cellular network infrastructure with a single hop. Normally, the communication in the cellular networks takes place through multi-hop paths. Each base station provides coverage for certain number of mobile handsets. When one mobile handset (also known as mobile station) communicates with another handset, the communication normally happens through the base stations and the mobile services switching centers which maintain the information of the mobile handsets and base stations. The transmission ranging from the base station to the mobile stations is quite small when compared with the total coverage area of a network (i.e., coverage area of an operator). Therefore, most of the communications in the multi-hop networks take place through multiple numbers of hops. The number of hops in the communication process is affected by the physical separation between the mobile stations and the algorithm used for their communication and routing.

In the present frameworks, in an MCN the mobile stations can directly communicate with each other, provided they are mutually reachable (commonly known as the device-to-device communication). This provision too tends to multi-hop routing. For multi-hop within the cell, the source and destination need to be present in the same cell and the mobile stations are used to relay packets to the destination. If they are not present in the same cell, it is sent to the nearest base station, and with the help of multiple hops, the packet is forwarded to the final destination. Figure 1 depicts the above cases, dotted line being the routing path of an MCN, solid line being the routing path for single-hop paths, and the thick solid lines as the intercellular hop [7].

3 Measurement of Link Lengths

Link length is the distance between two nodes. The exact location of a node and its neighbors are specified in terms of its latitudes and longitudes. We measured the link lengths of the cellular networks used for this study using the latitudes and longitudes obtained from the information in [12]. The process of finding the link lengths is elaborated in the following parts.
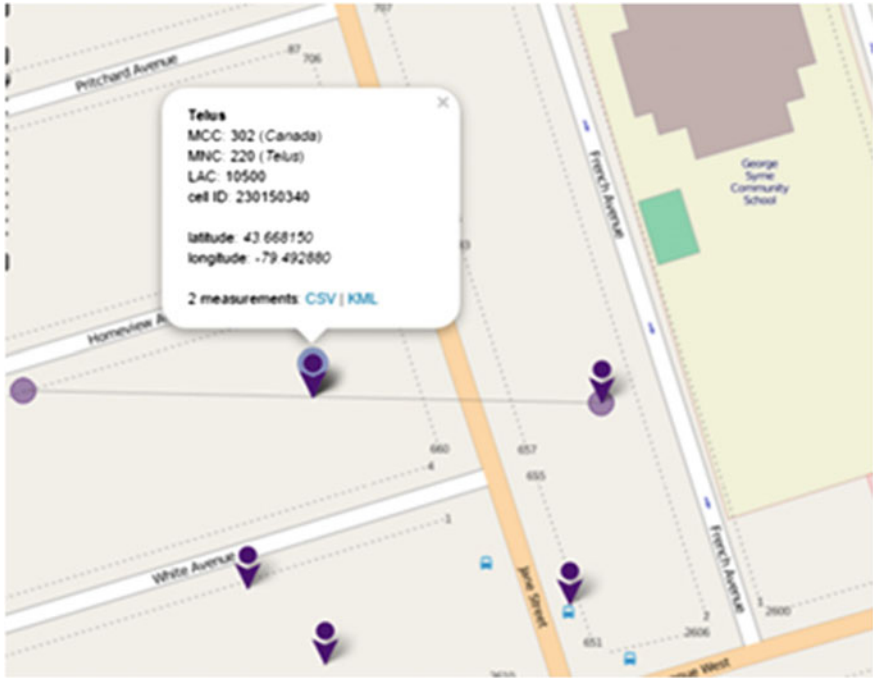
Fig. 2 Location of the primal node and its neighbors

The geographical locations of the primal nodes and their neighbors are first identified, as shown in Fig. 2. Then, these locations are fed into Sun–Earth tools (a software used for several location-related estimations including the distances between two points [13]) in order to derive the distance between the said nodes, as shown in Fig. 3. We repeat this process to find all the link lengths in the given network. The same is done for the other networks as well. These data are then analyzed statistically.

4 Statistical Analysis

Statistical analysis of networks is helpful in several aspects such as estimation and characterization of the networks. Also, network-related problems such as estimation of shortest paths, diameter, and clustering coefficients are important when it comes to analyzing the behavior of communication networks and their applications [2]. In this case, we use PDF to model the multi-hop networks. PDF-based models are directly helpful in finding the network parameters and network characteristics. They also help in the study of the random behaviors of a network.

PDFs of link lengths of all networks used for this study are obtained and analyzed using EasyFit (a mathematical tool for statistical analysis). We then arrange

Fig. 3 Link length measurement using Sun–Earth tools

the distributions increasing order of their Kolmogorov–Smirnov statistic (KSS) value. The distribution with the least average KSS value is considered as the best fitting.

5 Results

From the analysis of the link lengths, we found the following order of distributions, as shown in Table 1. Since log–logistic (3P) has the least average KSS value, it is considered to be the best-fitting PDF. We show the PDF of this distribution in Fig. 4.

Log–logistic distribution has three parameters: the continuous shape parameter (α), the continuous scale parameter (β), and the continuous location parameter (γ).

Table 1 The list of top 5 distributions with best-fitting PDFs

#	Distribution	No of I/P	Avg. KSS	Max. KSS	Min. KSS
1	Log–logistic (3P)	3	0.04918	0.11004	0.02739
2	Gen. Pareto	3	0.04967	0.08618	0.02317
3	Log-Pearson 3	3	0.05112	0.1064	0.03170
4	Lognormal (3P)	3	0.05136	0.09871	0.02813
5	Inv. Gaussian (3P)	3	0.05400	0.09790	0.03066

In this table, we also show the number of parameter of the PDF, average, maximum, and minimum KSS values
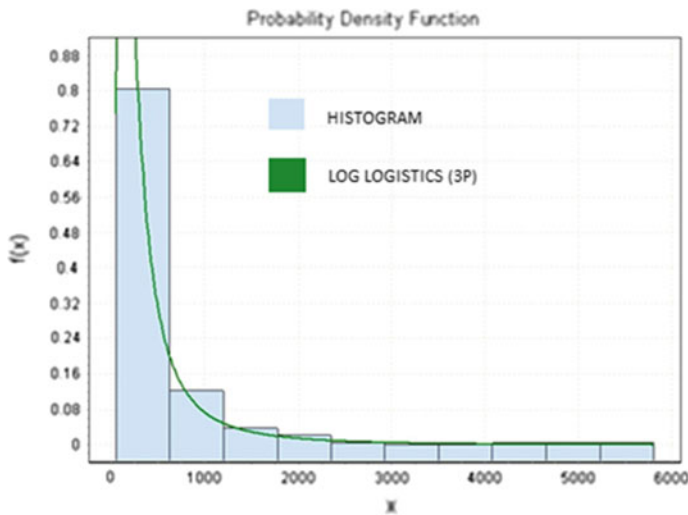
Fig. 4 PDF of log–logistic (3P) distribution fitted with the link lengths of a cellular network topology in the city of Vancouver

The first two parameters always take positive values. The PDF used for the modeling of the link lengths has been shown in (1)

$$f(x) = \frac{\alpha}{\beta} \left(\frac{x - \gamma}{\beta} \right)^{\alpha - 1} \left(1 + \left(\frac{x - \gamma}{\beta} \right)^{\alpha} \right)^{-2}. \tag{1}$$

6   Conclusion

After analyzing the link lengths of all the networks, we obtained log–logistic (3P) as the best-fitting distribution. Thus, it shows a pattern followed by the link lengths of these networks. Thus, a generalized model for the link lengths of all multi-hop

cellular networks can be estimated. This model can be used for the estimation of link-dependent parameters of these networks.

References

1. Albert R, Barabási A-L (2002) Statistical mechanics of complex networks. Rev Modern Phys 74:47–91
2. Wang XF, Chen G (2003) Complex networks: small-world, scale-free and beyond. IEEE Circuits Syst Mag 3(1):6–20
3. Strogatz SH (2001) Exploring complex networks. Nature 410:268–276
4. Barabási A-L, Albert R, Jeong H (2000) Scale-free characteristics of random networks: the topology of the world-wide web. Phys A 281(1):69–77
5. Routray SK, Morais R, da Rocha JRF, Pinto AN (2013) Statistical model for the link lengths in optical transport networks. IEEE/OSA J Opt Commun Netw 6(7):762–773
6. Routray SK, Pavan C, Sahin G, da Rocha JRF, Pinto AN (2014) Estimation of link-dependent parameters in optical transport networks using statistical models. IEEE/OSA J Opt Commun Netw 6(7):601–609
7. Routray SK, Pavan C, Sahin G, da Rocha JRF, Pinto AN (2015) Statistical analysis and modeling in optical transport networks: characterization and application. In: Conference on telecommunication, Aveiro, Portugal
8. Routray SK, da Rocha JRF, Pinto AN (2014) Estimating the parameters of optical transport networks from their circumferential ellipses. In: 21st IEEE International conference on telecommunications (ICT), Lisbon, Portugal, pp 119–123
9. Routray SK, Pavan C, Sahin G, da Rocha JRF, Pinto AN (2015) Statistical analysis and modeling of shortest path lengths in optical transport networks. IEEE/OSA J Lightw Technol 33(13):2791–2801
10. Ulm M, Widhalm P (2015) Characterization of mobile phone location errors with Opencellid data. In: 4th International conference on advanced logistics and transport (ICALT), pp 100–104 (2015)
11. Chou NT, Lin SH, Cheng SM, Chang SH (2013) Performance evaluation of self-configured two-tier heterogeneous cellular networks. In: IEEE International conference on systems, man, and cybernetics, pp 2968–2972
12. Open Cell Identification Network. http://opencellid.org/
13. Sun Earth Tools. http://www.sunearthtools.com/

Implementation of Dual 16 QAM Modulators Combined with 2 × 2 STBC Block

**B.K.V. Prasad, D. Mazumdar, G. Narendra,
G. Chaitanya and T.V. Mani Kanta**

Abstract Quadrature amplitude modulation (QAM) is a common adaptive modulation technique for satellite communications and high-speed wireless networking. In satellite communication and wireless networking, 16 QAM offers high data rates combined with less power and better spectral efficiency than QPSK and BPSK. 16 QAM easily affected by noise; hence, one can use Space Time Block Coding (STBC) to improve performance under noise. The objective of this paper was to design and implement the transmitter part of communication system with 16-QAM modulation cascaded with a 2 × 2 Space Time Block Coding (STBC) block. A design has been implemented using System Generator (Sysgen™) tool from MATLAB. The simulations of STBC were performed in Xilinx ISE Design Suite 14.2 using Verilog for STBC blocks. The complete HW is implemented in a Xilinx Virtex 6 FPGA.

1 Introduction

Modulation is defined as converting digital or analog information into a waveform suitable for transmission over a wireless medium. QAM modulation is commonly used adaptive modulation technique for its power efficiency and bandwidth [1]. The adaptive modulation technique increases transmission rate by matching the modulation schemes to a data transmission requirement. The adaptive modulation can also overcome the interference problems. In QAM modulation, two AM signals are combined to form a single channel, therefore doubling the bandwidth. STBC is performed using Alamouti Scheme [2], where the same signal is transmitted in two

B.K.V. Prasad (✉) · D. Mazumdar · G. Narendra · G. Chaitanya · T.V. Mani Kanta
Department of ECE, KL University, Guntur, India
e-mail: baditakali@gmail.com

G. Narendra
e-mail: narendra220595@gmail.com

time slots, one with signal that to be transmitted and another with negative conjugate of the signal. STBC is designed to achieve the maximum diversity order for the given number of transmit antennas and receive antennas [3]. It is common to classify MIMO systems by using the number of antennas involved as an index. In this case, (2 × 2) MIMO system is considered to implement STBC. This requires two transmit side antennas, two receive side antennas, and the consideration of an odd slot or even slot. There are many papers on QAM modulation and on different STBC schemes, but to our knowledge, they are simulation oriented and many are performed in Simulink tool in MATLAB. This paper presents the FPGA implementation of a 16 QAM transmitter and a 2 × 2 STBC transmit side block. The complete block diagram of the system is shown in Fig. 1.

The detailed explanation of QAM along with (2 × 2) STBC block is discussed on the very next sections, and the discussion of design and implementation and results are followed.

1.1 16 Quadrature Amplitude Modulation(QAM)

Quadrature amplitude modulation (QAM) uses amplitude modulation as well as phase modulation for the transmission of the data. If we transmit a symbol of 'N' bits which means '2N,' different symbols are possible. The Simulink model of 16 QAM is given in Fig. 2. Figure 2 illustrates a complete 16 QAM-based communication system comprising a 16 QAM modulator, additive white Gaussian channel (AWGN), raised cosine filter at the end of the channel, and a 16 QAM demodulator.

16 QAM is more tolerant of channel contortion when contrasted with 64 QAM, and it additionally includes a lower BER [4]. The QAM provides effective transmission to radio transmission framework using stage varieties and abundances and has disadvantages too [5]. 16-QAM has spectral efficiency of 4 bits/Hz. Its spectral efficiency is better than QAM and BPSK. It is not as good as 64QAM. But it has a
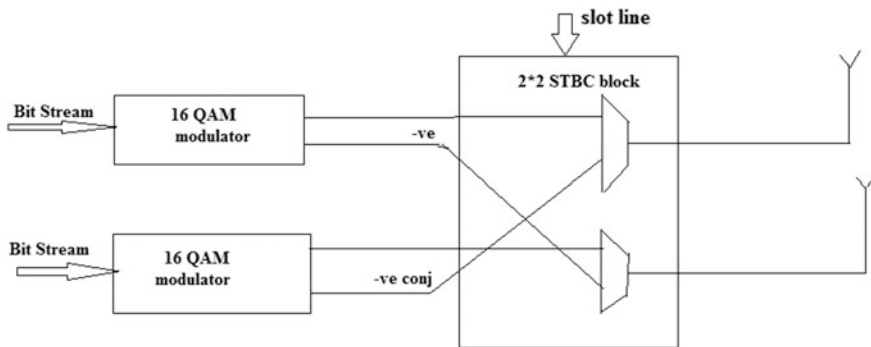
Fig. 1 Block diagram of dual 16 QAM modulation with single 2 × 2 STBC block
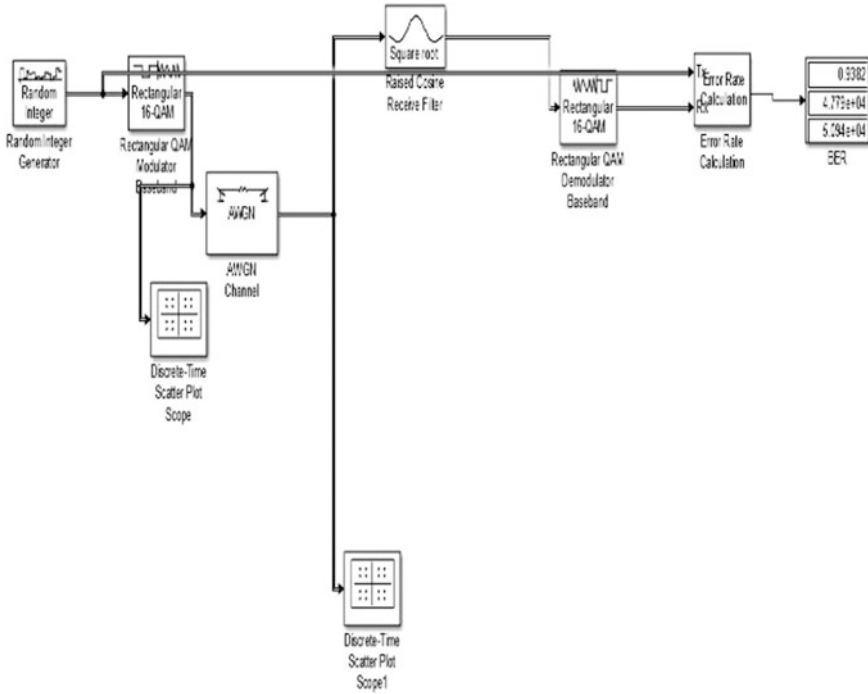
Fig. 2 Simulink model of 16 QAM modulator and demodulator

better BER for a given E_b/N_0 (bit energy divided by noise level) ratio. The BER versus SNR curve of 16 QAM is illustrated in Fig. 3.

$$P_E = 3Q\left(\sqrt{\frac{4}{5}\frac{E_b}{N_0}}\right) \tag{1}$$

2 STBC Architecture

The illustrative diagram of Alamouti STBC for 2 input–output channels connected to two transmit antennas which communicate to two different receive antennas. The block diagram of STBC block is represented in Fig. 4. Each symbol and its conjugate must be transmitted on succeeding time slots. The receiver looks at the received signal for each odd and even time slot and reconstructs the signals which have been received on succeeding time slots. The RTL schematic and output waveforms of the STBC block is given in Figs. 5 and 6.
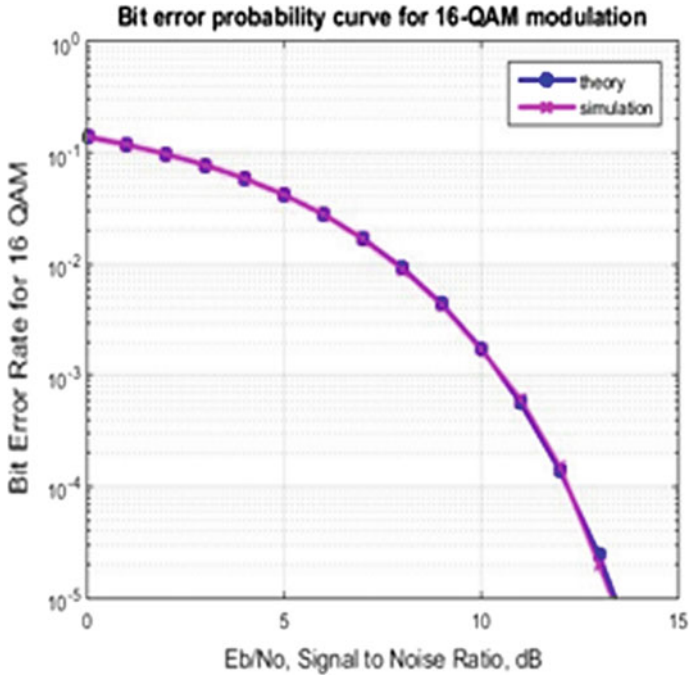
Fig. 3 BER versus E_b/N_0 for 16 QAM

Fig. 4 STBC block diagram
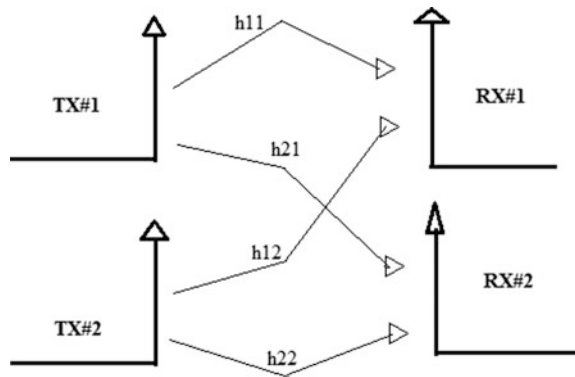
$$\begin{bmatrix} y_1^1 \\ y_2^1 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1^1 \\ n_2^1 \end{bmatrix} \tag{2}$$

where y_1^1, y_2^1 are received information at first time slot for first and second antennas respectively, h_{ij} is channel transfer function for ith, jth receive and transmit
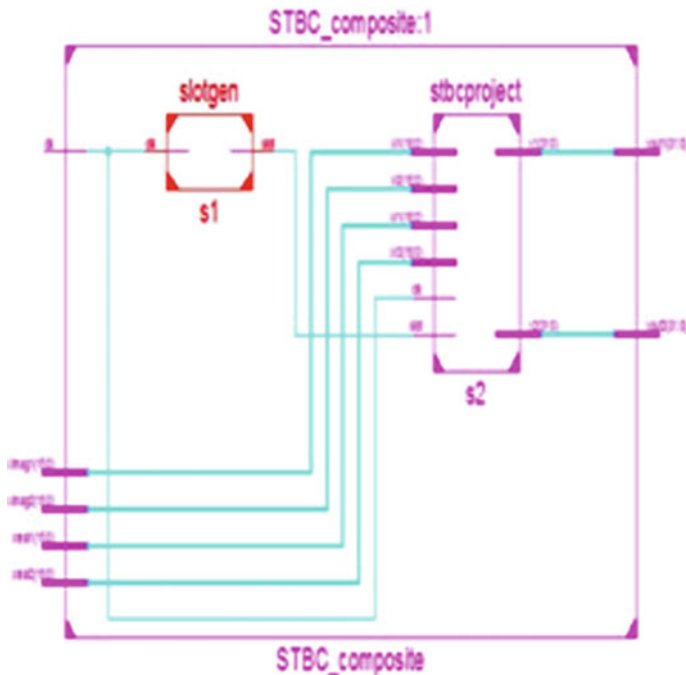
Fig. 5 Schematic of STBC block
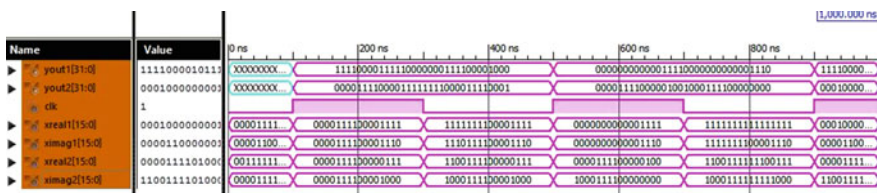
Fig. 6 Waveforms of STBC block

antennas, x_1^*, $-x_2^*$ are transmitted information at first time slot, n_1^1, n_2^1 are the noise at two time slots of first and second antennas.

$$\begin{bmatrix} y_1^2 \\ y_2^2 \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \end{bmatrix} \begin{bmatrix} -x_2^* \\ x_1^* \end{bmatrix} + \begin{bmatrix} n_1^2 \\ n_2^2 \end{bmatrix} \tag{3}$$

where y_1^2, y_2^2 are received information at second time slot for first and second antennas, respectively; h_{ij} is channel transfer function for ith, jth receive and transmit antenna; x_1, x_2 are transmitted information; n_1^2, n_2^2 are the noise at two time slots of first and second antennas. After combining Eqs. 2 and 3, we obtain

$$
\begin{bmatrix} y_1^1 \\ y_2^1 \\ y_1^{2*} \\ y_2^{2*} \end{bmatrix} = \begin{bmatrix} h_{11} & h_{12} \\ h_{21} & h_{22} \\ h_{12}^* & h_{11}^* \\ h_{22}^* & h_{21}^* \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} n_1^1 \\ n_2^1 \\ n_1^{2*} \\ n_2^{2*} \end{bmatrix} \tag{4}
$$

The waveform in Figs. 9 and 10 represents the two time slots, each at the posedge of the clock. At first posedge of the clock, the signal is sent at first time slot and its conjugate is send at the second time slot, and both signals are combined. At second posedge clock, the signal is sent at first time slot and negative of its conjugate is sent in second time slot, and both are combined and are represented.

3 Hardware of 16 QAM

LFSR generates a random signal, and it is given as input to the time-division multiplexer and Bit Basher. The LFSR generates a random string of bits which are converted into symbols 4 bits wide using a Bit Basher block. The Bit Basher output is given to the select line of the multiplexer. Constant along with counter is given to every ROM, and the ROMs are loaded with the sine values according to their position. In sine wave conjugation, if x is always real, then the complex conjugate of $\sin(x)$ is $\sin(x)$.
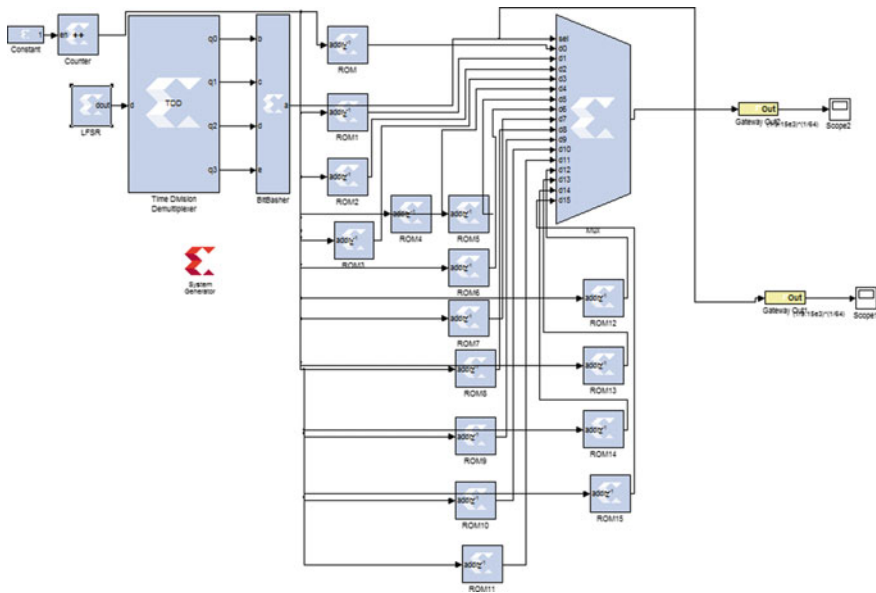
Fig. 7 Single 16 QAM

$$\sin(x) = \frac{1}{2} - i * \left[e^{(-ia)} - e^{(ia)} \right] = \frac{-1}{2} i * \left[e^{(-ia)} - e^{(ia)} \right] = \frac{1}{2} i * \left[e^{(ia)} - e^{(-ia)} \right]$$
$$= \sin(x)$$

The generation of 16 QAM signals requires 16 ROMs in this case. Depending on the bit basher output, one ROM output is connected to the MUX output at a time. All ROMs are connected to the multiplexer, and the ROMs are selected according to the select line. The output of the multiplexer is given to the scope to give the sinusoidal waveform of the modulated signal. The hardware design of single and dual 16 QAM model along with STBC are represented in Figs. 7 and 8, respectively. Figure 8 is the final FPGA block which combines a first modulator using 16 QAM, and it features a second modulator using 16 QAM. The separate outputs of the first and second modulators are combined in the STBC block, and the STBC implementation comprises two multiplexers and a conjugation generation logic. Figure 7 illustrates the implementation of a single 16 QAM modulator in Sysgen™.
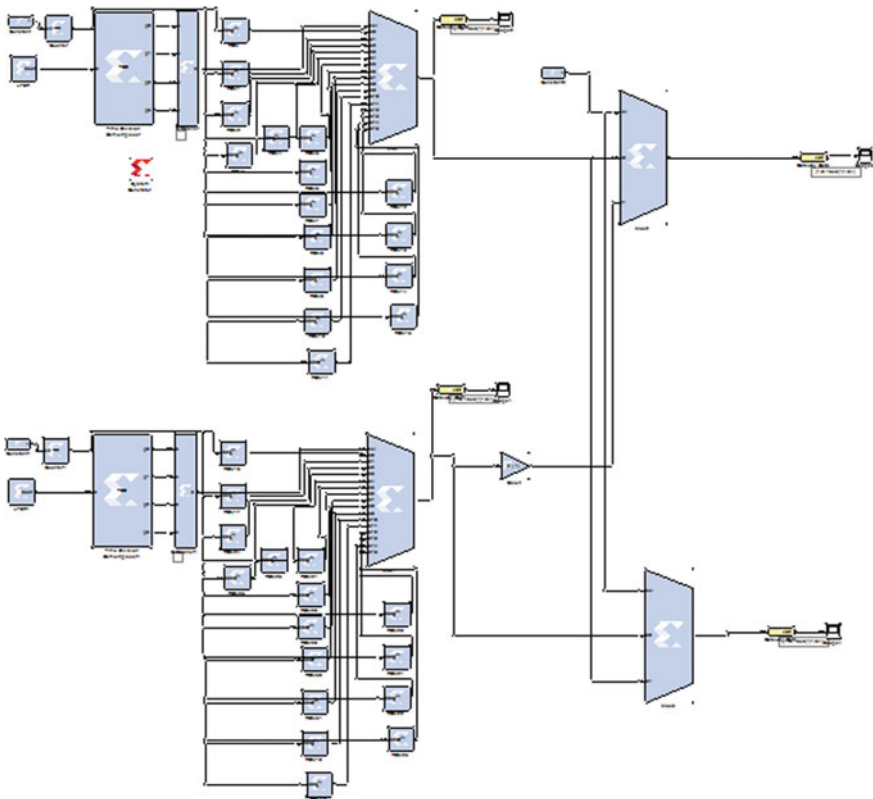
Fig. 8   Dual 16 QAM along with STBC

4 Results

The outcomes of two output of STBC block are shown in Figs. 9 and 10
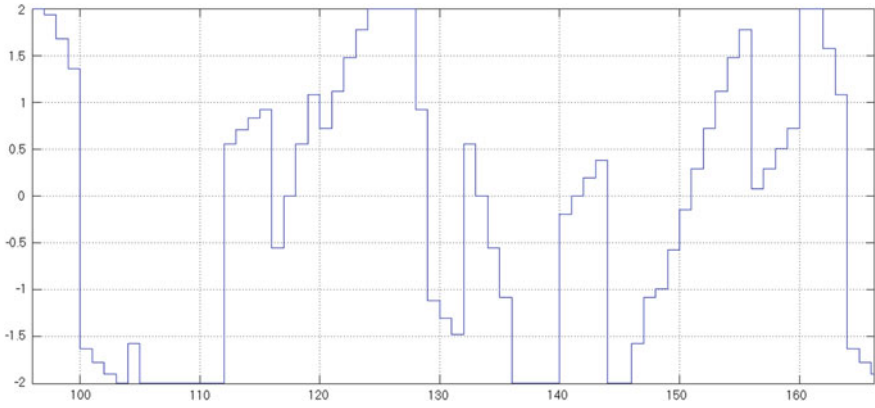
Fig. 9 First slot output of STBC
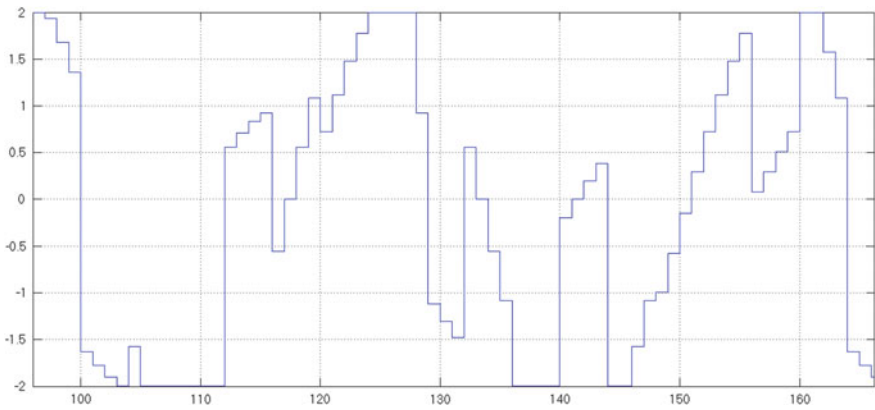
Fig. 10 Second slot output of STBC

5 Conclusion

The given signal is modulated using 16 QAM modulation and encoded with Alamouti scheme of 2 input–output channels of STBC block. Through this scheme, efficient transmission is done with high bit rate with less power. The given signal will have less effect on the distortion of the channel.

References

1. Gupta RK, Jain A, Singodiya P (2015) Bit error rate simulation using 16 QAM technique in MATLAB. Int J Multidiscip Res Dev 2(5):59–64
2. Kiran T, Sundar Rajan B (2005) STBC-schemes with nonvanishing determinant for certain number of transmit antennas. IEEE Trans Inf Theory 51(8):2985
3. Pathak P, Pandey R (2014) A novel Alamouti STBC technique for MIMO system using 16-QAM modulation and moving average filter. Int J Eng Res Appl 4(8):49–55
4. Huang G, Miyoshi Y, Maruta A, Yoshida Y (2012) All optical OOK to 16 QAM modulation format conversion employing non linear optical loop mirror. J Light Wave Technol 30 (9):1342–1350
5. Sano A, Masuda H, Kobayashi T, Fujiwara M (2011) Ultra-high capacity WDM transmission using spectrally-efficient PDM 16-QAM modulation and C- and extended L-band wideband optical amplification. J Light Wave Technol 29(4):578–586
6. Senaratne D, Tellambura C (2012) Spatial multipath resolution with space time block codes. IEEE Wireless Commun Lett 1(3):249–252
7. Li X, Wang L (2014) High rate space-time block coded spatial modulation with cyclic structure. IEEE Commun Lett 18(4):532–535
8. Sherratt RS, Linfoot SL (2003) Deterministic equalization and results of a DVB-T multipath equalizer for both 16-QAM and 64-QAM operation. IEEE Trans Consum Electron 49(1): 21–26
9. Pfau T, Noe R (2010) Phase-noise-tolerant two-stage carrier recovery concept for higher order QAM formats. IEEE J Sel Top Quantum Electron 16(5):1210–1216
10. Hwang C-S, Nam SH, Chung J, Tarokh V (2003) Differential space time block codes using non constant modules constellations. IEEE Trans Signal Process 51(11):2955–2964

A Meandered Transmission Line-based Microstrip Filter for Multiband OFDM Applications

Samiappan Vinothkumar, Shanmugasundaram Piramasubramanian and Muthu Ganesh Madhan

Abstract A meandered transmission line-based filter design is reported for MB-OFDM-based UWB communications. The filter is realized for a center frequency of 3.96 GHz and bandwidth of 1.6 GHz. The filter relies on transversal signal interference concept and characterized by low insertion loss and sharp roll-off. The basic filter design is made compact by meandering the lines along with vias to suppress undesirable signal coupling. An insertion loss of 0.5 dB and return loss better than −14 dB are measured in the pass band. Also, a roll-off rate of 188 dB/GHz is achieved in this design.

Keywords MB-OFDM · Meandered lines · Microstrip filters
Stepped impedance resonator · Vias

1 Introduction

Multiband Orthogonal Frequency Division Multiplexing (MB-OFDM) is an effective scheme to implement ultrawide band (UWB) communications for short distances. UWB technology has received considerable attention in recent days, due to its wide bandwidth (3.1–10.6 GHz). In this approach, the complete UWB range (3.1–10.6 GHz) is sliced into fourteen slots of 528 MHz bandwidth [1]. These slots are clubbed into four groups, in which group A has a bandwidth around 1.6 GHz

S. Vinothkumar (✉)
Amphenol Omni Connect India Pvt. Ltd, CMDA Industrial Area,
Maraimalai Nagar, Chennai, India
e-mail: vinodhanhere@gmail.com

S. Piramasubramanian · M.G. Madhan
Electronics Engineering Department, MIT Campus,
Anna University, Chennai, India
e-mail: spsnanthan@gmail.com

M.G. Madhan
e-mail: mganesh@annauniv.edu

and a centre frequency of 3.96 GHz. There has been an increased interest for the development of wideband and UWB filters in the recent days, due to its huge application potential [2–8]. Since UWB is a low power and broad band technology, these filters should have sharp roll-off to reject any interference from adjacent bands. In general, filters with higher order provide better roll-off characteristics. However, other schemes such as cross coupling between the resonators and the use of shunt-open stubs on the input and output feed lines have also been reported [2]. But many schemes do not provide sharp roll of characteristics, required for MB-OFDM-based UWB applications. Alternatively, signal interference technique is investigated due to its good roll-off characteristics [9]. A number of authors have reported band stop filters design based on this scheme. However, in the case of wide band-pass filter, the reports are not many. Low impedance transmission line-based technique is reported for the design of band-pass filters [10, 11]. In this paper, we have utilized the signal interference technique to implement a two-stage filter for MB-OFDM-based UWB communications. Further, the basic design is made compact by meandered transmission line with vias in between. The filter provided a bandwidth of 1.6 GHz at a centre frequency of 3.96 GHz. Experimental results of the proposed design is found to agree well with the simulations.

2 Filter Design

2.1 Basic Design

A two-stage transmission line-based design is considered for this work. The design includes shunt-open stubs in both input and output feed. A basic transmission line stage with open stubs is developed as a first step as illustrated in Fig. 1. The interference of the signals from the parallel paths leads to the desired filter response [10]. This approach provides low insertion loss and sharp rejection characteristics.

The electrical length (θ_i) of the transmission line sections are fixed as θ_{10} and θ_{20} at f_0, then θ_1, θ_2 at any arbitrary frequency, f, are given by Mandal et al. [10]

$$\theta_i = f\left(\frac{\theta_{i0}}{f_0}\right), \quad i = 1, 2 \tag{1}$$

Fig. 1 Transmission line circuit of Cascaded sections [10]
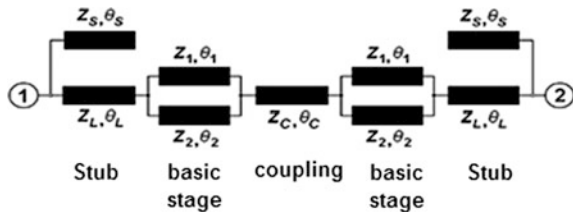
For the basic filter section, the ABCD matrix is derived and conditions for symmetric responses are set. The variations of zero positions with impedance ratio (Z_1/Z_2) are evaluated. To improve the rejection level, cascading two or more basic stages are carried out. This enables the improvement in the number of pass band poles, and hence, the filter selectivity improves. Further, open stubs of equal characteristic impedance Z_s and electrical length θ_s are connected at the input and output feed lines. The filter response is symmetric if the connecting length between the two basic configurations θ_c is taken as 90° at f_0. The value of Z_1 and θ_1, Z_2, and θ_2 are fixed as 33 Ω and 90°, 95 Ω, and 270°, respectively, for the design. The solution of Z_c is found to be 56.5 Ω, for $Z_s = 50$ and the impedances, $Z_L = Z_0 = 50$ Ω. The design procedure follows the approach of Mandal et al. [10]. The centre frequency, the substrate details, impedances, and the corresponding electrical lengths are provided as data to the ADS software, which develops the transmission line segment with appropriate length and width. The layout corresponding to a two-stage filter is depicted in Fig. 2. This structure provides the required bandwidth along with good rejection in the stop bands. The substrate used for the design is RT Duroid 5880 with dielectric constant of 2.2 and thickness of 0.381 mm. In this structure, the input and output stubs are folded in order to minimize the filter area.

2.2 Meander-Based Design

The basic design explained in the previous section is a straight forward implementation; however, it occupies a large area. If the traces can be meandered, the area required could be reduced, which makes the filter a compact one. Hence, the cascaded filter transmission lines and the open stubs (basic filter) are meandered to reduce the size. The filter layout with meandered transmission line and stubs is
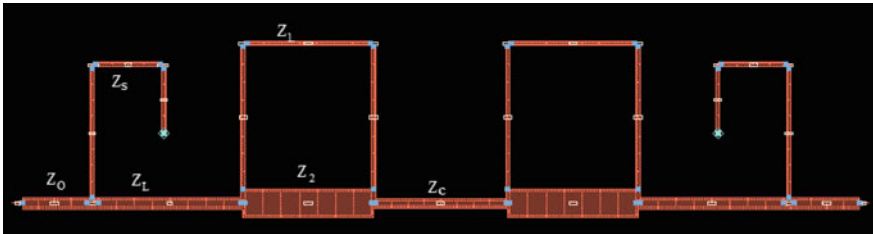
Fig. 2 Basic filter layout with dual stage and stubs

shown in Fig. 3. However, the coupling of signals between the traces, due to meandering, has to be avoided, in order to obtain the desired response. If we could block the signal coupling between traces by some means, a compact filter can be realized. It is always better to have 40 dB isolation between the lines, so that the signal would follow the intended traces without coupling unintentionally. A simple simulation is carried out to study the effect of coupling between two lines in the desired substrate. From the results, it is clear that when the transmission lines are separated by 2.5 mm distance and with two traces grounded in between them give better isolation. This is realized by providing via structures in the traces in between the transmission lines, to avoid signal coupling in the meandered lines. The layout of the single and dual trace in between the signal lines are shown in Fig. 4a, b, respectively. Figure 5 shows the layout of the complete filter with reduced area using vias. The vias are provided in areas where coupling between lines are possible. A 10 mm distance is left for the connector soldering feasibility. The total reduction achieved by meandering is approximately 45% of the original design.
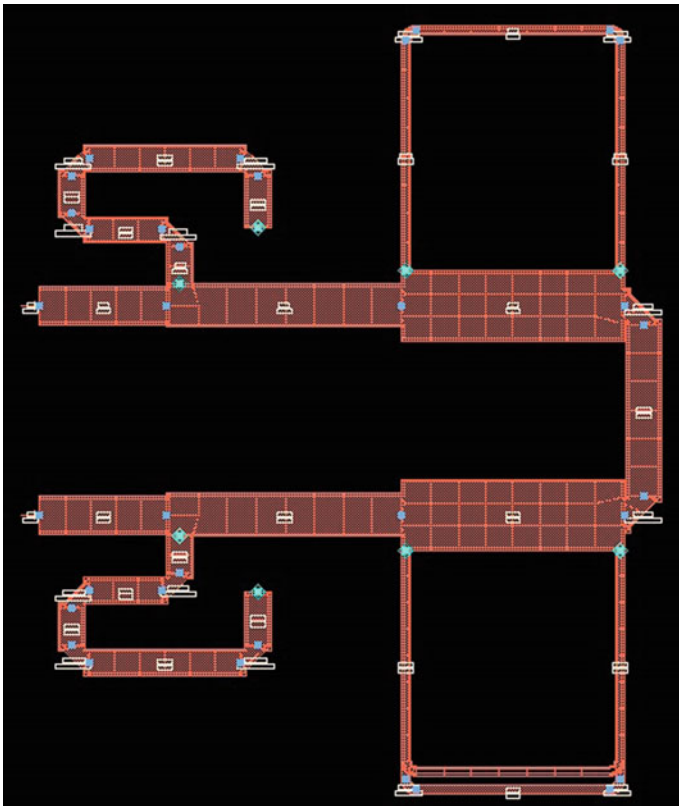
Fig. 3 Filter layout with meandered structures

Fig. 4 Layout of two transmission lines with vias on **a** Single trace. **b** Two

3   Results

Simulation of the final layout is carried out in ADS software for obtaining the filter characteristics. As shown in the Figs. 6 and 7, the insertion loss ranges from 0.4 to 0.6 dB within the pass band and return loss better than 17 dB are observed. Figure 8 shows the photograph of the fabricated meandered filter. The measured insertion losses are 0.5 dB, while the return losses are better than −14 dB at the center frequency of 3.75 GHz. Figures 9 and 10 illustrate the return loss (S_{11}) and transmission (S_{21}) characteristics, respectively. Measurements are carried out using Rohde and Schwarz ZVH-8, cable and antenna analyzer. The measured and simulated values are found to agree well. Moreover, the rejection characteristics in the adjacent lower band are greater than 30 dB as shown in Fig. 10. On the higher band
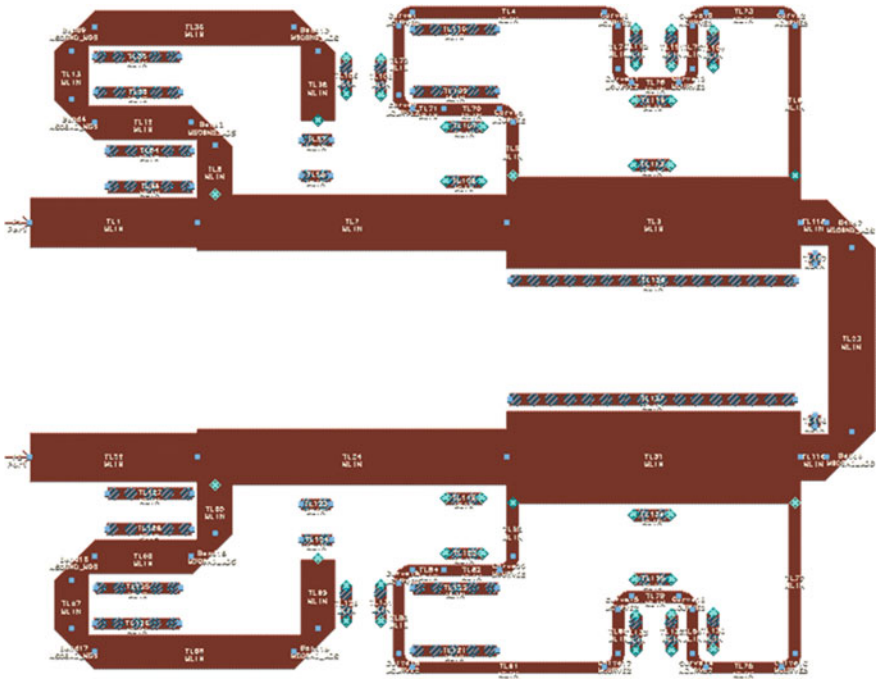
Fig. 5 Filter layout with reduced area using vias
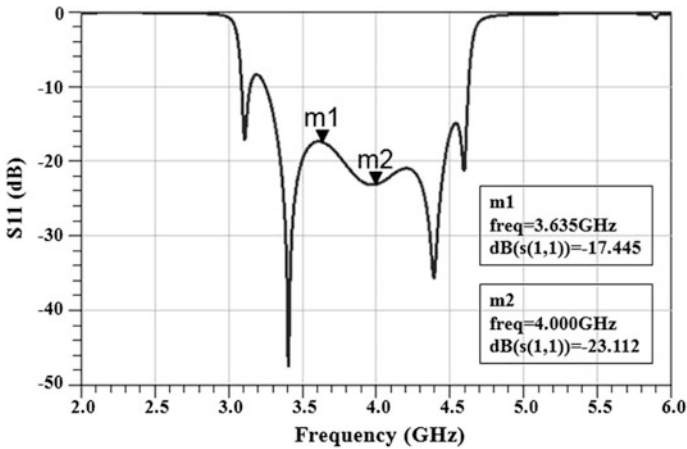
Fig. 6 Simulated return (S_{11}) loss characteristics

(4.75–6 GHz), the rejection is around 22 dB. Figure 11 shows that the minimum group delay is 0.8 ns and the variation in group delay is up to 5 ns. The impedance of the filter is also measured and found to be around 50 Ω in the pass band.
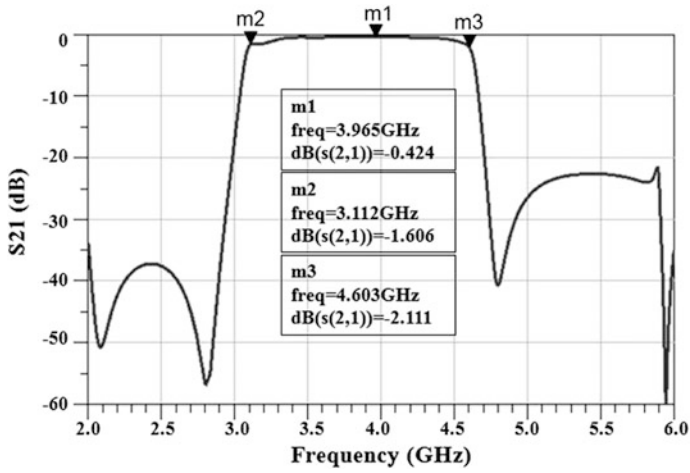
Fig. 7 Simulated transmission characteristics (S_{21})
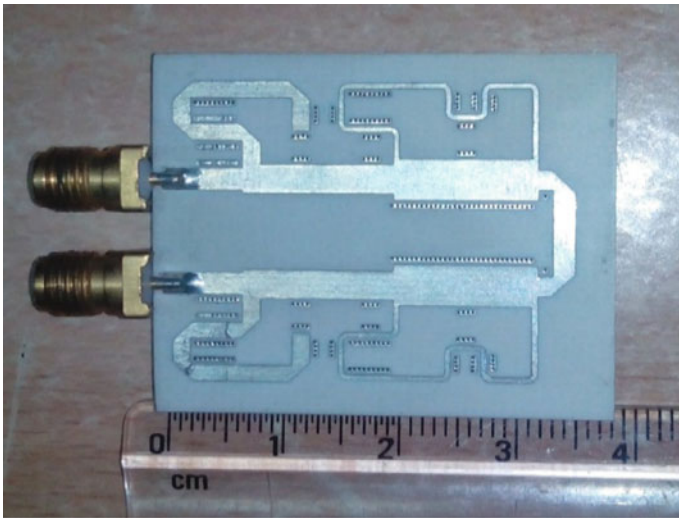
Fig. 8 Photograph of the meandered filter
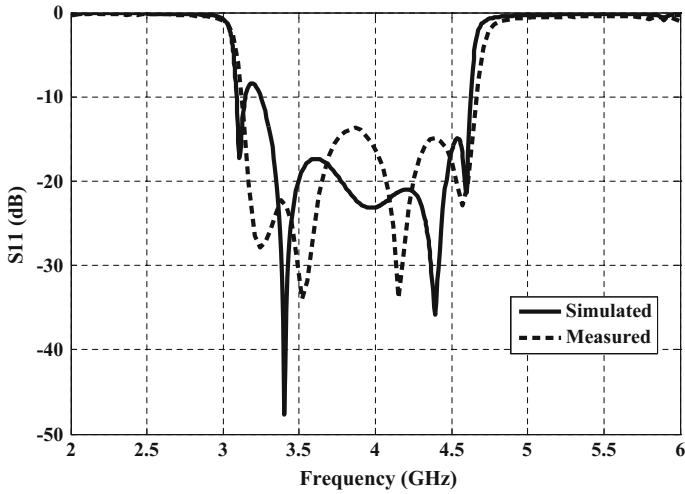
Fig. 9 Measured return loss (S_{11})
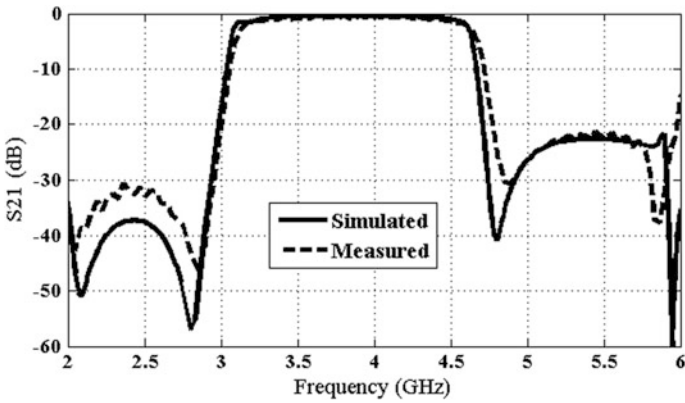
Fig. 10 Measured transmission characteristics (S_{21})

In frequencies outside the pass band, the impedance is higher ($\approx 100\,\Omega$ at 3.15 GHz). The impedance in smith chart format is depicted in Fig. 12. Further, a comparison of the roll-off characteristics, which indicate the rejection of adjacent band signals, is also carried out for a number of UWB filter designs.

From Table 1, it can be inferred that the proposed filter provides approximately 2.4 times higher roll-off rate, compared to the best reported [3] UWB filter.
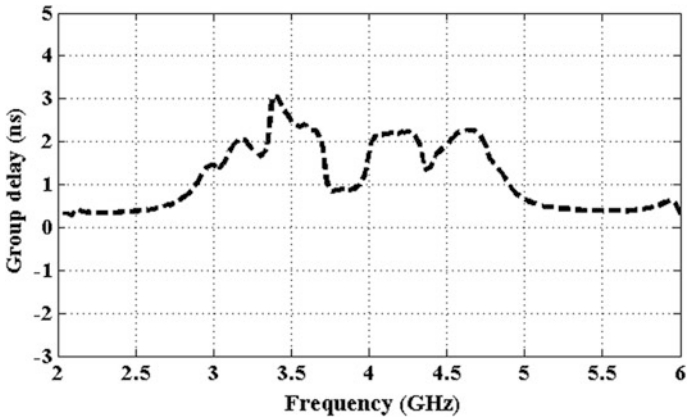
Fig. 11 Measured group delay

m3
freq=3.148GHz
trfin35_mom_a..S(1,1)=0.325 / -6.279
impedance = Z0 * (1.945 - j0.154)

m4
freq=4.574GHz
trfin35_mom_a..S(1,1)=0.139 / -136.736
impedance = Z0 * (0.803 - j0.156)

freq (2.000GHz to 6.000GHz)
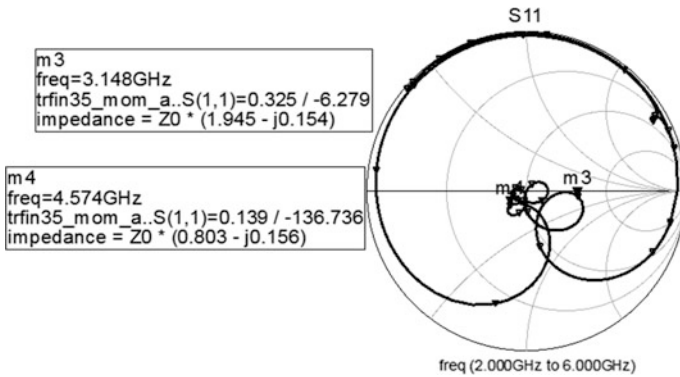
Fig. 12 Impedance characteristics

Table 1 Comparison with the reported UWB filters

References	Roll-off rate [a](dB/GHz)
Tu [2]	60
Chu et al. [3]	78
Kim and Chang [4]	36
Shaman and Hong [5]	29
Hammed and Mirshekar Syahkal [6]	29
Kuo et al. [7]	52
Jin et al. [8]	32
Proposed filter	188

[a]Roll-off rate $= \alpha_{max} - \alpha_{min}/f_s - f_c$, where α_{max} is the 30 dB attenuation point and α_{min} is 3-dB attenuation point; f_s is the 30-dB stopband frequency and f_c is the 3-dB cutoff frequency

4   Conclusion

A basic filter configuration using two transmission line sections and two shunt-open stubs is adopted for MB-OFDM-based ultra-wideband communications. The basic structure is meandered and vias are introduced between lines, to realize a compact filter. The main advantages of the filters are low-pass band insertion loss and sharp rejection characteristics. The filter structures are simple and easy to realize. The measured results show that the designed filter has 40% fractional bandwidth with a center frequency of 3.96 GHz, insertion loss of 0.5 dB, and a return loss better than of −14 dB. The roll-off rate of the proposed filter is 188 dB/GHz, which is quite higher than other designs in the same class.

References

1. Federal Communications Commission (2002) Revision of Part 15 of the commission's rules regarding ultra-wideband transmission systems. First Report and Order, FCC 02. 48
2. Tu WH (2010) Broadband microstrip bandpass filters using triple-mode resonator. IET Microw Antennas Propag 4(9):1275–1282
3. Chu Q-X, Wu X-H, Tian X-K (2011) Novel UWB bandpass filter using stub-loaded multiple-mode resonator. IEEE Microw Wirel Compon Lett 21(8):403–405
4. Kim CH, Chang K (2011) Ultra-wideband (UWB) ring resonator bandpassfilter with a notched band. IEEE Microw Wirel Compon Lett 21(4):206–208
5. Shaman H, Hong J-S (2007) Ultra-wideband (UWB) bandpass filter with embedded band notch structures. IEEE Microw Wirel Compon Lett 17(3):193–195
6. Hammed RT, Mirshekar Syahkal D (2012) A compact high selectivity seventh-order UWB bandpass filter with ultra-stopband attenuation. In: Radio and wireless symposium (RWS), pp 135–138
7. Kuo TN, Wang C-H, Chen CH (2007) A compact ultra-wideband bandpass filter based on split-moderesonator. IEEE Microw Wirel Compon Lett 17(12):852–854
8. Xu J, Wu W, Kang W, Miao C (2012) Compact UWB bandpass filter with a notched band using radial stub loadedresonator. IEEE Microw Wirel Compon Lett 22(7):351–353
9. Gomez-Garcia R, Alonso JI (2005) Design of sharp-rejection and lowloss wide-band planar filters using signal-interference techniques. IEEE Microw Wirel Compon Lett 15(8):530–532
10. Mandal MK, Mondal P, Sanyal S (2008) Low insertion loss, wideband bandpass filters with sharp rejection characteristics. IET Microw Antennas Propag 4(1):99–105
11. Ganesh Madhan M, Fatima Rani GA, Sridhar K, Sathis Kumar J (2012) Design and fabrication of transmission line based wide band bandpass filter. Proc Eng 30:646–653

Vulnerable Network Analysis Using War Driving and Security Intelligence

Stuti Gupta, Bharat S. Chaudhari and Boudhayan Chakrabarty

Abstract Wireless network is growing explosively not only in the corporate environment, but also at the consumer space. If the network is left open and unsecured, anyone can not only use the network for downloading illegal content, but also such a network can be used as a hacking medium to bring down other networks. Such open networks needs to be protected against such bad people who might take advantage of such an insecure network. Wireless networks can be vulnerable to various types of attacks such as eavesdropping, hacking, and freeloaders if there are no protections present in such wireless networks. IBM QRadar is security information, and event management (SIEM) solution is used by security professionals to protect their networks and themselves. In this paper, we propose a system that uses war driving for collecting access point information and QRadar to analyze vulnerable networks by correlating real-time traffic with the information present within the network. The work undertaken integrates information collected by war driving with IBM QRadar and then used to refine its correlation using the network information from open networks. After detecting vulnerable attacks and users, alerts are sent out to the security operation center.

Keywords War driving · Vistumbler · Security · QRadar

S. Gupta (✉) · B.S. Chaudhari
Maharashtra Institute of Technology, Pune, India
e-mail: stutigupta.230492@gmail.com

B.S. Chaudhari
e-mail: bharat.chaudhari@mitpune.edu.in

B. Chakrabarty
IBM Software Labs, Pune, India
e-mail: bochakra@in.ibm.com

1 Introduction

Nowadays. every office, residence, coffee houses, libraries etc. has wireless access point for accessing internet. The main concern regarding these access points are the security issues related with it. Wireless security prevents unauthorized damage or access to computer using wireless networks; some access points which are not protected via security standards are known as open access points. These access points are vulnerable to different types of security threats. Anyone within the range of an open wireless network can "sniff," or capture and record, the traffic, gain illegitimate access to internal network resources and also to the internet, and then use the information and resources to perform illegal work. Such security breaches issues are important for both enterprise and home networks. Hence, there is a need of analyzing these open access points using security solutions to protect them against misuse. SIEM technology provides real-time analysis of security alerts generated by applications. QRadar is a SIEM solution. This paper makes use of QRadar which is a product of IBM. QRadar is used for real-time monitoring, correlation of events, notifications as well as analysis and reporting of log data.

A process of finding access points while driving in a vehicle is known as war driving. Basic requirements of war driving are laptop and software used for war driving. War driving was invented by Peter Shipley in 1999 or 2000. This paper makes use of Vistumbler as software for war driving.

Vistumbler is war driving software used for scanning Wi-fi access points which are within the range of Wi-fi adapter. After finding wireless access point, Vistumbler displays various information about scanned access points such as networks SSID, MAC address, encryption being used, signal strength, and the networks channel. This paper makes use of Vistumbler for gathering information about the access points.

SIEM is a system used for analyzing the information detected by Vistumbler. SIEM provides long-term analysis of security events and real-time reporting. SEM provides real-time monitoring of security events whereas SIM provides log management and reporting for security events. Two products, namely SIM (Security information Management) and SEM (Security Event management), are combined to form SIEM. A security information system has been developed to address two requirements that emerged in the late 1990s and early 2000s. Mark Nicolette and Amrit Williams invented SIEM in 2005 describing products ability to analyze and represent network information and security devices.

Section 1 consists of introduction of proposed theory followed by Sect. 2 gives the brief explanation about QRadar. Section 3 gives the details of proposed work and presents the results, whereas Sect. 4 concludes the paper.

2 Study of QRadar

This paper makes use of IBM Security QRadar is a SIEM solution. QRadar is used in several small and large organizations for their security operation center for collecting, normalizing, and correlating network data. Collection and analysis of collected data, correlation of data across various system that include network and security solutions, server, operating system, host, and applications are provided by QRadar. QRadar has highly scalable database. It combines logsource event data from various devices distributed among the network. Every activity is stored in raw format within QRadar, and it correlates these activities for finding real threats. QRadar collects various events such as security events and network events. Security events are the events from virtual private networks, firewalls, intrusion detection system, and intrusion prevention system. Network events are from router, hosts, switches, and server. User or asset context is the contingent data from identity and access management products and vulnerability scanners. Operating system information is the information about vendor name and version number specifics for network assets.

In this paper, various access points are detected using war driving. Vistumbler is the war driving software which has been used for detecting access points. This information regarding access point is sent to QRadar with the help of WinCollect agent. QRadar will analyze this information by correlating it with the information received from different log source devices with the help of rules. This will be explained in next section of this paper.

The architecture of QRadar is shown in Fig. 1. employs multiple models of event processor appliances, event collector appliances, a central console, and flow processor appliances, all available as hardware-based, software-only, or as virtual software appliances. Smaller installations can be started with a single all-in-one solution and easily be upgraded to console deployments, adding flow and event processor appliances as needed.

Fig. 1 Architecture of QRadar
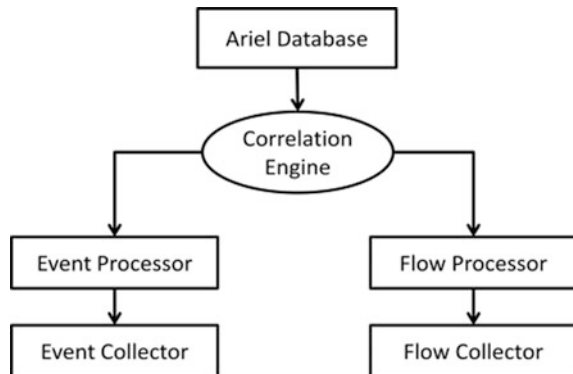
3   Proposed Work and Results

We have developed a system for analyzing vulnerable network which is based on a flowchart as shown in Fig. 2. The detection of access point has been achieved using Vistumbler software which is a software used for war driving. Once Vistumbler finds a wireless network, it will display the networks SSID, signal strength, encryption being used, Mac address, the networks channel, and much more. Since the output is from Windows machine, there is a requirement of an agent to send Windows-based events. This agent is known as WinCollect agent. WinCollect is a scalable IBM Security QRadar feature that collects Windows-based events. These events would appear as unknown because device support module (DSM) is not able to parse events properly. The events which are coming as unknown can parsed with the help of Logsource extension. Logsource extension is based on java regular

Fig. 2 Flowchart of proposed work
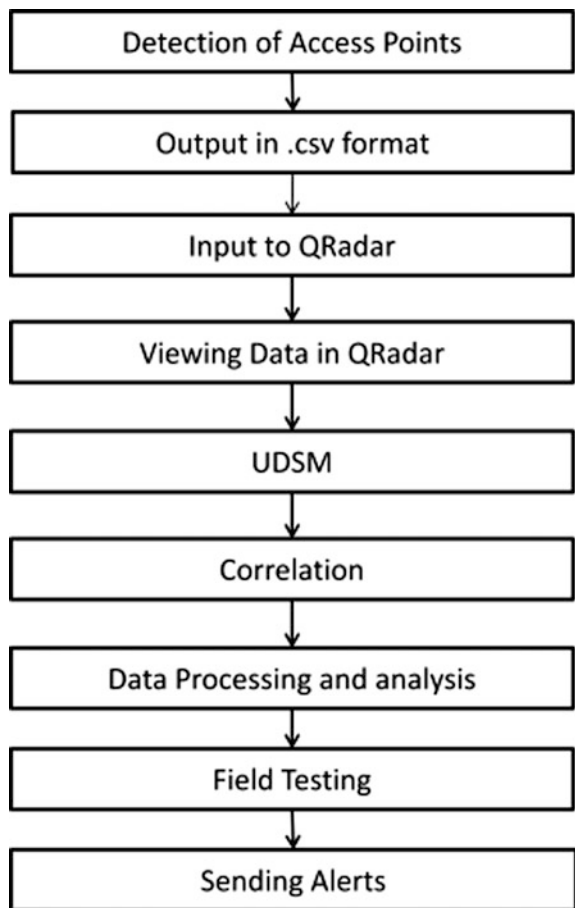
expression which is used for extracting the information present in event payload and mapping to QRadar fields.

The system for analyzing vulnerable network is used to check if any packets matches the alerting rule and if any of the packets matches the condition present in rule, then alert will be generated. Based on certain use cases, these rules have been created. This has been shown in Fig. 3.

Use Case 1 Suppose a user is using an open access point and accessing enterprise network, then there will be a policy violation since that user is trying to access a secured network with an unsecured access point. On the basis of this use case, a rule is created such that if payload of packet consists the information about an open and unsecured network and also if those packets of data are coming from the logsource WinCollect or any other logsource device present within the network, then the username present in the payload will be added to a reference set 1. Also, if the payload of the packet consists the information about the open access point and the user name present in the payload matches the username present in the reference set 1, then there will a policy violation and alert will be generated. This has been shown in logical diagram of rules in Fig. 3.

Use Case 2 In this scenario, if a device is accessing the secured network with an unsecured access point, then it is policy violation since that device is using unsecured access point for communication within the network. On the basis of this use case, a rule has been created such that if the packet of data consists the information about open access point and also the information about the MAC address of the device which is using that open access point, then that MAC address will listed in reference set 2. Also, the events or we can say that the packets of data are coming from different logsource devices present within the network and information about the source or destination. If MAC present in the payload of those events matches
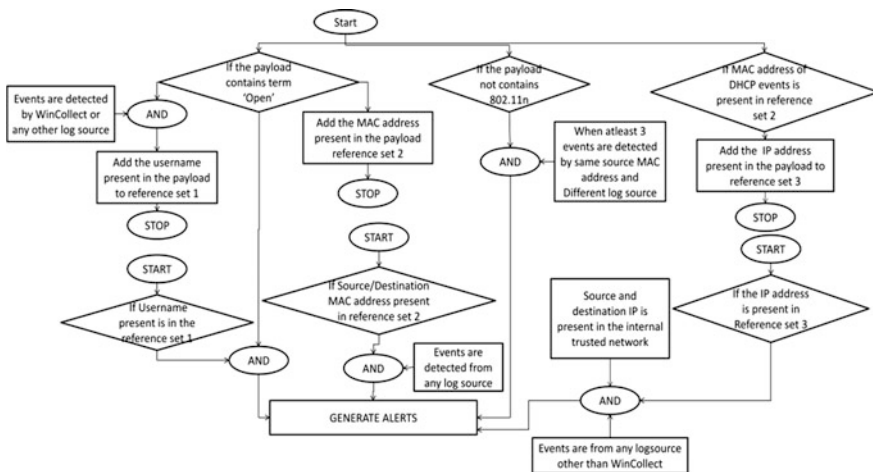
Fig. 3 Logical diagram of rules

the MAC address that is present in the reference set 2, then an alert will be
generated. This has been shown in Fig. 3.

Use Case 3 In this scenario, the channel used for secure communication has been
considered. If the user is using a channel other than 802.11n, then there will be
security breach. Based on this use case, a rule has been created which states that if
events are coming from different logsource device present in the network and there
are three events detected from the same MAC address and the information present
in the payload does not contains the 802.11n as communication channel, then an
alert will be generated.

Use Case 4 In this scenario, if a device is trying to access the secured network with
an unsecured access point and that device is present within the organization, then
there will be a policy violation. On the basis of this use case, a rule has been created
such that if the events are coming from DHCP device and the MAC address that is
present in the payload of those event are one of the MAC address that is present in
the reference set 2, then the IP address present in the payload of DHCP events
would be added to reference set 3. Now if the events are coming from any device
and IP address present in the payload of such events is one of the IP address that are
present in the reference set 3 and if these IP address are from internal trusted
network then an alert will be generated.

   These alerts are termed as offense in QRadar terminology. Whenever a rule
triggers, it generates offense as a response. On the basis of above mentioned use
cases and rules, alerts that are offenses have been generated which are shown in the
Fig. 4. Event named ExploitBackdoor are the events that are generating alerts since
someone is trying to access the secured enterprise network with open access points.
Another event that is Clint IP lease expired DNS deleted is generating offense since
device was trying to access the secure network and that device was present within
the network and was using open access point, so their IP address has been deleted
and those device would not be able to access the secured network.
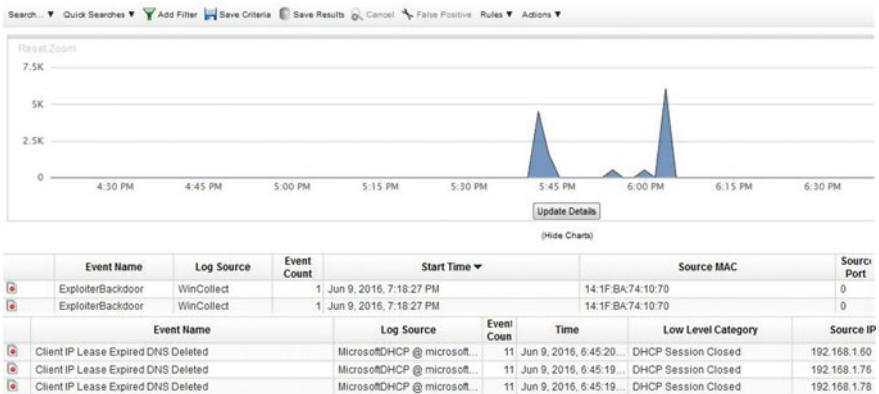
Fig. 4 Generated alerts in QRadar

4 Conclusion

This paper makes use of war driving for detection of open access points. Detection of access points has been achieved using war driving software called Vistumbler. The information about access points are given as input to QRadar. WinCollect agent has been used to send Windows-based events to QRadar. These events are correlated and normalized and stored in internal database providing real-time analysis of events based on certain set of rules. These rules have been created on the basis of four use cases and hence generating alerts when events triggers rule. The main contribution of this paper is it will help to find out security breaches within the network caused by using vulnerable access points and different devices present with the enterprise network.

References

1. Badea A, Croitoru V, Gheorghica D (2015) Computer networks security based on the detection of user's behavior. In: 9th international symposium on advanced topics in electrical engineering (ATEE), pp 55–60
2. Hermanowski D (2015) Open source security information management system supporting IT security audit. In: 2nd IEEE international conference on cybernetics (CYBCONF), pp 336–341
3. Said H, Guimaraes M, Al Mutawa N, Al Awadhi I (2011) Forensics and war-driving on unsecured wireless network. In: International conference on internet technology and secured transactions (ICITST), pp 19–24
4. Priya CS, Umar S, Sirisha T (2013) The impact of war driving on wireless networks. Int J Comput Sci Eng Technol: IJCSET 3(6):230–235
5. Datasheet of IBM Security QRadar SIEM. http://www.ibm.com/in-en/

Optimal Channel Estimation Using DFT-Based Interpolation with Comb-Type Pilots for OFDM Systems

Sireesha Biyyam and Anuradha Bhuma

Abstract Orthogonal frequency division multiplexing (OFDM) is the multicarrier modulation scheme of choice in most of modern digital communication systems. For coherent detection of OFDM signals that maximize the system bit error rate (BER) performance, we need the knowledge of channel frequency response (CFR) coefficients at the receiver. Therefore, channel estimation becomes a critical task. In this paper, we present a novel DFT-based optimal channel estimation procedure that has very low computational complexity. The assumption made in this work is that the number of pilot symbols transmitted is at least equal to the number of channel impulse response coefficients. This assumption is quite reasonable in practice. The estimation is optimal, in the sense that the interpolation of the CFR from pilot subcarriers to all subcarriers is exact, in the absence of noise at pilot subcarriers. Simulation results demonstrate the advantage of our approach over the other prevalent methods.

Keywords OFDM · Channel estimation · Interpolation · Error analysis

1 Introduction

Orthogonal frequency division multiplexing (OFDM) is a popular multicarrier modulation scheme in wireless communications due to its high bandwidth efficiency and ability to reject multipath interference [1]. Owing to its implementation using IFFT and FFT blocks, it does not require oscillators to generate the carriers, due to which it uses minimum hardware. The performance of OFDM depends on the extent to which it is time and frequency synchronized as well the quality of

S. Biyyam (✉)
Department of ECE, PESIT-Bangalore South Campus, Bangalore 560100, India
e-mail: bsiree@gmail.com

A. Bhuma
Department of ECE, S.V.U. College of Engineering, Tirupati, AP 517502, India
e-mail: anubhuma@yahoo.com

channel estimation. Several pilot-based channel estimation algorithms are proposed in the literature. In all the pilot-based estimation procedures, the estimated frequency response coefficients will have two types of errors in which one is the error due to interpolation and the other is the error caused by the noise at the front end of the receiver.

The methods for channel estimation in an OFDM-based system with frequency-selective Rayleigh fading can be broadly classified into two categories: blind and pilot-based. Since blind channel estimation techniques are computationally demanding and suffer from slow convergence, pilot-based approaches are typically preferred in practical systems [2].

Pilot-based channel estimation techniques for OFDM can be of two kinds: block-type and comb-type [2]. Block-type approaches are ideal in the case of a slow fading channel. Here, all the subcarriers of an OFDM frame carry known pilot symbols that are used to obtain the estimate of the CFR across all subcarriers. This estimate will be valid for a specified number of subsequent OFDM frames, after which a pilot frame is once again transmitted. Comb-type methods are preferred in a fast-fading scenario, so that channel estimation becomes necessary in every OFDM frame [3–5]. In this case, pilot symbols are transmitted over specified subcarriers of every OFDM frame. These pilot symbols are used to obtain the estimate of the CFR across all subcarriers.

The process of estimating the CFR across all subcarriers in a comb-type system involves two steps. The first task is to estimate the CFR at the pilot locations. Least square and minimum mean-squared error estimators have been proposed for accomplishing this task [3]. Once the CFR at pilot locations has been estimated, the CFR at the other subcarriers is obtained by interpolation. Different interpolators have been proposed toward this end. Some of the prominent ones are piecewise linear interpolation, second-order interpolation, low-pass interpolation, and cubic spline interpolation [5–8].

In this paper, we consider the problem of obtaining the CFR at all the subcarriers from the CFR to the pilot subcarriers, using interpolation. In all the methods proposed in the literature, the final estimate of the CFR contains errors from two sources: error due to interpolation and error due to noise. We propose an impulse response-based interpolation scheme that provides CFR estimate with zero interpolation error. We also show that with the proper choice of pilot locations, noise enhancement can be avoided.

Impulse response-based channel estimation methods have been studied in the literature [7, 9]. These schemes concentrate on performing smoothing/estimation in time domain and then transforming the result back to the frequency domain. In our approach, the focus is on the optimality aspect of the impulse response-based estimation due to the absence of interpolation error.

The rest of the paper is organized as follows. In Sect. 2, we present the OFDM system model and DFT-based optimal interpolation in Sect. 3. Numerical results and discussion are presented in Sect. 4, and conclusions are given in Sect. 5.

2 System Model

We consider an OFDM system with N subcarriers, in which P of them are allocated to transmit pilot symbols. The channel is assumed to be a frequency-selective Rayleigh fading one, with L independent channel impulse response taps. We assume $L \leq P$ and the variance of each tap equal to $1/L$. The channel impulse response vector is $h = [h_0, h_1, h_{L-1}]^T$. Let S be the symbol vector of size $N - P$. By interleaving the P pilot symbols with this vector at specified locations, we obtain the N-length vector X. Let $p_0, p_1, \ldots p_{P-1}$ be the pilot positions.

The transmitted vector y is the N-point IFFT of the vector X, after adding the cyclic prefix of length G_1 at least equal to the delay spread of the channel $L - 1$. The transmitted symbols are

$$y(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} X(k) \exp^{\frac{j2\pi kn}{N}}, \quad -G_1 \leq n \leq N-1 \tag{1}$$

At the receiver, after removing the cyclic prefix and taking the N-point FFT, the output symbols are

$$Y(k) = H(k)X(k) + Z(k), \quad 0 \leq k \leq N - 1 \tag{2}$$

Here, $H(k)$ are the CFR coefficients given by

$$H(k) = \frac{1}{\sqrt{N}} \sum_{l=0}^{N-1} h(l) \exp^{\frac{-j2\pi lk}{N}}, \quad 0 \leq k \leq N-1 \tag{3}$$

and $Z(k)$ are the noise output samples given by

$$Z(k) = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} z(n) \exp^{\frac{-j2\pi nk}{N}}, \quad 0 \leq k \leq N-1 \tag{4}$$

with $z(n)$ as i.i.d complex Gaussian noise samples, each of zero mean and variance σ_n^2.

3 DFT-Based Optimal Interpolation

From Eq. (3), it follows that the CFR vector H can be expressed in terms of the impulse response vector h, as

$$H = D_T h \tag{5}$$

where D_T is the truncated DFT matrix of size $N \times L$, obtained by selecting the first L columns of the $N \times N$ DFT matrix.

Now, consider the $P \times 1$ vector H_P that contains the CFR values at the pilot locations $p_0, p_1, p_2, \ldots, p_{P-1}$. We see that H_P can be expressed in terms of h as

$$H_P = D_P h \tag{6}$$

where D_P is the $P \times L$ matrix obtained by selecting the rows p_0, p_1, p_2, $\ldots$, p_{P-1} out of the N rows of D_T. Since it is assumed that $P \geq L$, D_P is either tall or square. Therefore, it follows that in the absence of noise, we will be able to recover the channel impulse response from the CFR at pilot locations if the matrix D_P is full rank.

Lemma 1 *The matrix D_P is full rank.*

Proof D_P is constructed by selecting first L columns and rows p_0, p_1, p_2, $\ldots$, p_{P-1} of the $N \times N$ DFT matrix. Hence, it is a Vandermonde matrix, with a distinct generator element for each row. Therefore, it is full rank.

Hence, the channel impulse response can be recovered as

$$h = D_P^\dagger H_P \tag{7}$$

where $D_P^\dagger$ is the pseudoinverse of D_P, given by $D_P^\dagger = \left(D_P^H D_P\right)^{-1} D_P^H$. Here, D_P^H is the Hermitian of the matrix D_P.

Once the impulse response coefficients are known, the CFR values at all subcarriers can be obtained by using Eq. (5). Therefore, we can finally express the CFR at all subcarriers in terms of that at pilot subcarriers as

$$H = D_T D_P^\dagger H_P \tag{8}$$

3.1 Practical Analysis

In the previous section, it was seen that we can estimate exactly the CFR values at all subcarriers from the CFR values at pilot locations if there is no noise, and if $P \geq L$. Here, we consider the question of optimal CFR estimation in the presence of noise. Recall that the received symbols are given by

$$Y(k) = H(k)X(k) + Z(k), \quad 0 \leq k \leq N - 1 \tag{9}$$

The symbols $X(k)$ at $k = p_0, p_1, p_2, \ldots, p_{P-1}$ are the known pilot symbols. Therefore, from the knowledge of $Y(k)$ at the pilot locations, the CFR values $H(k)$ at those locations can be estimated, using either ZF or MMSE criteria. Let $\hat{H}(k)$, $k = p_0, p_1, p_2, \ldots, p_{P-1}$ be the estimated CFR at pilot locations. Therefore, we have

$$\hat{H}(k) = H(k) + e(k), \quad k = p_0, \, p_1, \, p_2, \, \ldots, p_{P-1} \tag{10}$$

where $e(k)$ is the estimation error. If the CFR values at pilot locations are assumed to be estimated symbolwise, then the sequence $e(k)$ will be a zero mean white Gaussian sequence. Let its variance be σ_e^2. In vector notation, we can write

$$\hat{H}_{\mathrm{p}} = H_{\mathrm{P}} + e \tag{11}$$

Let $R_e = E[ee^H] = \sigma_e^2 I$ be the autocorrelation matrix of the noise vector. Applying $D_{\mathrm{P}}^{\dagger}$ on $\hat{H}_{\mathrm{P}}$, we get

$$\hat{h} = D_{\mathrm{P}}^{\dagger} \hat{H}_{\mathrm{p}} = D_{\mathrm{P}}^{\dagger} H_{\mathrm{P}} + D_{\mathrm{P}}^{\dagger} e \tag{12}$$

where $\hat{h}$ is the estimated channel impulse response. Now, applying D_{T} on $\hat{h}$, we have

$$\hat{H} = D_{\mathrm{T}} \hat{h} = D_{\mathrm{T}} D_{\mathrm{P}}^{\dagger} H_{\mathrm{P}} + D_{\mathrm{T}} D_{\mathrm{P}}^{\dagger} e \tag{13}$$

From Eq. (8), this reduces to

$$\hat{H} = H + D_{\mathrm{T}} D_{\mathrm{P}}^{\dagger} e \tag{14}$$

$$\hat{H} = H + w \tag{15}$$

Therefore, in the absence of estimation error, we obtain the exact CFR at all subcarriers. This means that we can perfectly interpolate the CFR at pilot subcarriers to all subcarriers, with zero interpolation error. The $N \times P$ matrix $D_{\mathrm{T}} D_{\mathrm{P}}^{\dagger}$ is the ideal interpolator from the CFR at pilot subcarriers to all subcarriers. The only remaining error component $w = D_{\mathrm{T}} D_{\mathrm{P}}^{\dagger} e$ is due to the estimation error. The variance of this component depends on $D_{\mathrm{P}}^{\dagger}$, which in turn depends on the choice of pilot positions. In the next section, we show that under some benign assumptions, the matrix D_{P} can be made unitary by a suitable choice of pilot locations, ensuring that the variance of the error component w remains σ_e^2.

3.2 Pilots on Harmonic Subcarriers

Assume that the number of subcarriers N is a power of 2, i.e., $N = 2^R$, R being a positive integer. Further, assume that the number of pilot subcarriers P is also a power of 2, i.e., $P = 2^{R_0}$, where $R_0 < R$ is a positive integer. These assumptions are seen to be satisfied in most practical systems. Define $M = N/P$. Let the P pilots be

located at the subcarriers $0, M, 2M, \ldots, (P - 1)M$. (In this section, all the indices start at 0.) Therefore, the matrix D_P is constructed by selecting the first L columns and the rows $0, M, 2M, \ldots, (P - 1)M$. Thus,

$$[D_P]_{kl} = \frac{1}{\sqrt{N}} \exp^{\frac{-j2\pi kMl}{N}} \tag{16}$$

where $k = 0, 1, \ldots, P - 1$ and $l = 0, 1, \ldots, L - 1$. Simplifying, we have

$$[D_P]_{kl} = \frac{1}{\sqrt{N}} \exp^{\frac{-j2\pi kl}{N/M}} \tag{17}$$

$$= \frac{1}{\sqrt{N}} \exp^{\frac{-j2\pi kl}{P}} \tag{18}$$

where we have used the fact that $P = N/M$. We now recognize D_P as the first L columns of the $P \times P$ DFT matrix (with the scaling factor $1/\sqrt{N}$) and hence unitary. It is easily seen that $D_P^H D_P = \frac{P}{N} I$. Therefore, we have

$$D_P^\dagger = \sqrt{\frac{N}{P}} D_P^H \tag{19}$$

Now, consider the variance of the error component w. Let $R_w = E[ww^H]$

$$\sigma_w^2 = \frac{1}{N} tr\ (R_w) \tag{20}$$

$$= \frac{1}{N} tr\left(D_T D_P^\dagger E[ee^H] \left(D_P^\dagger \right)^H D_T^H \right) \tag{21}$$

$$= \frac{1}{N} tr\left(\sigma_e^2 \sqrt{\frac{N}{P}} D_P^H \sqrt{\frac{N}{P}} D_P D_T^H \right) \tag{22}$$

Noting that $D_P^H D_P = \frac{P}{N} I$ and $D_T^H D_T = I$, we get

$$\sigma_w^2 = \sigma_e^2 \tag{23}$$

Thus, we see that selecting harmonic subcarriers as pilot locations ensures that the error in estimating the CFR at pilot subcarriers does not get enhanced when those noisy pilot CFR values are used to estimate the CFR at all subcarriers.

To summarize,

- If the CFR estimate at the pilot locations is noise-free, then our DFT-based interpolation results in exact CFR at all subcarriers.

- If the CFR estimate at the pilot locations has nonzero estimation error, then positioning the pilots at harmonic subcarriers ensures that this estimation error component is not enhanced during interpolation.

3.3 Other Interpolation Methods

In this section, we consider two popular approaches for interpolation.

3.3.1 Linear Interpolation

In this approach, a piecewise linear approximation is used to reconstruct the CFR values. The CFR at any subcarrier is approximated as a linear combination of the CFR values at the two neighboring pilots. Specifically, for the subcarriers between the pilots p_i and p_{i+1}, we have

$$\hat{H}(k) = \hat{H}(p_i) + \frac{k - p_i}{p_{i+1} - p_i} \left[\hat{H}(p_{i+1}) - \hat{H}(p_i) \right], \quad p_i \leq k < p_{i+1} \qquad (24)$$

Applying Eq. (10) and rearranging, we get

$$\hat{H}(k) = H(p_i) + \frac{k - p_i}{p_{i+1} - p_i} \left[H(p_{i+1}) - H(p_i) \right]$$
$$+ e(p_i) + \frac{k - p_i}{p_{i+1} - p_i} \left[e(p_{i+1}) - e(p_i) \right] \qquad (25)$$

$$= H(pi) + w_{\text{int}}(k) + w_{\text{e}}(k) \qquad (26)$$

where $w_{\text{int}}(k)$ is the interpolation error and $w_{\text{e}}(k)$ is error due to the noise at pilot locations. Thus, in contrast to our proposed DFT-based optimal interpolation, linear interpolation suffers from two kinds of errors.

3.3.2 Spline Interpolation

Spline interpolation is a piecewise polynomial interpolation technique that offers a better performance than polynomial interpolation. In this paper, we used the MATLAB function spline() to implement cubic spline interpolation.

4   Numerical Results and Discussion

In this section, we present the simulation results for an OFDM system with $N = 64$ subcarriers. The channel considered is a frequency-selective Rayleigh fading channel with length $L = 3$ and uniform power delay profile. The number of pilots used is $P = 4$. The pilot symbols are all equal to 1, and the pilots are placed at harmonic subcarriers. For the cubic spline interpolation, spline() function of MATLAB was used. For the BER simulations, BPSK modulation is considered.

- Figure 1 shows the mean-squared error (MSE) of channel estimation for the proposed optimal interpolation scheme, along with the linear and spline interpolation schemes. We see that the MSE for our proposed optimal interpolation scheme decreases linearly with E_b/N_0, in the log–log scale. This follows from the fact that the only error in optimal estimation is that due to noise at the receiver front end, the error variance equals that of the receiver front-end noise when the pilots are placed at harmonic subcarriers.

  The linear and spline interpolation approaches, on the other hand, exhibit MSE performances that degrade significantly as the E_b/N_0 increases. The performance gap between the optimal scheme and the other two schemes keeps increasing with increasing E_b/N_0. This phenomenon can be understood by considering Fig. 2.
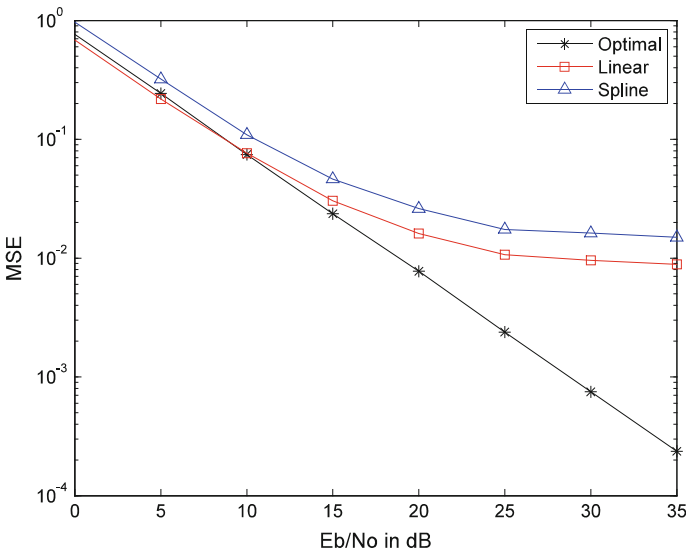
Fig. 1 Comparison of the mean-squared error (MSE) of the proposed optimal interpolation scheme with linear and spline interpolations
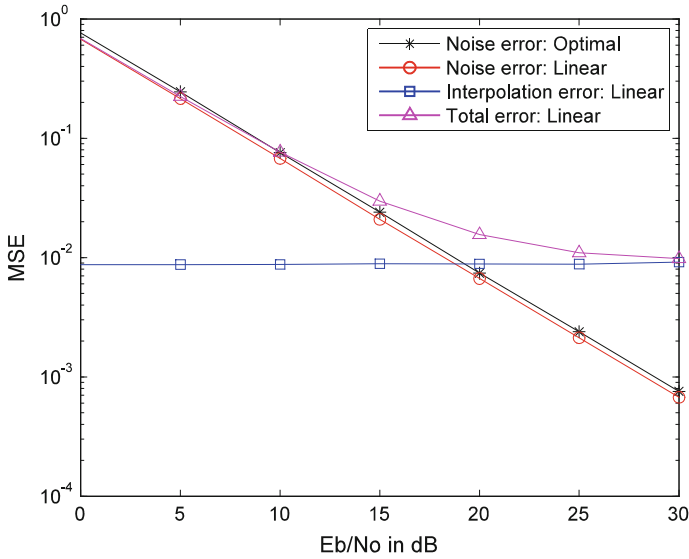
Fig. 2 Comparison of MSE of the proposed optimal interpolation scheme with linear interpolation, taking the noise error and interpolation error components of linear interpolation into account

- Figure 2 shows the breakup of the components of the MSE in the linear interpolation scheme, in comparison with the MSE of the optimal interpolation scheme. As can be observed from the figure, the contribution of the noise error component to the MSE of the linear interpolation scheme decreases linearly with E_b/N_0, in a manner similar to that of the optimal interpolation scheme. But in contrast to the optimal interpolation scheme, the linear scheme also has an interpolation error component that is constant and independent of E_b/N_0.

This component dominates at high SNR and makes sure that no performance improvement is possible by increasing E_b/N_0. The only way to reduce this component is to increase the number of pilot symbols, thus reducing spectral efficiency. Even with an increased number of pilot symbols, the interpolation error cannot be made zero with linear interpolation, whereas the optimal interpolation scheme achieves zero interpolation error with the minimal number of pilots $P \geq L$. The case of spline interpolation in this context is similar to that of linear interpolation.

- Figure 3 shows magnitude of the actual frequency response of a randomly generated channel, along with the different estimated responses. The figure also shows the noisy version of the actual frequency response, whose values at pilot
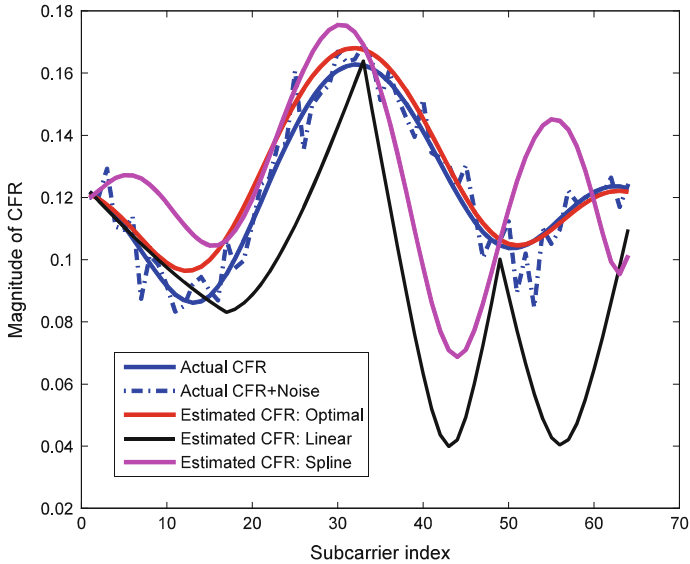
Fig. 3 Comparison of the magnitude of the actual frequency response of a random channel with that of different channel estimation schemes

locations are used by the different interpolation schemes. We see that the proposed optimal interpolation provides the closest match to the actual CFR.

Both linear and spline interpolation schemes show considerable deviation from the actual CFR especially at the subcarriers that are distant from pilot locations. An interesting aspect that can be noticed is that the magnitude of the linear interpolation is not piecewise linear between the subcarriers. But it can be seen that the real part of the CFR for the linear interpolation scheme is indeed piecewise linear. Similar comments hold for the imaginary part also.

- Figure 4 compares the BER performance of the OFDM system with $N = 64$ and BPSK input, for the actual and estimated CFRs. It is seen that the BER for optimal interpolation scheme has the same slope as the one using the actual CFR. This is in contrast to the linear interpolation, whose BER curve flattens at high SNR, resulting in significant performance degradation. As discussed before, this happens due to the interpolation error component that remains constant across SNRs.
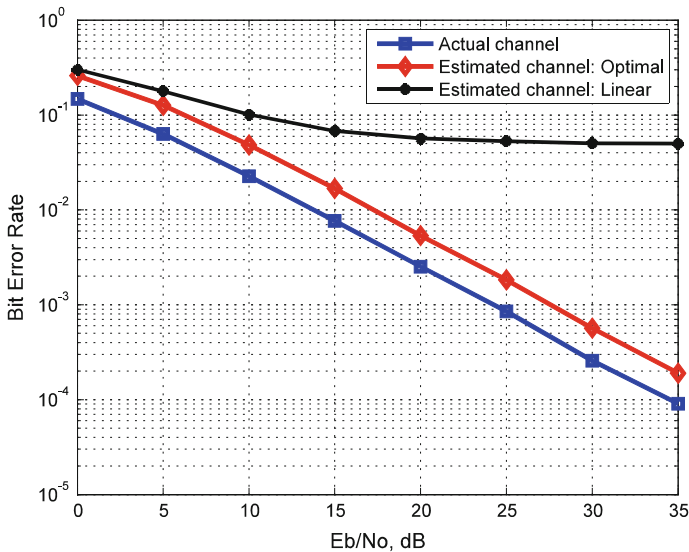
Fig. 4 Comparison of BER performance of OFDM with actual and estimated channel frequency response

5 Conclusions

In this paper, we analyzed a DFT-based channel estimation technique for comb-type pilot arrangement in an OFDM system. The CFR at pilot locations was used to obtain an estimate of the channel impulse response, using which the CFR at all subcarriers was estimated. We showed that this approach provides ideal interpolation, with zero interpolation error. The CFR values at all subcarriers are estimated exactly provided the CFR estimates at pilot locations are noise-free. We also showed that the effect of noise on the final CFR estimation can be minimized by the proper selection of pilot locations. Simulation results indicate that our approach provides significant performance gains over other interpolation techniques used in channel estimation, like linear and spline interpolations.

References

1. Pietrzyk S (2006) OFDMA for broadband wireless access. Artech House, USA
2. Shen Y, Martinez E (2006) Channel estimation in OFDM systems. In: Freescale semiconductor application note (2006)
3. Hsieh M-H, Wei C-H (1998) Channel estimation for OFDM systems based on comb-type pilot arrangement in frequency selective fading channels. IEEE Trans Consum Electron 44:217–225
4. Morelli M, Mengali U (2001) Comparison of pilot-aided channel estimation methods for OFDM systems. IEEE Trans Signal Process 49:3065–3073

5. Dong X, Lu W-S, Soong ACK (2007) Linear interpolation in pilot symbol assisted channel estimation for OFDM. IEEE Trans Wirel Commun 6:1910–1920
6. Coleri S et al (2002) Channel estimation techniques based on pilot arrangement in OFDM systems. IEEE Trans Broadcast 48(3):223–229
7. Tsai P-Y, Chiueh T-D (2004) Frequency-domain interpolation based channel estimation in pilot-aided OFDM systems. In: 2004 IEEE 59th vehicular technology conference, 2004. VTC 2004-Spring, vol 1. IEEE
8. Li Y (2000) Pilot-symbol-aided channel estimation for OFDM in wireless systems. IEEE Trans Veh Technol 49(4):1207–1215
9. Edfors O, Sandell M, Van De Beek JJ, Wilson SK, Brjesson PO (2000) Analysis of DFT-based channel estimators for OFDM. Wirel Pers Commun 12(1):55–70

Part IV
Microwave, Electromagnetics & Antennas

Investigation of Rectangular Dielectric Resonator Antenna with Varying Probe Length

Bidisha Biswas, Durjoy Roy and Manotosh Biswas

Abstract In this article, we have thoroughly investigated the change of resonant frequency, input impedance, quality factor, bandwidth, and radiation characteristics for a rectangular dielectric resonator antenna (RDRA) due to the change of probe length. Due to non-availability of theoretical and experimental results for a RDRA with varying probe length, we have employed Ansoft's high-frequency structure simulator (HFSS).

1 Introduction

Dielectric resonator antennas (DRAs), with the support of a ground plane, have received increased interest in recent years for their potential applications in microwave and millimeter wave communication systems. Analysis and studies on characteristic equation, radiation patterns, and excitation methodology made DRAs popular. Both the limitations of low gain and low bandwidth in patch antennas can be eliminated by the use of a rectangular dielectric resonator antenna (RDRA) operating in resonant modes. Several researchers have investigated the RDRA [1–15]. But in these studies, the effect of probe length on the characteristics of

B. Biswas (✉)
Bhairab Ganguly College, Belghoria, Kolkata 700056, India
e-mail: bidisha005@gmail.com

D. Roy
Rishi Bankim Chandra College, Naihati, 24 PGS (N), India
e-mail: roy.durjoy@gmail.com

M. Biswas
Department of Electronics & Tele-Communication Engineering, Jadavpur University, 188, Raja Subodh Chandra Mullick Road, Kolkata 700032, India
e-mail: mbiswas@ieee.org

RDRA has not been reported. Due to the change of probe length, the resonant frequency, input impedance, bandwidth, gain, and radiation pattern are changed. The RDRA designed without consideration of this probe-length effect will not perform as expected once installed in a microwave and millimeter wave communication systems. So, the investigation of resonant frequency, quality factor, bandwidth, input impedance, and gain of a RDRA is very important. To the best of our knowledge, neither any theoretical nor any experimental results for resonant frequency, quality factor, bandwidth, input impedance and gain of a RDRA are available in open literature.

We have addressed this problem and provide the results for resonant frequency, input impedance, bandwidth, gain, and radiation pattern of a RDRA with varying probe length using an electromagnetic software (HFSS).

2   Design of RDRA

Figure 1 shows the grounded rectangular DRA with three-dimensional parameters (a, d, h), and it is excited by a coaxial probe whose length (L_P) is varying inside the RDRA. The relative permittivity of the DRA is assumed to be ε_{rd}. For the coaxial probe structure where the probe is along the z-axis, the field for the fundamental mode of concern here is TE_{111}^x. In the conventional waveguide model, the calculation of the resonant frequency is based on a waveguide structure with a perfect electric conductor.

(PEC) at $z = 0$, and the three PMC walls are located at $y = \pm a/2$ and $z = h$ planes, respectively. We have previously stated that no design guideline is available to compute the resonant frequency of a RDRA with varying probe length. In order to obtain the dimension of the RDRA, initially we have employed the expression without probe-length effect. Applying the boundary conditions of the continuity of the tangential electric and magnetic field at $x = \pm d$ planes, the eigenvalue equation determines the resonant frequency of the structure as

$$k_x \sin\,(k_x d/2) = \gamma_x \cos(k_x d/2) \tag{1}$$

where

$$k_x = \left[\in_{rd} k_0^2 - k_y^2 - k_z^2 \right]^{1/2} \tag{2}$$

and

$$\gamma_x = \left[k_y^2 + k_z^2 - k_0^2 \right]^{1/2} \tag{3}$$

$$k_x^2 + k_y^2 + k_z^2 = k_0^2 \in_{rd} \tag{4}$$
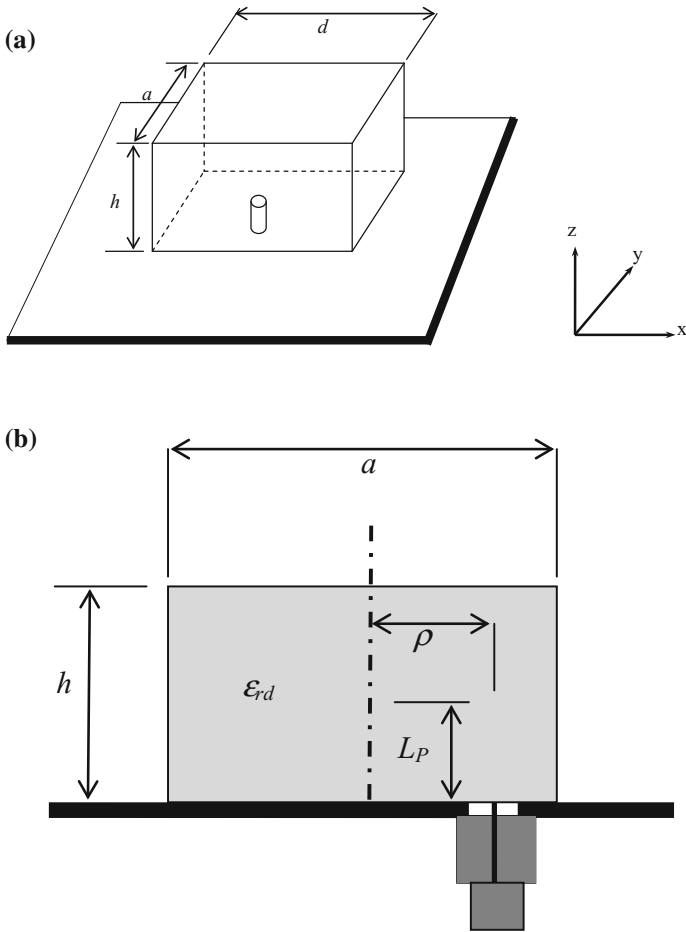
Fig. 1 Antenna geometry and feed mechanism

$$k_y = \Pi/a \tag{5}$$

$$k_z = \Pi/2h \tag{6}$$

The resonant frequency of the TE^x_{111} mode of this antenna (Fig. 1) can be obtained using the relation [7]

$$f^{(x)}_{111} = \frac{c}{2\Pi\sqrt{\epsilon_{rd}}} \left[k_x^2 + k_y^2 + k_z^2 \right]^{1/2} \tag{7}$$

In this analysis, we have taken $a = d = h = 15$ mm, $\varepsilon_{rd} = 8.9$, $\tan \delta = 0.002$, probe dia $g = 1.24$ mm, the feed is located at a distance $\rho = 6.0$ mm from the

center of the RDRA toward the edge, and the probe length (L_P) is varying inside the RDRA.

3   Results and Discussions

Figures 2 and 3 show the variation of resonant frequency (f_r) and input impedance (R & X) as a function of probe length (L_P). The f_r decreases with the increase of L_P, and the input resistance at resonance (R_r) increases with the increase of L_P. Good impedance matching is seen for $L_P = 7.0$ mm.

Fig. 2 Variation of input impedance with the variation of probe length (L_P) for a rectangular dielectric resonator antenna. $a = d = h = 15$ mm, $\varepsilon_{rd} = 8.9$, tan $\delta = 0.002$, $\rho = 6$ mm
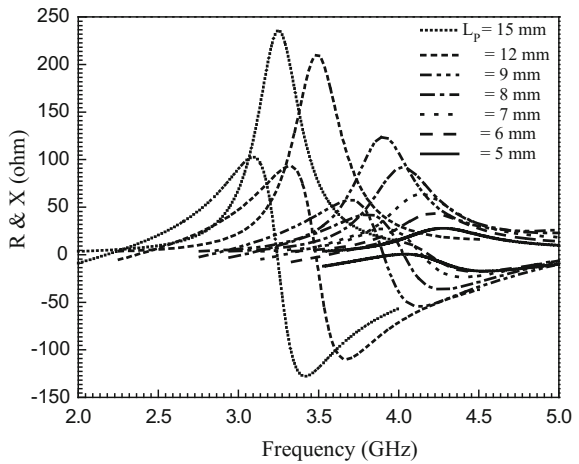
Fig. 3 Variation of resonant frequency (f_r) and resonant resistance (R_r) as a function of probe length (L_P) for a rectangular dielectric resonator antenna. $a = d = h = 15$ mm, $\varepsilon_{rd} = 8.9$, tan $\delta = 0.002$, $\rho = 6$ mm
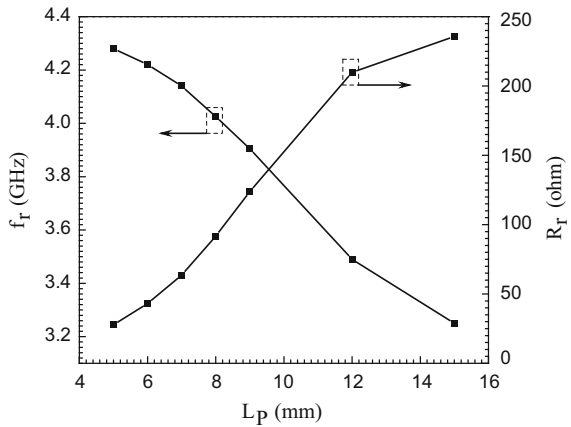
Fig. 4 Variation of bandwidth (B. W. %) and total quality factor (Q_T) as a function of probe length (L_P) for a rectangular dielectric resonator antenna. $a = d = h = 15$ mm, $\varepsilon_{rd} = 8.9$, $\tan \delta = 0.002$, $\rho = 6$ mm
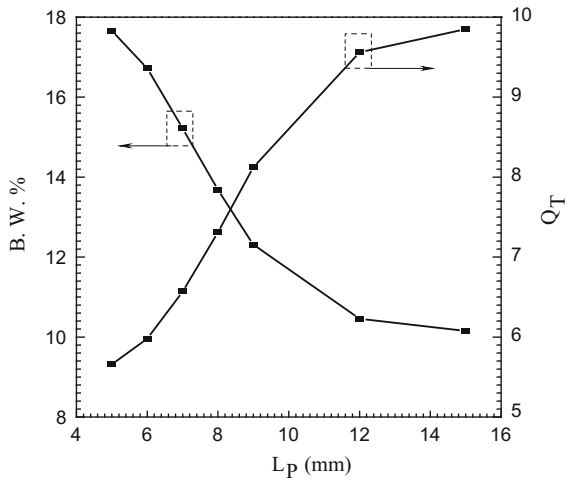
The variation of total quality factor (Q_T) and percentage bandwidth with the variation of L_P is depicted in Fig. 4. The Q_T increases with the increase of L_P, whereas the percentage bandwidth is decreased with the increase of L_P.

The variation of E-plane co-polarized and H-plane co- and cross-polarized gain pattern with the variation of L_P is depicted in Figs. 5 and 6, respectively. The E- and H-plane co-polarized gain decreases with the increase of L_P. The large value of L_P causes the distortion in E-plane co-polarized gain pattern. The H-plane cross-polarized gain increases with the increase of L_P (Fig. 7).

Fig. 5 Variation of E-plane co-polarized gain pattern with the variation of probe length (L_P) for a rectangular dielectric resonator antenna. $a = d = h = 15$ mm, $\varepsilon_{rd} = 8.9$, $\tan \delta = 0.002$, $\rho = 6$ mm
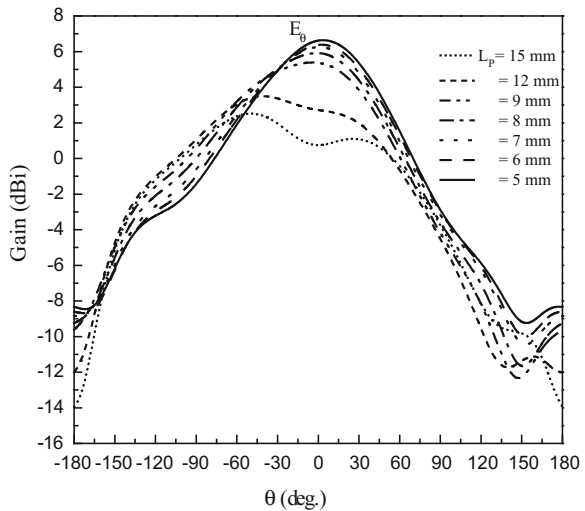
Fig. 6 Variation of H-plane co-polarized and cross-polarized gain pattern as a function of probe length (L_P) for a rectangular dielectric resonator antenna. $a = d = h = 15$ mm, $\varepsilon_{rd} = 8.9$, $\tan \delta = 0.002$, $\rho = 6$ mm
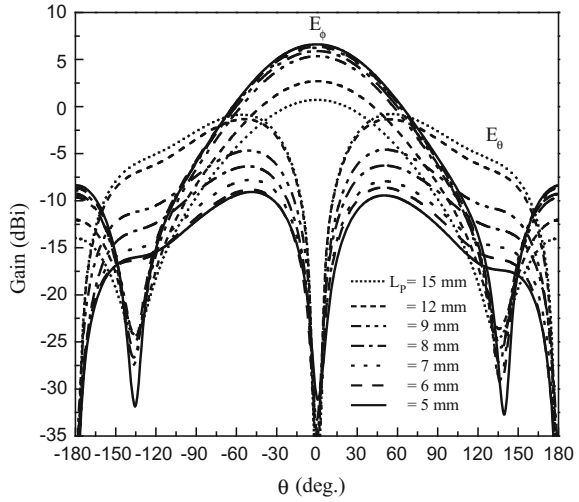
Fig. 7 Variation of resonant frequency (f_r) and peak gain with the variation of probe length (L_P) for a rectangular dielectric resonator antenna. $a = d = h = 15$ mm, $\varepsilon_{rd} = 8.9$, $\tan \delta = 0.002$, $\rho = 6$ mm
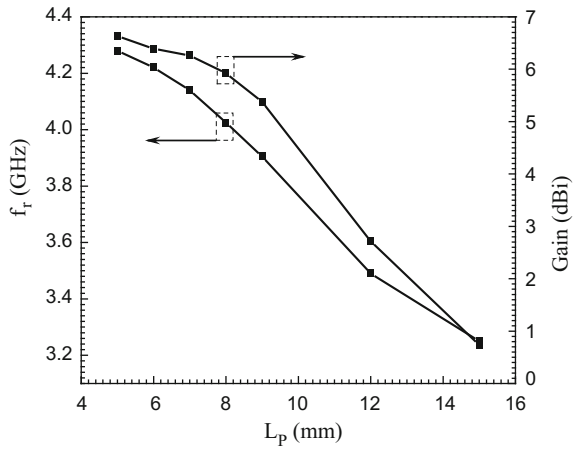
In Fig. 6, authors represent the effect of change of L_P on peak gain. The peak gain decreases with the increase of L_P.

The numerical values for f_r, R_r, Q_T, B. W. %, and gain of a RDRA with variable L_P are presented in Table 1.

Table 1 Values of f_r, input resistance at resonance (R_r), total quality factor (Q_T), bandwidth %, and gain of a RDRA with varying L_P; $a = d = h = 15$ mm, $\varepsilon_{rd} = 8.9$, tan $\delta = 0.002$, $\rho = 6$ mm

L_P (mm)	f_r (GHz)	R_r (ohm)	Q_T	B. W. %	Gain (dBi)
5	4.275	27.79	5.66	17.66	6.63
6	4.215	43.14	5.98	16.72	6.39
7	4.140	63.53	6.57	15.22	6.26
8	4.020	91.66	7.31	13.68	5.92
9	3.90	123.87	8.12	12.31	5.375
12	3.49	209.84	9.56	10.46	2.715
15	3.25	235.59	9.85	10.153	0.794

4 Conclusion

In this article, we have thoroughly investigated the effect of change of probe length on the resonant frequency, input impedance, bandwidth, and gain of a rectangular dielectric resonator antenna. The resonant frequency decreases, input impedance increases, peak gain decreases, cross-polarized level increases with the increase of probe length. The E-plane co-polarized gain pattern is distorted when the probe length becomes larger.

References

1. Van Bladel J (1975) On the resonances of a dielectric resonator of very high permittivity. IEEE Trans Microw Theory Tech MTT-23:199–208
2. St. Martine JTH, Antar YMM, Kishk AA, Ittipiboon A, Cuhaci M (1990) Dielectric resonator antenna using aperture coupling. Electron Lett 26:2015–2016
3. Mongia RK (1992) Theoretical and experimental resonant frequencies of rectangular dielectric resonators. IEE Proc H 139:98–104
4. Mongia RK, Ittipiboon A, Cuhaci M (1994) Low profile dielectric resonator antennas using a very high permittivity material. Electron Lett 30:1362–1363
5. Mongia RK, Bhartia P (1994) Dielectric resonator antennas—a review and general design relations for resonant frequency and bandwidth. Int J Microw Millim Wave Comput Aided Eng 4(3):230–247
6. Mongia RK, Ittipibon A, Cuhaci M (1994) Measurement of radiation efficiency of dielectric resonator antennas. IEEE Microw Guided Wave Lett 4(3):80–82
7. Mongia RK, Ittipiboon A (1997) Theoretical and experimental investigations on rectangular dielectric resonator antennas. IEEE Trans Antennas Propag 45:1348–1356
8. Petosa A, Ittipibon A, Antar YMM, Roscoe D, Cuhaci M (1998) Recent advances in dielectric resonator antenna technology. IEEE Antennas Propag Mag 40(3):35–48
9. Antar YMM, Cheng D, Seguin G, Henry B, Keller MG (1998) Modified waveguide model (MWGN) for rectangular dielectric resonator antenna (DRA). Microw Opt Technol Lett 19:158–160
10. Petosa A, Simons N, Siushansian R, Ittipibon A, Cuhaci M (2000) Design and analysis of multisegment dielectric resonator antennas. IEEE Trans Antennas Propag 48:738–742
11. Baghaee RM, Neshati MH, Mohassel JR (2008) Rigorous analysis of rectangular dielectric resonator antenna with a finite ground plane. IEEE Trans Antennas Propag 56:2801–2809

12. Yau D, Shuley MV (1999) Numerical analysis of an aperture coupled rectangular dielectric resonator antenna using a surface formulation and the method of moments. Proc Inst Elect Eng Antennas Propag 146:105–110
13. Baghaee RM, Neshati MH, Lashkari D, Koochakzadeh M, Mohassel JR (2005) Analysis of probe-fed rectangular dielectric resonator antennas on a finite ground plane using method of moment. In: Proceedings of ANTEM, Saint Malo, France, pp 392–393
14. Baghaee RM, Neshati MH, Mohassel JR (2005) Moment method analysis of probe-fed rectangular dielectric resonator antennas with a rigorous source modeling and finite ground plane. In: Proceedings of APMC, Suzhou, China, pp 2927–2930
15. Kishk AA, Glisson AW, Kajfez D (2003) Numerical analysis of stacked dielectric resonator antennas excited by a coaxial probe for wideband applications. IEEE Trans Antennas Propag 51:1996–2006
16. HFSS 13: Ansoft's Corp.

Design, Simulation and Performance Comparison of 1, 2, 4 and 8 Elements Quarter-Wave Transformer-Fed Circular Patch Antenna Array at L Band for Airborne Applications

U. Srinivasa Rao and P. Siddaiah

Abstract Of late, microstrip patch antenna in an airborne application is gaining importance as one of the most powerful technological trends. It has the immense potential, in airborne applications due to its low weight, low cost, its conformal and low profile. Hence, the present technological trends focus on microstrip patch antenna. However, it has disadvantage too; low gain, low efficiency, low directivity and narrow bandwidth. Nevertheless, these disadvantages can be overcome by implementation of many patch antennas in array configuration. Research studies reveal that as we increase number of patch elements to form an array, an increment in gain and decrement in beamwidth are noticed. However, this chapter elucidates modeling and simulation of a microstrip line quarter-wave transformer-fed 1, 2, 4 and 8 circular patch antenna array. The substrate material used for these antennas is RT/Duroid 5880 with a thickness of 1.588 mm, and dielectric constant (ε_r) is 2.2 and has a design frequency of 2 GHz and VSWR ≤ 2. The proposed antennas are thus modeled and simulated using ANSOFT HFSS. The radiation characteristics of the above arrays are compared.

Keywords Quarter-wave transformer (QWT) · Circular patch antenna (CPA)
Gain · Return loss · Beamwidth

U. Srinivasa Rao (✉)
Department of ECE, Vignan's Lara Institute of Technology & Science,
Guntur, AP, India
e-mail: usrao75@gmail.com

P. Siddaiah
University College of Engineering and Technology,
Acharya Nagarjuna University, Guntur, AP, India
e-mail: siddaiah_p@yahoo.com

1 Introduction

Microstrip antennas are one of the most popular geometries inexpensive to fabricate and can be easily made conformal to the host body. The attractive features of the microstrip enhanced their application in the recent past and stimulated an ever-increasing attention from all around to investigate their performance further. The circular micro strip radiator is the smaller configuration of all the other shapes. As, feed can be connected at any point along the periphery of circular micro strip, circular geometries have certain advantages over others, in few applications like arrays. This chapter presents the results of circular patch antenna fed with QWT. The QWT is used to match a 50 Ω feed transmission line with the circular microstrip patch. This model has been explained earlier in [1, 2]. This chapter also discusses the design of 2, 4 and 8 element of circular patch array antenna useful for airborne applications.

2 Design of QWT Fed Circular Patch Antenna

The geometry of QWT fed circular microstrip patch antenna is shown in Fig. 1 [3]. The design of the circular patch antenna is divided in three steps as given below.

A. **Calculation of radius of Circular Patch**

The expression for radius 'R' of circular patch is given by [4, 5]:

$$R = \frac{F}{\left[1 + \frac{2h}{\pi \varepsilon_r F \left[ln\left(\frac{F\pi}{2h}\right) + 1.7726 \right]}\right]^{1/2}} \tag{1}$$

where $F = \frac{8.791 \times 10^9}{f_r \sqrt{\varepsilon_r}}$

R = Circular patch radius (mm);
h = Substrate Height (mm);
f_r = Resonant Frequency (Hz);
ε_r = Relative Dielectric Constant of the Substrate.

Using the above expression, the calculated radius is 28.52 mm for 2 GHz operating frequency, dielectric constant is 2.2 and the height of the RT/Duroid 5880 substrate is 1.588 mm.
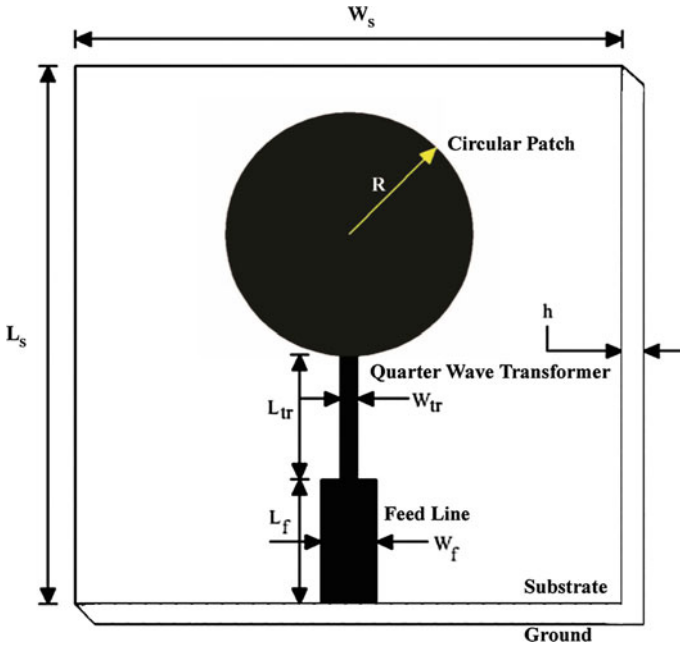
Fig. 1 Geometry of the circular patch antenna

B. **Design of a Microstrip line feed**

To excite the circular microstrip patch, a 50 Ω microstrip line is used. The width of the microstrip line calculated with the known values of the characteristic impedance Z_o (50 Ω) and dielectric constant of the substrate ε_r (2.2) using the standard equations is given below [6].

$$\frac{W_f}{h} = \begin{cases} \dfrac{8e^A}{e^{2A}-2} & \text{for } \frac{W_f}{h} < 2 \\ \dfrac{2}{\pi}\left[B - 1 - \ln(2B - 1) + \frac{\varepsilon_r-1}{\varepsilon_r+1}\left\{\ln(B - 1) + 0.39 - \frac{0.61}{\varepsilon_r}\right\}\right] & \text{for } \frac{W_f}{h} > 2 \end{cases}$$

(2)

where

$$A = \frac{Z_o}{60}\sqrt{\frac{\varepsilon_r + 1}{2}} + \frac{\varepsilon_r - 1}{\varepsilon_r + 1}\left\langle 0.23 + \frac{0.11}{\varepsilon_r}\right\rangle$$

(3)

and

$$B = \frac{377\pi}{2Z_o\sqrt{\varepsilon_r}} \tag{4}$$

The calculated width of 50 Ω microstrip line is 4 mm.

C. Design of the (QWT)

To match load impedance with source impedance, the simple device is QWT. From the known values of characteristic impedance of transmission Z_c is given by [7]:

where Z_0 the characteristic impedance of the 50 Ω is line Z_o and input impedance of the circular patch antennas Z_{in}, the impedance of the QWT is calculated as given by [7]:

$$Z_c = \sqrt{Z_o Z_{in}} \tag{5}$$

With the above Eq. (5), the Z_c of the impedance transformer is 130 Ω when Z_o is 50 Ω and Z_{in} is 340 Ω. The width of the 130 Ω QWT is calculated using (3), and the calculated value is 0.74 mm. The geometry of the proposed circular patch antenna is shown in Fig. 1.

Dimensions of proposed single circular patch antenna are given in Table 1.

The modeled structure of single circular patch antenna is as shown in Fig. 2 with a maximum size of antenna as 200 mm × 350 mm × 24 mm.

The performance characteristics like return loss, bandwidth, VSWR, elevation beamwidth, azimuth beamwidth, gain and efficiency of single element circular patch antenna, respectively, are shown in Figs. 3, 4, 5, 6, 7, 8, 9.
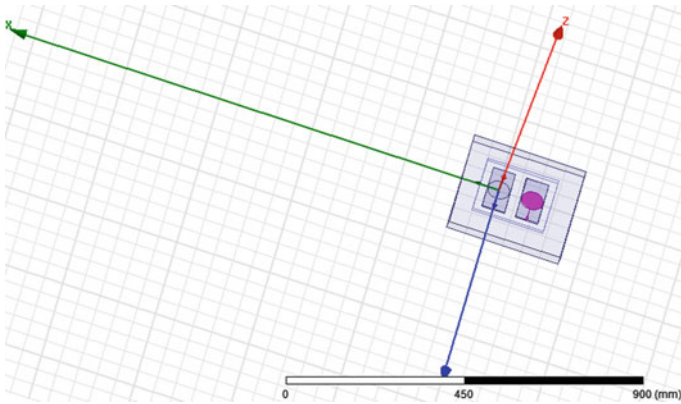
The simulated result of return loss of the single CPA is obtained as −19.3 dB.

Bandwidth is the range of frequencies over which antenna works satisfactorily. The simulated result of bandwidth of the single CPA is obtained as 436 MHz.

The simulated result of VSWR of single CPA is obtained as 1.8.

Beamwidth is in general angular separation between half power points. The simulated result of elevation beamwidth of the single CPA is 86.6°.

The simulated result of azimuth beamwidth of the single CPA is obtained as 74.8°.

Table 1 Calculated dimensions of proposed single CPA

Parameter	Value (mm)
Radius of CPA (R)	28.52
Substrate height (h)	1.588
Substrate length (L_s)	138.52
Substrate width (W_s)	60
length (L_{tr})	28.62
width (W_{tr})	0.74
Feed line length (L_f)	10
Feed line width (W_f)	4
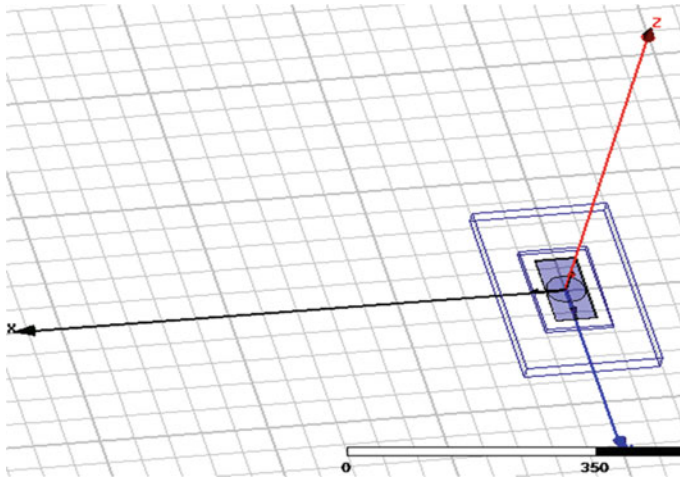
Fig. 2 Structure of modeled single circular patch antenna
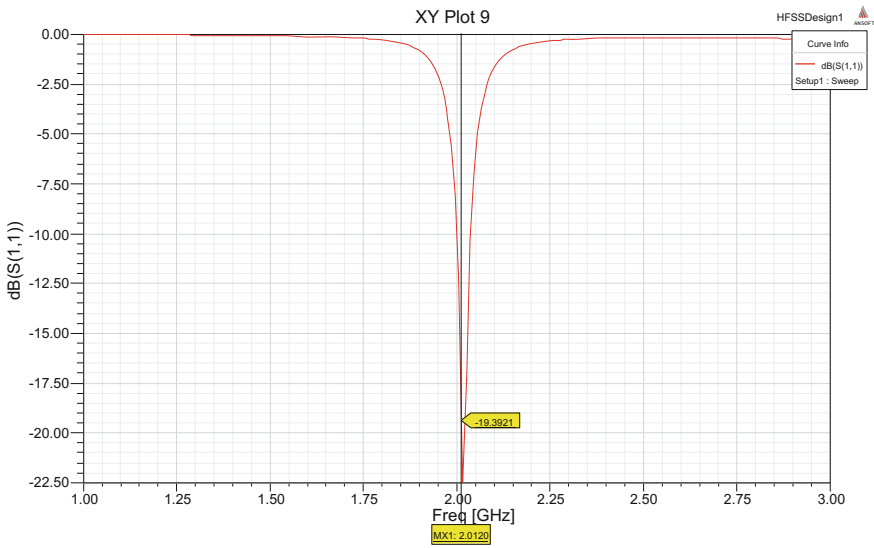
Fig. 3 Return loss of Single Antenna

Gain is the maximum power directed in the desired direction. The simulated result of gain of the single CPA is obtained as 5.78 dB.

Efficiency is the antenna parameter which indicates how effectively antenna radiates. The simulated result of efficiency of single CPA is obtained as 98%.
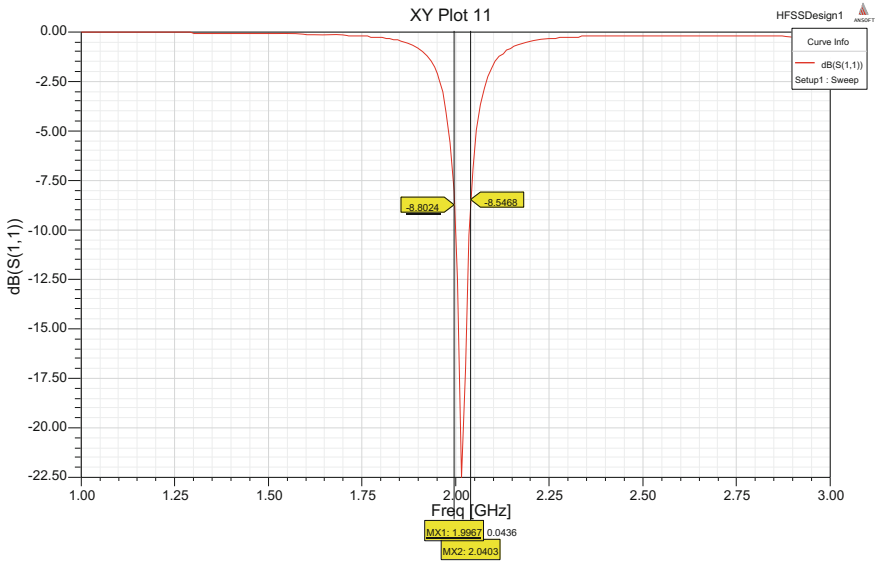
Fig. 4 Bandwidth calculation of Single Antenna
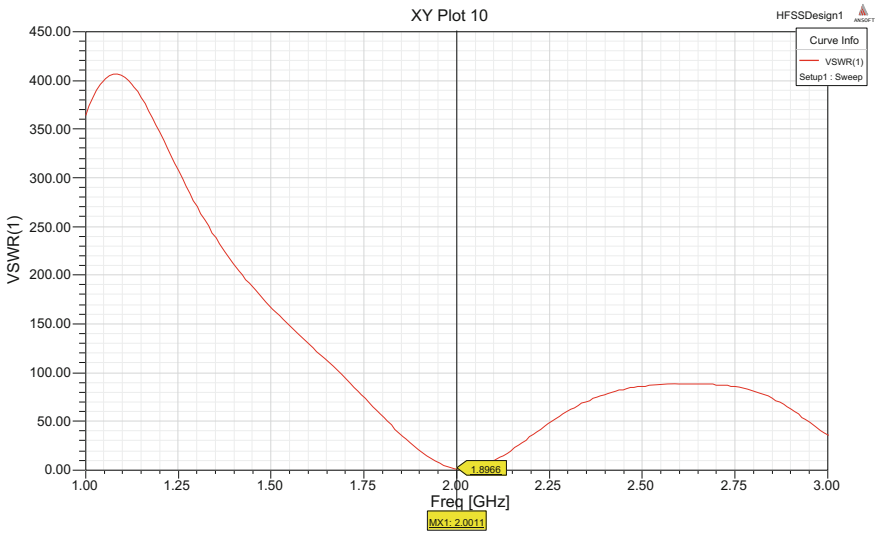
Fig. 5 VSWR of Single Antenna
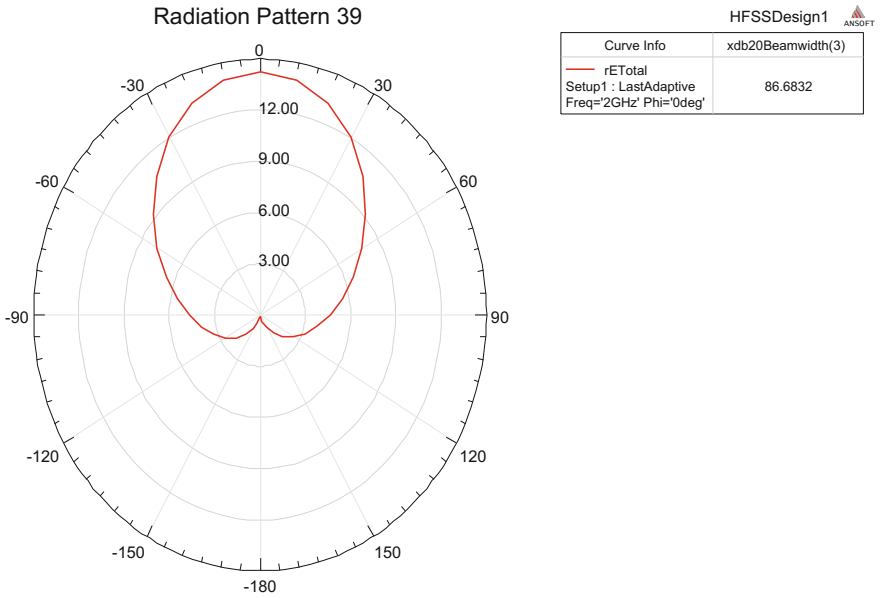
Fig. 6 Elevation beamwidth of Single Antenna
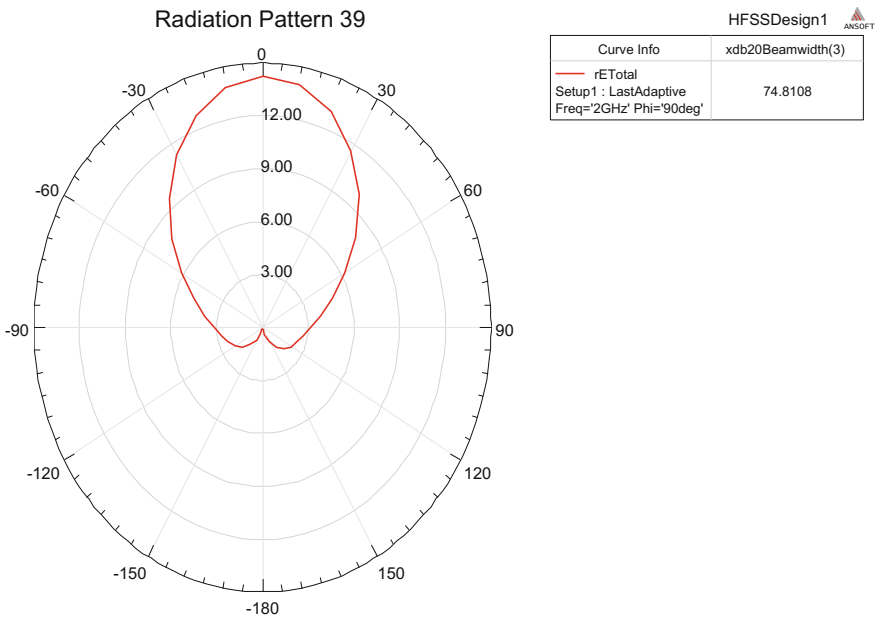
Fig. 7 Azimuth beamwidth of Single Antenna
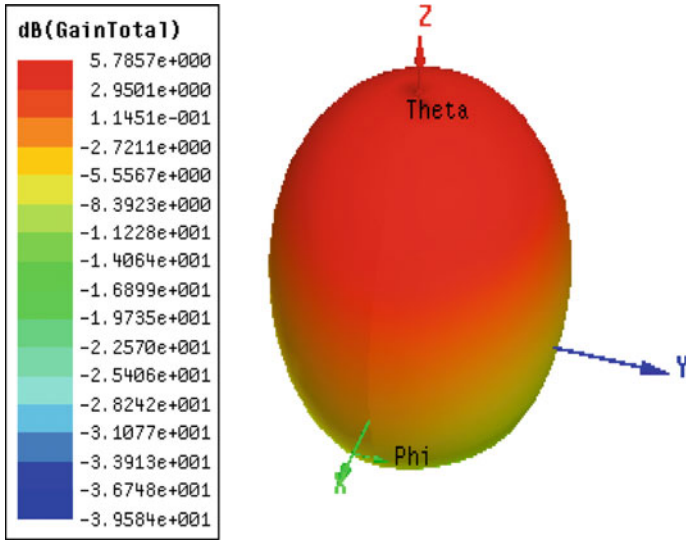
Fig. 8 Gain of Single Antenna
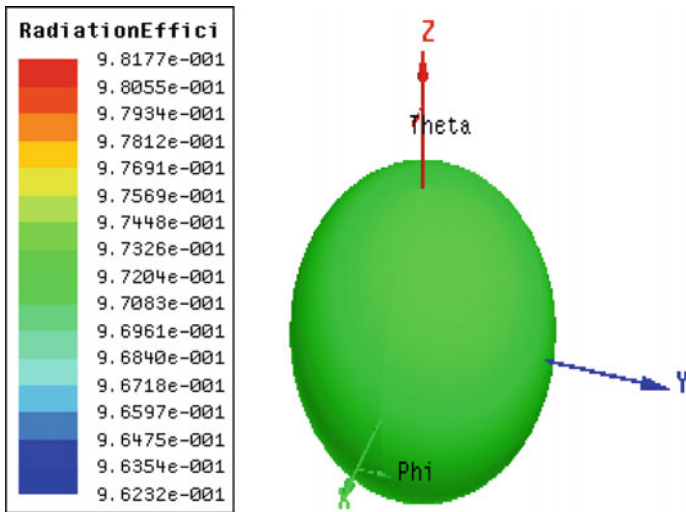
Fig. 9 Efficiency of Single Antenna

3 Performance of 2 Elements Array Antenna

The structure of the 2 elements array is shown in Fig. 10, and the spacing between the two elements is $\lambda/2$, with a maximum size of 2 element array antenna as 300 mm $\times$ 300 mm $\times$ 24 mm. All the elements in the array are fed with same amplitude and phase.

The performance characteristics like return loss, bandwidth, VSWR, elevation beamwidth, azimuth beamwidth, gain and efficiency of 2 element circular patch array antenna, respectively, are shown in Figs. 11, 12, 13, 14, 15, 16, 17.

Fig. 10 Structure of the 2 elements array
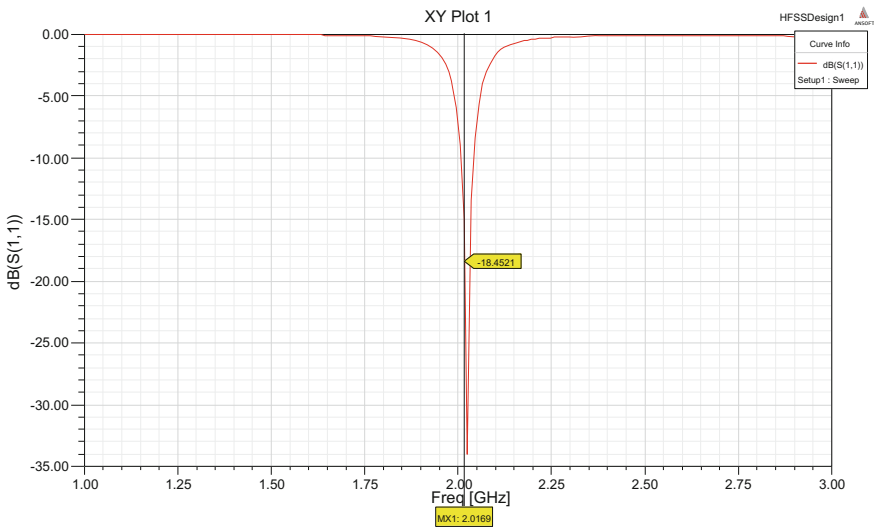
Fig. 11 Return loss
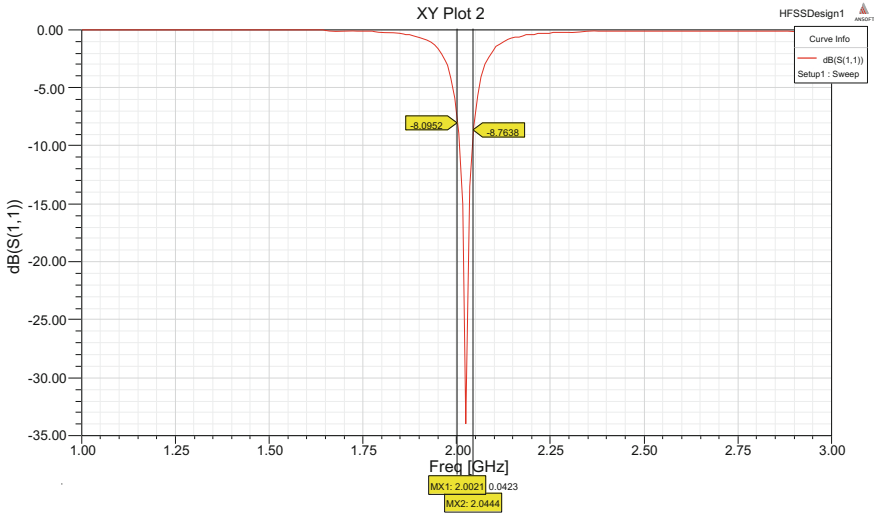
Fig. 12 Bandwidth calculation of 2 Elements Array Antenna
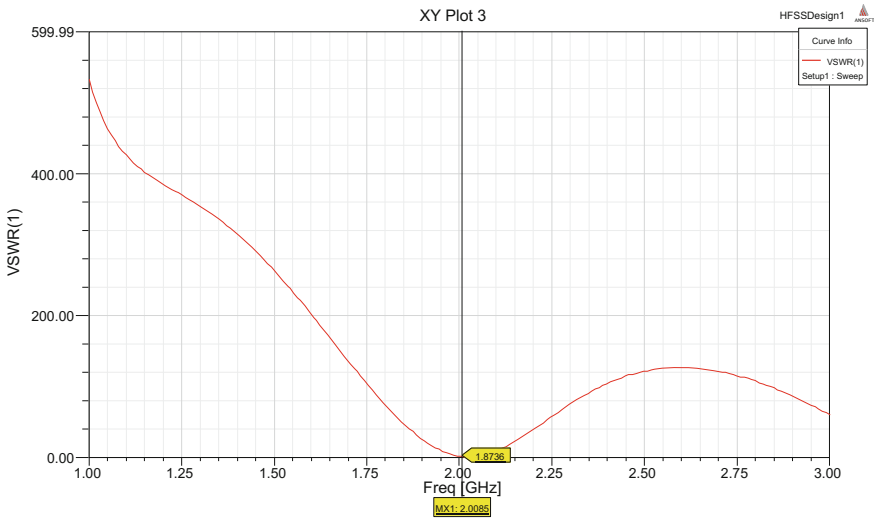
Fig. 13 VSWR of 2 Elements Array Antenna

The simulated result of return loss of the 2 element CPA array is obtained as 18.4 dB.

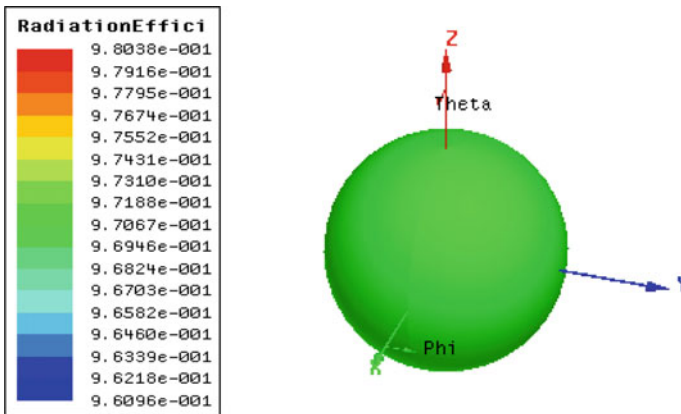The simulated result of bandwidth of the 2 element CPA array is obtained as 423 MHz.

The simulated result of VSWR of the 2 element CPA array is obtained as 1.8.
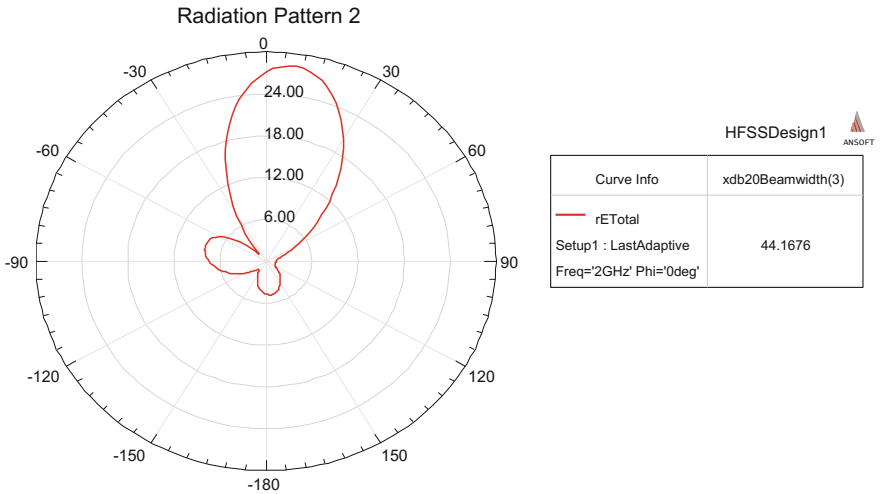
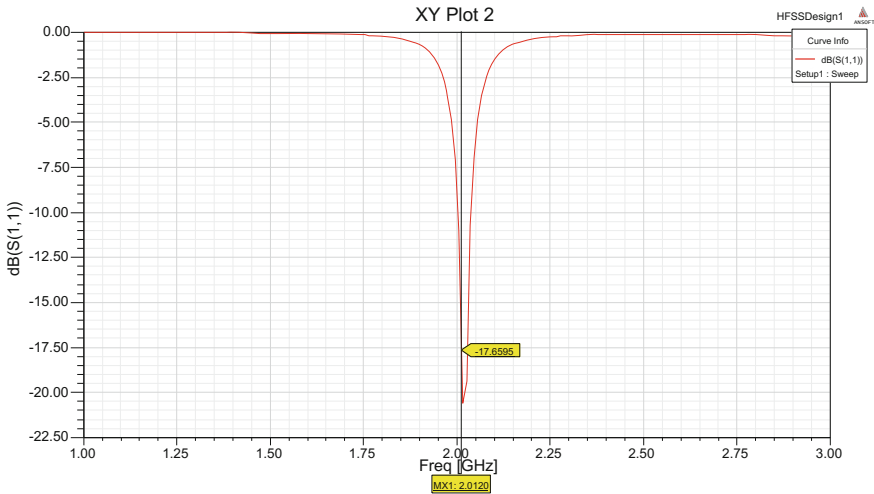**Fig. 14** Elevation beamwidth of 2 Elements Array Antenna
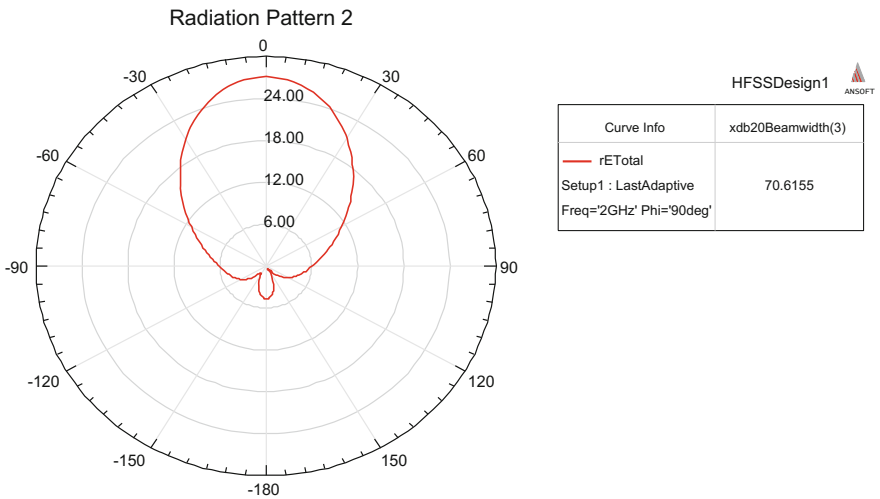
Fig. 15 Azimuth beamwidth of 2 Elements Array Antenna

The simulated result of azimuth beamwidth of the 2 element CPA array is obtained as 44.1°.

The simulated result of azimuth beamwidth of the 2 element CPA array is obtained as 70.6°.

The simulated result of gain of the 2 element CPA array is obtained as 8.9 dB.

The simulated result of efficiency of the 2 element CPA array is obtained as 97%.
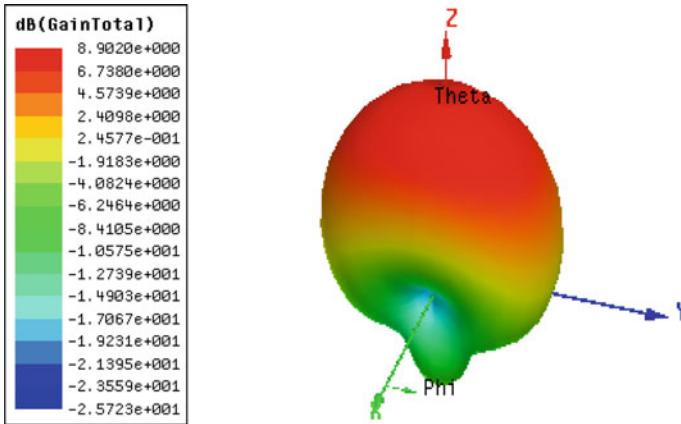
Fig. 16 Gain of 2 Elements Array Antenna

Fig. 17 Efficiency of 2 Elements Array Antenna

4 Performance of 4 Elements Array Antenna

The structure of the 4 elements array is shown below Fig. 18, and the spacing between the two elements is $\lambda/2$, with a maximum size of 4 element array antenna as 500 mm $\times$ 300 mm $\times$ 24 mm. All the elements in the array are fed with same amplitude and phase.

The performance characteristics like return loss, bandwidth, VSWR, elevation beamwidth, azimuth beamwidth, gain and efficiency of 4 element circular patch array antenna, respectively, are shown in Figs. 19, 20, 21, 22, 23, 24, 25.

The simulated result of return loss of the 4 element CPA array is obtained as −17.6 dB.
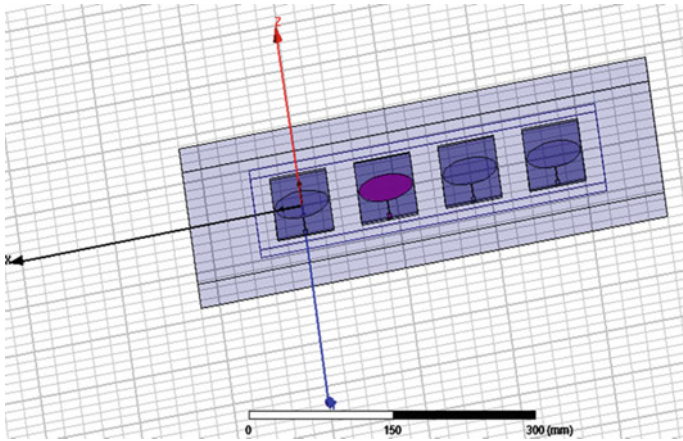
Fig. 18 Structure of the 4 element array

Fig. 19 Return loss of 4 Elements Array Antenna

The simulated result of bandwidth of the 4 element CPA array is obtained as 414 MHz.

The simulated result of VSWR of the 4 element CPA array is obtained as 1.9.

The simulated result of elevation beamwidth of the 4 element CPA array is obtained as 20.5°.

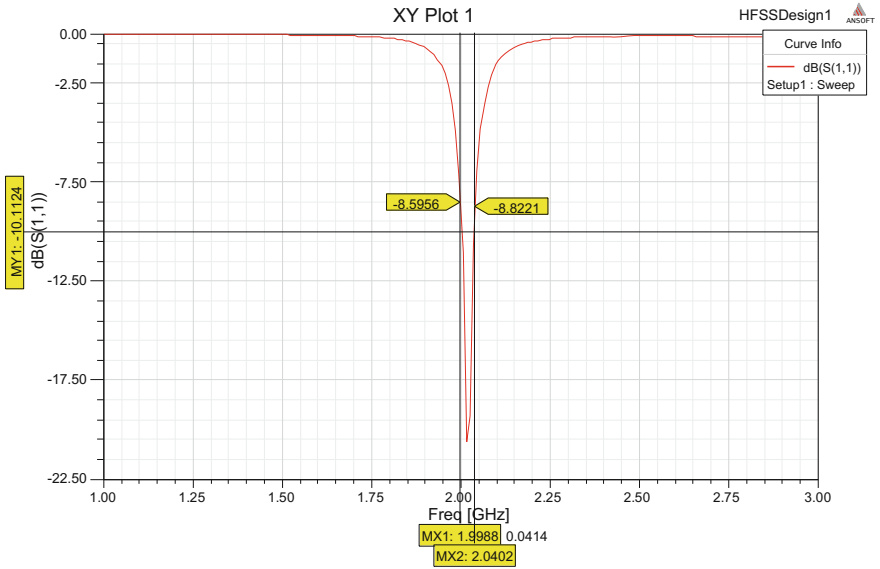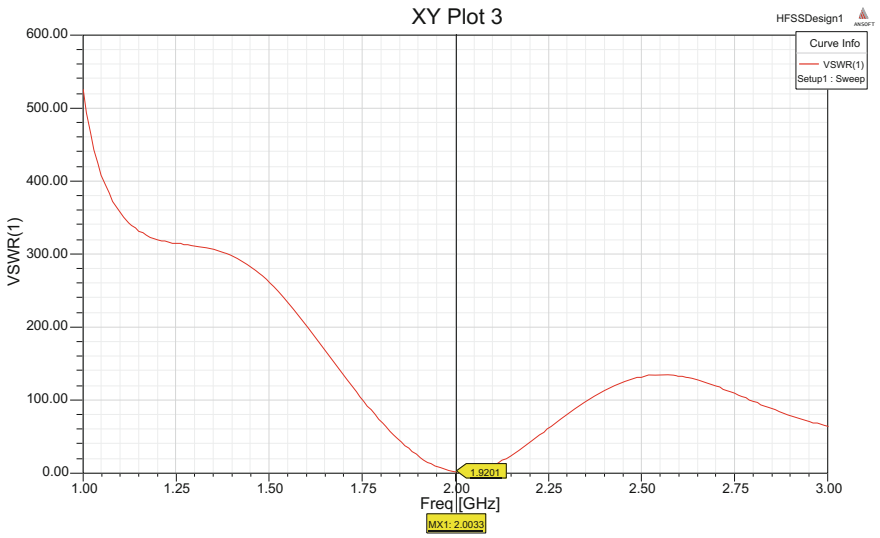The simulated result of azimuth beamwidth of the 4 element CPA array is obtained as 20.5°.

Fig. 20 Bandwidth calculation of 4 Elements Array Antenna

Fig. 21 VSWR of 4 Elements Array Antenna

The simulated result of gain of the 4 element CPA array is obtained as 9.8 dB.

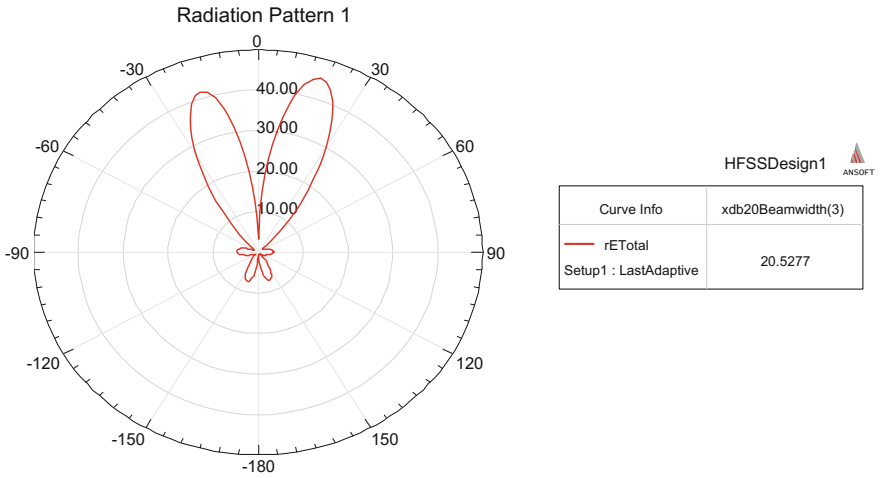The simulated result of efficiency of the 4 element CPA array is obtained as 97%.

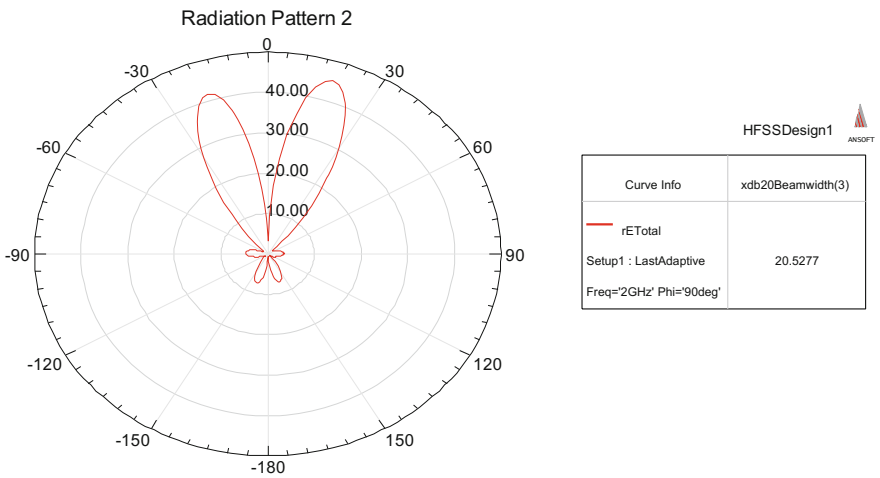**Fig. 22** Elevation beamwidth of 4 Elements Array Antenna

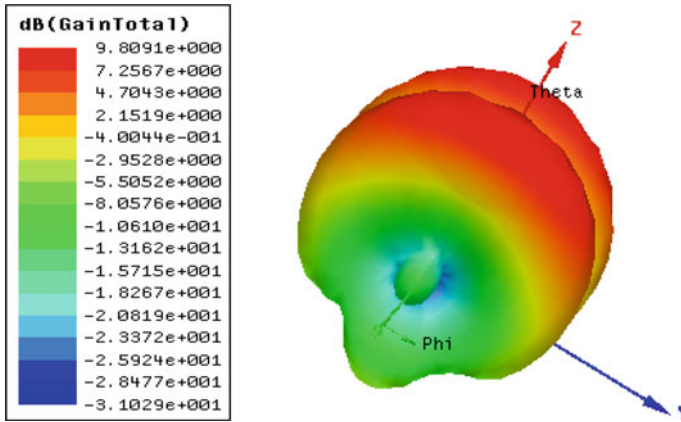**Fig. 23** Azimuth beamwidth of 4 Elements Array Antenna

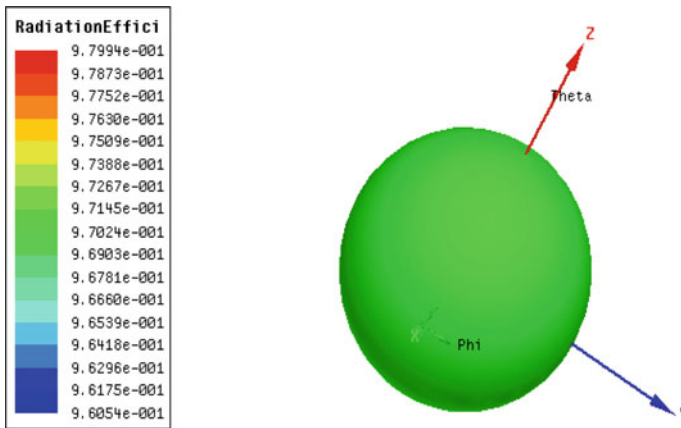**Fig. 24** Gain of 4 Elements Array Antenna

Fig. 25 Efficiency of 4 Elements Array Antenna

5 Performance of 8 Element Array Antenna

The structure of the 8 element array is shown below Fig. 26, and the spacing between the two elements is $\lambda/2$, with a maximum size of 8 element array antenna as 900 mm $\times$ 300 mm $\times$ 24 mm. All the elements in the array are fed with same amplitude and phase.

The performance characteristics like return loss, bandwidth, VSWR, elevation beamwidth, azimuth beamwidth, gain and efficiency of 8 element circular patch array antenna, respectively, are shown in Figs. 27, 28, 29, 30, 31, 32.
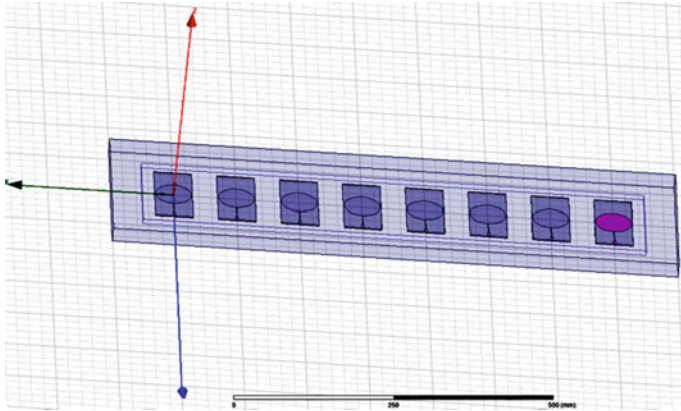
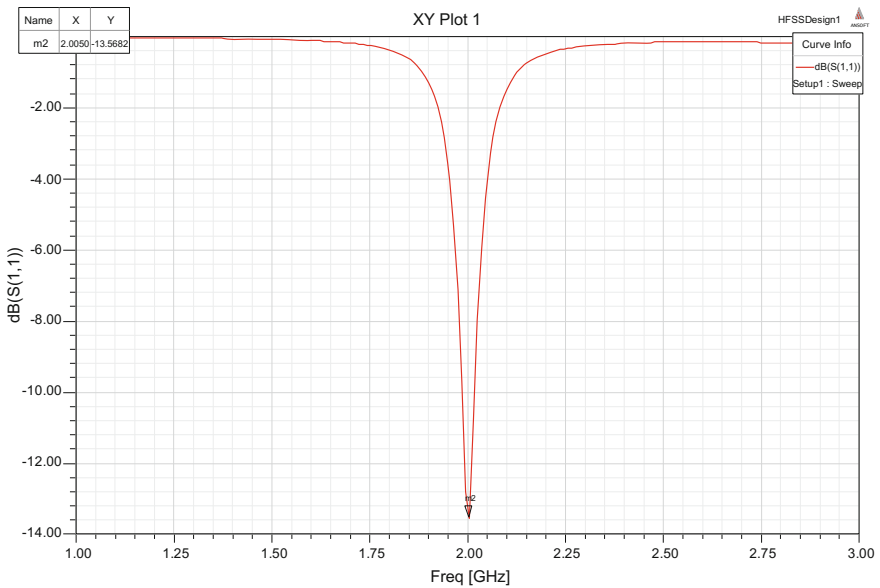**Fig. 26** Structure of the 8 element array

Fig. 27 Return loss of 8 Elements Array Antenna

The simulated result of return loss of the 8 element CPA array is obtained as −13.5 dB.

The simulated result of bandwidth of the 8 element CPA array is obtained as 433 MHz.

The simulated result of VSWR of the 8 element CPA array is obtained as 1.5.

The simulated result of elevation and azimuth beamwidths of the 8 element CPA array are obtained as 13.6° and 70.3°.
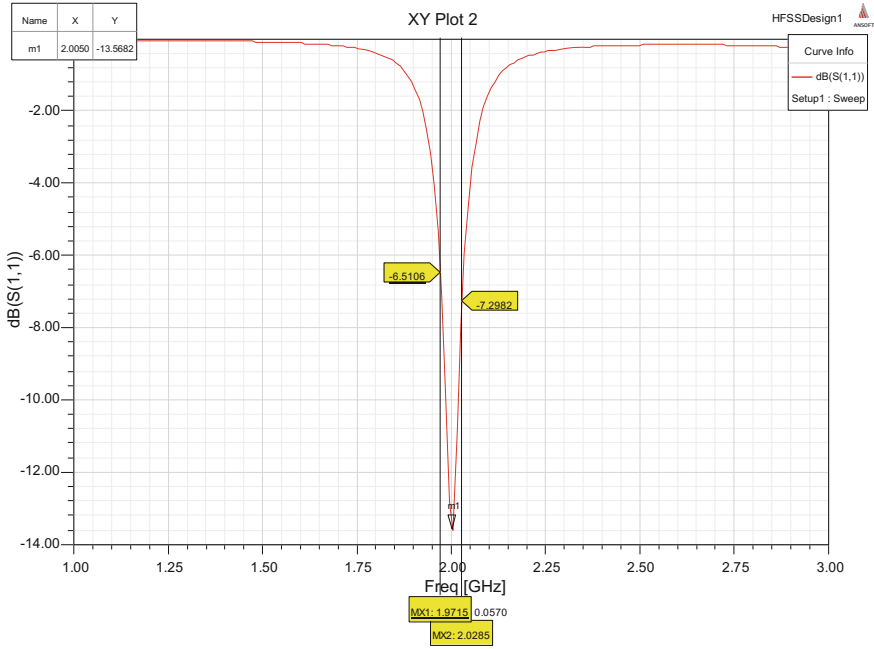
Fig. 28 Bandwidth of 8 Elements Array Antenna
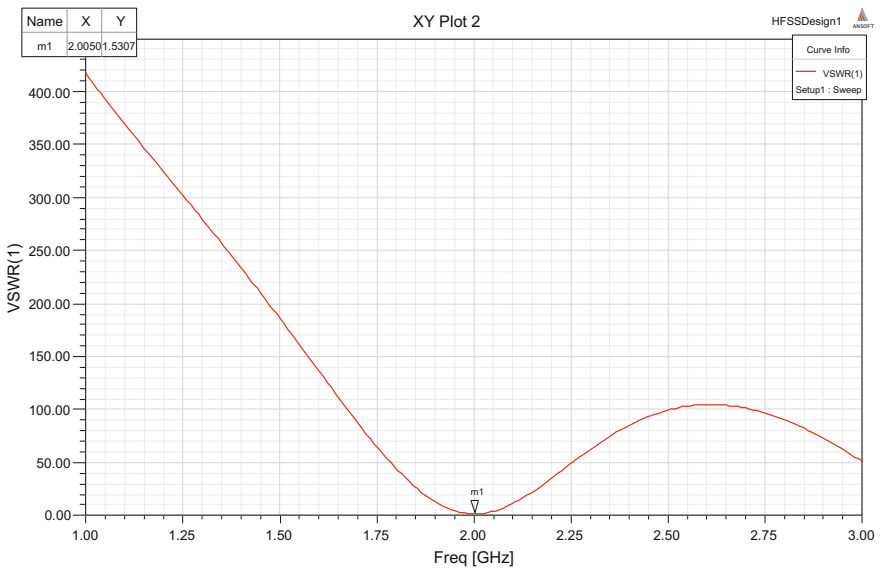

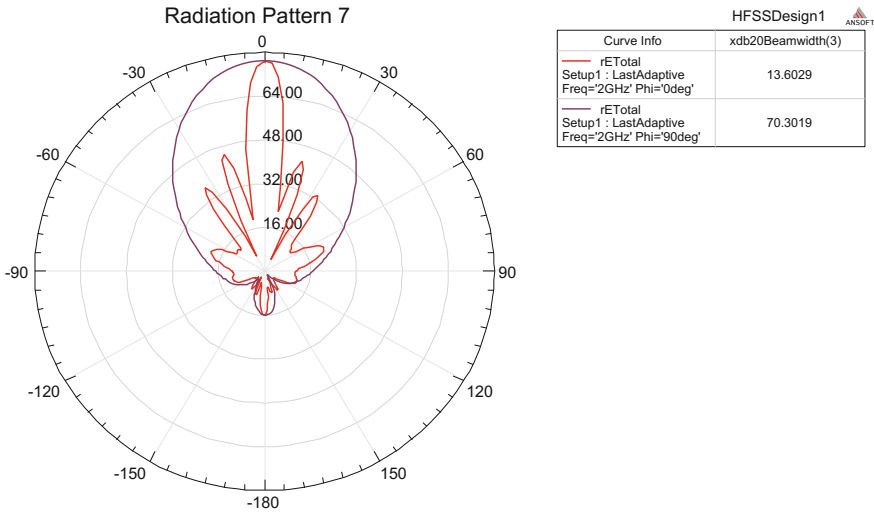
Fig. 29 VSWR of 8 Elements Array Antenna

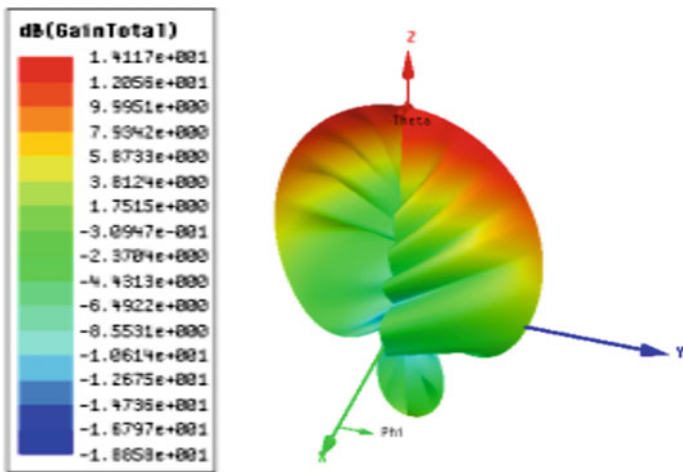**Fig. 30** The beamwidths of 8 Elements Array Antenna

Fig. 31 Gain of 8 Elements Array Antenna

The simulated result of gain of the 8 element CPA array is obtained as 14.1 dB.

The simulated result of efficiency of the 8 element CPA array is obtained as 97%.
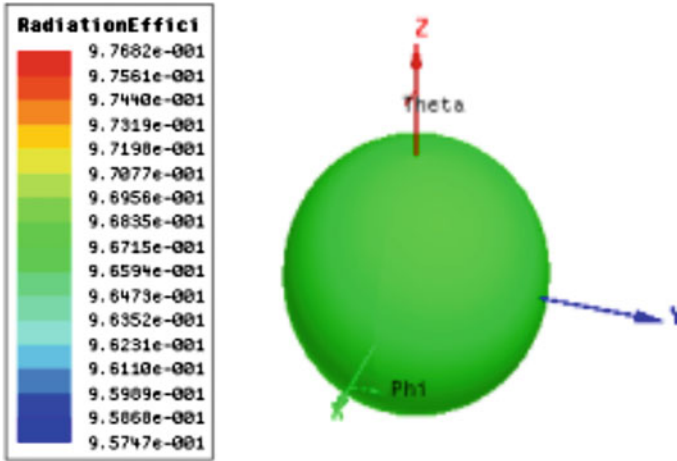
Fig. 32 Efficiency of 8 Elements Array Antenna

Table 2 Performance comparison of the proposed CPA arrays

S. No.	Parameter	No. of Elements			
		1	2	4	8
1	Frequency (GHz)	2	2	2	2
2	VSWR	1.8	1.8	1.9	1.53
3	Return loss (dB)	−19.3	−18.4	−17.6	−13.5
4	Gain (dB)	5.7	8.9	9.8	14.1
5	E Theta (°)	86.6	44.1	20.5	13.6
6	E Pi (°)	74.8	70.6	20.5	70.3
7	Efficiency (%)	98	97	97	97
8	Bandwidth (MHz)	436	423	414	433

6 Discussion

The radiation characteristics of the above arrays are compared. The proposed array antenna is successfully modeled and simulated using ANSOFT HFSS, and the radiation characteristics of the proposed array antennas are summarized in the Table 2.

7 Conclusion

The proposed array antennas are successfully modeled and simulated at 2 GHz in ANSOFT HFSS. The gain of 1, 2, 4, and 8 element CPA arrays are 5.7, 8.9, 9.8, 14.1 dB and elevation beamwidth are, respectively, 86.6, 44.1, 20.5 and 13.6. Thus

from the results, we conclude that as the number of elements increases, gain increases while beamwidth decreases. The radiation characteristics obtained by this proposed antenna are very much useful for airborne applications.

References

1. Shen LC et al (1977) Resonant frequency of a circular disk printed-circuit antenna. IEEE Trans Antennas Propag AP-25:595–596
2. Watkins J (1969) Circular resonant structures in microstrip. Electron Lett 5:524–525
3. Kumar R, Chand G, Gupta M, Gupta DK (2010) Circular patch antenna with enhanced bandwidth using narrow rectangular slit for Wi-max application. IJECT 1(1)
4. Singh M, Basu A, Koul SK (2006) Circular patch antenna with quarter-wave transformer feed for wireless communications. IEEE 1-4244-0370-7/06/$20.00 C 2006 IEEE
5. Balanis CA (1997) Antenna theory analysis and design. Wiley, New York
6. Pozar DM (1998) Microwave engineering. Wiley, New York
7. Bhal IJ, Bhartia P (1980) Microstrip antenna. Artech House, Dedgham

Author Biographies

U. Srinivasa Rao obtained his B.Tech degree in Electronics and Communication Engineering from RVR&JC College of Engineering in the year 1997. He received his M.E degree from Osmania University, Hyderabad in 2005. At present, he is pursuing his Ph.D in Acharya Nagarjuna University, Guntur, Andhra Pradesh, India. He is currently working as Associate Professor and Head, Department of ECE in Vignan's Lara institute of Technology and Science, Vadlamudi, Andhra Pradesh, India. He teaches for UG students and is guiding several projects. He is a life-time member for MISTE. His interested research areas are Microwave antennas, radar and optical communications.

P. Siddaiah obtained B.Tech degree in Electronics and Communication Engineering from JNTUA College of engineering in 1988. He received his M.Tech degree from SV University, Tirupathi. He did his Ph.D program in JNTU Hyderabad. He is the Chief Investigator for several outstanding Projects sponsored by Defense Organizations, AICTE, UGC and ISRO. He is currently working as Principal, University College of Engineering and Technology, Acharya Nagarjuna University, Guntur, India. He taught a wide variety of courses for UG and PG students and guided several projects. Several scholars successfully completed their Ph.D under his guidance and many are pursuing their Ph.D degree. He has published several papers in National and International Journals and Conferences. He is a life member of FIETE, IE and MISTE.

Effect of Loading Dielectric Made of Silicone Rubber on Microstrip Antenna

Ishmeet Singh Riar, Suman Wadkar and Avinash Vaidya

Abstract The design of dielectric loaded microstrip antenna (MSA) is presented in this paper. The substrate and dielectric load of MSA are made up of silicone rubber polymer, which has measured dielectric constant of 5.01 at 1 GHz. The resonant frequency of MSA is 24.35 GHz (K Band). Significant gain improvement of over 3 dB is obtained when the MSA is loaded with dielectric. The value of VSWR is also less than 2 in frequency range from 24.25 to 24.45 GHz. Measured results indicate that efficient directive radiation pattern is generated by the antenna in the frequency band of interest and is suitable candidate for Digital Electronic Message Service (DEMS) applications such as two-way high-speed communication and transfer of graphic images. The design and simulation of the MSA is carried using Computer Simulation Technology studio suite.

Keywords Silicone rubber · MSA · Dielectric cover

1 Introduction

There are many applications such as aircraft, satellite, mobile radio, and wireless communication, where antenna size, cost, and ease of installation play a very vital role and MSA is the best candidate which fulfills all these requirements effectively. But it also suffers from some disadvantages such as low gain and narrow bandwidth. But it is important to know that in long-range wireless communication, high-gain antennas are required [1].

I.S. Riar (✉) · S. Wadkar · A. Vaidya
PIIT, Mumbai University, Panvel, India
e-mail: ishmeetsingh@mes.ac.in

S. Wadkar
e-mail: swadkar@mes.ac.in

A. Vaidya
e-mail: avinashvaidya@mes.ac.in

Gain improvement technique based on Fabry–Perot Cavity (FPC) has also been studied [2]. Between the ground plane and partially reflecting surface (PRS), multiple reflections will take place when a PRS is placed above a ground plane at approximately $0.5\lambda_0$ and fed by an antenna. The improvement in the gain will take place when partial rays emanating through the PRS into space have same phases in the normal direction resulting in constructive interference [3]. To obtain improvement in gain of MSA, initially, the PRS will be loaded on the MSA having air gap between the PRSs. PRS reflection coefficient and feed antenna radiation are the factors which decide the gain of antenna. Gain can be increased by using high permittivity layer or $\lambda/4$ thick layer, but it may not be always practical to have high permittivity layer or $\lambda/4$ thick layer [4].

The substrate and dielectric load are made up of silicone rubber polymer. It is very important to know that to keep silicone rubber suspended in air is very tedious task as it is very flexible. Therefore, instead of suspending silicone rubber in air, it is directly loaded on MSA. The thickness of dielectric cover is varied, and its effect on antenna characteristics is then studied. Because of dielectric cover, there is change in the effective dielectric constant, and hence, resonant frequency will change [5]. The thickness of the dielectric cover will decide the frequency at which the MSA will resonate. It is studied that to achieve good efficiency, the dielectric cover should have thickness of about half the operating wavelength [6]. In [7], different antenna characteristics were studied, by varying the thickness of the dielectric cover. Both radiation efficiency and resonant frequency decrease when the antenna is loaded with dielectric.

To enhance the gain of MSA, first, the dielectric load of size as $\lambda \times \lambda$, where λ is wavelength in dielectric, is considered. After loading the MSA with dielectric, the thickness of dielectric cover is optimized so as to obtain improvement in the value of gain. The size of dielectric cover is then incremented by λ, i.e., the dielectric load of size $2\lambda \times 2\lambda$ is considered and again the height of the dielectric load is varied. Same steps are followed when the size of dielectric cover is $3\lambda \times 3\lambda$. The results show that when the thickness of dielectric cover is approximately half the wavelength in dielectric, maximum enhancement in gain is obtained. To make the antenna resonate at 24.35 GHz, the length of the patch is optimized.

2   Proposed Antenna Design

The 24.25–24.45 GHz frequency band is allocated newly to DEMS by Federal Communications Commission (FCC) which is known as the "24 GHz band" [8]. The resonant frequency of the antenna is 24.35 GHz.

To design the MSA, it is important to know the dielectric constant of the silicone rubber which is used as substrate and dielectric cover. The marker was kept at 1 GHz since the impedance analyzer was able to measure dielectric constant up to maximum frequency of 1 GHz only. From the results, it is observed that silicone rubber has dielectric constant of 5.01 and loss tangent of 0.01 at 1 GHz. But the

frequency of interest is 24.35 GHz. Therefore, it is important to find dielectric constant and loss tangent value at that frequency. These values are obtained by using Computer Simulation Technology (CST) studio suite. Figures 1 and 2 show that value of dielectric constant and loss tangent at 24.35 GHz is 4.96 and 0.0094, respectively. The MSA is fed using quarter wave transmission line. The height of the silicone rubber polymer substrate is optimized, and efficient radiation pattern is obtained when the thickness of silicone rubber is 1 mm.

With the help of design equations of MSA [1], first, the simple MSA is designed and then it is simulated using CST studio suite. It is observed that the gain of antenna is around 4.09 dB with return loss of −51.49 dB. The MSA has the impedance bandwidth of around 12.83%, and radiation efficiency is 74.4%. To enhance the gain, MSA is first loaded with the dielectric cover of size $\lambda \times \lambda$. The dimension of the substrate is reduced to $\lambda \times \lambda$, so that it is possible to load the dielectric cover of the same size. It has been shown that good results will be obtained from MSA if the size of the ground plane is greater than six times the substrate height all around the periphery of the patch which is not possible when
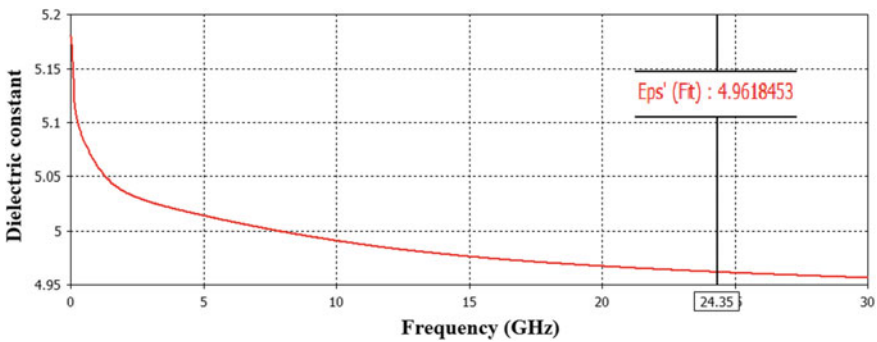

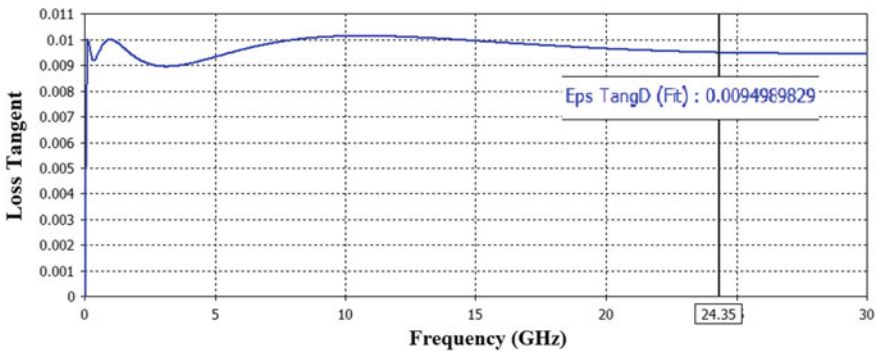**Fig. 1** Variation of dielectric constant with frequency

Fig. 2 Variation of loss tangent with frequency

the size of substrate is $\lambda \times \lambda$ as shown in Fig. 3a [1]. Figure 3b shows dielectric loaded MSA. The gain of the structure is around 4 dB, and VSWR is 1.53 which is less than 2.

Figure 4a shows the design of MSA when the size of substrate is $2\lambda \times 2\lambda$. The structure is optimized to make it resonate at 24.35 GHz. The dielectric cover of the same size is then loaded on MSA as shown in Fig. 4b. The thickness of the dielectric cover is optimized to obtain significant improvement in gain. Simulation results show that resonant frequency reduces to 23.88 GHz when the MSA is loaded with silicone rubber dielectric cover. The length of the patch is then optimized to make the antenna resonate at 24.35 GHz. The value of gain improves to 5.66 dB, and VSWR is less than 2 over frequency band of interest.

When the MSA is loaded with dielectric of size $3\lambda \times 3\lambda$ as shown in Fig. 5, the resonant frequency reduces further to 23.61 GHz. After optimizing the design, the gain of the antenna is 5.94 dB at 24.35 GHz and front-to-back ratio (F/B) is around 16 dB. Further gain can be enhanced if we increase ground plane size. When the size of the ground plane is incremented by λ, the gain improves from 5.94 to 6.84 dB. Also, the F/B ratio is increased to 25.67 dB.

(a) Withoutdielectric loading   (b) With dielectric loading

Fig. 3 MSA with substrate of size $\lambda \times \lambda$

(a) Without dielectric loading   (b) With dielectric loading
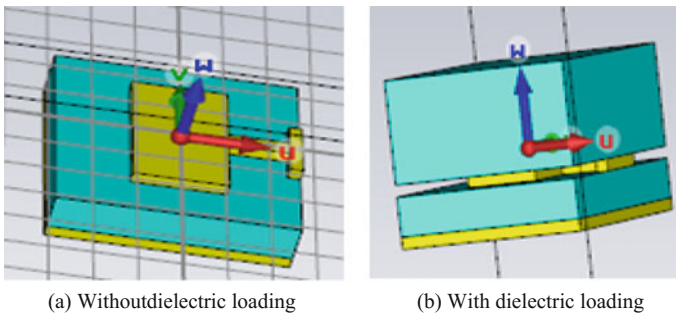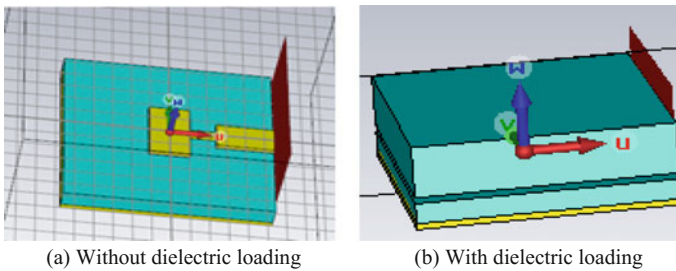
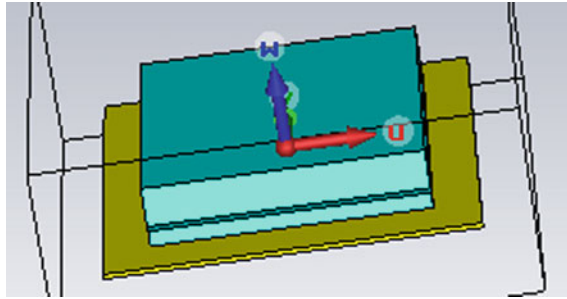**Fig. 4** MSA with substrate of size $2\lambda \times 2\lambda$

Fig. 5 MSA with substrate of size $3\lambda \times 3\lambda$

Table 1 Effect of loading silicone rubber on resonant frequency and radiation efficiency

Structures	$\lambda \times \lambda$ dielectric loaded MSA	$2\lambda \times 2\lambda$ dielectric loaded MSA	$3\lambda \times 3\lambda$ dielectric loaded MSA
Resonant frequency (GHz)	24	23.88	23.61
Radiation efficiency (%)	47	66	70
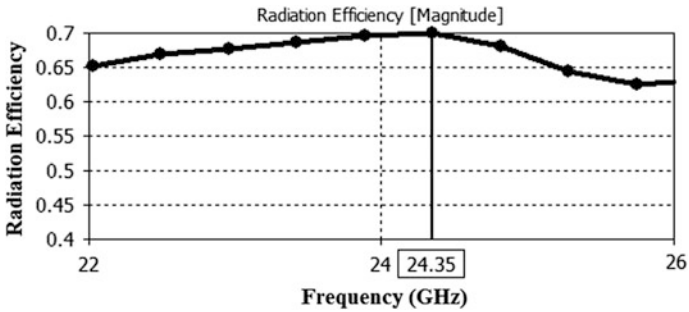
Fig. 6 Radiation efficiency versus frequency for MSA loaded with dielectric of size $3\lambda \times 3\lambda$

Table 2 Comparison of different structures resonating at 24.35 GHz

Different techniques		VSWR	Gain (dB)	% Bandwidth
Dimension of substrate is $\lambda \times \lambda$	Without dielectric loading	1.5	3.7	7.35
	With dielectric loading	1.53	4	6.49
Dimension of substrate is $2\lambda \times 2\lambda$	Without dielectric loading	1.095	5.41	12.2
	With dielectric loading	1.176	5.66	6.37
Dimension of substrate is $3\lambda \times 3\lambda$	Without dielectric loading	1.132	5.43	14.03
	With dielectric loading	1.66	5.94	6.21
	Dielectric loaded MSA with ground plane of size $4\lambda \times 4\lambda$	1.57	6.84	5.49

Fig. 7 VSWR plot for MSA loaded with dielectric of size $3\lambda \times 3\lambda$

Fig. 8 Radiation pattern for MSA loaded with dielectric of size $3\lambda \times 3\lambda$

3 Results and Discussion

All the structures are optimized to resonate at 24.35 GHz before they are loaded with dielectric cover. Table 1 explains the effect on resonate frequency when MSA is loaded with silicone rubber. It is observed that there is reduction in resonant frequency when MSA is loaded with dielectric. The radiation efficiency decreases when MSA is loaded with dielectric cover of size $\lambda \times \lambda$ as compared to simple MSA, but then it increases as the size of substrate and dielectric cover is

incremented by λ. Figure 6 shows that the radiation efficiency of antenna is about 70% at 24.35 GHz.

The comparative study of all the designed structures is shown in Table 2. With all structures resonating at 24.35 GHz, the improvement in the gain is highest when silicone rubber of size $3\lambda \times 3\lambda$ is loaded on MSA. The impedance bandwidth decreases on loading silicone rubber on MSA for every design.

From the Fig. 7, it can be seen that VSWR is less than 2 over the frequency range of interest. The radiation pattern of antenna is directive as shown in Fig. 8.

References

1. Balanis CA (2005) Antenna theory and design, 3rd edn. Wiley, New York
2. Trentini GV (1956) Partially reflecting sheet arrays. IRE Trans Antennas Propag 4:666–671
3. Vaidya A (2015) Gain and bandwidth improvement of linearly and circularly polarized microstrip antennas, Ph.D. thesis, IIT Bombay, February 2015
4. Gupta RK, Mukherjee J (2010) Efficient high gain with low SLL antenna structures using circular array of square parasitic patches on a superstrate layer. Microw Opt Technol Lett 52:2812–2817
5. Bahl IJ, Bhartia P, Stuchly SS (1982) Design of a microstrip antenna covered with a dielectric layer. IEEE Trans Antennas Propag AP-30(2):314–318
6. Alexopoulos NG, Jackson DR (1984) Fundamental superstrate (cover) effect on printed circuit antennas. IEEE Trans Antennas Propag AP-32:807–816
7. Yadav RK, Yadava RL (2011) Effect on performance characteristics of rectangular patch antenna with varying height of dielectric cover. Int J Power Control Signal Comput (IJPCSC) 2 (1). ISSN: 0976-268X
8. https://apps.fcc.gov/edocs_public/attachmatch/FCC-14-154A1.pdf. Assessed 15 Nov 2015

Part V
Miscellaneous

3D Smart Map

Yash Jagtap, Hitesh Shewale, Dinesh Bhadane and M.V. Rao

Abstract 3D Smart Maps are an initiative of Adikara firm aimed to provide a navigation tool which will help navigate even illiterate people in a temporary town (Sadhugram) made for the sadhus and pilgrims who will come to take a holy bath in river Godavari during Nashik-KUMBH 2015 (Kumbh is largest peaceful gathering in the World). Fifty-two Maps were provided throughout Sadhugram (i.e. in 375 acres) for this purpose, and the Maps were designed so that it can even be read by illiterate people and were very easy to understand and be interpreted.

1 Introduction

Kumbh is a Sanskrit word for pitcher, and Mela means fair. Crores of devotees gather in a bathing ritual once in 3 years.

Mythological story about Kumbh Mela which revolves around the story of the ocean churning which was done by gods and demons to obtain the amrita to live forever. According to the divine counting (God's Calendar), one day of Gods is equal to the one year of humans and the God Jayanta ran for 12 days to avoid amrit kalasha to fall in the wrong hands of demons. There are four places where Jayanta had put down the amrit kalasha with jewels in these twelve human years which are Haridwar, Prayag, Nashik-Trimbakeshwar and Ujjain, and at these four places at the time the sun, moon and planets is in a unique astrological alignment, during which kumbhmela is held here. This festival is considered to be one of the world's largest peaceful gatherings. The number of pilgrims is usually in millions or also crores [1].

Y. Jagtap (✉) · H. Shewale · D. Bhadane · M.V. Rao
Sandip Polytechnic, Nashik, India
e-mail: yashsuhasjagtap1@gmail.com

Mostly, the crowd which is in millions coming to the KumbhMela is illiterate ones; thus, it becomes very difficult for the crowd to navigate. So, we identified this problem and realized how many consequences it may result and took it as a challenge and decided to eliminate this problem; thus, we decided to make 3D SMART MAPS which can be readable by anyone it does not matter if he is literate or not. He/she can read it. For this purpose, we (Yash Jagtap, Hitesh Shewale, Saish Bhalerao, Rushabh Rambhia, and Shubhangi Kachhwa) decided to place infographics for a presentation of information. We also identified the most places where people are going to visit and set up maps right over there, so we set up total of 52 3D smart maps of which 4 are 15 feet by 20 feet, 12 maps are 10 feet by 10 feet and 36 maps are 6 feet by 6 feet, (The 3D smart maps were like Google street view but all Animated) very effectively but we mainly focused on the Flex which was set up in Sadhugram. We also thought that most of the public did not have smartphones also and those who have they did have the Sadhugram map on it as it was made by government. We saw that Sadhugram had become a small town, and the map of it was not updated on Google Maps yet. So we made maps for all the 12 sectors of Sadhugram and all sector offices had 15 feet by 20 feet maps and in sectors main point where the most crowd is going to visit had 10 feet by 10 feet map and we identified places surveying a bit and set up 3 maps per sector of 6 feet by 6 feet accordingly in each of the sector. We were the only people who were allowed to put up commercial maps or commercial flexes in Sadhugram. We then decided to make the maps from the sponsorships of the Bhumi Project and the Sapat International Pvt. Ltd.

Project's planning started from January 2014 and successfully completed it by September 2015. This project is done under MIT Kumbhthon. MIT Kumbhthon is an innovation centre which is setup to promote social entrepreneurship in Nashik [2]. Also, the intention was just to make Maps smarter and better.

2 Technical Aspects of 3D Smart Maps

3D Smart Map is effectively designed which made sure that proper message is communicated to the users of the map. A viewer can easily distinguish between the main figure on a map and the background information by using the concept of figure-ground with differentiation, closed form, centrality, articulation and good contour. Map designers developed a visual hierarchy on the map that corresponds accordingly by considering the intended intellectual hierarchy [3].

Viewers interpret the figure objects in the map that are closed so incorporation of the closure on a map is important [4]. The shape and location of figures is adjusted by centring and alignment [5].

Articulation is art of differentiating figure from ground [5]. Like to pop outline the continent pops out by texturing water [6].

When viewer's can trace the line throughout the map, it can be described as good contour [7]. A simple black contour line can be drawn to differentiate a figure entirely from the ground if it is not separated [8].

Intellectual hierarchy is also known as scale of concept [9], and it states that some of the features of maps are more important then the other. Visual hierarchy draws attention to the most important objects first [10].

The 3D Smart Map comes under the category of thematic Map, aimed at delivering cultural and sociological aspects of the Sadhugram area.

A particular theme connected with a specific area is shown by a thematic map. *These maps "can portray physical, social, political, cultural, economic, sociological, agricultural aspects"* [11].

These maps have mainly focused on the accessibility and friendlessness of the matter and audience and using it allows a person to be conscious about its surrounding and makes him/her known to the place also delivers all the necessary data which the cryptographer wishes to convey.

3 Problem Statement

Everyone is aware of the major events such as Kumbhmela, sunburn, pilgrimage places where a lot of population is engaged which has many literates and illiterates as well. Where there is population there are problems. Problems may arise like heavy traffic at some areas as people wont be aware or familiar with the area which would lead to confusion and irritation. Major events need proper management and trained volunteers to guide the people. Now, in the case of trained volunteers, the happening drawback is guiding a short group at a time and hence increasing traffic on that way. Management plays an important role in this things, and hence, everything should be well planned. It becomes a challenging situation of guiding the population in lakhs and also to give comfort to the people avoiding problems. Which? Where? How? These questions are the most frequently asked questions from the people regarding the routes. Wrong direction takes people to wrong places which create a mess mostly for the illiterate people. So by brainstorming on this, we came out with a revolutionary idea and we called it as "3D SMART MAPS".

Here in India, the number of people getting lost in usually in thousands and the main reason behind it is lack of knowledge of streets and locality. Pilgrims find it difficult to navigate themselves in places they visit.

Usually the childrens get lost at big festivals where are huge gatherings of large number of people, thus there is a need to take special care for people's navigation, and make people familiar with locality as early as possible.

4 Remedies

As per the above stated problem statement, there are many ways to tackle the problem but every solution has its own advantages and disadvantages. So, we proposed here a solution to use the 3D Maps efficiently to prove for the future enhancements.

Prior to our 3D Map structure, defined by our team, the maps available (existing) are tedious to visualize and also difficult to understand by illiterate or common people. So, in an attempt to minimize confusion and to make event a huge success we proposed this technique after studying limitations of the other solutions in market.

Some of the solutions are as follows:

1. Nashik Municipal Corporation (NMC) (Fig. 1).
2. Google Maps (Fig. 2).
3. GPS.
4. Handouts of Sandhugram.
5. Trained volunteers

Fig. 1 Map made by Nashik Municipal Corporation in Sadhugram

Fig. 2 Sadhugram Google Map during Kumbh

5 Proposed Technique

So for the above problem statement, our solution (Fig. 3) was to prepare 3D maps of Sadhugram. But first we were not clear on our idea. We first started with the map of Nashik, then we got to know that it is quite confusing and this will not help people. So we started searching for another option to help people of KumbhaMela. Then, we came with different ideas such as mobile application, handouts, colour coding Sadhugram and map having data of various sectors. But every idea had its own pros and cons. Then, combining all these ideas, the team came with 3D map which was a great idea.

Government allowed us to post the maps in all 12 sectors. And the sizes given by them were in following manner:

Fig. 3 3D smart map

4 maps of size 20′ × 15′.
12 maps of size 10′ × 10′.
36 maps of size 6′ × 6′.

6 Raw Data Required

1. AutoCAD File of Sadhugram.
1. Infographics related to kumbh mela.
2. Data of toilets and drinking water, etc.

7 Softwares Used

We completed project with the help of following softwares:

1. Autodesk 3Ds MAX.
2. Autodesk Autocad.
3. Corel DRAW.
4. Adobe Photoshop.

8 Key Features

1. Compass point.
2. Full Sadhugram miniature map.
3. Index.
4. All amenities representation done in Sadhuram such as dispensary, water supply and ration shop
5. Helpline numbers.
6. Wide audience reach due to presence of Map at every crossroads.
7. Use of colour coding system.

9 Conclusion

In the end after many ups and downs in the project, it was very successfully completed with great feedback. We also got covered in newspaper several times. We introduced a new technology in the market which leads to that not only the literate could read the maps but now even illiterate people who cannot even read or write can read the maps, and navigate. These 3D elevated smart maps proved to be of much useful, and many people were benefited from this even the work of government officials was tremendously reduced as the maps were so big that even at a time 50 people were able to read the maps. So it was useful a lot! We also made the small maps—the miniature version of full Sadhugram and highlighting only the sector shown in map. Also, the close camera views of 6 feet by 6 feet gave sponsors the wide area which they wished to cover and all liked the idea.

For now, the team targeting only kumbh but the team is planning to collaborate with Cubix and Nashik city connect so as to make each and every gathering place in 3D. As it may help from even small children getting lost and even elderly people who cannot see much can be able to navigate through towns. As for the technology loving people, the team is planning to launch an App in near future in which a user can get experience of actually in game movements and can move freely through and

around the buildings and thus is planning to make a whole new system of navigation and a new industry in travelling field (Prototype is ready).

References

1. https://kumbhmela2015.maharashtra.gov.in/1191/Kumbhamela-Introduction
2. http://web.archive.org/web/20100403014350/
3. http://cameraculture.media.mit.edu/kumbhathon-innovating-the-kumbh-mela/
4. Dent BD (1972) Visual organization and thematic map design, pp 79–9. http://onlinelibrary.wiley.com/doi/10.1111/j.1467-8306.1972.tb00845.x/full
5. Robinson A (1995) Elements of cartography. https://books.google.co.in/books/about/Elements_of_Cartography.html?id=mUyAAAAAMAAJ&redir_esc=y
6. Slocum TA (2005) Thematic cartography and geographic visualization. Pearson Hall, Upper Saddle River
7. Head G (1972) Land-water differentiation in black and white cartography. Can Cartogr 9(1):25–38
8. Mac Eachren A (1992) The role of brightness differences in figure-ground: is darker figure? Cartogr J. https://books.google.co.in/books?id=5gMmDwAAQBAJ&lpg=PA280&ots=UWY9AnKOWH&dq=Mac%20Eachren%20A%20(1992)%20The%20role%20of%20brightness%20differences%20in%20figureground%3A%20is%20darker%20figure%3F%20Cartogr%20J&pg=PA280#v=onepage&q=Mac%20Eachren%20A%20(1992)%20The%20role%20of%20brightness%20differences%20in%20figureground:%20is%20darker%20figure?%20Cartogr%20J&f=false
9. Monmonier M (1993) Mapping it out: expository cartography for the humanities and social sciences. University of Chicago Press, Chicago
10. Geospatial Visualisation. True-3D in cartography: autostereoscopic and solid visualisation of geodata relief holographic map production, p 378. In: Part of the series Lecture Notes in Geoinformation and Cartography, pp 41–65, 2 October 2012
11. (1992) Thematic map collection & cartographic information services unit. University of Washington. https://books.google.co.in/books?id=8FRP2ZbvL0gC&lpg=PA67&ots=ufHTIvlTco&dq=(1992)%20Thematic%20map%20collection%20%26%20cartographic%20information%20services%20unit.%20University%20of%20Washington&pg=PA67#v=onepage&q=(1992)%20Thematic%20map%20collection%20&%20cartographic%20information%20services%20unit.%20University%20of%20Washington&f=false

Software-Defined Network-Based Intrusion Detection System

Yogita Hande, Akkalashmi Muddana and Santosh Darade

Abstract The traditional networks conflate control and data plane on a physical device, normally comprise of software and hardware. The new emerging technology software-defined networks (SDNs) detach control plane from the data plane. SDN control plane exits controller has ability to control the entire network which is a possible security concern as compared to traditional network with a distributed control plane. In software-defined network, malicious flow exploits vulnerability of the controller through SDN switches, which demands essential SDN-based security model. This paper introduces the SDN architecture with their threat vectors. We proposed the SDN-based intrusion detection system (IDS) which identifies malicious behaviors or attacks and reports to network administrators as intrusion events.

Keywords Software-defined networking · Intrusion detection system · Security

1 Introduction

Software-defined networking (SDN) is one of the most innovative network technologies. The main idea of SDN is to decouple the control plane from the physical infrastructure, which enables network administrator to define and configure network through programming or using the software to define it. Software-defined

Y. Hande (✉) · A. Muddana
Department of Information Technology, GITAM School of Technology,
Hyderabad, AP, India
e-mail: yhande18@gmail.com

A. Muddana
e-mail: lakshmi.muddana@gitam.edu

S. Darade
Department of Computer Science and Engineering, GITAM School of Technology,
Hyderabad, AP, India
e-mail: darade.santosh@gmail.com

networking [1] incites network operator to manage and configure the network essential by enhancing the OpenFlow in switch product.

Considering the environment of SDN, security is the prime challenge in SDN architecture. Centralized controller may exploit vulnerability due to continuous update from network administrator or operator [2]. There is a need to analyze SDN traffic which can degrade the SDN network performance and require effective implementation of security model. The network assets and communication transactions across the network must be protected from malicious attacks and unintentional damage through some security professionals.

An intrusion detection system is defined as one of the solution for network security to identify the abnormal activities. An IDS enables us to secure the system from unauthorized users. It is considered as the backbone for securing the SDN architecture. Here, we proposed the SDN-based intrusion detection system (IDS) uses methods to monitor the network traffic for intrusion detection which helps to provide the security toward SDN.

1.1 Software-Defined Networking

Software-defined network (SDN) is a dynamic, manageable, adaptable, and cost-effective network. This SDN is encompassing several network technologies aimed to make directly programmable, agile, centrally managed network infrastructure. The goal of SDN is to allow network administrator or operator to quickly handle the business needs.

The SDN idea was proposed by Nick et al. [1]. They proposed an idea that separates out the control plane from the data plane (forwarding plane) from network device, where a control plane controls several devices (Switch's). In software-defined network, administrators can shape traffic by programming the controller at control plane without interrupting network devices (switches) at data plane. Control plane has a centralized controller, which sets forwarding rules in switch to route traffic from source to destination. Data plane that consists of network devices handles all packets according the flow entries set by the controller. All flow entries at the switch level are managed by centralized logical controller.

With this ability, SDN technique is introduced to handle the cloud service providers challenges such as changing load traffic, more bandwidth requirement, and security and scalability issues. Enterprise and organization use OpenFlow-based SDN to balance the traffic load; redirect the traffic; manage on demand, bandwidth requirement; and execute polices to scale the network.

The software-defined networking (SDN) approaches have been earlier adopted by many companies for their business particularly cloud and telecom services. The traditional technologies are not capable to drive their current business challenges, so modern network techniques are based on software. Service provider companies are using SDN concepts such as Amazon, Rackspace, IBM SoftLayer also telecoms field are also competitive in developing their software network strategies [3].

1.2 Software-Defined Network Architecture

A software-defined network distributes the network into the control layer (control plane), infrastructure layer (data plane), and application layer as shown in Fig. 1 The main concepts of SDN are to separate out the control plane from the data plane from every network device [4–6].

On the top of SDN architecture, application plane exits many applications such as load balancing, traffic management, and security. The application plane communicates their requirements and desired network behavior to control plane through northbound interface (NBI).
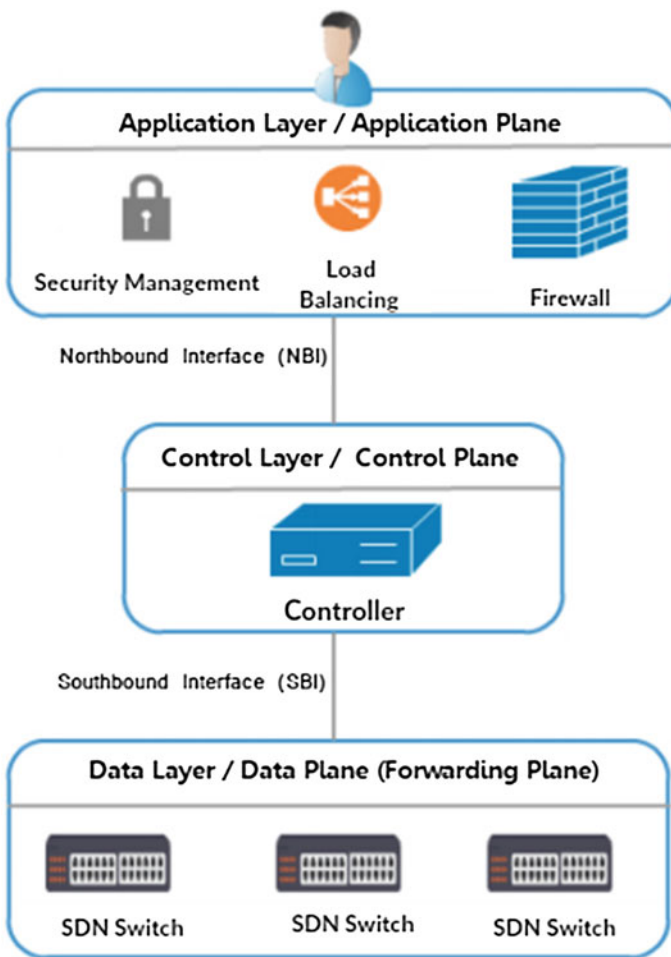

Fig. 1 Software-defined networking architecture

The control plane of SDN involves centralized logical controller which translates received application layer requirements to SDN datapaths and provides SDN application with the abstract view of network. This controller implements control logic and allows controlling the data layer resources.

Data layer of SDN also called forwarding plane comprises of many network elements (switches). Traditional devices expose datapath and forward the network traffic over the network. In SDN, switches become simple forwarding devices that forward the traffic as per rule set by the controller. The southbound interface (SBI) is defined between an SDN control layer and data layer which attempt to programmatically control the all forwarding.

Control plane and data plane separation allow the network administrator to easily change the network policies, which enable network administrator to construct flexible, scalable network by handling the business need through software rather than hardware.

2 SDN Security Literature Survey

SDN is a programmable and agile network [9]; the network administrator, researchers, and engineers programmatically generate strategies to simplify the management of data centers. SDN model assists campus networks, enterprises, wireless networks, and DCNs. IT organizations and network enterprises should be aware of these challenges such as reliability, performance under latency constraints, security, introduced in [7].

In [8], the author has made a comparative survey related to security in SDN and believes that the architecture introduced by SDN for data centers must be assessed to ensure their network security.

The author in [1] has proposed OpenFlow; it is a way for researchers to run their experimental protocols on their heterogeneous network switches. They encourage vendors to add OpenFlow to their network switches product that allows researchers to evaluate their ideas in real-world traffic settings; however, the experiment may not have shown any security features implementations with OpenFlow.

In [2], the author has presented many SDN challenges such as security. Considering security as challenges for SDN, the need to develop new security model for centralized controller may be vulnerable due to constant updates from network administrator.

The author in [10] has proposed architecture for intrusion prevention using SDN. He has defined two ways for the detection and prevention of intrusion, such as on path and off path. They preferred off path intrusion detection system to increase the network performance but have not mentioned the IDS functionality.

The author in [11] has proposed a concept of new openflow switch consists of IDS for SDN security. They proposed a secure openflow protocol. An IDS uses the database for IP verification and packet verification. Such type of IDS implementation may be complex, and it degrades the network performance.

The proposed scalable intrusion detection system (IDS) architecture [12] on SDN environment, using a virtualization infrastructure, inspects malicious flow open source Suricata IDS which enables us to sample data packets belonging to suspicious traffic flows. They propose sampling rate adjustment method in order to order to maximize the detection rate.

3 Threat Vector to SDN Architecture

The trend may be toward implementing software-defined networking in enterprises and organization to configure network and modify the network as per the business needs which creates security challenge to SDN developer to secure the SDN architecture. Enterprise and organization would be ready to replace their traditional network with SDN network with security mechanisms that assure them that their network could not be venerable.

Here, we are focusing on the threat vector [13, 14] on SDN with the security countermeasures currently being considered. Each layer of SDN exploits the vulnerability by attacker [15].

3.1 Data Layer (Southbound)

Attacker could gain unauthorized access to network elements within network itself and try to destabilize the network by performing denial of service attack (DOS) on network elements (switches) connected to SDN network. Controller communicates to network elements using southbound APIs and protocols. These southbound communications could use OpenFlow (OF), Open vSwitch Database Management Protocol (OVSDB), etc. Many of these protocols under the development have their own way to secure the network element, but many of the protocols are not able to cope with current security challenges. The network elements maintain flow table having flow entries updated by controller used to route traffic over the network. The attacker could initiate new flow entries which misguide the network elements. If the attacker steers the traffic in their direction, it sniffs the traffic and performs man-in-the-middle attack by eavesdropping southbound communication between controller and network elements.

3.2 Controller Layer

SDN controller is the main and core component of a SDN network which manages whole networks. The single point of failure of this controller can cost more in the SDN environment, because it manages the global view of a network. If attacker

targets it by performing DDOS attack, then it leads to complete networks failure. Some potential attackers over the network pretending as a real controller harm the controller itself. Few of the resource utilization attack inject by the attacker on a SDN controller tries to utilize maximum SDN resources.

3.3 Application Layer

Application layer of SDN runs many applications may communicate with controller through northbound APIs and protocols. The attacker could exploit vulnerabilities of controller by using northbound APIs such as Java, JSON, and Python.

To detect and identify the attacks which we explained earlier, the security mechanisms such as antivirus, firewall, and intrusion detection system will be highly preferable. Also SDN can configure some authentication policies like role-based access control (RBAC) to avoid unauthorized access of SDN controller. Some security policies must be implementing at application layer to provide better security.

4 Proposed SDN-Based Intrusion Detection System

The SDN architecture highly supports network monitoring and analysis mechanism due to programming environment of SDN controller. SDN adapts security service insertion [3] by developing intrusion detection system [IDS] to monitor incoming traffic to SDN network. We propose SDN-based intrusion detection system (IDS) which implement on SDN controller shown in Fig. 2. The role of this IDS analyzes the traffic intended to the SDN network to identify the malicious flow. The traffic coming from outside networks are received by SDN switches.

After receiving packets, SDN switches check action defined by the controller for packets transmission through checking of flow table entries. If periodically updated flow table does not contain flow entry for respective packets, those packets will be forwarded to controller to set the datapath to corresponding packets.

The IDS designed on controller defines two methods which apply on the traffic coming from switch to detect the intrusion that try to degrade the SDN network performance. Following are the IDS methods.

4.1 Packet Countermethod

Packet countermethod of IDS maintains the count of IP packets received within particular amount of period. Our IDS defines the threshold value which states the maximum number IP packets handled by the network in defined period.
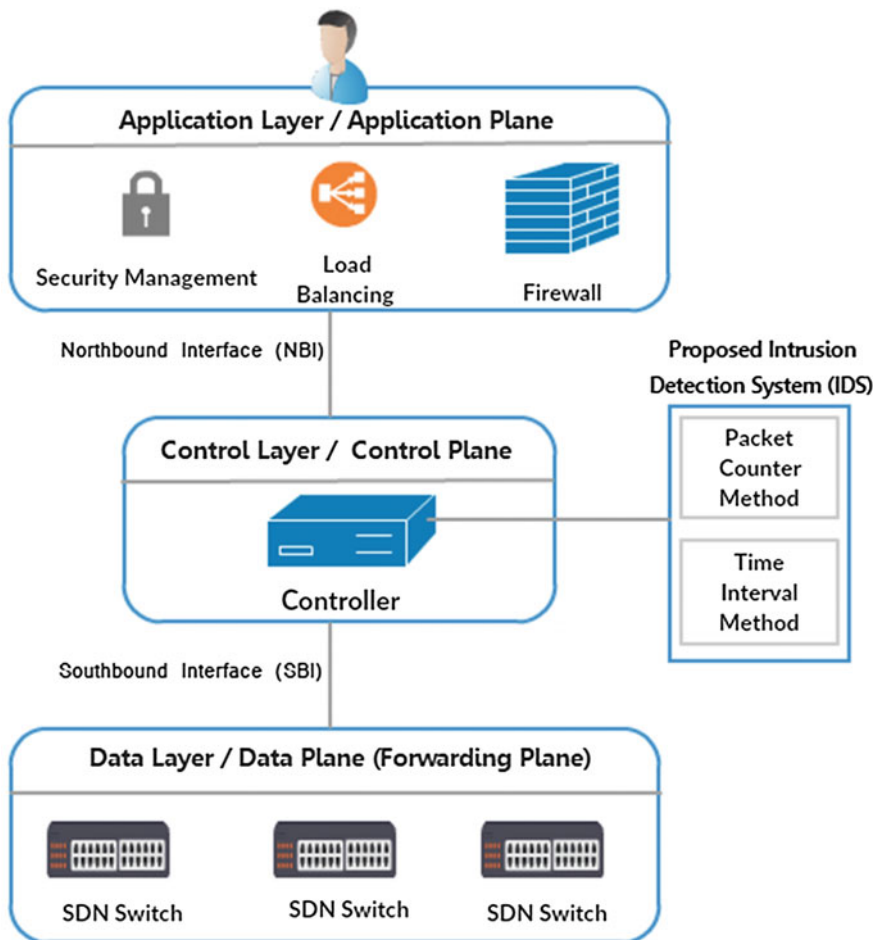
Fig. 2 SDN-based intrusion detection system

The countervalue of IP packet greater than threshold value will be considered as intrusion.

Tp is threshold value. Pc is packet counter, its counts packet received within defined time period t.

$$Pc = \int_{t1}^{t2} \sum_{i=1}^{n} Pi \tag{1}$$

$$Pc \geq Tp \tag{2}$$

Calculated packet value Pc is compared with Tp. If the packets countervalue Pc is greater than equal to threshold value Tp, then intrusion detected.

4.2 Time Interval Method

In time interval method, when the packets arrive to SDN network, IDS calculates the interval time of packets. Interval is calculated by considering the arrival time of previous packet and current packet. If the interval time of receiving packets is very less than threshold value, then IDS perceives this situation as intrusion.

Ti is threshold value, and Cpa and Ppa are defining the arrival times of current packet and previous packet, respectively. I is calculated interval time

$$I = \mathrm{Cpa} - \mathrm{Ppa}. \tag{3}$$

$$I \leq \mathrm{Ti} \tag{4}$$

If calculated interval I is less than equal to defined threshold value Ti, then intrusion detected.

Our proposed SDN-based IDS detect the malicious flow by relating the countervalue and time interval value to threshold values. The threshold values can be defined according to network behaviors, or it will vary as per the network type. Initially, our IDS considers these two methods to identify the intrusion over SDN. In future, we will try to expose the more operative methods to detect and prevent the intrusion over SDN network by focusing different problems like scalability.

5 Conclusion

Security is a main challenge in software-defined network (SDN) due to centralize controlling of the network through programmable controller. Here, we studied SDN architecture with different SDN threats which are a major security issues in SDN. We proposed SDN-based intrusion detection system (IDS) methods which is useful to detect malicious flow of the network traffic.

In future, we will try to simulate our proposed IDS methods over mininet emulator with Python-based POX controller. Also in future, we will represent some more effective intrusion detection and prevention methods and algorithm to handle malicious network traffic competently.

References

1. Nick M, Tom A, Hari B, Guru P, Larry P, Jennifer R, Scott S, Jonathan T (2008) OpenFlow: enabling innovation in computer networks, ACM SIGCOMM Comput Commun Rev 38(2):69–74
2. Angle VC, Lorena BL, Luise CV (2013) Evolution and challenges of software defined networking. In: IEEE on SDN Future Network and Services (SDN4 FNS), pp 1–7
3. Latest Technology News and Reviews. http://www.in.techradar.com
4. Open Networking Foundation: SDN architecture, Issue 1 June, 2014
5. Open Networking Foundation: SDN architecture overview, Issue 12, Dec 2013
6. Open Networking Foundation (ONF) White Paper (2012) Software-defined networking: the new norm for networks
7. Manar J, Taranpreet S, Abdallah S, Rasool A, Yming L (2014) Software-defined networking: state of the art and research challenges. Department of Electrical and Computer Engineering, Western University, Canada
8. Sandra SH, Gemma C, Sakir S (2013) SDN security—a survey. In: IEEE SDN for Future Networks and Services (SDN4FNS), pp 1–7
9. Open Networking Foundation. https://www.opennetworking.org
10. Antonio PL, Ulisses RF, Otto CD (2013) An architecture for intrusion prevention using software defined networks. Universidade Federal do Rio de Janeiro—GTA/COPPE—Rio de Janeiro, Brazil
11. Suresh K, Tarun K, Ganesh S, Maninder N (2012) Open flow switch with intrusion detection system. J Sci Res Eng Technol 1(7):001–004. ISSN 2278-0882
12. Chiwook J, Taejin H, Jargalsaikhan N, Hyuk L, JongWon K (2014) Scalable network intrusion detection on virtual SDN environment. In: 3rd IEEE international conference on cloud networking, pp 264–265
13. http://www.computerweekly.com
14. http://www.networkworld.com
15. Sakir S, Sandra SH, Pushpinder KC, Barbara F, David L, Jim F, Niel V, Marc M, Naveent R (2013) Are we ready for SDN? Implementation challenges for software-defined networks. IEEE Commun Mag 51:36–43

CFIM: Toward Building New Cloud Forensics Investigation Model

Ezz El-Din Hemdan and D.H. Manjaiah

Abstract In recent times, cybercrime investigation in cloud computing poses complex challenges due to virtualization, volatile data, deleted data, and dynamic and distributing nature of cloud computing. Performing cybercrime investigation in cloud environment is called Cloud Forensics. With the intention of overcoming these challenges, this paper introduces a Cloud Forensics investigation model (CFIM) that can help to investigate cybercrimes in the cloud in forensically sound and timely fashion. The proposed model is an intelligent system that is able to take a snapshot periodically for each virtual machine running in the cloud, sends it automatically to trusted center server (TCS) that is responsible for monitoring and recording the status of the virtual machine and finally, sends it to the forensic server (FS) to perform forensic analysis. This model can increase probability of tracking attackers, determining weaknesses of virtual machines for future use, and also can support in the process of extraction and collection of digital evidence.

Keywords Cloud computing · Digital forensics · Cloud forensics
Digital evidence and cybercrimes

1 Introduction

Recently, cloud computing has become one of the most popular and important computing paradigms which has been adopted by several companies and organizations because it provides and offers various benefits and advantages such as optimizing the general usage of IT infrastructures, high degree of scalability and availability of enormous computing, and storage resources. With the appearance of cloud computing technology that depend on a theory of distributing of datacenters

E.E.-D. Hemdan (✉) · D.H. Manjaiah
Department of Computer Science, Mangalore University, Mangalore, Karnataka, India
e-mail: ezzvip@yahoo.com

D.H. Manjaiah
e-mail: manju@mangaloreuniversity.ac.in

which are allocated around the world to provide numerous cloud services with cost-per-use method, the cybercrimes are increased against cloud systems. This new technology has changed the thinking of criminals and attackers through utilizing massive capabilities because old-fashioned cybercrime techniques and methods have to be modified to adapt with the new features of the cloud computing infrastructure.

In the other side, there is one of the important sciences that can help to find and trace criminals which is called "Digital Forensics." The digital forensics science is used to identify, extract, and analyze digital evidence to reconstruct crime events. Digital investigators and practitioners have to change and expand their tools, methods, and techniques to deal with these new types of crimes against the cloud infrastructure and to reconstruct the crime events that occurred.

The process of performing digital forensics investigation in cloud environment is known as "Cloud Forensics." Cloud Forensics is challenging at best, but can be performed in a way steady with law through proposing new methods, techniques, and tools. Several researchers have identified and explored security problems relating to storing data in the cloud. Few number of research, however, have fully explained or developed new methods, techniques, and tools to carry out forensics investigation of cybercrimes in the cloud environment. In addition to this, there are little hypothetical scenarios and case studies to illustrate and explain the complex challenging issues related to Cloud Forensics investigation of cybercrimes in real cases.

Although cloud computing can offer various opportunities to provide and improve digital investigation process in the cloud through utilizing massive capabilities of the cloud computing like processing, computing, and storage resources but still digital investigators are facing many legal, organizational, and technical challenges in performing Cloud Forensics. These challenges specially include dependence on the cloud service provider (CSP), less control in cloud infrastructure, physical inaccessibility of digital evidence, volatile and deleted data, legal and trust issues, large bandwidth and multi-tenancy.

This paper presents a cloud forensic investigation model (CFIM) to investigate cybercrimes which are in the cloud computing environment. The proposed model is an intelligent system that can take a snapshot periodically for the state of each virtual machine (VM) which is running in the cloud and sends it automatically to trusted center server (TCS) that works as storage for VM snapshots. The VM snapshot is very useful to digital investigators and examiners because it helps in analyzing the VM in order to extract digital evidence related to crime in virtualized environments such as cloud computing. The TCS server is responsible for monitoring and recording the status of the VMs and therefore sends it to the forensic server (FS) in order to be used to perform the forensic analysis process remotely by the investigator. The use of this model can to be mention as one of the terms in service level agreement (SLA) between cloud user (CU) and cloud service provider (CSP) to guarantee the success of the investigation process in the cloud using this model integrated with cloud infrastructure in the future.

The proposed model can provide several benefits for carrying out digital forensic in cloud in an effective manner as follows:

- It will introduce a real example of Forensic as a Service (FaaS) by including the FS in the CSP side for performing the digital investigation process and remotely through using the enormous capabilities of cloud computing.
- One big problem that may face digital investigators is when the VM instance is deallocated, the digital investigators may not know whether evidential data or the entire VM instance could be recovered or not. The proposed model can provide a solution to this problem by taking snapshot of the VM which contains the status of VM during the malicious activities.
- It may also help to unravel the problem of volatile data that cannot sustain without power where, if the user turn off his/her virtual machine, then all the information is lost so there is the need to have the image of the VM instance. This encourages attackers and criminals to exploit this vulnerability (i.e., loss of volatile data) to do illegal activities and therefore makes the investigation process almost difficult to extract digital evidence about the crime (i.e., anti-forensics). However, this can be solved using virtual machine snapshots.
- It would also support the problem requiring large bandwidth to download VM instance where the size of the VM instance depends on the size of data inside it. This can be solved through installing the FS in the cloud side.
- The model may also apply the concept of proactive measures by preserving regular snapshots in running of virtual machines that can significantly help response to handlers and forensic investigators if such types of incidents happen.
- The proposed model can be used within several organizations which will use cloud services such as financial and healthcare systems.

The remainder of the paper is structured as follows. Section 2 provides a brief overview about basics of cloud computing, digital forensics, and Cloud Forensics, while previous and related works are presented in Sect. 3. Section 4 introduces the proposed Cloud Forensics investigation model (CFIM), while a malicious scenario is discussed in Sect. 5. The paper conclusion and future work on this innovative research are presented in Sect. 6.

2 Overview

This section introduces a brief overview about fundamentals of cloud computing, digital forensics, and Cloud Forensics.

2.1 Cloud Computing

Cloud computing is a revolutionary technology that offers individuals, governments, and organizations with massive resources as services accessible through networks either internally or externally over the Internet. National Institute of

Standards and Technology (NIST) defined cloud computing as "a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction" [1].

From the aforementioned NIST definition of cloud computing, there are five essential characteristics such as on-demand self-service, broad network access, resource pooling, rapid elasticity, and measured service. Having said about the characteristics so far, we are able to extract two cloud computing models which are deployment models and service models. Each of them have different types under it. Deployment models consist of four types, and these are private cloud, public cloud, community cloud, and hybrid cloud. On the other hand, service models consist of three types which are Software as a Service (SaaS), Platform as a Service (PaaS), and Infrastructure as a Service (IaaS).

2.2   Digital Forensics

Digital forensics science is a sub-discipline of forensics science that concerns identifying, collecting, extracting, preserving, and analyzing digital evidence that are extracted from crime scenes in order to assist digital investigators and practitioners to trace and catch criminals along with determining vulnerabilities in digital systems such as personal computer (PC), mobile phones, tablets, networks, and cloud systems. Digital forensics science was defined in 1999 by McKemmish [2] as "The process of identifying, preserving, analyzing and presenting digital evidence in a manner that is legally acceptable by court of law." From this definition, it can be concluded that the digital forensics consists of four crucial phases in performing the digital investigation as shown in Fig. 1 as follows:

- **Identification:** It is the process of identification of a crime and digital evidence that could be needed to verify the committed incident.
- **Preservation:** Here, the digital investigators and examiners preserve the collected and extracted evidence as well as virtual hard disks, log files, mobile phones, and any related proofs that are in the crime scene.
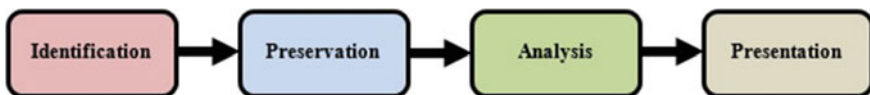

Fig. 1 Digital forensic investigation process

- **Analysis:** This phase involves in understanding, explaining, and correlating digital evidence data to reach a summary of findings, in order to help in proving or disproving illegal activities.
- **Presentation:** In this phase, the investigators and experts prepare a forensic report as an official document on summary/conclusions about the incident that occurred. This report should be suitable to show as legal and admissible proof in the court of law.

2.3   Cloud Forensics

Cloud Forensics is an emerging area that refers to the applications of concepts, principles, and methods of digital forensics in cloud. Ruan et al. [3] considered Cloud Forensics as subset of network forensics, and it is defined as "The application of digital forensic science in cloud computing environments. Technically, it consists of a hybrid forensic approach (e.g., remote, virtual, network, live, large-scale, thin-client, thick-client) toward the generation of digital evidence. Organizationally it involves interactions among cloud actors (i.e., cloud provider, cloud consumer, cloud broker, cloud carrier, cloud auditor) for the purpose of facilitating both internal and external investigations. Legally, it often implies multi- jurisdictional and multi-tenant situations" [4]. From the above definition, it can be noticed that the investigators will face sundry legal, organizational, and technical challenges such as physical inaccessibility of digital evidence, dependence on CSP, volatile and deleted data, less control in cloud infrastructure, trust issues and large bandwidth. In cloud, strategies of digital forensic analysis ought to diverge according to each cloud model. For instance, the control over network monitoring or process by customers is more limited in SaaS and PaaS than IaaS where the customer has more privilege over computing resources as given in Table 1. Table 1 displays how CSP can accomplish and control cloud services. The traditional digital forensics for computer systems would vary for the cloud models in a cloud environment.

Table 1 Cloud service provider's control over services models [8]

Layer	Model		
	IaaS	PaaS	SaaS
Applications	✕	✕	✓
Data	✕	✕	✓
Runtime	✕	✓	✓
Middleware	✕	✓	✓
Operating system	✕	✓	✓
Virtualization	✓	✓	✓
Servers	✓	✓	✓
Storage	✓	✓	✓
Networking	✓	✓	✓

Collection process in the SaaS and IaaS models will not be similar where, in SaaS, the CSP have control over application data, while in IaaS, the customer would have control over data generated by the virtual machines. On the other side, in private cloud, investigators have physical access to the evidence data, but they merely can get physical access to the data in public cloud [8].

3   Related Work

Rani and Geethakumari [5] proposed an efficient approach for performing digital forensic investigation in cloud based on virtual machine (VM) snapshots. The basic idea of this approach is that the Cloud Service Provider stores snapshots for virtual machines identified as criminal using an intrusion detection system (IDS). Concurrently, the CSP ought to be demanded for log files of the suspected VM and the investigator collects and analyzes the log files to find any evidence related to the malicious activities. Dykstra et al. [6] suggested a cloud management plane for using in the IaaS model where the CSP can make a vital task in data collection by providing a web-based management console. Using this console panel, users along with digital investigators extract related digital evidence from virtual machine image, network, and process in an effective manner. There tricky aspect of this solution is that it needs an additional level of trust in the management plane as well as evidence collection methods; however, where investigators have physical access to the system, this level of trust is not necessary.

Simou et al. [7] proposed a model for supporting the process of Cloud Forensics and moved existing research one step more through recognizing the major ideas, actors, and their interactions that take part in a Cloud Forensics process by presenting a means of new meta-models. They also presented a running example as well for well understanding the proposed concepts related to Cloud Forensics investigation scenario. Povar and Geethakumari [8] emphasized on the procedures of identifying and analyzing digital evidence in the cloud environment with regard to the cloud customer as well as services provider. They proposed and introduced a heuristic model for performing Cloud Forensics process. Their proposed model emphasized on the necessities of modifications required in data collection, preservation, and analysis in the digital investigation process. Zawoad et al. [9] introduced open Cloud Forensics (OCF) model which considered the new role of the cloud services provider (CSP) to support reliable and effective digital forensics in the cloud. They suggested that the proposed model can be used by cloud architect to design clouds that support trustworthy cloud forensic investigation.

4 Proposed Cloud Forensics Investigation Model

This section explains architecture and description of the proposed model.

4.1 Model Architecture

The proposed model consists of the following components as shown in Fig. 2 as follows:

1. **Cloud User (CU):** Request a cloud services provider (CSP) to create a virtual machine (VM) with specific descriptions.
2. **Cloud Services Provider (CSP)/Cloud Server:** Responses to the CU and creates the VM; hence, a direct connection between the CU and the required VM would be established.
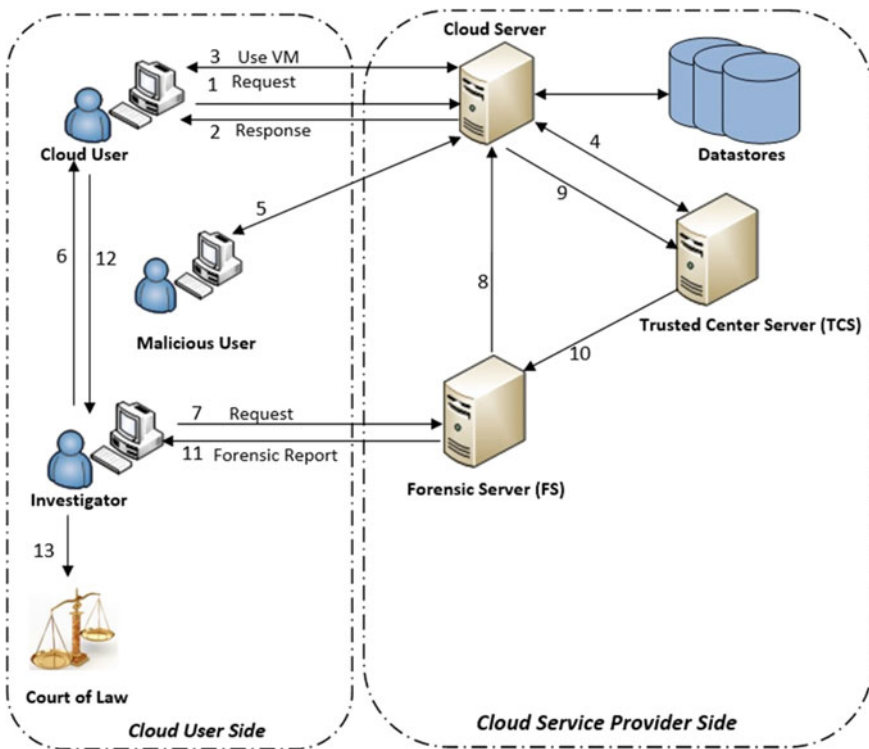
Fig. 2 Architecture of proposed model

3. **Trusted Center Server (TCS):** Working in the background, it takes periodically snapshots for the VM for the purpose of forensic investigation based on a request from the CSP.
4. **Cloud Investigator (CI):** Remotely controls the investigation process in forensic server (FS) and prepares a final report about a committed crime that occurred in the cloud environment.
5. **Forensic Server (FS):** Manage, handle, and process digital evidences.
6. **Malicious User (MU):** Perform and launch malicious and illegal activities.

4.2 Model Description

The description of the proposed model which is shown in Fig. 2 is as follows:

1. CU requests CSP for VM with specific descriptions.
2. The CSP responses to the CU's request and the VM is created.
3. The CU starts to use his/her VM in the cloud side.
4. TCS starts to take periodically snapshots for the running VM.
5. Imagine an incident occurs against the CU's VM (i.e., hacking scenario by the malicious user).
6. The CU requests a cloud investigator (CI) to investigate the crime.
7. The CI connects to the forensic server (FS) in CSP side to request about the CU's VM.
8. The FS requests the CSP about the CU's VM then the CSP requests the TCS to communicate with FS.
9. The TCS sends the required VM's snapshots and Base Files to the FS for performing the investigation process phases which are identification, preservation, analysis, and presentation to reconstruct the committed crime events.
10. A final report is then generated about the committed crime.
11. The CI sends the report to the CU or the court of law as admissible evidence about the committed crime.

5 Malicious Scenario

This section discusses a hypothetical malicious scenario and the importance of using the proposed model in cloud environment. The malicious scenario can happen in the cloud as follows [10]: Bucky is a cloud user who runs a virtual machine (VM) in a cloud. John is a malicious user. He rented also another VM in the cloud to enable him to access cloud infrastructure. John decided to use Bucky's VM to launch many types of attacks to other VMs running in the cloud to steal their data. One of the attacked VM's owner stores important data which are stolen by the

malicious attacker. Consequently, the owner requested a forensic investigator to investigate the crime. The investigator discovers that the Owner's VM records each visiting activity information like IP address. Examining the visiting user records, the investigator found that the attacking is done through Bucky's VM. Eventually, the investigator issued a subpoena to the corresponding cloud provider to provide him the network logs for this particular IP addresses. On the other hand, there are two scenarios might have occurred. Firstly, John managed to collude with the cloud provider after the attack. Therefore, while providing the logs to the investigator, the cloud provider supplied a tampered log to the investigator, who had no way to verify the correctness of the logs. Under this circumstance, John would remain undetected. Secondly, Even if the cloud provider was honest, John could terminate her rented VM and leave no trace of the attack. Hence, the cloud provider could not give any useful logs to the investigator. Finally, if the investigator can't found the truth, Bucky would be responsible for the malicious attack.

From the above malicious scenario, there are many challenges that may face digital investigators ability to investigate this crime that occurred in the cloud environment. There are two situations facing the digital investigators during the investigation of the crime as:

1. John succeeded to conspire with the cloud service provider after the attack. And so, while providing the logs to the investigator, the cloud provider provided a modified log to the investigator, who had no way to verify the correctness of the logs. Under this situation, John will continue unnoticed.
2. Although the cloud provider was sincere, John could lay off her rented VM and not keep any traces related to the crime. Therefore, the cloud provider could not provide any valuable logs to the investigator.

In the end of the investigation, if the investigators did not discover the truth, in this case the cloud user, Bucky, would be responsible for the committed crime in the cloud. To identify who makes the malicious attack, there is the need to introduce new methods, techniques, or models to help digital investigators and examiners in their mission in the investigation of cybercrimes to be able to implement digital forensics procedures successfully in the cloud. From this scenario, we can conclude that the benefits and advantages of using the proposed model in real cloud scenarios are that it can save time and costs in performing digital investigation process through recording and monitoring the running virtual machines. This will help to identify malicious users in the cloud and discover any vulnerabilities and weaknesses inside them.

6   Conclusion and Future Work

In the last years, cloud computing adopted by several users, companies, and organizations where the number crimes against it increased so that it is become very urgent for providing cloud services to them with ensuring that the cloud providers support and help to provide the digital investigation process in forensically sound and timely fashion manner. This paper therefore introduces a cloud forensic investigation model (CFIM) to investigate cybercrimes in the cloud computing environment. This model will increase probability of tracking attackers, determine weaknesses of virtual machines in the cloud for future use, and help in performing the process of extraction and collection of digital evidence about committed cybercrimes through providing vital and useful information. A hypothetical scenario is discussed to clarify the value of using the proposed system in real forensic investigation in the cloud. In future work, we scheduled to carry out the proposed model in real cloud computing environment.

References

1. Mell P, Grance T (2011) "The NIST definition of cloud computing." 20–23
2. McKemmish R (1999) What is forensic computing? Australian Institute of Criminology, Canberra
3. Ruan K et al (2011) Cloud forensics. Advances in digital forensics VII. Springer, Berlin, pp 35–46
4. Ruan K et al (2013) Cloud forensics definitions and critical criteria for cloud forensic capability: an overview of survey results. Digit Invest 10(1):34–43
5. Rani DR, Geethakumari G (2015) An efficient approach to forensic investigation in cloud using VM snapshots. In: IEEE international conference on pervasive computing (ICPC)
6. Dykstra J, Sherman A (2012) Acquiring forensic evidence from infrastructure-as-a-service cloud computing: exploring and evaluating tools, trust, and techniques. In: DoD Cyber Crime Conference
7. Simou S et al (2015) A meta-model for assisting a cloud forensics process. Risks and security of Internet and systems. Springer, Berlin, pp 177–187
8. Povar D, Geethakumari G (2014) A heuristic model for performing digital forensics in cloud computing environment. Security in computing and communications. Springer, Berlin, pp 341–352
9. Zawoad S, Hasan R, Skjellum A (2015) OCF: an open cloud forensics model for reliable digital forensics. In: 2015 IEEE 8th international conference on cloud computing (CLOUD). IEEE
10. El-Din Hemdan E, Manjaiah DH (2015) Exploring digital forensic investigation issues for cyber crimes in cloud computing environment. In: Proceeding of international conference on computer communication and networks (i3CN), on 8th May 2015

TS Fuzzy Controller of Maglev System Based on Relaxed Stabilization Conditions

Munna Khan and Amged Sayed Abdelmageed Mahmoud

Abstract This chapter addresses the synthesizing of modified TS fuzzy controllers for magnetic levitation (Maglev) system which is a complex nonlinear system. The applied TS fuzzy controller is based on piecewise Lyapunov function in modified form to relax stabilization condition of the model. The proposed technique exploits PDC controller to robustly stabilize the position of the iron ball of the Maglev system in the existence of disturbances. Furthermore, this technique will ensure the robustness, maximize the disturbance tolerance, improve the performance and reduce H_∞ performance index γ. The simulation results for control of Maglev will compare the effect of the proposed controller with other PDC controller to indicate that the new algorithms give better results under external disturbances.

Keywords Takagi–Sugeno (TS) fuzzy system · Magnetic levitation (Maglev) system · Parallel distributed compensator (PDC) · Linear matrix inequality (LMI)

1 Introduction

The stabilization of complex nonlinear systems has gained considerable attentions in the automatic control research topic in past years. Recently, the Takagi–Sugeno (TS) fuzzy approach [1] has been intensely considered for the stability analysis problem of nonlinear dynamical systems [2–7]. It is simple and effective mathematical platform to handle the nonlinear systems as it allows nonlinear model to be presented by blending a series of linear sets via the "IF-THEN" fuzzy rule. Based on TS fuzzy model, the technique of PDC is used to control and stabilize complex nonlinear systems by using state feedback controller for each rule in TS fuzzy model [8].

M. Khan (✉) · A.S.A. Mahmoud
Department of Electrical Engineering, Faculty of Engineering and Technology,
Jamia Millia Islamia, New Delhi 110025, India
e-mail: mkhan4@jmi.ac.in

A.S.A. Mahmoud
e-mail: amged1983@gmail.com

Lately, stabilization issue of TS fuzzy approach is a principal point in this regard. Therefore, significant results have been performed in order to attain and get the strict and straightforward algorithms using Lyapunov function strategy. Also, using linear matrix inequality (LMI) technique for the performance conditions and stability criteria is very effective, because the optimal procedure of the LMI approach can solve the difficulties in finding a common positive-definite matrix of quadratic Lyapunov function for stability conditions. Despite the efficiency of LMI-based approach, there is a significant issue that the solutions related to LMIs are quite conservative. Hence, considerable researches on relaxed criteria and stability conditions for TS fuzzy system have been carried out to reduce the conservatism. For example, the utilizing of order relation information of MFs to relax the conservativness is given in [9]. Where an improvement technique has been reported to relax the conservative of TS fuzzy systems by exploiting Kronecker products for quadratic Lyapunov function in [10]. Whereas, in [11], a multi-sample approach is used to propose a new relaxed algorithm for TS fuzzy. A new relaxed condition based on delay Lyapunov function is addressed in [12]. Moreover, such different kind of techniques to reduce the conservatives like non-PDC scheme [13] and switched PDC techniques [14–17] are frequently used to reduce the conservatism. The relaxed conditions and schemes are worthy addressed not only for ensuring the stability of TSF system, but also for improving other requirements such as speed performance, the domain of attraction and input–output constraint.

Maglev trains, wind turbines, launching rockets, electrodynamics' suspension and the centrifuge of nuclear reactor are examples of important applications of magnetic levitation systems. These systems are complex nonlinear and unstable open-loop systems, so numerous results have been done for building the high-performance feedback controllers to stabilize and control the Maglev systems. Up until now, designing a robust controller to manage the magnetic levitation is a challenging task, and there are various researches try to handle the Maglev systems [7, 18–20].

The scope of this work is to propose TS fuzzy controller for Maglev system. The proposed technique has a slack matrix which provides less conservative conditions, therefore, yields good results. The main purpose of modified PDC is to robustly stabilize the closed-loop system under different conditions. Based on LMI technique, the state feedback controller's gain will be obtained. Finally, simulation results are given to prove that the proposed technique guarantees the stability condition of the closed-loop nonlinear system and ensures robustness to external disturbance. The structure of this chapter is as follows. The Maglev model for current-controlled scheme is introduced in Sect. 2. The discrete-time TS fuzzy model is given in Sect. 3. The proposed PDC controller is reported in Sect. 4. Section 5 simulation results are shown to emphasize the efficiency and explore the effectiveness of the proposed method and compare it with PDC scheme. Finally, in Sect. 6, the conclusion is given. Notation: The superscript "T" represents the transpose of a matrix. An asterisk (*) located inside a matrix denotes the transpose of its symmetrical term.

2 Maglev Dynamics

The Maglev's dynamics equations are presented as [7]

$$m\ddot{h} + f = mg + E_d$$
$$f = \alpha \left(\frac{i}{h}\right)^2 \tag{1}$$

where m_d is the mass of the iron ball; h is the position of ball; E is the disturbance force; g is the acceleration gravity; α is the electromagnet constant; i is controlled current of the electromagnet. The control purpose is to stabilize the position of the ball by the force of the electromagnet.

Define the state variable as of (1) as:

$$x_1 = h, \quad x_2 = \dot{h}$$

Then the state equation

$$\dot{x}_1 = x_2$$
$$\dot{x}_2 = g - \frac{k}{m}\left(\frac{i}{x_1}\right)^2 + \frac{1}{m}f_d \tag{2}$$

3 TS Fuzzy Model

If δ_{1k} is M_{i1}… and δ_{Pk} is M_{iP},

$$\text{Then} \quad \begin{cases} x_{k+1} = A_i x_k + B_{1i} w_k + B_{2i} u_k \\ y_k = C_i x_k + D_{1i} w_k + D_{2i} u_k \end{cases} \tag{3}$$

Here $i = 1, 2, …, r$ where r is the number of model rules; suffix k in sample time; M_{ij} is the fuzzy set; $x_k \in R^n$ is the state vector; $u_k \in R^m$ is the control input; $y_k \in R^q$ is the output vector; $w_k \in R^s$ is the energy-bounded disturbance. A_i, B_{1i}, B_{2i}, C_i, D_{1i} and D_{2i} are of appropriate dimensions; δ_{1k}, …, δ_{Pk} are known premise variables at k sample.

The final outputs of the fuzzy systems are inferred as follows:

$$x_k = \sum_{i=1}^{r} \eta_i(\delta_k) A_i x_k + B_{1i} w_k + B_{2i} u_k) \tag{4}$$

$$y_k = \sum_{i=1}^{r} \eta_i(\delta_k) C_i x_k + D_{1i} w_k + D_{2i} u_k) \tag{5}$$

Where $\delta_k = [\delta_{1k} \quad \delta_{2k} \cdots \delta_{Pk}]^{\mathrm{T}}$

$$\eta_i(\delta_k) = \frac{\prod\limits_{j=1}^{P} M_{ij}(\delta_{jk})}{\sum\limits_{i=1}^{r} \prod\limits_{j=1}^{P} M_{ij}(\delta_{jk})} \tag{6}$$

For all k. The term $M_{ij}(\delta_{jk})$ is the grade of membership of (δ_{jk}) in M_{ij}. For brief expression, we will denote $\eta_{ik} = \eta_i(\delta_k)$. We have

$$\begin{cases} \sum\limits_{i=1}^{r} \eta_{ik} = 1 & \forall i \\ \eta_{ik} \geq 0 \end{cases} \tag{7}$$

For the above TS fuzzy model, Fuzzy controller can be constructed using the parallel distributed compensation (PDC) concept. The PDC control technique is synthesized by designing a linear compensator to control each of fuzzy rules. So the fuzzy controller uses the same fuzzy sets of TS fuzzy model.

Control rule i :   IF δ_{1k} is M_{i1} and $\cdots$ and δ_{rk} is M_{ir}, then

$$u_k = -F_i x_k, \quad i = 1, 2, \cdots, r$$

Where $F_i(i = 1, 2, \ldots, r)$ is the gain of feedback controller. Hence, the PDC controller can be obtained by

$$u_k = -\sum_{i=1}^{r} \eta_{ik} F_i x_k, \tag{8}$$

By combining (3), (4) and (8) then the closed-loop discrete fuzzy system is:

$$\begin{aligned} x_k &= (A(\eta_k) + B_1(\eta_k)F(\eta_k))x_k + B_2(\eta_k)w_k \\ y_k &= (C(\eta_k) + D_1(\eta_k))x_k + D_2(\eta_k)w_k \end{aligned} \tag{9}$$

Where

$$A(\eta_k) = \sum_{i=1}^{r} \eta_{ik}A_i, B_1(\eta_k) = \sum_{i=1}^{r} \eta_{ik}B_{1i}, F(\eta_k) = \sum_{i=1}^{r} \eta_{ik}F_i$$

$$B_2(\eta_k) = \sum_{i=1}^{r} \eta_{ik}B_{2i}, C(\eta_k) = \sum_{i=1}^{r} \eta_{ik}C_i,$$

$$D_1(\eta_k) = \sum_{i=1}^{r} \eta_{ik}D_{1i}, D_2(\eta_k) = \sum_{i=1}^{r} \eta_{ik}D_{2i}$$

4 Fuzzy Controller Design

In this part, state feedback controller is presented using PDC technique. The proposed controllers will be used to control and stabilize unstable system. The proposed PDC will exploit an LMI-based approach to calculate the controller's gain to solve the problem of state feedback controller for discrete-time TS fuzzy systems.

Theorem *For a given $\mathcal{H}_\infty$ performance level $\gamma > 0$ and for positive prescribed value δ, if there exists a symmetrical matrix $G = G^T > 0$, $\Omega_{ij} = \Omega_{ij}^T$, $E > 0$, Y_j, $1 \leq i, j \leq q$, satisfying the following LMIs:*

$$\Omega_{ii} \geq 0, \tag{10}$$

$$\Omega_{ij} > 0, \quad 1 \leq i, j \leq q, \tag{11}$$

$$\Delta_{ii} < 0, 1 \leq i \leq q, \tag{12}$$

$$-2\Delta_{ii} > \left(\Delta_{ji} + \Delta_{ij}\right), 1 \leq i \neq j \leq q \tag{13}$$

$$\Delta_{ij} = \begin{bmatrix} -G_j + E_j & \left(A_i E_j + B_{1i} Y_j\right)' & \left(C_i E_j + D_{1i} Y_j\right)' & 0 \\ * & G_i - E_i - E_i' & 0 & B_{2i} \\ * & * & -\gamma^2 I & D_{2i} \\ * & * & * & \boldsymbol{\delta} \end{bmatrix} + \Omega_{ij}, \qquad 1 \leq i, j \leq q \tag{14}$$

Then the controller (6) with

$$F_i = Y_i G^{-1}$$

Makes the TS fuzzy system asymptotically stable with guaranteed H_∞ performance level γ.

Remark If we choose $\Omega_{ij} = \Omega_{ii} = 0$, $1 \leq i, j \leq q$ in the theorem, and $E_j = 0$ then the LMIs will be the same as PDC scheme which is based on piece-wise Lyapunov function in [21].

5 Simulation Results

In this section, discrete-time TS fuzzy controller is employed to Maglev system to robustly stabilize the system. The proposed fuzzy controller is used to control the position of iron ball in Maglev exposed to disturbances with current-controlled scheme. The efficiency of the proposed algorithm in the disturbance attenuation will be addressed by changing the amplitude of the disturbance and compared the response of proposed PDC controller with other PDC scheme [18]. It should be

noted that the proposed controller has the same response or better than PDC scheme. The modified algorithm has relaxed conditions and less conservative. So it is more efficiency and effective to control the complex nonlinear system exposed to external disturbance than ordinary PDC controller.

Consider a following discrete-time magnetic levitation TS fuzzy system model in [7] with sampling time $T = 0.2$

$$x_{k+1} = \sum_{i=1}^{2} \delta_{ki}[A_i x_k + B_{wi} w_k + B_{ui} u_k]$$

$$y_k = \sum_{i=1}^{2} \delta_{ki}[C_i x_k + D_i u_k]$$

Where the grade of membership functions is:

$$M_i(x(k)) = \exp\left[-\frac{(x(k) - m_i)^2}{\sigma_i^2}\right] \quad i = 1, 2$$

$A_1 = \begin{bmatrix} -1.064 & -0.006442 \\ -20.64 & -1.064 \end{bmatrix}, B_{w1} = \begin{bmatrix} 0.04123 \\ 0.4123 \end{bmatrix}, A_2 = \begin{bmatrix} -1.058 & -0.005791 \\ -20.58 & -1.058 \end{bmatrix},$

$B_{w2} = \begin{bmatrix} 0.04155 \\ 0.4155 \end{bmatrix}, B_{u1} = \begin{bmatrix} -0.010 \\ -0.095 \end{bmatrix}, B_{u2} = \begin{bmatrix} -0.001 \\ -0.014 \end{bmatrix}, C_1 = \begin{bmatrix} -0.006442 & -0.006442 \\ 0 & 0 \end{bmatrix},$

$D_1 = \begin{bmatrix} 0 \\ 0.004123 \end{bmatrix}, C_2 = \begin{bmatrix} -0.005791 & -0.005791 \\ 0 & 0 \end{bmatrix}, D_2 = \begin{bmatrix} 0 \\ 0.00416 \end{bmatrix}$

$m_1 = 0.32 \quad m_2 = -0.18, \quad \sigma_1 = 0.21, \quad \sigma_2 = 1.76$

Consequently, by using a modelling language YALMIP [22] with SeDuMi solver [23], we solve the LMI of theorem, and the controller gain matrices for PDC (8) are obtained as follows

$$F_1 = [-\mathbf{156.2610} \quad -\mathbf{4.2142}], F_2 = 10^3 \times [-\mathbf{1.2197} \quad -\mathbf{0.0335}], \gamma = 0.004$$

With the initial position of ball at 1 cm, the response of the controlled system is depicted in the Fig. 1. It is noticed that the proposed PDC controller and PDC scheme stabilize and control the system. It is shown that the response obtained from proposed PDC is better than PDC and the obtained H_∞ performance via proposed PDC equal 0.004, whilst the value using PDC equal 0.0448.

Furthermore, The Maglev is exposed to exogenous disturbance $w = 20 \sin(10 k) \times \exp(-k/2)$ to notice the effect of disturbances on the controlled system and compare the results with other PDC scheme. So Fig. 2 shows the disturbance amplitude and depicts that the proposed PDC is more tolerable to disturbance than other PDC scheme.

Note that, notwithstanding of high disturbance amplitude, the system is quadratically stable as shown in Fig. 3. It is also shown that the position of ball in case of a new controller does not exceed more than 0.1 m and response of the
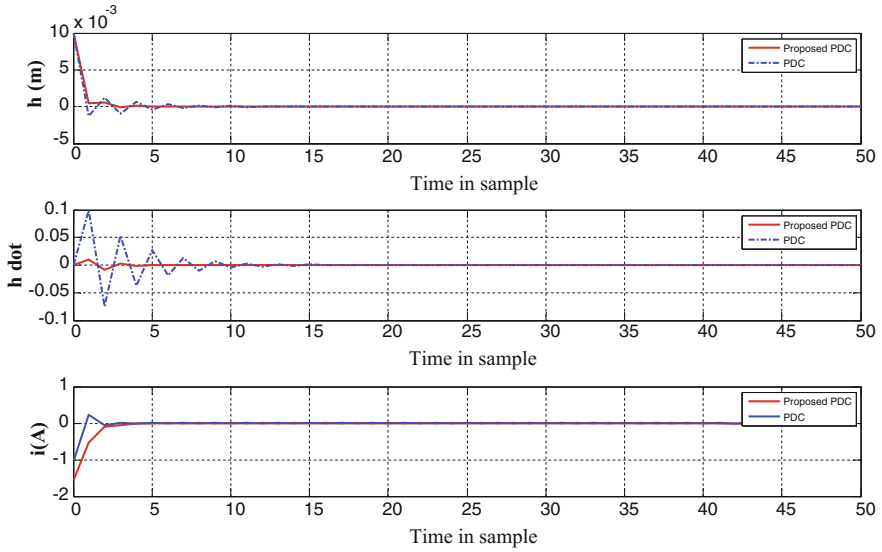
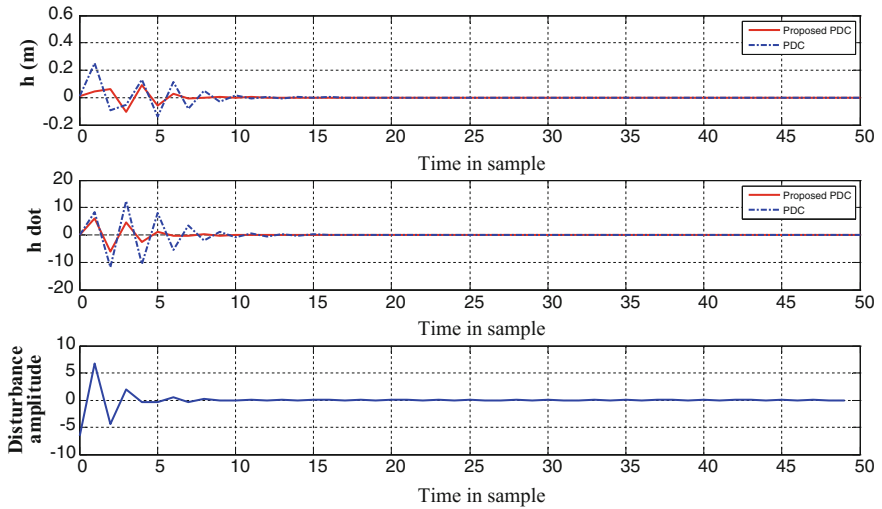Fig. 1 State response of controlled Maglev

Fig. 2 Zero initial condition of Maglev with disturbance

controlled system is decaying faster than PDC. So from the calculations and figures above, it is regarded that the modified technique presents better response and gives good results.
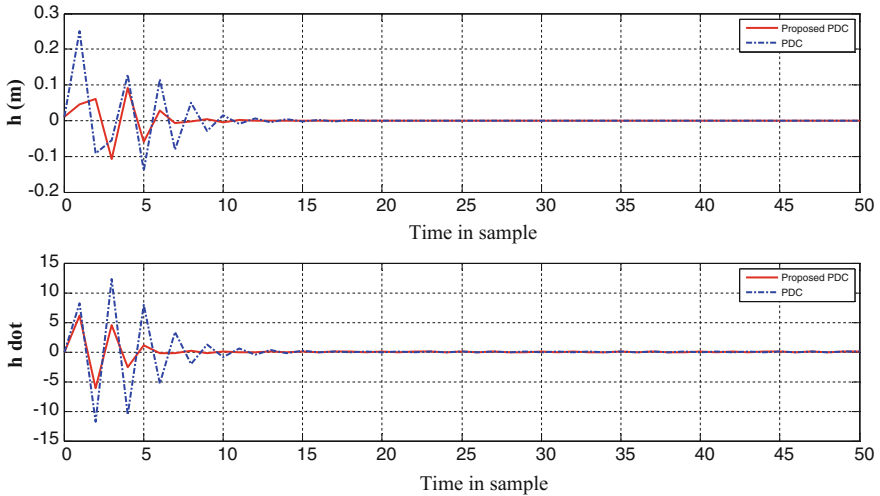
Fig. 3 State response of Maglev under large disturbance

6 Conclusions

A modified TS fuzzy controller for magnetic levitation (Maglev) system is addressed. The discrete TS fuzzy controller is based on piecewise Lyapunov function to reduce the conservativeness of the model. The proposed technique exploits PDC controller to robustly stabilize the position of iron ball of the Maglev system in the existence of disturbances. Furthermore, this technique will ensure the robustness, maximize the disturbance tolerance, improve the performance and reduce H_∞ performance index γ. The new algorithm has superior performance and has better results under external disturbances.

References

1. Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern 1:116–132
2. Wang L, Du H, Wu C, Li H (2016) A new compensation for fuzzy static output-feedback control of nonlinear networked discrete-time systems. Sig Process 120:255–265
3. Wang X, Yaz EE (2016) Robust multi-criteria optimal fuzzy control of continuous-time nonlinear systems. Syst Sci Control Eng 2583:1–23
4. Cao K, Lam H-K, Gao X, Vasilakos A (2016) H_∞ fuzzy PID control synthesis for Takagi–Sugeno fuzzy systems. IET Control Theory Appl 10:607–616
5. Tognettia ES, Oliveirab RCLF, Peres PLD (2015) H_∞ and H_2 nonquadratic stabilisation of discrete-time Takagi–Sugeno systems based on multi-instant fuzzy Lyapunov functions. Int J Syst Sci 46:76–87

6. Sadeghi MS, Farughian A (2014) Parallel distributed compensator design for process control based on fuzzy Takagi–Sugeno model. Appl Soft Comput 21:280–285
7. Mahmoud ASA, Khan M, Siddique AS (2015) Discrete-time control of Maglev system using switched fuzzy controller. India Conference (INDICON), 2015 Annual IEEE: 1–6
8. Kazuo T, Wang HO (2001) Fuzzy control systems design and analysis: a linear matrix inequality approach. Wiley
9. Chen J, Xu S, Zhang B, Chu Y, Zou Y (2016) New relaxed stability and stabilization conditions for continuous-time T–S fuzzy models. Inf Sci (NY) 329:447–460
10. Chen J, Xu S, Zhang B, Qi Z, Li Z (2016) Novel stability conditions for discrete-time T–S fuzzy systems: a Kronecker-product approach. Inf Sci (NY) 337–338:72–81
11. Deng S, Yang L, Yang D, Xie X (2016) Relaxed control design of discrete-time Takagi–Sugeno fuzzy systems: a multi-samples approach. Neurocomputing 171:106–112
12. Xie X-P, Weng S-X, Zhang H-F (2016) Reducing the conservatism of stability analysis for discrete-time T–S fuzzy systems based on a delayed Lyapunov function. Neurocomputing 171:1139–1145
13. Xie X, Ma H, Zhao Y, Ding D-W, Wang Y (2013) Control synthesis of discrete-time T–S fuzzy systems based on a novel non-PDC control scheme. IEEE Trans Fuzzy Syst 21:147–157
14. Xiang W, Xiao J, Han L (2014) Switching PDC control for discrete-time T–S fuzzy system: a membership function ranking approach. J Franklin Inst 351:3536–3558
15. Xiang W, Xiao J (2013) H∞ control synthesis of switched discrete-time fuzzy systems via hybrid approach. Optim Control Appl Methods 34:298–312
16. Chen YJ, Ohtake H, Tanaka K, Wang WJ, Wang H (2012) Relaxed stabilization criterion for T–S fuzzy systems by minimum-type piecewise Lyapunov function based switching fuzzy controller. IEEE Trans Fuzzy Syst 20:1166–1173
17. Xiang W, Xiao J, Iqbal MN (2012) control for switched fuzzy systems via dynamic output feedback: hybrid and switched approaches. Commun Nonlinear Sci Numer Simul 18:1499–1514
18. Khan M, Siddique AS, Mahmoud ASA (2016) Robust H_∞ control of magnetic levitation system based on parallel distributed compensator. Ain Shams Eng J
19. He G, Li J, Cui P, Li Y (2015) T–S fuzzy model based control strategy for the networked suspension control system of Maglev train. Math Probl Eng 2015:1–11
20. Kole A (2015) Design and stability analysis of adaptive fuzzy feedback controller for nonlinear systems by Takagi–Sugeno model-based adaptation scheme. Soft Comput 19:1747–1763
21. Wang L, Feng G (2004) Piecewise H infinity: controller design of discrete time fuzzy systems. IEEE Trans Syst Man Cybern B Cybern 34:682–686
22. Lofberg J (2004) YALMIP: a toolbox for modeling and optimization in MATLAB. In: Proceedings IEEE international symposium on computer aided control systems design, pp 284–289
23. Sturm JF (1999) Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones. Optim Methods Softw 11:625–653

Firefighting Robot

**Udayagiri R. Pranava, Vishal Guruprasad,
Preetham S. Nag and S. Suraj**

Abstract The goal here is to design and construct the prototype of a remotely controlled robot capable of extinguishing the fire. The robot consists of DC motors to facilitate the movement of the robot. Two types of extinguishing actions have been employed here. It incorporates a water pump to extinguish the fire. The water pump is immersed in a water tank that is on board. Another type of extinguisher contains a DC motor that actuates a water spray. This is used to mimic a traditional carbon dioxide-based fire extinguisher. The robot can be controlled using radio frequency transmitter. Based on which key is pressed the encode ion the transmitter sends a command to the receiver. The receiver has a decoder with it, which decodes the command and initiates the action to be taken, like rotating the DC motor in a particular direction. The robot works on direct current supplied via a switched mode power supply (SMPS).

Keywords Robot · Microcontroller · Relay DC motors · Extinguishing units

1 Introduction

In the event of a fire, there is a huge risk to the lives of both civilians and firefighters. The primary motivation for this project is to reduce the chance of life. Firefighters lay their lives on the line to protect people. The inescapable fact of the

U.R. Pranava (✉) · V. Guruprasad · P.S. Nag · S. Suraj
Department of Electronics and Communication, B.N.M. Institute of Technology,
Bangalore, India
e-mail: pran.mathura@gmail.com

V. Guruprasad
e-mail: vishalguruprasad1994@gmail.com

S. Suraj
e-mail: suraj.ssss4@gmail.com
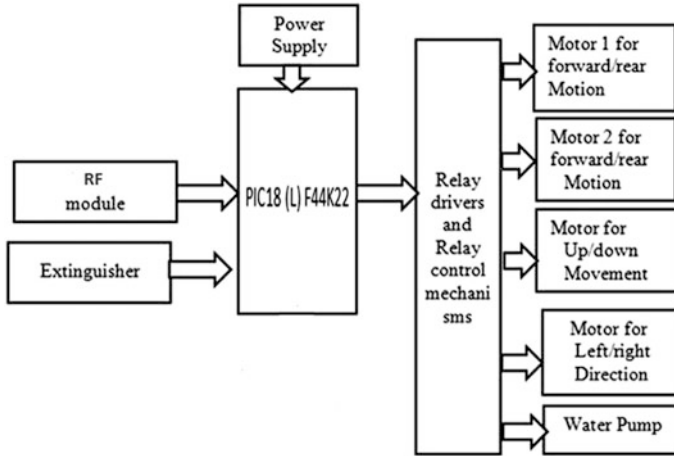
Fig. 1 Description of the proposed system

job of firefighters is a danger that they face while dousing flames. However, there may be situations where it is too difficult for the firefighters to reach and extinguish the fire. Also, a fire will cause significant loss to property. This is mainly caused by spreading of fire rather than the fire itself. It is critical to containing the fire and try to keep damages to a minimum. In situations where it is challenging and risky for human beings to be involved, we can take the help of robots.

A robot is an electronic machine with mechanical capabilities, whose actions are controlled by a computer code. Robotics deals with building, operating, and use of robots in real life [1]. The research and development in the field of artificial intelligence have given rise to robotics. Robotics is the emerging solution to problems that cannot be solved by only humans. Robots are nowadays implemented in various areas ranging from manufacturing toys to synthesizing medicines.

The firefighting robot constructed here is capable of being operated remotely. It consists of a PIC18 series microcontroller as the heart of the device. The description of the system is depicted (Fig. 1).

The microcontroller derives its power from an SMPS. An RF receiver and decoder are connected as inputs to the microcontroller. Two basic operations are done by the microcontroller. One is to control the DC motors and in turn actuate the robot's movement and the water sprayer. The DC motors are controlled via a series of relays and motor driver. Another operation is to operate the submersible water pump.

2 Hardware

2.1 Microcontroller

A microcontroller is a single integrated circuit containing a processing core, memory for storage, and other programmable peripherals. The robot constructed here consists of PIC18 (L) 44K22 microcontroller. The PIC18 series microcontroller has several advantages over other microcontrollers. It has program memory of 16 kilobytes, RAM of 768 bytes, and data EEPROM of 256 bytes. It has execution speed of about 16 million instructions per second. The most important feature is the presence of two UARTs for serial communication. It also comes with a module that can directly be interfaced with a computer via USB port [4].

2.2 Relay

A relay is an electrical switch. They are used to control several circuits using a single low-power signal. The robot includes SC 12 V DC relay. It has a coil voltage of 6–48 V [6].

2.3 Motor Driver

A motor driver amplifies small current signals to higher values to help operate a DC motor. The ULN2003A motor driver has been incorporated here. The ULN2003 is a high voltage, high current Darlington arrays. Each of these contains seven open collector CE Darlington pairs [7].

2.4 DC Motors

A DC motor is used to convert electricity into mechanical power. They use magnetic fields to achieve the conversion. There are several different types of DC motors. Heavy-geared motors are the ones employed in this robot. The largely geared motor has a rated voltage of 12 V, speed of 84 rpm, and current of 1 A all at no load. At maximum efficiency, the speed is about 75 rpm and current is 4 A.

Fig. 2 Remote used to control the robot
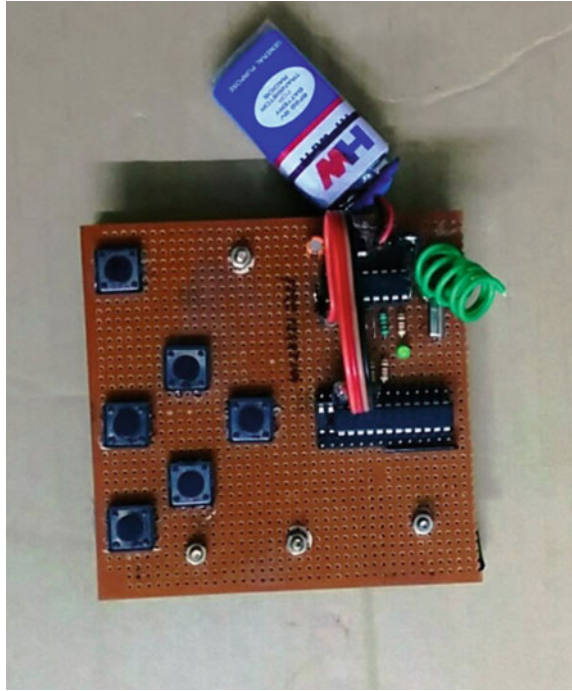
2.5 RF Transmitter and Receiver

The transmitter part consists of an HT12E encoder and TLP434A RF transmitter. HT12E encoder IC has been used for transmission of data. They are employed in systems incorporating radio signals and infrared signals. 12-bit parallel data are encoded and transmitted via a radio transmitter in serial form. Out of the 12 bits, address bits are eight and data bits are four. The output of the encoder is connected to TLP434A which uses ASK modulation and transmits data via an antenna [8] (Fig. 2).

There are ten possible movements of the robot, and six keys have been used to achieve these. Four keys are used to move the robot. When these four keys are used along with the 'shift' key, four different movements happen.

The receiver part consists of an HT12D decoder and RLP434A RF receiver. HT12D is an IC used to decode data. The encoder and decoder pairs have the same number addresses and data format. HT12D does serial to parallel data conversion and decodes them. Three comparisons are made between serial input data and local

addresses. The data code gets decoded after favorable comparison [9]. The RF signal is received by RLP434A receiver. Both TLP and RLP operate at a frequency of 434 MHz.

2.6 Extinguishing Units

A submersible water pump has been used in this robot. The pump has a power of 0.4–1.5 W and can pump out 80–120 L of water in an hour. There is one more type of extinguisher used that mimics a carbon dioxide-based fire extinguisher. Figure 3 shows the robotic arm that can be controlled by the remote. This arm has a pipe running along its length through which the water pump spurts out water. Figure 4 shows the sprayer. The sprayer is filled with water instead of carbon dioxide to depict its operation.

2.7 Power Supply

The robot gets continuous power from an SMPS. Switched mode power supply converts the available unregulated AC or dc input voltage to a regulated dc output voltage. If input supply is drawn from the AC mains, the input voltage is first rectified and filtered using a capacitor at the rectifier output. The unregulated dc voltage across the capacitor is then fed to a high-frequency dc-to-dc converter.
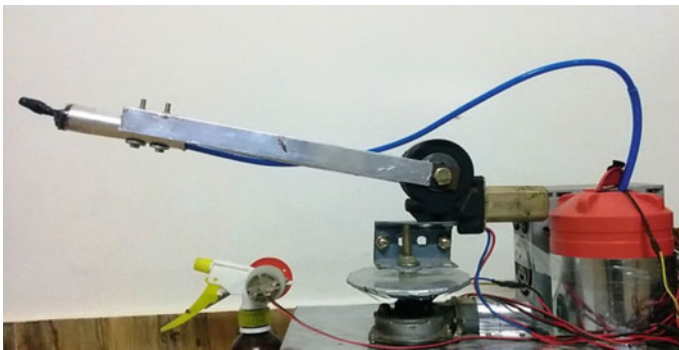
Fig. 3 Robotic arm

Fig. 4 Water sprayer

3   Software

3.1   Algorithm

In the beginning, all the input and output ports of the PIC18 microcontroller are initialized. Any command is sent through the TLP434A transmitter. Data is received via the RLP434A. Depending on what the data is intended for, individual actions take place. Motors 1 and 2 are for forward, backward, left and right movements of the robot. Rotating them clockwise will move the robot forward, Rotating them anti-clockwise will move the robot back, and stopping either of them will turn the robot left or right. Motors 3 and 4 are for angle and height adjustments of the robotic arm, respectively. If data are received for switching on of the pump or the extinguisher, own actions take place. If no data are removed, all the motors are stopped. The flowchart is as shown. Coding is done in Embedded C in MPLAB IDE platform.
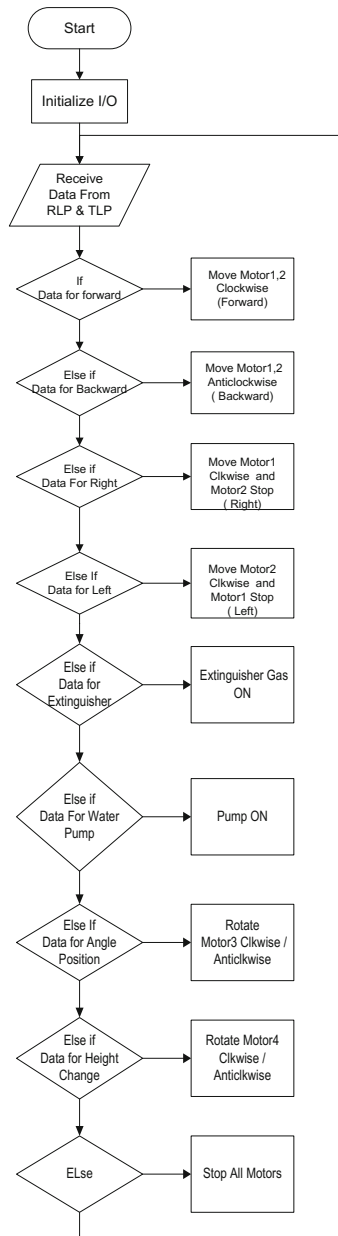
Figure 5 shows the completed robot.
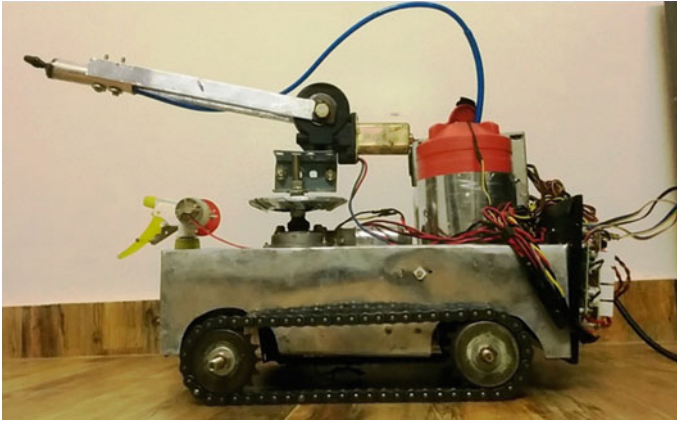
Fig. 5 Firefighting robot

4 Conclusion and Future Scope

This project has been inspired by the motive to design a system that can fight fire in situations that put the lives of firefighters in danger. This project has presented an example as to how technology can be used to achieve practical solutions to real-life problems. It aims to promote technology innovation to achieve a reliable and efficient outcome from the various instruments. Through this, it can conclude that robot could be placed where human lives are at risk. The robot can operate in the environment which is out of human reach in a relatively short time. These robots should be controlled by remote operators who have located far away from the fire site using remote communication systems like RF. The conclusion is to minimize the threat to human life and provide residential safety areas, labs, office establishments, industries and other buildings through an intelligent system that makes use of personal discretion. The robot, in its present condition, can be run by a continuous supply of power. It can be improved by replacing the power supply with a rechargeable battery. A water sprayer has been used here to imitate a carbon dioxide-based fire extinguisher. An actual carbon dioxide-based extinguisher can also be included. The robot can be improved by adding a water tank of larger size. To increase the extinguishing rate, a pump of a higher power may be added. Firefighting robot can also be automated. This provides the robot the abilities like detecting fire and maneuvering around obstacles. An independently working collection of firefighting robots communicate and work in tandem to achieve the desired result.

References

1. https://en.wikipedia.org/wiki/Robot
2. International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) Volume 2, Issue 4, July–August 2013, ISSN:2278-6856
3. FIRE FIGHTING ROBOT M.Nithiya & E.Muthamizh, II Yr - ECE IFET College of Engineering
4. Singh HP, Mahajan-d A, Sukavanam-b N, Budhraja-c V, Singh-a S, Kumara A, Vashisht-c A (xxxx) Control of an autonomous industrial fire fighting mobile robot. DU J Undergrad Res Innov
5. http://www.microchip.com/ParamChartSearch/chart.aspx?branchID=1009&mid=10&lang=en&pageId=74
6. http://www.shelfkey.com/DataSheets/6348203.pdf
7. http://www.ti.com/lit/ds/symlink/uln2003a.pdf
8. http://www.engineersgarage.com/electronic-components/ht12e
9. http://www.engineersgarage.com/electronic-components/ht12d

A Nearest Centroid Classifier-Based Clustering Algorithm for Solving Vehicle Routing Problem

V. Praveen, V. Hemalatha and P. Gomathi

Abstract A solution is designed for the vehicles to minimize the cost of distribution by which it can supply the goods to the customers with its known capacity that can be named as a vehicle routing problem. In Clarke and Wright saving matrix method and Chopra and Meindl saving matrix method, mainly an efficient vehicle routing can be achieved by calculating the distance matrix and saving matrix values based on the customers' location or the path where the customer's resides. The main objectives of this paper are to reduce the total distance and the total number of vehicles which is used to deliver the goods to the customers. The proposed algorithm is based on K-means clustering algorithm technique which is used in the data mining scenario effectively. The proposed algorithm decreases the total distance and the number of vehicles assigning to each route. The important thing needs to consider is that this new algorithm can enhance the Chopra and Meindl saving matrix method and Clarke and Wright saving matrix method.

Keywords K-means clustering · Centroid · Cluster · Saving matrix
Vehicle routing problem

1 Introduction

In general, there are many practical applications which can provide efficient distribution of goods to the customers. Goods in the sense it can be any home appliance products which are used daily. The vehicle scheduling problem was first

V. Praveen (✉) · V. Hemalatha
CSE, N.S.N College of Engineering & Technology, Manalmedu, Karur 639003, India
e-mail: praveen.sfj@gmail.com

V. Hemalatha
e-mail: hemalathav_cse@nsn.ac.in

P. Gomathi
EEE, N.S.N College of Engineering & Technology, Manalmedu, Karur 639003, India
e-mail: gomathip_eee@nsn.ac.in

formulated in the year 1959 [1] and in that set of customers with each of its known locations and known demand for any commodity, and that required goods can be delivered to the customer from a single depot by some calculated amount of delivery vehicles with some basic conditions and constraints are specified [2]:

  (i)   The demands of all customers are met;
 (ii)   Each customer is served by exactly one vehicle;
(iii)   For each route, the total demands must not exceed the capacity of the vehicle which is already defined.

From a depot, different products must be distributed to several retailers. An efficient collection (or) distribution of goods keeps transport inventories low, and it saves resources and energy. Therefore, vehicle routing is one of the important topics for this kind of problems.

The vehicle routing problem is a common name given to a whole class of problems involving the visiting of customers by using vehicles. These problems derive their name from the basic practical problem of supplying geographically dispersed customers with goods using a number of vehicles operating from a common goods depot (or) warehouse.

An example for a single-depot-based vehicle routing problem is shown in the Fig. 1. For a classical vehicle routing problem, the best solution is to serve the goods to the customers exactly once by starting from and ending to the depot.

The main objective is to reduce the overall transportation cost by satisfying all the constraints. The cost for transporting the goods can be reduced by minimizing the total distance travelled and as well as the total number of vehicles. While comparable to the classical vehicle routing problem, the majority of the real-world problems are much more complex to solve. In general, the classical vehicle routing
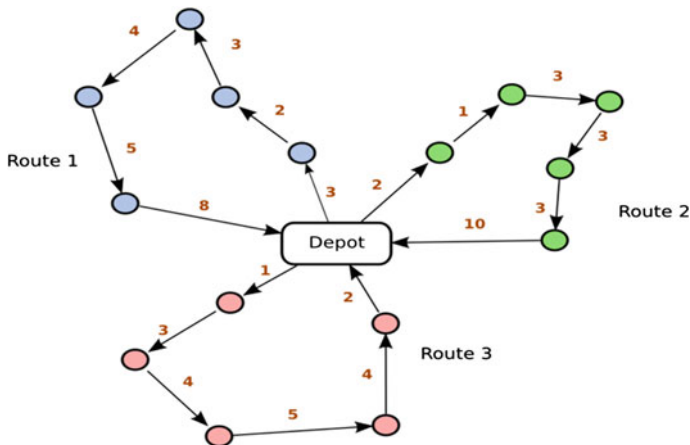

Fig. 1 Example for vehicle routing problem

problem is based on some constraints like the total vehicle capacity or some time interval to reach the customers.

A single-depot vehicle routing problem uses a single depot (or) warehouse for delivering the goods to the customers, and several algorithms and saving methods are proposed for solving the single-depot-based real-time problems. In general, the VRP is a Combinatorial Optimization Problem [3] and it consists of two main things that are depot and destinations. A formal example for this kind of problem is Soft Drink Company, in that they are traveling from the company to all the retail stores to distribute the products and again came back to the company. The main constraint followed here is to visit the customers exactly once.

Vehicle routing problem is also known as vehicle scheduling or delivery problem. For example, garbage must be collected from households and industries to a distracting place, so for transportation that we need an efficient route to travel from one place to another. It is very much useful for the daily transportation because it reduces the cost of forming the routes based on the capacity of the vehicle.

The rest of the paper is organized as follows. Section 2 presents the literature review about the existing methodologies. Section 3 presents a numerical example, using Clark and Wright saving matrix method. Section 4 describes the proposed methodology-based K-means clustering method. Section 5 describes the comparison results. The final conclusion is presented in Sect. 6.

2 Literature Survey

The first article for the 'truck dispatching problem' was published by Dantzig and Ramser who presented a larger truck dispatching problem, that is referred to as D&R problem, and many more publications have been made which is completely relevant to this article after it was published.

The Clark and Wright algorithm [4] is one of the most popular heuristic algorithms in the vehicle routing problem. Cordeau et al. [5] present that parallel version is much better because the merge yielding the largest saving is always implemented, but the sequential version keeps expanding the same route until there is no longer feasible route.

Chopra and Meindl [6] provide a solution for transport planning, in that they present a routing and scheduling, transportation problem for a company in which they solve with a method called saving matrix method. It can be classified into four steps which are as follows: (1) identify the distance matrix for the given location, (2) calculate the saving matrix using the distance matrix values, (3) assign customers to vehicles or routes, and (4) sequence the customers within the routes. The first two steps are explained clearly. The third step is that assign the customers to vehicles and routes by which initially each customer is assigned to a separate route. If the two routes can provide a feasible solution by which it does not cross the limited capacity means, it can be combined. The procedure is continued until no more feasible combinations are possible.

For a transportation problem, Lumsden (2006) and Jonsson (2008) present a similar explanation, but it is not clear. Rand [7] made an analysis and presents an article about the different saving methods for the vehicle routing problems, in that he argues about the parallel version, because it is not always better than the sequential version. Parallel version is a heuristic, and there is no assurance from the obtained results that it produces the optimal solution or near-optimal solution.

Mainly Chopra and Meindl saving matrix method [8] decreases the total length of the trip about 7%, and even a better solution is obtained against this method by decreasing the total distance of about 3%. Chopra and Meindl description presents a clearer idea about saving methods than the original Clarke and Wright algorithm (1964), even though so many wonder that how a vehicle routing is done effectively. The method presented here is widely used in the teaching section since 2008, because it provides an excellent result and is easy to implement when compared to the Clarke and Wright saving matrix method. An important advantage for saving methods in vehicle routing problem is its simplicity and robustness. So it is a best suited one for vehicle routing problem.

Anders Segerstedt [9] provides an enhanced solution for Clarke and Wright saving matrix method by introducing the search procedure method. It is same as that of Clarke and Wright algorithm, but in addition to that, the final routes are arranged in different orders to reduce the total distance. This method does not provide solution for vehicle routing problems because it only produces an enhanced solution for truck dispatching problem which was already solved by Dantzig and Ramser in the year 1959.

3 Numeric Example

The ultimate base for the vehicle routing problem is travelling salesman problem because the constraint 'visit all the customers exactly once' used in the vehicle routing problem is as same as that of travelling salesman problem. Imagine that a delivery boy must visit some 'n' number of customers and returned to the starting point after visiting all the customers exactly once and the total cost for visiting all the customers is the major problem.

Laporte [10] describes the Clark and Wright algorithm with the following steps:

Step 1   Select the warehouse as the central city and a distance matrix calculated.
Step 2   Calculate the saving $S_{ij} = c_{i0} + c_{0j} - c_{ij}$ sigh for all pairs of cities (customers) i, j $(i = 1, 2, \ldots n; j = 1, 2, \ldots n; ij)$.
Step 3   Order the savings, S_{ij} from largest to smallest.
Step 4   Starting with the largest savings, do the following:

   (a) If linking cities I and j results in a feasible route, then add this link to the route;
   (b) Try the next savings on the list and repeat (a).

Do not break any links which are formed earlier. Start new routes when necessary and stop when all the cities are on a route.

The solution is to obtain a minimum cost route to visit all the customers exactly once. Suppose when the cost for traveling from city x to city y is equal to the cost of city y to city x, then the problem is symmetric. If $C_{xy} \neq C_{yx}$, then the problem is asymmetric. Starting from the central depot (or warehouse), goods or items are delivered to the customers: 0–8. Initially, the distance for each customer is presented in the Table 1 as locations and the demands for each customer are also given.

According to the existing scenario, both the Clarke and Wright saving matrix method and Chopra and Meindl saving matrix method use the symmetric cost for returning to the depot, i.e., the distance from 1 to 5 is the same as the distance from 5 to 1.

The location for the depot (or warehouse) is (40, 40) (x-axis and y-axis values), and obviously, the demand is zero. Based on the customer location, initially distance matrix is calculated using the Eq. 1 and its cost are symmetric.

Equation 1, Distance matrix formula

$$D(c_i, k) = \sqrt{(x_{c_i} - x_k)^2 + (y_{c_i} - y_k)^2} \tag{1}$$

Equation 1 represents the distance between the customer c_i and the depot k. The calculated distance matrix values are shown in the Table 2. Based on the calculated distance matrix values, the cost savings are calculated using the Eq. 2, and here also, the values are symmetric which are shown in the Table 3.

Equation 2, Saving matrix formula,

$$S(c_i, c_j) = D(k, c_i) + D(k, c_j) - D(c_i, c_j) \tag{2}$$

Once the values are found, it is arranged in a non-increasing order (from largest to smallest), and then, the routes are combined one by one up to which the total capacity is reached.

Table 1 Distance and demands for customers

Customer	Location	Demand
1	(22, 22)	18
2	(36, 26)	26
3	(21, 45)	11
4	(45, 35)	30
5	(55, 20)	21
6	(55, 45)	16
7	(26, 59)	29
8	(55, 65)	37

Table 2 Distance matrix calculation

C_{ij}	0	1	2	3	4	5	6	7	8
0	–	26	15	20	7	25	16	24	29
1		–	15	23	26	33	40	38	54
2			–	24	13	20	27	35	43
3				–	26	42	34	15	39
4					–	18	14	31	32
5						–	25	49	45
6							–	32	20
7								–	30

Table 3 Saving matrix calculation

S_{ij}	1	2	3	4	5	6	7	8
1	–	26	23	7	18	2	12	1
2		–	11	9	20	4	4	1
3			–	1	3	2	29	10
4				–	14	9	0	4
5					–	16	0	9
6						–	8	25
7							–	23
8								–

Table 4 Solution for the problem with four routes

	Trip	Total distance	Total demands
Route 1	0-1-2-5-0	86	65
Route 2	0-3-7-0	59	40
Route 3	0-4-0	14	30
Route 4	0-6-8-0	65	53

If the newly added route exceeds the total capacity, then the new link is discarded and the previously formed links are undisturbed. Based on the total capacity defined earlier, the number of vehicles may found along with the feasible route for each vehicle.

The maximum vehicle capacity defined in this example is 70. Table 4 shows that the result with the total distance of 224 and four vehicles is needed for that transportation. Finally, the search procedure applies to the final routes, and it reduces the total distance of 1% compared to the Clarke and Wright saving matrix method, but the total number of vehicles is increased by one.

The search procedure method does not produce a better solution, but Chopra and Meindl saving matrix method produces a result of about 3% decrease in the total cost of the trip. This decrease is not enough when the vehicle is used daily for delivering the goods to the customers, so for these kind of methods are proposed to solve all kinds of vehicle routing problems.

4 Proposed Method

The first parameter is the total distance travelled to deliver the goods or some products to different distribution points. Because of routing problems, the main thing that all need to focus is on the total distance. If the total distance is reduced at some amount of range, definitely the total cost is reduced in parallel.

The second parameter is the total number of vehicles. This type of parameter will help for a large instance set of problems which are mainly related to real-world applications. When the total distance is reduced to a somewhat minimum, then the time consumption is reduced because the time consumption is calculated based on the total distance traveled to distribute the goods to the customers.

In existing methods, initially the distance matrix values are calculated using the Eq. 1, and using that value, saving matrix values are calculated to find the most efficient route for the vehicles which are going to deliver the goods to the customers, and simultaneously, the total load does not exceed the maximum vehicle capacity.

Some of the methods used to solve the vehicle routing problem are min–min and max–min methods [11] which are effectively used to solve the task scheduling problems with cloud computing. Using these two methods, these methods will reduce the total cost when compared to the Chopra and Meindl saving matrix method. Initially, the locations (*x*-axis and *y*-axis) are marked, and based on the capacity (or limit) of the vehicle defined, the total number of vehicles needed to travel can be found.

In min–min method, the distance matrix values are found as like the Clark and Wright saving matrix and Chopra and Meindl saving matrix methods. The distance from depot to all customers is found in the distance matrix calculation, and from those set of values, the highest minimum values are selected and routed to that customer. After reaching those highest minimum distance customers, again the distance is calculated from that point to all the customers who are all unvisited. The process is carried out by all the customers are reached by the vehicle. The return path for reaching the depot is same because it follows symmetric method.

Max–min method is an exact opposite to the min–min method. Initially, the first three largest maximum values are selected and that three customers are routed. Once the routing is finished from the depot, again the distance matrix is calculated from the present customer to all the customers who are all unvisited. This process is carried out again and again until all the vehicles reached the depot.

These two methods can work well for large set of problems with ease, but the results are same. In order to yield a better result compared to all these methods, a new methodology is needed.

4.1 Enhanced K-Means Clustering Algorithm (EKCA)

The third proposed methodology is based on k-means clustering algorithm method for solving the vehicle routing problem with multiple depots [12]. By following the k-means algorithm, the centroid values which can satisfy the termination condition can act as a cluster points or depots.

4.1.1 Algorithm for Enhanced K-Means Clustering

The steps for enhanced K-means clustering algorithm are as follows:

1. Choose the number of clusters, k.
2. Randomly generate k clusters and determine the cluster centers, or directly generate k random points as cluster centers.
3. Assign each point to the nearest cluster center.
4. Recompute the new cluster centers.
5. Repeat the two previous steps until some convergence criterion is met or the assignment has not changed.
6. Apply search procedure.

4.1.2 Flowchart for Enhanced K-Means Clustering Algorithm

The flowchart for the enhanced K-means clustering algorithm (EKCA) is shown in the Fig. 2.

4.1.3 Numeric Example

A set of 15 customers with their locations and demands are shown in the Table 5. Initially, the customers are divided into two set of points. Each set of points refers to one cluster point. When the clustering is performed for a set of values, there may be two or more values that can be formed. Based on the algorithm of k-means clustering, two centroid values are found with two sets of different customers. The final cluster points and their values which satisfy the termination condition are shown in the Table 6.

Each customer belongs to one centroid point, otherwise called as depot. Using the Euclidean distance formula, the distance between the customers to particular
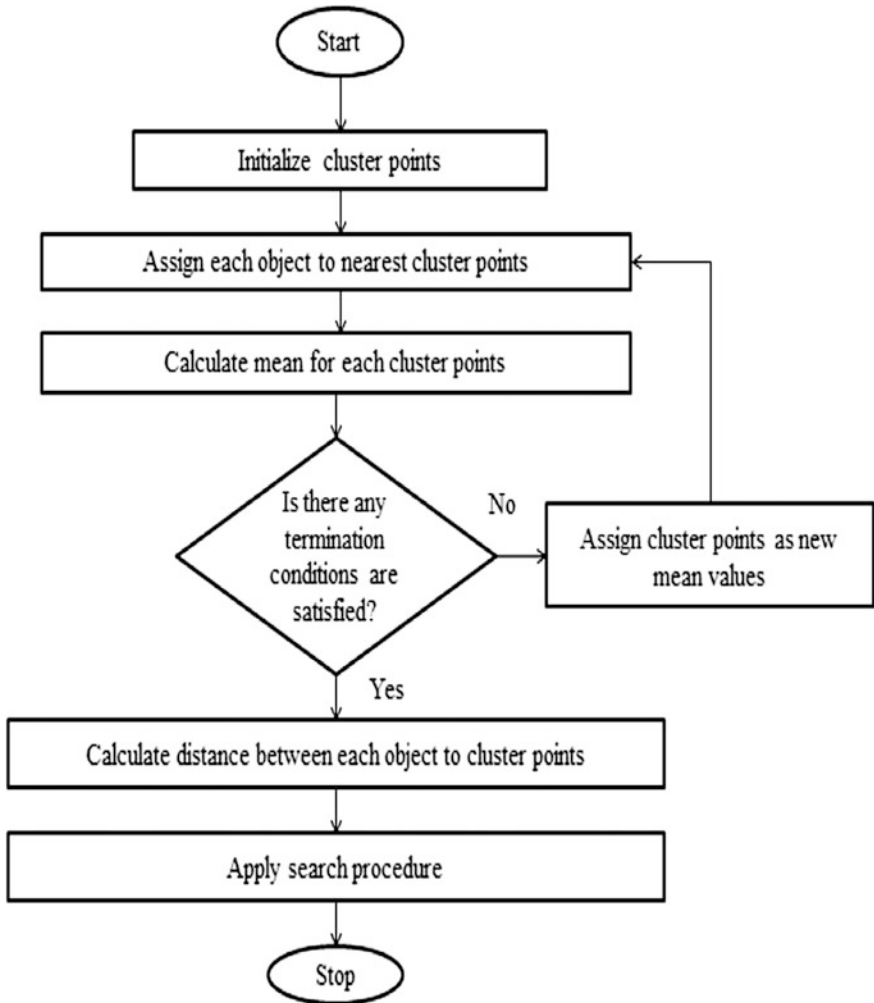
Fig. 2 Enhanced K-means clustering algorithm

depots is calculated, and finally, the total distance and the total number of vehicles needed to perform efficient transportation are found. This method definitely pro-vides a good solution compared to the previous proposals by different authors.

Table 5 Customers with location and demands

Customers	A (*x*-axis)	B (*y*-axis)	Demands
1	40	40	12
2	36	26	21
3	21	45	25
4	45	35	15
5	10	10	16
6	55	45	24
7	26	59	12
8	55	15	17
9	40	30	20
10	20	14	25
11	44	11	15
12	64	11	12
13	19	24	16
14	33	26	21
15	15	12	13

Table 6 Final centroid points with satisfied values

Cluster points	Customers	(*A*, *B*)	Centroid value 1 (37, 40)	Centroid value 2 (33, 15)
1	1	(40, 40)	3	25.961
1	3	(21, 45)	6.32	28.071
1	4	(45, 35)	2	25.079
1	6	(55, 45)	3.162	26.925
1	7	(26, 59)	5	30.265
1	9	(40, 30)	2.236	27.459
1	14	(33, 26)	0	25.317
2	2	(36, 26)	14.035	11.40
2	5	(10, 10)	26.076	10.44
2	8	(55, 15)	23.194	2.236
2	10	(20, 14)	25	3.162
2	11	(44, 11)	25.317	0
2	12	(64, 14)	27.01	5.385
2	13	(19, 24)	25.079	2
2	15	(15, 12)	25.317	0

5   Comparison Results

The sample problems were solved by NetBeans IDE, and the proposed enhanced k-means clustering algorithm shows the better result compared to Clarke and Wright saving matrix method. The results for difference in total number of vehicles and total distance are shown in the Figs. 3 and 4, respectively.

Fig. 3 Number of vehicles
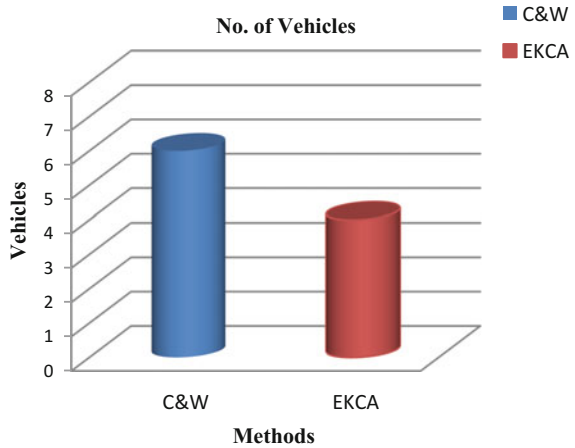
Fig. 4 Total distance
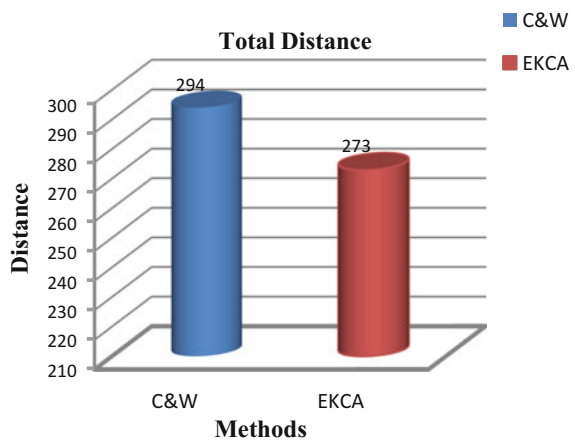
6   Conclusion

While calculating the saving matrix approach for the Clark and Wright method, it provides a good solution for the small instance set, but for a large instance set, it does not yield a better result. The proposed method shows a better solution against the previously proposed methods and k-means clustering methods which are mainly used in the data mining concepts. Among all the existing methods, enhanced K-means-based clustering method can reduce the total number of vehicles of about 40% (small set of problems) and total distance of about 7% while using multidepots for delivering the products to customers.

References

1. Dantzig GB, Ramser JH (1959) The truck dispatching problem. Manag Sci 6(1):80–91
2. Afshar-Nadjafi B, Afshar-Nadjafi A (2014) A constructive heuristic for time dependent multidepot vehicle routing problem with time-windows and heterogeneous fleet. J King Saud Univ Eng Sci
3. Christofides N, Mingozzi A, Toth P (1979) The vehicle routing problem. In: Christofides N, Mingozzi A, Toth P, Sand C (eds) Combinatorial optimization. Wiley, Chichester, pp 315–338
4. Clarke G, Wright J (1964) Scheduling of vehicles from a central depot to a number of delivery points. Oper Res 12(4):568–581
5. Cordeau J-E, Gendreau M, Laporte G, Potvin J-Y, Semet F (2002) A guide to vehicle routing heuristics. J Oper Res Soc 53(5):512–522
6. Chopra S, Meindl P (2004) Supply chain management—strategy, planning and operation, 2nd edn. Pearson Prentice hall, Upper Saddle River
7. Rand GK (2009) The life and times of savings method for vehicle routing problems 25 (2):125–145. http://www.orssa.org.za
8. Chopra S, Meindl P (2010) Supply chain management—strategy, planning and operation, 2nd edn. Pearson Prentice hall, Upper Saddle River
9. Segerstedt A (2013) A simple heuristic for vehicle routing—a variant of Clarke and Wright's saving method. Lulea University of Technology, Lulea
10. Laporte G (1992) The vehicle routing problem: an overview of exact and approximate algorithms. Eur J Oper Res 59(3):345–358
11. Sharma G, Bansal P (2012) Min–min approach for scheduling in grid environment. Int J Latest Trends Eng Technol (IJLTET) 1(1)
12. Prins C (2009) Two memetic algorithms for heterogeneous fleet vehicle routing problems. Eng Appl Artif Intell 22:916–928

Design and Evaluation of Turbine Spirometer: A New Approach

Manisha Mhetre, Yogesh Patil and H.K. Abhyankar

Abstract Respiratory disorders are increasing day to day due to air pollution. Early detection of the disease is very important from diagnostic point of view. Spirometry is the technique to assess the lung disorder, asthma. Exhaled air flow rate and volume is measured by exhalation through the spirometric pipe using sensor. This paper presents a development of the turbine type of spirometer and its testing. Out of many sensors, turbine type of sensor is one of the cost-effective and accurate methods of air flow measurement, which is addressed in this paper. Turbine blades are designed using simulation software for maximum Cp (efficiency), i.e., for wind power conversion. After a number of iterations, blade angle, number of blades, and pitch angle are optimized for this application. Turbine is manufactured by rapid prototyping for the given designed parameters. Turbine efficiency obtained is good for wind power conversion. Electrical power is measured by connecting generator to the turbine and converted appropriately to the spirometric parameter. Effectiveness of the turbine spirometer is tested by carrying out measurement using Labview software with human begins with different BMI and gender. Values such as forced vital capacity (FVC) and forced expiratory volume in one second (FEV_1) are measured and compared with standard spirometer. Results are encouraging. Further designing will have development of cost-effective spirometer with a new approach.

Keywords Lung disorders · Spirometry · Turbine spirometer
Turbine efficiency · Labview

M. Mhetre (✉) · Y. Patil · H.K. Abhyankar
Department of Instrumentation Engineering, Vishwakarma Institute
of Technology,
Pune, India
e-mail: manisha.mhetre@gmail.com

Y. Patil
e-mail: ypatil7867@gmail.com

587

1 Introduction

Now days, environmental conditions are affecting adversely human health conditions. Air pollution has significant effect on respiratory system. Many communicable lung diseases are increasing and need to be diagnosed in their early stages. Unfortunately, there is a need of a cost-effective and easy device for the lung assessment. There are many methods available for lung parameter measurements such as body box plethsymography, spirometry, and X ray imaging. Out of these spirometry is one of the important and less tedious methods of measurement. It measures exhaled air volume and/or speed of air flow for analysis. Different sensors are used for the measurement of air flow [1]. Out of these, turbine type is one of the prime methods used in air flow detection. But due to inertia of the turbine, errors occur in counting the number of revolutions for parameter calculations. This issue is addressed by measuring generator power connected to the turbine instead of its revolutions. Spirometer development is presented in this paper.

1.1 Spirometry

In a spirometry test exhaled air flow rate in a special spirometric pipe is measured as it passes through the mouthpiece. Sensor signal is analyzed to calculate different parameters and indicated by plotting flow versus volume (FV) or volume versus time (VT) graphs of expiration. Figure 1 shows the two graphs.

Two important parameters forced vital capacity (FVC) and forced expiratory volume in one second (FEV_1) indicate severity of the disease as: FVC in between 60 and 79% of the standard value, then mild lung dysfunction. When FVC is in between 40 and 59% of the standard value, moderate lung dysfunction and if the FVC is below 40%, subject needs immediate medication. These values are having diagnostic importance to indicate the lung conditions.

Fig. 1 Flow versus volume
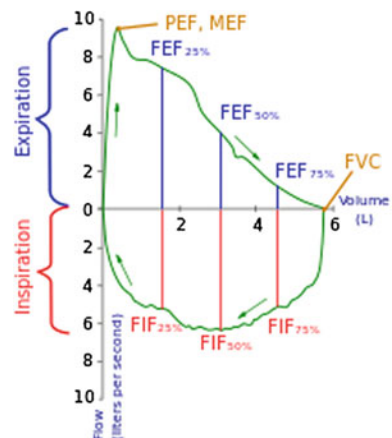
1.2 Different Types of Spirometers

Spirometers are classified on the basis of different types of sensors used for the extraction of signal generated by the forceful exhalation of the subject. Volume and flow type of spirometers are available with methods such as bell jar, ultrasonic, turbine, venturi, wedge, and differential pressure. each having some advantages and disadvantages [2, 3]. The most commonly used devices for spirometry test are KOKO Legend and three-ball spirometer. The main barrier for the use of these devices is cost. Also they are not portable. Three-ball spirometer can measure only PEF without any electronics, which is considered to be a low indicator of lung function. However, some devices are available on android and Windows platforms which indicate functions of lung. But they are not trustworthy. The work presented here is related to development of turbine type of spirometer. Up till now in turbine type of spirometers, revolutions are generated in the turbine blades placed in the spirometric pipe. These revolutions are detected and converted into the pulses outside the pipe by either magnetic pick up or photometric principle. Moment of inertia of the blades prevents sudden stopping of turbine when exhalation stops, generating false reading. This drawback is removed by converting revolutions into the power by connecting turbine to generator. After proper conversion parameters are calculated. First of all, blade designing is carried out and presented next.

2 Experimental Setup

This test requires the subject to hold the mouthpiece in his mouth and breathe in their full lung volume, and forcefully exhale at the turbine as long as possible. Generator generates the electrical signal and sends the data to DAQ card, which is given to the LABVIEW which will calculate exhaled flow rate and total exhaled volume by estimating total power generated by generator. Peak slope of power graph will give the peak exhalation flow.

2.1 Hardware Design

Figure 2 shows basic building block of turbine type of spirometer.

Turbine blades will convert the linear flow of forcefully exhaled air stream into rotational movement. Turbines are classified on the basis of the alignment of axis

Fig. 2 Block diagram
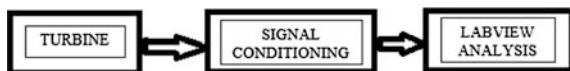
with respect to air flow as HAWT and VAWT. Due to high efficiency and low drag, HAWT approach is used in this design.

2.1.1 Turbine Design

According to Betz limit maximum power that can be extracted from wind cannot exceed 59.3% of the total power. In practice, the actual Betz Limit for a turbine is between 0.35 and 0.45. This applies to all wind turbines regardless of size. To extract maximum energy from incoming air steam the blade design should have maximum efficiency. It is observed that lift force is directly proportional to the angle of attack. At very large angle of attack the blade stalls and the lift decreases again. Due to large angle of attack, drag increases. Twist of 10°–20° from root to tip is applied to the blades. Best lift to drag characteristic is obtained by an airfoil that is fairly thin (thickness might be only 10–15% of chord length).

2.1.2 Design Consideration for the Application

1. Outer Diameter: The human mouth can approximately hold maximum 80-mm diameter mouth piece. Hence, mouth piece diameter can be maximum 80 mm.
2. Blade Thickness: Thickness is the most important issue during design. The efficiency of the blade is related to thickness. Thickness provides strength to the blade. If the thickness is more than the required power to rotate the blade is more and the overall efficiency is decreased.
3. Blade Length: As the outer diameter is fixed hence the blade length and the hub diameter are the two things that need to be decided within the diameter. The combination of blade length and hub diameter should be selected such that the efficiency should be higher.
4. Chord length: Base chord length is selected as 2.5 cm, and tip chord length is selected as 1.25 cm.
5. Angle of Attack: Research has shown that the best range for the angle of attack is from 20° to 30°. After the combination angle of attack was selected as 20° with the twist of 10°. Hence, the angle of attack for base is 20° and for tip it is 30°.

By using simulation software, following parameters are calculated and turbine is manufactured with rapid prototyping. Turbine with generator is placed in a 40 mm pipe and shown in Fig. 3.

Length of blade = 1.4 cm
Profile used at base and tip are same that are NACA5510
Chord length at base = 1.2 cm
Chord length at tip = 1.2 cm

Fig. 3 3-D image of turbine
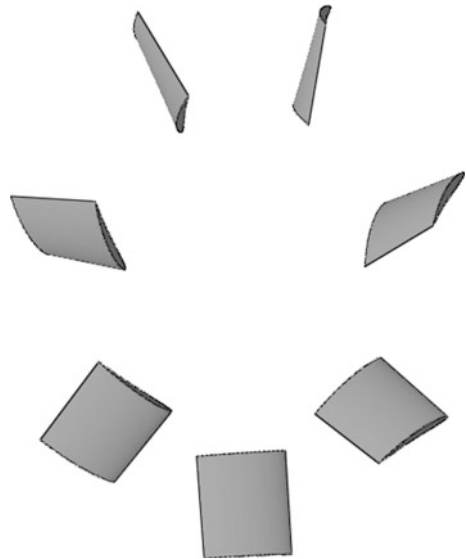
Fig. 4 Power versus speed
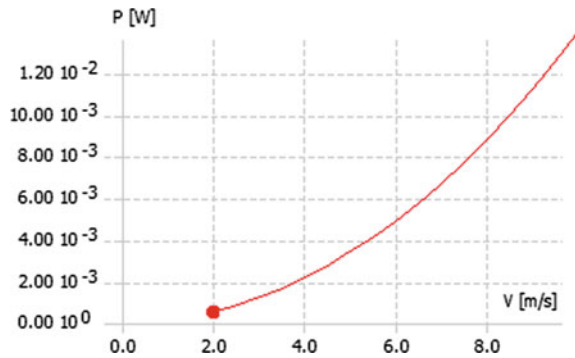
Lift coefficient = 0.699
Drag coefficient = 0.122
Thrust coefficient = −0.093

From Fig. 4, power coefficient in this case turns out to be 0.08256 at tip speed ratio of 3. Hence power generated by this turbine at the wind speed of 12 m/s is 396 mW.

2.1.3 Power Calculations

Power generated by the kinetic energy of a free flowing wind stream is calculated by using theoretical formulae and it is turns out to be:

$$P = \frac{1}{2}\rho s\omega^3 \tag{1}$$

where

$$s = \text{Cross Sectional Area} = \pi R^2 \rho = 1.22\,\text{kg m}^{-3}$$
$$R = \text{Blade Radius} = 2.1\,\text{cm}\,\omega = 9\,\text{ms}^{-1}$$

Hence, the total power generated by the kinetic energy of a free flowing wind stream in this case is $615.75\,\text{mW}$.

2.1.4 Power Coefficient (Cp)

Power extracted by the turbine relative to that available in the wind stream is calculated with the help of theoretical expression. Maximum efficiency is.

For this case $C_p = 0.4497$. Hence maximum power extracted through turbine is:

$$P_{\text{max}} = 276.90\,\text{mW}$$

2.2 Generator

Simple two-wire brushless DC motor is used as a generator here. It has usually 4 magnets around the perimeter. The stator of the motor is having the electromagnets, generally 4 of them, placed in a cross pattern with 90° angle between them. The advantage is that, the rotor carries only the permanent magnets; hence, it needs of NO power at all. As brushes are absent in brushless DC motors, they are more reliable and requires less maintenance. Brushless motor requires very less power for its operation.

2.3 Labview Analysis

Total signal analysis is carried out with the help of LABVIEW [3]. Basic block diagram for the flow of signal is shown (Fig. 5).

Total lung capacity can be compared to the total power generated by turbine type spirometer. Hence for calculation of total power, mean square value of input signal generated through forceful exhalation has to be calculated. The peak slope of power graph is nothing but peak exhalation flow which can be calculated by taking derivative of power graph. Frequency analysis is carried out to determine the power
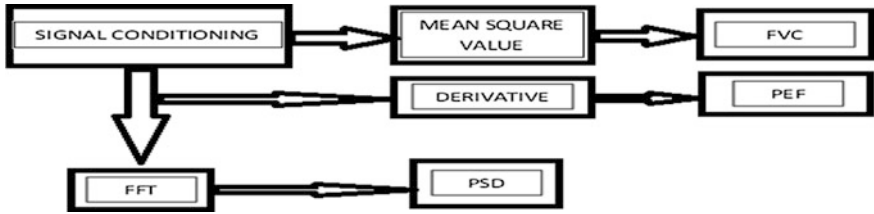
Fig. 5 Signal flow

calculations. FFT is used for the detection of frequency components present in signal. Further analysis is carried out to calculate power spectral density.

3 Results

Following graphs (Fig. 6) are obtained during analysis. First graph shows the original signal acquired by turbine. Because of environmental noise, original signal should be filtered. Second graph shows the filtered signal. And third graph shows total power generated by turbine during test.

A pilot study has been performed with the participation of different subjects to evaluate the performance of developed spirometer. Total five subjects (three males, two females) with age varying from 20 to 30, weight varying from 55 to 60, height varying from 5.1 to 5.9 were asked to exhale through the spirometer. Standard procedure has been followed while taking the reading. Results are shown in a Table 1.

Table 1 shows the results in detail.

Body mass index versus power and FVC shows the dependency of physical parameter on lung capacity. These results are compared with standard spirometer
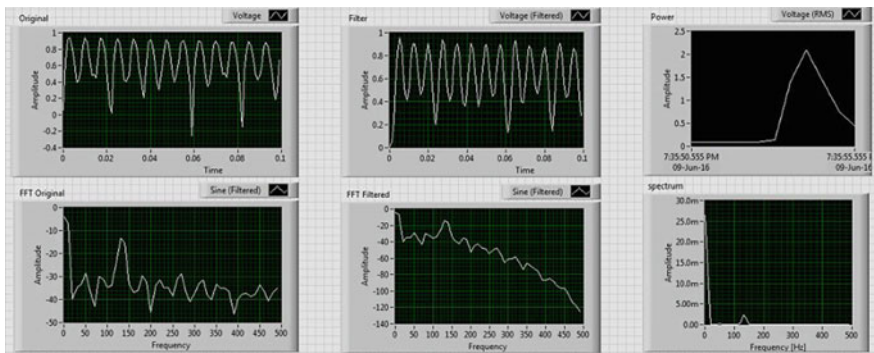
Fig. 6 Output waveforms

Table 1 Experimental results

No.	Name	Sex	Age	Weight	Height	FVC (ml)	PEF (Lit/Min)
1	Subject 1	M	23	60	5.6	607	555.67
2	Subject 2	M	27	67	5.9	855	609.23
3	Subject 3	F	21	48	5.2	560	440.89
4	Subject 4	F	30	58	5.1	596	506.06
5	Subject 5	M	19	55	5.4	798	507.11

and found in close relation. But it needs proper calibration and implementation using microcontroller.

4 Conclusion

Turbine type of spirometer with new approach is successfully implemented. Simulation and rapid prototyping of blade shows good wind power conversion efficiency. Use of off the shelf generator for power generation proves low-cost development of turbine. Errors due to sensor are minimized in the system which is helpful in further spirometer development. Basic generator design for maximum exhaled air power efficiency is tested and simulated. Thus, prototype developed is found satisfactory.

References

1. Webster JG (2009) Medical instrumentation application and design, 4th edn. Wiley, New York
2. Chiu N-F, Hsiao T-C, Lin C-W (2005) Low power consumption design of micro-machined thermal sensor for portable spirometer. Tamkang J Sci Eng 8(3):225–230
3. Alejos-Palomares R, Cortes JMR, Dominguez-Martinez N (2008) Digital spirometer with LabView interface. In: Proceedings of 18th international electronics, communication and computers, Puebla, Mexico
4. Gupta JK, Lin CH, Chen Q (2010) Characterizing exhaled airflow from breathing and talking. Indoor Air 20(1):31–39
5. Milos T, Bej A, Dobanda E, Manea A, Badarau R, Stroita D "Politehnica" University of Timisoara/MMUT/Hydraulic Machinary Department, Timisoara, Romania

Author Index