Lecture Notes in Electrical Engineering 423

Qilian Liang
Jiasong Mu
Wei Wang
Baoju Zhang
*Editors*

# Communications, Signal Processing, and Systems

## Proceedings of the 2016 International Conference on Communications, Signal Processing, and Systems

Springer

# Lecture Notes in Electrical Engineering

Volume 423

"Lecture Notes in Electrical Engineering (LNEE)" is a book series which reports the latest research and developments in Electrical Engineering, namely:

- Communication, Networks, and Information Theory
- Computer Engineering
- Signal, Image, Speech and Information Processing
- Circuits and Systems
- Bioengineering

LNEE publishes authored monographs and contributed volumes which present cutting edge research information as well as new perspectives on classical fields, while maintaining Springer's high standards of academic excellence. Also considered for publication are lecture materials, proceedings, and other related materials of exceptionally high quality and interest. The subject matter should be original and timely, reporting the latest research and developments in all areas of electrical engineering.

The audience for the books in LNEE consists of advanced level students, researchers, and industry professionals working at the forefront of their fields. Much like Springer's other Lecture Notes series, LNEE will be distributed through Springer's print and electronic publishing channels.

More information about this series at http://www.springer.com/series/7818

Qilian Liang · Jiasong Mu · Wei Wang
Baoju Zhang
Editors

# Communications, Signal Processing, and Systems

## Proceedings of the 2016 International Conference on Communications, Signal Processing, and Systems

Springer

*Editors*
Qilian Liang
University of Texas at Arlington
Arlington, TX
USA

Wei Wang
Tianjin Normal University
Tianjin
China

Jiasong Mu
Tianjin Normal University
Tianjin
China

Baoju Zhang
Tianjin Normal University
Tianjin
China

ISSN 1876-1100 ISSN 1876-1119 (electronic)
Lecture Notes in Electrical Engineering
ISBN 978-981-10-3228-8 ISBN 978-981-10-3229-5 (eBook)
https://doi.org/10.1007/978-981-10-3229-5

Library of Congress Control Number: 2017943834



Printed on acid-free paper

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

# Preface

The 5th International Conference on Communications, Signal Processing, and Systems (CSPS) is held in Chongqing, on October 21–22, 2016. Chongqing is the only inland municipality of all four in China, it is located in Southwest China, and the Yangtze River and Jialing River meet here. The city is known as the upper Yangtze River economic and financial center, shipping center, transportation hub in Southwest the industrialization of scientific research base, the State high-tech industrial center and the national historical and cultural city. CSPS 2016 brings together Chinese and international researchers and practitioners in communications, signal processing, and systems together.

The accepted papers of CSPS 2016 are from various regions around the world, which includes 10 different technical sessions such as 'Wireless Communication,' 'Radar Techniques,' 'Wireless Networks,' 'Coding, Encryption & Algorithm Design,' 'Mobile Communication, Positioning & Tracking,' 'Optical Communication,' 'Digital Signal Processing,' 'Patten Recognition, Deep Learning and Learning Automata,' 'Digital Image & Video Processing,' and 'Circuit Processing System & System Design.'

The technical program team did an excellent job in soliciting submissions, coordinating the review process, and promoting the technical program. We would like to thank everyone of them for taking leadership roles in organizing the various aspects of the technical program.

Also we would like to express our thanks to all members of the organizing committee and all the volunteer reviewers who have been working hard days and nights for this conference. We are grateful to the host institutions, Tianjin Normal University and Chongqing University of Posts and Telecommunications and sponsorships from IEEE Fort Worth Section, University of Texas at Arlington,

Beijing University of Posts and Telecommunications, University of Electronic Science & Technology of China, and Shanghai Jiaotong University, China. Finally, the publication support from Springer is deeply appreciated.

Arlington, USA — Qilian Liang
Tianjin, China — Jiasong Mu
Tianjin, China — Wei Wang
Tianjin, China — Baoju Zhang
October 2016

# Welcome Message from the General Chairs

It is our great honor and pleasure to welcome you to Chongqing for the 5th International Conference on Communications, Signal Processing, and Systems (CSPS) held during October 21–22, 2016. During this conference, scholars and practitioners from all over the world in communications, signal processing, and electronic systems will get together in Chongqing.

Chongqing has a humid subtropical monsoon climate; it is surrounded by the Yangtze and Jialing Rivers, so landscape is very beautiful. It is known as a 'Mountain city' and also a 'River city.' The South and North Hot Spring Park and Nanshan Park are situated in the near suburbs, while Jiangjin Simianshan forest, Wulong Furongdong Cave, and Xiannushan Grasslands are situated in distant suburbs. There are also many historical sites here, such as Fengdu's Ghost City and Ghost King Stone Carvings, Fuling's Baiheliang, Yunyang's Zhang Fei Temple, Fengjie's Baidicheng City, and Ba people's hanging coffins. This city is also famous for its delicious spicy foods.

CSPS 2016 is organized by an international team. The conference features 10 technical sessions and three keynote sessions. We invite you to join us by attending the technical and social events held in CSPS 2016.

On behalf of the organizing committee, the technical program committee, and all the volunteers that have been working hard for this conference, we warmly welcome you to CSPS 2016 and hope that you will enjoy the conference, and the beautiful city in which it takes place.

Tariq S. Durrani
Qilian Liang
Qianbin Chen
Yiming Pi
Baoju Zhang
General Co-Chairs, CSPS 2016

# Organization

The International Conference on Communications, Signal Processing, and Systems (CSPS) is organized by Tianjin Normal University, Chongqing University of Posts and Telecommunications, University of Texas at Arlington, Beijing University of Posts and Telecommunications, University of Electronic Science and Technology, China, and Shanghai Jiaotong University, China.

## General Co-Chairs

Tariq S. Durrani, University of Strathclyde, UK
Qilian Liang, University of Texas at Arlington, USA
Qianbin Chen, Chongqing University of Posts & Telecommunications, China
Yiming Pi, University of Electronic Science and Technology, China
Baoju Zhang, Tianjin Normal University, China

## TPC Co-Chairs

Habib M. Ammari, Norfolk State University, USA
Zengshan Tian, Chongqing University of Posts & Telecommunications, China
Jing Liang, University of Electronic Science and Technology, China
Wei Wang, Tianjin Normal University, China
Jiasong Mu, Tianjin Normal University, China

## International Advisory Committee

Tieniu Tan, Chinese Academy of Science, China
Gene Frantz, Texas Instruments, USA
Er Meng Hwa, Nanyang Technological University, Singapore
Sanjit Mitra, University of Southern California, USA
Zhisheng Niu, Tsinghua University, China
Wan-Chi Siu, Hong Kong Polytechnic University, Hong Kong
Xiaorong Wu, Tianjin Normal University, China
Zheng Zhou, Beijing University of Posts and Telecommunications, China

## Publication Co-Chairs

Yue Wang, Tianjin Normal University, China
Mu Zhou, Chongqing University of Posts and Telecommunications, China
Shenghong Li, Shanghai Jiaotong University, China

## Publicity Co-Chair

Zongjie Cao, University of Electronic Science and Technology, China
Fei Zhou, Chongqing University of Posts and Telecommunications, China
Weixia Zou, Beijing University of Posts and Telecommunications, China

## Keynote Co-Chairs

Xiaorong Wu, Tianjin Normal University, China
Jean X. Gao, University of Texas at Arlington, USA
Yong Li, Chongqing University of Posts and Telecommunications, China
Chenglin Zhao, Beijing University of Posts and Telecommunications, China

## Local Arrangement Co-Chairs

Hongqing Liu, Chongqing University of Posts and Telecommunications, China
Rui Min, University of Electronic Science and Technology, China
Shubin Wang, Inner Mongolia University, China

## Registration & Financial Co-Chairs

Yupeng Li, Tianjin Normal University, China
Ting Zhang, Chongqing University of Posts and Telecommunications, China

## Sponsoring Institutions

Tianjin Normal University, China
Chongqing University of Posts & Telecommunications, China
IEEE Fort Worth Section, USA
University of Texas at Arlington, USA
University of Electronic Science and Technology, China
Beijing University of Posts and Telecommunications, China
Shanghai Jiaotong University, China

# Contents

**Part VI Optical Communication**



**Part VII Digital Signal Processing**

# Part I
# Wireless Communication

# Mobility Management in AP Using SDN-NFV Technologies

**Syed Mushhad M. Gilani, Wenqiang Jin, Tang Hong, Guofeng Zhao and Chuan Xu**

**Abstract** Traditional WLAN typically relied on signal strength for handoff that is not sufficient for fair selection of physical access point ($_P$AP) and also causes the poor performance of network and load imbalance. To deal with these issues, we proposed a mobility management scheme on the basis of logical AP ($_L$AP) that keeps a connection with the mobile terminal (MT) during handoff either triggered by the user or software-defined network (SDN) controller for seamless mobility. The proposed scheme was implemented on a real testbed in WLAN environment, and the evaluation results demonstrated that it could provide a rather good seamless handover without throughput degradation and load imbalancing between $_P$APs, with an efficient handoff procedure to allocate the best AP in the neighbor region.

**Keywords** Software-defined network · NFV · Access point · Mobility management · WLAN

## 1 Introduction

WiFi has the powerful signal strength and able to support multimedia data in contrast to other networks (3G, 4G). More than half of mobile data traffic [1] can migrate to WiFi networks by the end of the decade. The cost of WiFi network is also very cheap, and that easily attracts consumers. However, the limited range of WiFi signals makes it threatening environment whenever the mobile user is performed handover (Fig. 1). The installation of APs in a closed region to overwhelm the signal gap between APs that aims to provide a continuous connection to the user. However, APs in an adjacent region may cause of signal interference [2]. In this circumstance, MT scanned multiple APs in range and connected only with

S.M.M. Gilani (✉) · W. Jin · T. Hong · G. Zhao · C. Xu
Research Center of Future Internet (RFI), Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: mushhad@hotmail.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_1

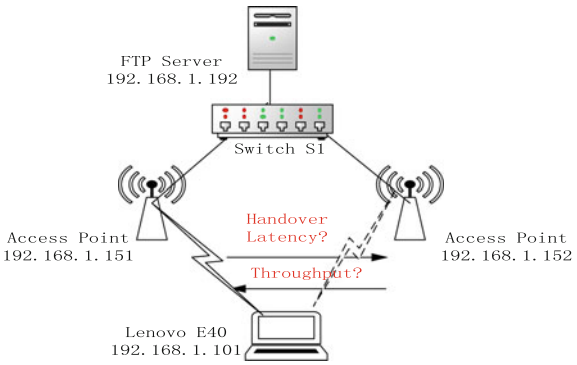



**Fig. 1** Traditional handover scenario

strongest received signal strength (RSS). If available AP was already congested, then, the user could experience low connection bandwidth.

Software-defined network (SDN) [3] is a promising solution in wireless network personification that may introduce programmable control plane and data plane into $_{P}$AP. OpenFlow [4] may incorporate handover parameters [5, 6] into existing applications for smooth performance of handover activities. Network function virtualization (NFV) [7] highly encourages implementing within SDN platform that can offer multiple benefits for service operators and mobile terminals.

In this article, we propose a logical access point based mobility management ($_{L}AP^{M}$) scheme for WLAN in the SDN environment with the further functionality of NFV. This research work examines the conventional handover approach with a comparison of the proposed algorithm and also focuses on load balancing between APs sharing the signal interference area ($_{P}AP^{s}$). We build a testing environment to analyze the functionality of $_{P}$AP after adopting SDN and NFV technologies.

## 2 $_{L}$AP Based Mobility Management ($_{L}AP^{M}$) Scheme

The proposed scheme enriches the performance of the APs and mobile terminals without requiring any change at the client side. It also offers seamless mobility in which $_{L}$AP can migrate from source $_{P}$AP to target $_{P}$AP during handoff. The traditional RSS-based association procedure could generate the load imbalance network environment. As depicted in Fig. 2a, $_{P}AP^{2}$ associated with four MTs with individual $_{L}$AP, $_{P}AP^{3}$ associates with three MTs, and while $_{P}AP^{1}$ associated with two mobile terminals. $_{L}AP^{2}$ also exists in the coverage area of $_{P}AP^{1}$, but does not associate because of less RSS rather than $_{P}AP^{2}$. In comparison, our proposed approach incorporates load information during the MT association that makes the balanced network. As illustrated in Fig. 2b, $_{L}AP^{2}$ can associate with $_{P}AP^{1}$ that can reduce the load of $_{P}AP^{2}$ and also enhance the throughput of associated MT.

$_{L}$AP is a logical entity which resides in AP. $_{L}$AP acts as SDN agent [8] in $_{P}$AP that is responsible for performing multiple tasks generated by either SDN controller

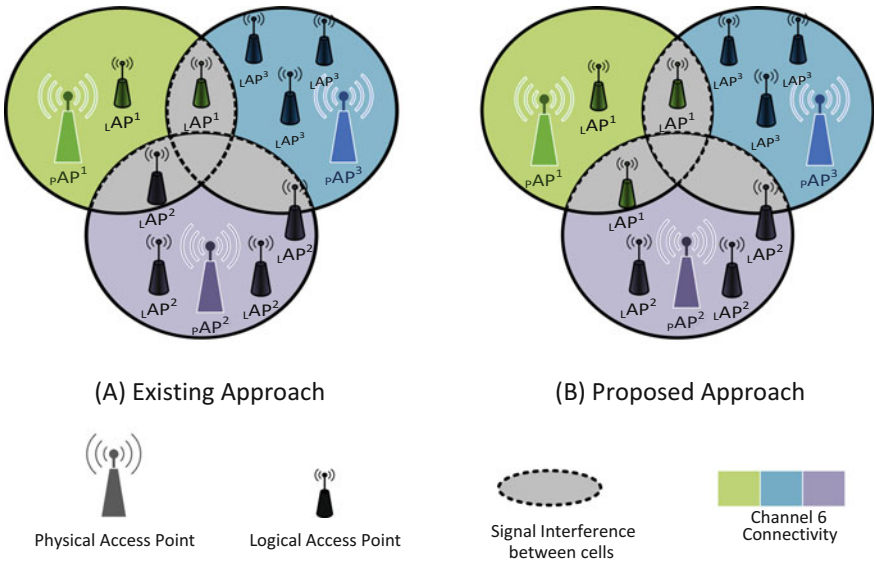



**Fig. 2** An example of mobility management within the load balancing scenario

or mobile terminal. The first task is to assign unique BSSID per-client that generates beacon and acknowledgment. Second, it sets initial configuration parameters such as IP address, MAC address, SSID for the MT and set up a connection with the controller. Third, it calculates the load $_{P}$APs that includes inbound and outbound traffic load with the number of $_{L}$AP ($N_{L}AP$) as described in Eq. 1.

$$N_L AP = \sum_{n=1}^{N} {}_{A}CS \Big/ \sum_{m=1}^{M} {}_{P}CS \tag{1}$$

where $_{A}$CS (Active Current State) contains the current traffic load (bytes) of MTs within specified time period. Meanwhile, $_{P}$CS (Passive Current State) denotes the connection interval time and sleeping time. Next, we present the method of threshold calculation.

$$ {}_{s}A0 = \{{}_{P}AP^{i}, {}_{P}AP^{j}\} \tag{2}$$

$$\{MT_{1j}, MT_{2j}, MT_{3j} \ldots MT_{nj}\} \in {}_{P}AP^{j} \tag{3}$$

$$\{MT_{1i}, MT_{2i}, MT_{3i} \ldots MT_{mi}\} \in {}_{P}AP^{i} \tag{4}$$

$$\{MT_{1i}, MT_{2i}, MT_{3i} \ldots MT_{mi}, MT_{1j}, MT_{2j}, MT_{3j} \ldots MT_{nj}\} \in {}_{s}A0 \tag{5}$$

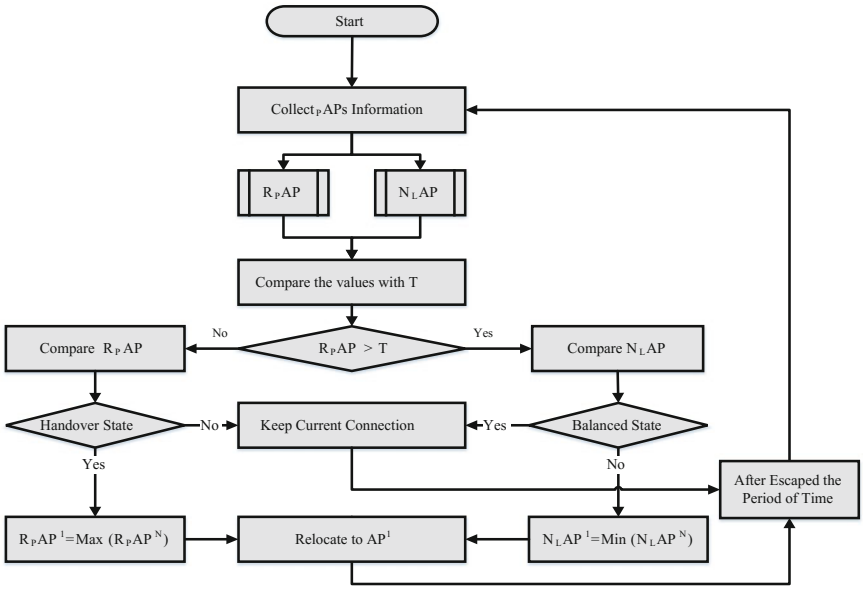



**Fig. 3** Flowchart of the mobility management scheme ($_{L}AP^{M}$)

$$T = \frac{1}{M+N}\left(\sum_{n=1}^{N} R_{MT_{nj}} + \sum_{m=1}^{M} R_{MT_{mi}}\right) \quad \begin{matrix} \forall M = 1, 2, 3 \ldots \\ \forall N = 1, 2, 3 \ldots \end{matrix} \tag{6}$$

Where $_{s}A0$: $_{P}$APs deployment area. *SpAP*: is the set of MTs, which connected to $_{P}$AP. $R_{MTn}$: is RSS value of $MT_n$.

Figure 3 presents the flowchart of the proposed $_{L}AP^{M}$ scheme. In the initialization phase, the controller gets information from the agent about MT connected to $_{P}$APs that belongs to A. The second stage is scanning and comparison between the received values with predefined parameters of the algorithm. If current RSS of MT is greater than the threshold T Eq. 6 then start a comparison of $N_{L}AP$ between $_{P}AP^{s}$ and if received value is False, then forward to mobility phase. Moreover, MT scans further $_{P}$AP with improving RSS and $N_{L}AP$. As illustrated in Fig. 2a MT within the signal interference zone can select one best AP. However, if the client performed handoff for more than three times, that means, it cannot find a better AP. Moreover, we set the handover state equals FALSE and stops the comparison process. Furthermore, if the handoff state is TRUE, it will continue with the next steps.

**Table 1** $_{P}$APs specifications

| | MAC | Standard | Bandwidth | SSID | OS | Tools |
|---|---|---|---|---|---|---|
| $_{P}AP^{1}$ | 00:00:30:b5: c2:88:63:1a | 802.11n | 40 MHz | Mobiletest1/ Mobiletest2 | OpenWRT [9] | Open vSwitch [10] |
| $_{P}AP^{2}$ | 00:00:62: a0:43:6e:8d:55 | | | | | |

## 3 Implementation of $_{L}AP^{M}$ Scheme

We have implemented the proposed scheme using the open source platform that makes feasible practically to adopt this scheme. The experimental devices briefly describe in this section. We select two $_{P}$APs for initial experiments as mentioned in Table 1.

**SDN Controller**: We deployed OpenFlow-based controller to provide an overall view of the underlying network. It can perform multiple tasks on the Floodlight controller [11] to manage $_{P}$AP such as disable $_{L}$AP or enable $_{L}$AP, mapping of $_{L}$AP to $_{P}$AP, migration of $_{L}$AP between $_{P}$AP. The controller operates a table containing information about $_{P}$APs that can support for MT selection and load balancing.

**DHCP Server**: During the experiment, the client received dynamic IP on a first connection to the network, during handoff does not require to change the address that saves the time of users for inquiring IP addresses.

**FTP Server**: We configure *Xlight* and select the large file size more than 1 GB to make the certain downloading process that should continue until handoff completion.

**Mobile Terminals** (**MT**): Mobile terminal having any operating system can connect with $_{L}$AP. However, the transfer rate of packets between cell phones and laptops can differ due to manufacturing hardware capacity.

## 4 Evaluation and Results

We have evaluated the $_{L}AP^{M}$ scheme in WLAN environment in Technology Building. A load of $_{P}$APs and MT's throughput measured by ping and *iPerf*. This section presents testbed evaluation results using real traffic traces. We have developed two cases to perform experiments: First, we compare the performance of $_{L}AP^{M}$ handover with traditional WLAN handover during accessing to the same file server. The second case is about experiments that were carried out on individual $_{P}$APs traffic load. The final results showed the balanced traffic load on $_{P}$APs and improvement of throughput on MT.

### 4.1 Case 1: Handover Between $_{P}$APs

In this experiment, we compare the performance of fixed RSS-based handover algorithm with proposed $_{L}AP^{M}$ algorithm. First, we evaluate $_{P}$AP handover performance in the traditional environment as shown in Fig. 1. The process of connecting to $_{P}$AP is a four-way handshake according to IEEE 802.11n standard. Lenovo E40 initially connected with $_{P}$AP (192.168.1.151) and activate downloading data from FTP server (192.168.1.192) at time t = 0 s with average throughput rate that is approximately 45 Mbits/s. Meanwhile, after 30 s MT has performed a handover to $_{P}$AP (192.168.1.152). Figure 5a indicates the throughput regarding elapsed time during the fixed RSS handover process. The handover process and connection interrupted during the downloading process at t = 30 s and decreased the throughput from 46.3 to 3.28 Mbits/s in 8 s. Moreover, the throughput becomes at zero level after elapsed time t = 43 s, because the connection between MT and network is lost for an extended time. It takes almost t = 10 s to recover in traditional WLAN environment caused by IEEE 802.11 standard handover mechanism that includes detection, probing and re-association processes. However, authentication time is not included that may take this process longer.

The second experiment performed in the SDN-based environment as illustrated in Fig. 4, two $_{P}$APs enabled with OpenFlow under SDN controller (192.168.1.221). Lenovo E40 initially connected with $_{P}AP^{2}$ through $_{L}AP^{2}$ at t = 0 s and started the file transferring process with average throughput 40 Mbits/s approximately. As

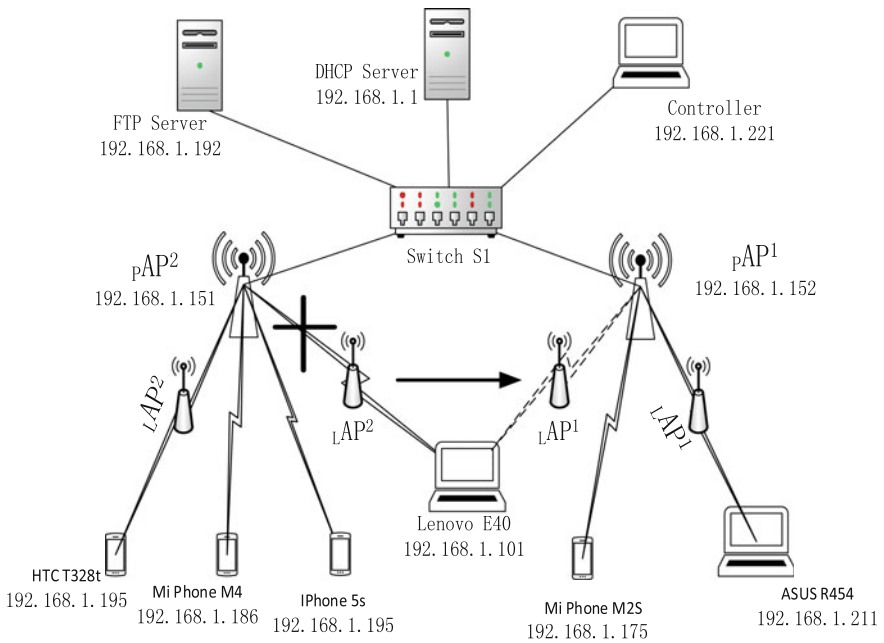



**Fig. 4** Mobility management testbed scenario

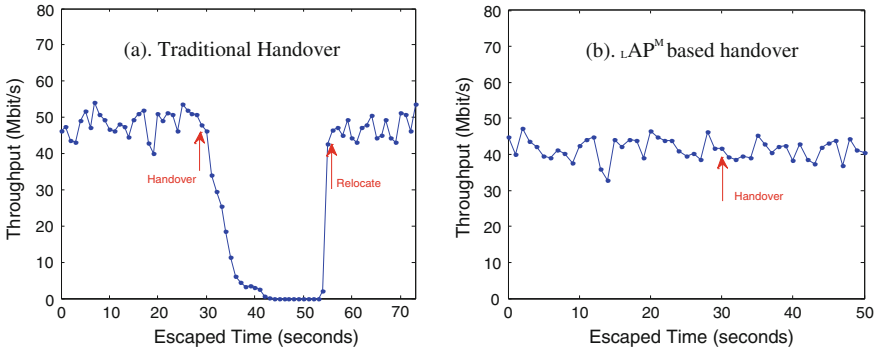



**Fig. 5** Performance analysis of handover cost

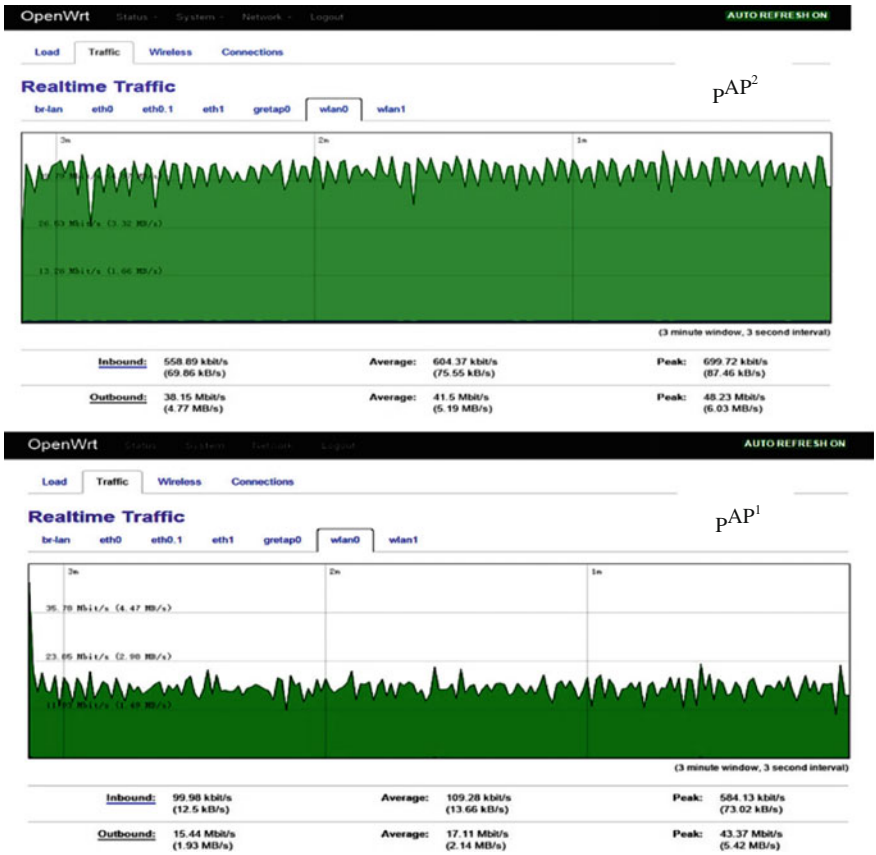



**Fig. 6** Real-time traffic examines of $_{P}$APs in area A

handover initiated at t = 30 s, the $_{L}AP^{2}$ migrates the credentials to target $_{P}AP^{1}$ for newly associated $_{L}AP^{1}$. As shown in Fig. 5b $_{L}AP$ handoff almost did not produce any delay in packet transmission that sustained the constant throughput average at 40 Mbits/s. In comparison with fixed RSS-based handover, our proposed handover scheme resulted in zero-like handover latency while remaining network throughput.

## 4.2 Case 2: $_{L}AP^{M}$ Scheme for Load Balancing

The goal of this experiment is to show how $_{L}AP^{M}$ scheme can enhance network throughput to those MTs, which were overlapping in $_{S}A$. We installed two $_{P}APs$ with the same overlapping region and selected six MTs for this experiment.

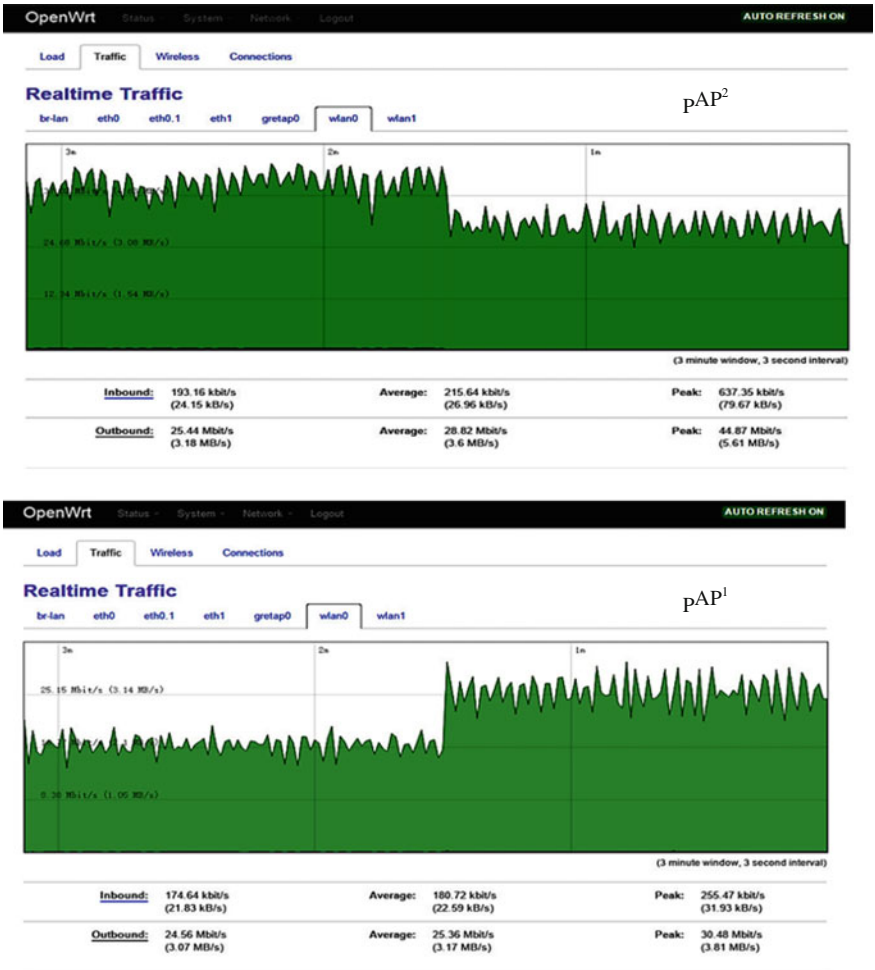



**Fig. 7** Demonstrate the real-time traffic analysis in a balanced network

**Imbalanced Situation**: Initially four mobile terminals connected with ${}_{P}AP^{2}$ because they received strong RSS from ${}_{P}AP^{2}$ and remaining 2 MT connected with ${}_{P}AP^{1}$ as described in Fig. 2a. We performed a test on real-time traffic and got results of 3 min data rate through OpenWRT interface as demonstrated in Fig. 6. The inbound traffic of ${}_{P}AP^{2}$ was 69.86 kB/s while ${}_{P}AP^{1}$ was 12.5 kB/s. The average traffic load of ${}_{P}AP^{2}$ was 41.5 Mb/s that was closest to peak value. Conversely, the bandwidth usage on neighbor ${}_{P}AP^{1}$ was 17.11 Mbits/s which far less than that on ${}_{P}AP^{2}$. This imbalanced situation of ${}_{P}APs$ reported in Fig. 6.

**Load Balancing**: After adopting ${}_{L}AP^{M}$ Scheme, the outbound traffic of ${}_{P}AP^{2}$ decreased 12.68 Mbits/s with the average data rate was 28.82 Mb/s as illustrated in Fig. 7. At the moment, the inbound traffic average rate of ${}_{P}AP^{2}$ reduced from 75.55 to 26.96 kB/s that was good enough for remaining connected MTs. Meanwhile, the outbound traffic load of ${}_{P}AP^{1}$ reported in Fig. 7 expanded from 17.11 to 25.36 Mbits/s and inbound traffic also boosted from 12.5 to 21.83 kB/s that diminished the wastage of resources. Hence, the network infrastructure was balanced and feasible to provide better potential for existing and new users.

## 5 Conclusions

In this research article, we have presented and evaluated the mobility management scheme for OpenFlow-enabled AP. The proposed scheme makes it feasible for users to connect with the best AP. Our experiment performed and demonstrated how to create ${}_{L}AP$, assign ${}_{L}AP$ and make migration of ${}_{L}AP$ according to run-time decision algorithm that was executing at the SDN controller. The evaluation has shown that ${}_{L}AP^{M}$ scheme improved the ${}_{P}AP$ performance and significantly minimized handoff delays. As a further study, the mobility management scheme will be investigated in dense WLAN, and a location-awareness module would be integrated to offer more true mobility in future Internet architecture.

**Acknowledgements** This work was supported by National Nature Science Foundation (61402065, 61501075), Prospective Research Project on Future Networks (BY2013095-2-03) fund by Jiangsu Future Networks Innovation Institute and Project Foundation of Chongqing Municipal Education Committee (No. KJ1500429).

## References

1. Mobile Europe, Mobile data to hit 197 KPB by 2019, mainly offloaded to Wi-Fi (2015), http://www.mobileeurope.co.uk/Press-Wire/mobile-data-traffic-to-hit-197-000-petabytes-by-2019-mainly-offloaded-to-wi-fi. Accessed May 2015
2. M.A. Ergin, K. Ramachandran, M. Gruteser, Understanding the effect of access point density on wireless LAN performance, in *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking* (ACM, 2007)

3. N.A. Jagadeesan, B. Krishnamachari, Software-defined networking paradigms in wireless networks: a survey. ACM Comput. Surv. **47**, 1–11 (2014). doi:10.1145/2655690
4. Open network foundation (2012), https://www.opennetworking.org/. Accessed May 2015
5. P. Dely, A. Kassler, L. Chow, N. Bambos, N. Bayer, H. Einsiedler, C. Peylo, D. Mellado, M. Sanchez, A software-defined networking approach for handover management with real-time video in WLANs. J. Mod. Transp. **21**, 58–65 (2013). doi:10.1007/s40534-013-0007-x
6. D. Zhao, M. Zhu, M. Xu, SDWLAN: a flexible architecture of enterprise WLAN for client-unaware fast AP handoff, in *Computing, Communication and Networking Technologies (ICCCNT), 2014 International Conference on IEEE*, 1–6 2014
7. C. Cui, H. Deng, D. Telekom, U. Michel, H. Damker, in *Network Functions Virtualisation*, An introduction, benefits, enablers, challenges & call for action (2012), http://portal.etsi.org/NFV/NFV_White_Paper.pdf. Accessed Sep 2015
8. L. Suresh, J. Schulz-Zander, R. Merz, A. Feldmann, T. Vazao, Towards programmable enterprise WLANs with Odin, in *HotSDN'12 August 13 2012 Hels. Finl. Copyr. ACM*. (ACM, 2012), pp. 115–120
9. OpenWrt (2004), http://wiki.openwrt.org/doc/start
10. Open vSwitch (2013), http://openvswitch.org/
11. Floodlight Project (2013), http://www.projectfloodlight.org/floodlight/

# Algorithm Research of Congestion Control in Wireless Router

**Xiang Li and Xirong Ma**

**Abstract** With the continuous improvement of social economy as well as science and technology, along the continuous development of Internet technology, Internet has become an indispensable and important part of the learning life and social work of people; moreover, with the increase of network ports, the giant network traffic will lead to network congestion, that is also the main reason which affects the network performance and restricts the development of the network. In this paper, the author puts forward that a kind of routing algorithm which may optimize the routing could be obtained by means of combining with correlation function, calculating the hop number of nodes as well as occupation situation of cache, which is able to reduce the network transmission delay and network congestion via combining with alternate routing.

**Keywords** Wireless network · Network congestion · Wireless network routing algorithm

## 1 Introduction

On account of the coverage of terminal wireless network is limited, those further nodes are incapable of having direct communication with each other and they need to retransmit the communication content in virtue of intermediate node, which will occupy the cache space of middle node and need to adopt corresponding routing protocols [1]. The above-mentioned content is the main factor which causes the network congestion; therefore, it is of great significance to research the routing algorithm of the optimized path.

Following are mainly several kinds of existing optimization algorithms: first, alleviating the congestion status of network by changing the topology structure,

X. Li · X. Ma (✉)
College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China
e-mail: tjmaxirong@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_2

namely, by adding or deleting nodes and link [2, 3], but its feasibility is not high; second, by increasing the path to forward the data package from different next hop; however, this kind of method only considers the static network feature when it optimizes the algorithm [4, 5], and fails to achieve the global optimization in the aspect of routing selection; third, by calculating the routing protocol of shortest-path first algorithm: on the one hand, its advantage is that it is loop-free and has fast convergence speed [6, 7]; on the other hand, its disadvantage is that data packets are always sent in accordance with the optimal path during transmission, therefore, when multiple optimal paths overlap together, the link will be frequently utilized, which will lead to excessive load of local network so as to cause congestion [8, 9]; fourth, serving node between-ness as connection weights, and reducing the computation complexity of between-ness by means of weighting the network shortest-path routing algorithm; fifth, by detecting the occupancy rate of buffer zone to predict whether congestion will occur. When the occupancy rate reaches a certain threshold, the backup routing should be activated [10]; due to it needs to compute the global routing table, this method is more suitable for the wired network.

In order to solve the shortage of the above methods, the author, on the basis of the research about existing routing algorithm, puts forwards that a kind of routing algorithm which may optimize the routing could be obtained by means of associating with correlation function, calculating the hop number of nodes as well as occupation situation of cache [11, 12], which is able to reduce the network transmission delay and network congestion via combining with alternate routing.

## 2 Wireless Routing Node Load

Nowadays, wireless routing protocol calculates the link cost only in virtue of time delay and hop number and without considering congestion state; therefore, it is unable to avoid congestion nodes, and finally parts of links would occur congestion [13, 14].

Detecting the occupancy rate of cache (ROC) (ratio between number of packets which are ready to send and buffer capacity), anticipation node as well as link state may be helpful to avoid congestion during transmission and improve the success rate packet delivery, For this purpose, this paper calculating net-into rate (NR) (the ratio between difference value of inbound and outbound node packets at per unit time and maximum capacity of buffer zone, which reflects the occupation trend of cache; the bigger the ratio, the larger the congestion risk) and cache occupancy rate (which reflect the occupancy status of node cache) to measure the congestion rate, thus predicting congestion degree; it periodically detects the congestion rate (CR), then combines it with the default cache threshold and congestion unit (CU) (divided into free stage, normal stage, warning stage and congestion) and data to measure the node and link load. The formula is as follows:

$$CR_i(t) = ORC_i(t) + NR_i(t) \tag{1}$$

Constraint conditions:

$$\begin{cases} CR_i(t) = 0, CR_i(t) < 0 \\ CR_i(t) = CR_i(t), 0 \le CR_i(t) \le 1 \\ CR_i(t) = 1, CR_i(t) > 1 \end{cases} \tag{2}$$

$$\begin{cases} CU = 0, CR < \frac{1}{4} \\ CU = 8, (CR - \frac{1}{4}), \frac{1}{4} \le CR \le \frac{3}{4} \\ CU = 1 + 5^{4*\left(CR - \frac{3}{4}\right)}, CR > \frac{3}{4} \end{cases} \tag{3}$$

## 3 The Link Cost

Link cost (LC) refers to the cost that needs to pay while transmitting packets from one node to another node, and the formula is

$$LC_{ab}(t) = metric_{ab} + CU_b * CR_b(t) \tag{4}$$

If the path from node A to node B is expressed as P $\rightarrow$ S, then its path cost if as follows:

$$LC_{ps}(t) = \sum_{i=1}^{n-1} LC_{i,i+1}(t) = metric_{ps} + \sum_{i=2}^{n} CU_i * CR_i(t) \tag{5}$$

In the formula, $\text{metric}_{\text{ps}}$ represents the hop number from node A to node B. According to the formulas 4 and 5, when the node takes place congestion, it can determine the shortest path according to congestion status in the network, so that to ensure that the algorithm may always choose global optimal path with minimum cost value to serve as updated routing.

## 4 The Improved Algorithm Description About Congestion Control Routing

According to the load of network node and link cost model, counting the global routing of nodes for sending the data packets; at the same time, when the nodes take place congestion, the main line should be replaced by alternate route so as to obtain the congestion control routing algorithm that proposed in this paper.

## 5 Simulation Experiment

In this paper, the author conducts the experiment by utilizing the NS2.35; among which, this simulation network is equipped with 100 nodes which randomly distributed in the plane region of 1000 * 1000; the communication patterns adopts 20 data sources, and the starting time of the data source is uniformly distributed between 0 and 5 s; MAC layer employs a kind of protocol that with channel bandwidth of 2 Mb/s of IEEE 802.11 MAC. The transmission range of node 250 m and its simulation time is 400 s. At the beginning of the simulation, each node keeps static p seconds, and then randomly selects destination and to constantly move towards the destination at a speed between 0 and max; after arriving the destination, keeps static p seconds again, then randomly selects another destination and to move on according to the aforementioned methods. This process will be circulating all the time until the end of the simulation.

According to Figs. 1 and 2 shows that when the source node sends a packet per second without congestion, when the Pause Time $\leq 150$ s algorithm optimized packet delivery success rate increased significantly. Because of the same fracture mechanism to link, so the results are close to the after. Because of the moving time of nodes significantly reduce the network load and smaller, link fault impact on the network data transmission reduced significantly, and the algorithm itself has the ability to update cycle route, so when the Pause Time $\geq 200$ s, the calculation result is close to the after.

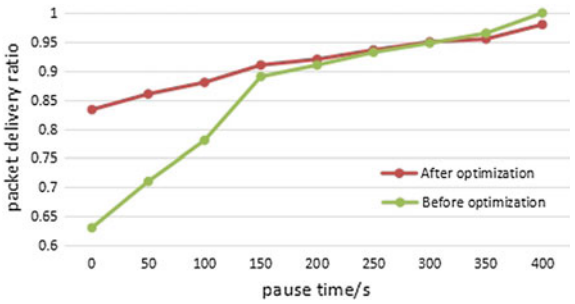



**Fig. 1** Optimized packet delivery success rate change with mobile rate curve

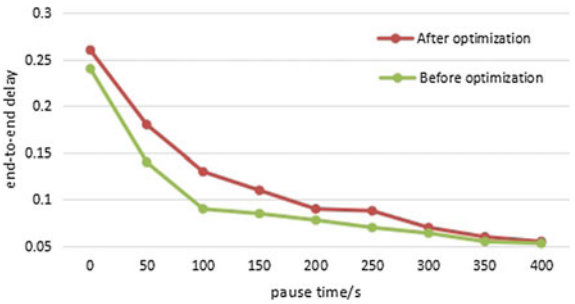



**Fig. 2** The end-to-end delay after optimization algorithm with the curves of node mobility

## 6 Conclusion

In this paper, by studying the node load, link cost, local routing and global routing, the author puts forward congestion degree periodic-detected node and link on the basis of network congestion phenomenon; in addition, according to the congestion status of node buffer zone, this paper adopts the processing strategy by combining local routing with global routing, so that to reduce the packet loss caused by the link fracture and node congestion. The simulation results show that this kind of algorithm can reduce the packet loss caused by congestion nodes and improve the success rate of packet delivery.

## References

1. B. Gedik, L. Liu, Protecting location privacy with personalized k-anonymity: architecture and algorithms. IEEE Trans. Mob. Comput. **7**(1), 1–18 (2008)
2. K. Sun, D. Wu, MPC-based delay-aware fountain codes for live video streaming, in *IEEE International Conference on Communications*, *ICC 2016* (2016)
3. Z. Lu, X. Lin, A data privacy-oriented multi-parities location collect scheme in location based services, in *Proceedings of 2009 the Fourth International Conference on Computer Sciences and Convergence Information Technology* (2009), pp. 964–969
4. X. Liu, Z. Li, Information-centric mobile ad hoc networks and content routing: a survey (2016). Article in press
5. K. Sun, H. Zhang, D. Wu, H. Zhuang. MPC-based delay-aware fountain codes for real-time video communication. IEEE Internet Thing J. (IoT) (2016)
6. Z. Li, Y. Chen, H. Shi, K. Liu, NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. (2016)
7. K. Sun, D. Wu, Video rate control strategies for cloud gaming. J. Vis. Commun. Image Represent. (2013)
8. R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Trans. Acoust. Speech Signal Process. **37**(7), 984–995 (1989)
9. Z. Li, K. Liu, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), November (2014)
10. K. Sun, B. Yan, H. Gharavi, Low complexity content-aware image retar-geting, in *IEEE International Conference on Image Processing 2012* (ICIP' 2012)
11. R.O. Schmidt, Multiple emitter location and signal parameter estimation. IEEE Trans. Antenna Propag. **34**(3), 276–280 (1986)
12. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. (2016) Article in press
13. C.M.E Gallery, Trusted mobile platform. Comput. Sci. (2007)
14. K. Sun, B. Yan, Efficient p-frame complexity estimation for frame layer rate control of H.264/AVC", in *IEEE International Conference on Image Processing 2011* (ICIP' 2011)

# A Novel Method to Generate Wi-Fi Fingerprint Database Based on MEMS

**Zengshan Tian, Zipeng Wu, Mu Zhou, Ze Li and Yue Jin**

**Abstract** Fingerprint-based positioning in Wi-Fi environment has caught much attention recently. One key issue is about the radio map construction, which generally requires significant effort to collect enough Wi-Fi Received Signal Strength (RSS) measurements. Based on the observation that the Micro Electromechanical System (MEMS) can automatically calibrate the target locations without complex equipment, we propose an efficient radio map construction method based on the technology of multi-sensor. Different from the conventional methods, the proposed one first relies on the gait detection approach and quaternion-based extend Kalman filter algorithm to estimate the velocity and heading of the target. Second, the Pedestrian Dead Reckoning (PDR) algorithm is used to calculate the current location of the target in a real-time manner, and meanwhile the data from Wi-Fi module are collected to generate the fingerprint database. The experimental results show that the proposed method is effective in positioning accuracy and efficient by saving the time and energy.

**Keywords** Indoor localization · Wi-Fi fingerprint database · MEMS · PDR · Extend Kalman filter

Z. Tian · Z. Wu (✉) · M. Zhou · Z. Li (✉) · Y. Jin (✉)
Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, People's Republic of China
e-mail: wuzipeng2010@foxmail.com

Z. Li
e-mail: 574625020@qq.com

Y. Jin
e-mail: 702814793@qq.com

Z. Tian
e-mail: tianzs@cqupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_3

## 1 Introduction

Nowadays, it is usually not a problem to locate a person or mobile device outside the building since the satellite navigation systems are available in outdoor environment. As we know, the propagation property becomes more complicated inside the building, where the LOS may not exist. To solve this problem, many other methods, e.g. the Wi-Fi Received Signal Strength (RSS) [1], the Wi-Fi fingerprinting is considered as one of the best techniques used for indoor localization due to this method have been significantly studied. The process of fingerprinting involves two phases, namely offline training phase and online positioning phase [2]. The purpose of offline phase is to construct a database. The existing works normally record the location fingerprints point by point, which consumes a large amount of time and labor cost. In this paper, we put forward a new method by integrating the Wi-Fi RSS measurements with the data from multi-sensor to reduce the workload and ensure positioning accuracy at the same time. The rest of this paper is organized as follows. We discuss system framework in Sect. 2. Section 3 summarizes the database building based on PDR. In Sect. 4, we design some experiments to examine the performance of the proposed method. Finally, Sect. 5 concludes the paper.

## 2 System Framework

The system architecture is shown in Fig. 1. Our work is based on the PDR algorithm [3] which is considered as a relative positioning algorithm. Specifically, first of all, we smooth the data from the 3-axis accelerometer by using a low-pass filter [4]. Second, we use the filtered accelerometer data to estimate the stride length of the pedestrian by the empirical step estimation model [5] and calculate the velocity by the estimated stride length and step detection. Third, using the quaternion-based extend Kalman filter [6], we merge the data collected from the 9-axis sensors to calculate the heading. Fourth, the PDR algorithm is applied to compute the real-time locations of the pedestrian. Finally, we save the locations of the pedestrian with the corresponding RSS measurements into the fingerprint database.

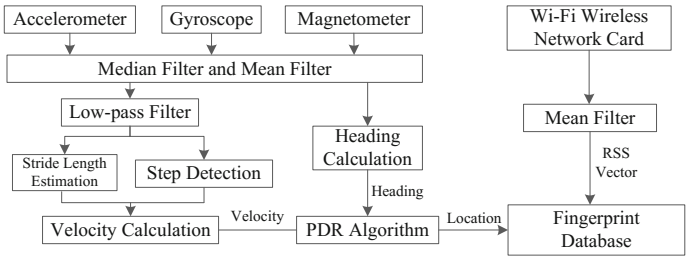



**Fig. 1** System architecture

# 3 Database Building Based on PDR

## 3.1 Step Detection

Based on the physiological characteristics of the pedestrian, it can be proved that the acceleration value changes in the form of sine wave during the walking. To avoid the false detection, there are generally two types of filters designed to filter the signal noise, namely windowed mean and median filters. In concrete terms, we calculate the 3-axis acceleration values by

$$Acc_{norm} = \sqrt{a_x^2 + a_y^2 + a_z^2} - G \tag{1}$$

where $a_x$, $a_y$ and $a_z$ are the output of 3-axis accelerometer after the windowed mean and median filter in $x$, $y$, and $z$ directions respectively. $G$ is the gravitational acceleration. After the windowed mean and median filter, it is still possible that the pseudo peak appears. To eliminate these peaks, we conduct the finite impulse response low-pass filter to smooth the 3-axis acceleration values.

## 3.2 Velocity Estimation

By assuming that the frequency of output from the MEMS sensors is $f_s$, the time interval of the $k$-th step, $t_k = \frac{\Delta N_k}{f_s}$ where $\Delta N_k$ is the number of sample points from the MEMS sensors during the $k$-th step. For simplicity, we assume that each step of the pedestrian is with the same walking speed and heading. Thus, the velocity of the $k$-th step is calculated by

$$v_k = \frac{P_k}{t_k} = \frac{P_k f_s}{\Delta N_k} \tag{2}$$

$$P_k = C_1 \sqrt[C]{Acc_{\max}^k - Acc_{\min}^k} \tag{3}$$

where $P_k$ is the length of the $k$-th step, $Acc_{\max}^k$ and $Acc_{\min}^k$ stand for the maximum and minimum of acceleration values for the $k$-th step. $C$ and $C_1$ are the personalized coefficients for different pedestrians.

## 3.3 Heading Estimation

### 3.3.1 Calculation of Initial Heading

By assuming that the pedestrian stands still at the beginning, we calculate the initial attitude of the pedestrian by using the accelerometer and magnetometer in (5).

$$\begin{cases} \left[a_x^b \; a_y^b \; a_z^b\right]^{\mathrm{T}} = \mathrm{T}_n^b \cdot \left[0 \; 0 \; g\right]^{\mathrm{T}} \\ \left[m_x^b \; m_y^b \; m_z^b\right]^{\mathrm{T}} = \mathrm{T}_n^b \cdot \left[0 \; b_y^n \; b_z^n\right]^{\mathrm{T}} \end{cases} \tag{4}$$

where $\left[a_x^b \; a_y^b \; a_z^b\right]$ and $\left[m_x^b \; m_y^b \; m_z^b\right]$ are the measurement values of the 3-axis accelerometer and magnetometer respectively. $g$ is the local gravitational acceleration. $b_y^n$ and $b_z^n$ are the northward and perpendicular components of the local magnetic field under the ENU reference frame. $T_n^b$ is the attitude rotation matrix. Owing to the advantages of low computation cost and wide applications of the quaternion algorithm, we use it to estimate the heading of the target. Thus, we use the attitude quaternion $\mathbf{Q} = \left[q_1 \; q_2 \; q_3 \; q_3\right]^{\mathrm{T}}$ to describe the attitude rotation matrix, such that

$$\mathrm{T}_n^b = \begin{bmatrix} q_1^2 + q_2^2 - q_3^2 - q_4^2 & 2(q_2q_3 + q_1q_4) & 2(q_2q_4 - q_1q_3) \\ 2(q_2q_3 - q_1q_4) & q_1^2 - q_2^2 + q_3^2 - q_4^2 & 2(q_3q_4 + q_1q_2) \\ 2(q_2q_4 + q_1q_3) & 2(q_3q_4 - q_1q_2) & q_1^2 - q_2^2 - q_3^2 + q_4^2 \end{bmatrix} \tag{5}$$

Based on (5) and (6), the initial heading of the pedestrian can be calculated by

$$\varphi = \arctan(-\frac{2(q_1q_2 + q_0q_3)}{q_0^2 + q_1^2 - q_2^2 - q_3^2}) \tag{6}$$

### 3.3.2 Design of Extended Kalman Filter

To achieve the accurate estimation of attitude angles, we rely on the Extended Kalman Filter (EKF) to integrate the measurement values from the gyroscopes, accelerometers, and magnetometers [4]. The EKF model is constructed as

$$\begin{cases} \mathbf{Q}(k) = \left(\mathbf{U} + \frac{1}{2}\Omega\left(\omega(k)\,T_s\right)\right)\mathbf{Q}(k-1) + \mathbf{w}(k) \\ \begin{bmatrix} \overrightarrow{a^b}(k) \\ \overrightarrow{m^b}(k) \end{bmatrix} = \begin{bmatrix} \mathrm{T}_n^b(Q(k)) & 0 \\ 0 & \mathrm{T}_n^b(Q(k)) \end{bmatrix} \begin{bmatrix} \mathbf{G} \\ \mathbf{b} \end{bmatrix} + \boldsymbol{\nu}(k) \end{cases} \tag{7}$$

where $\mathbf{U}$ is a unit matrix. $T_s$ is the sampling interval. $\boldsymbol{\omega}(k) = \left[\omega_x^b(k) \; \omega_y^b(k) \; \omega_z^b(k)\right]$ is the angular rate in the $k$-th iteration. $\mathbf{w}(k)$ is a vector of Gaussian noise. $\frac{1}{2}\Omega\left(\omega(k)\,T_s\right)$ is the state transition matrix described in (9). $\overrightarrow{a^b}(k)$ and $\overrightarrow{m^b}(k)$ are the measurement values from the 3-axis accelerometer and magnetometer respectively. $\mathbf{b} = \left[0 \; b_y^n \; b_z^n\right]^{\mathrm{T}}$ is the vector of the normalized magnetic field. $\nu(k)$ is a vector of Gaussian observation noise.

$$\frac{1}{2}\Omega\left(\omega(k)\,T_s\right) = \frac{1}{2}\begin{bmatrix} 0 & -\omega_x^bT_S & -\omega_y^bT_S & -\omega_z^bT_S \\ \omega_x^bT_S & 0 & \omega_z^bT_S & -\omega_y^bT_S \\ \omega_y^bT_S & -\omega_z^bT_S & 0 & \omega_x^bT_S \\ \omega_z^bT_S & \omega_y^bT_S & -\omega_x^bT_S & 0 \end{bmatrix} \tag{8}$$

### 3.4 *Fingerprint Data Acquisition*

In the conventional data acquisition method, position coordinate of reference points are measured in advance, the smartphones with the digitized map of the target environment are used to scan the signal from all the available Wi-Fi APs. In the proposed method, the smartphones are used to acquire not only the Wi-Fi signal, but also the data from the sensors including the accelerometers, magnetometers, and gyroscopes. A fingerprint is created by proposed method and position coordinate of reference points are calculated by sensors without measuring, which save time and are convenient to be applied widely.

## 4 Experimental Results

### 4.1 *Experiment Setup*

The experiments are conducted on the same floor in a building with the dimensions of 64.6 m × 18.5 m, as shown in Fig. 2. The red dot indicates the origin of the relative geographical coordinate frame, and the X- and Y-axis points rightwards and upwards respectively. The five D-Link 2310 APs are labeled with red starts in target environment, namely *Orange*, *Pear*, *Apple*, *Banana*, and *Coco*. The receiver used in the experiment is the Samsung Galaxy S3 smartphone with the Android 4.1.2 operation system which is integrated with the Wi-Fi module and sensors including the 3-axis accelerometer, gyroscope, and magnetometer. We use an own-developed App, namely *WifiSensors*, to collect the data from both the Wi-Fi network and sensors, as shown in Fig. 3.

In Fig. 3, we begin to collect the data by clicking the start button, while ending the data acquisition by clicking the end button. The four rows of the data illustrate the measurement values from the 3-axis accelerometer, gyroscope, magnetometer, and Wi-Fi network respectively. The measurement values from the Wi-Fi network is

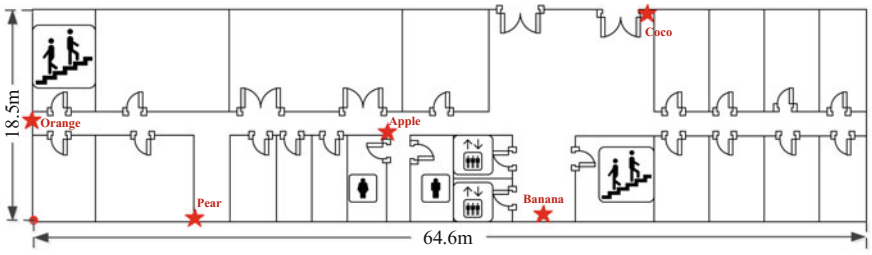



**Fig. 2** Floor plan of the target environment

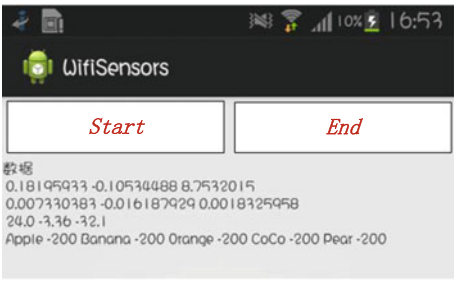



**Fig. 3** Interface of *WifiSensors*

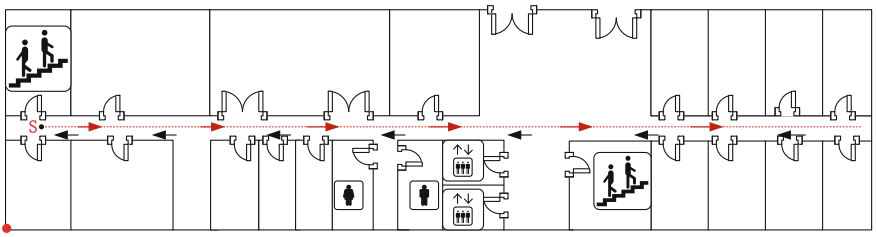



**Fig. 4** Straight line for the testing

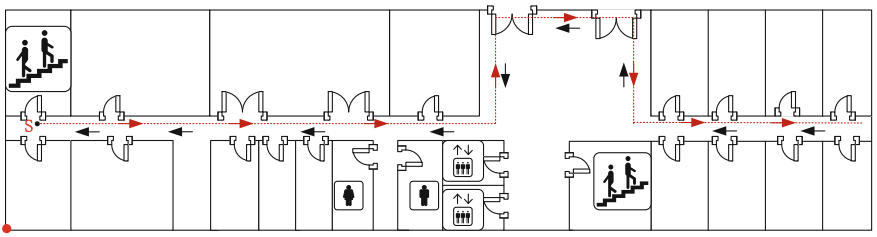



**Fig. 5** Polygonal line for the testing

composed of the Service Set Identifier (SSID), e.g. *Orange*, and RSS. In our system, we set −200 dBm as the lower limit of the received RSS value.

## *4.2 Accuracy of Location Estimation*

Based on the previous discussion, since the PDR is a relative and independent positioning method, it is required to start the positioning from a known location. We select two trajectories, i.e., straight and polygonal lines, for the testing. Both of them start from a given point S, as shown in Figs. 4 and 5.

Figures 6 and 7 compare the real and estimated trajectories for the straight and polygonal lines by using the PDR for positioning, in which red lines represent the

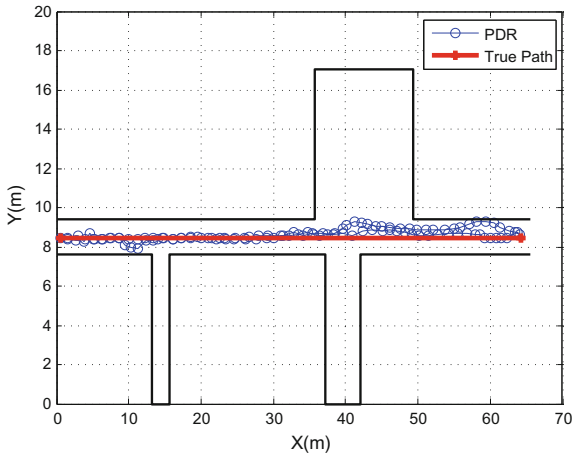



**Fig. 6** The trajectory of straight line

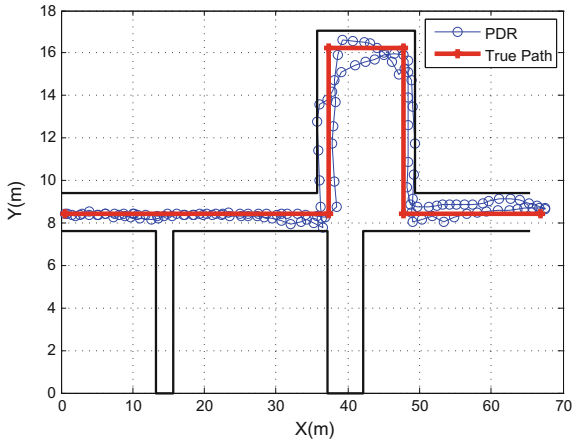



**Fig. 7** The trajectory of polygonal

**Table 1** The statistical results of positioning errors under different testing time

| Types of path (s) | 5 | 10 | 15 | 20 | 25 | 30 | 40 | 60 | 100 |
|---|---|---|---|---|---|---|---|---|---|
| Straight line (m) | 0.14 | 0.23 | 0.25 | 0.31 | 0.45 | 0.51 | 0.60 | 0.76 | 1.2 |
| Polygonal line (m) | 0.12 | 0.20 | 0.32 | 0.40 | 0.51 | 0.81 | 0.90 | 1.2 | 1.79 |

true path in Figs. 4 and 5, and blue lines represent the PDR trajectory. Table 1 shows the statistical results of localization errors under different testing time. As can be seen from these results, the PDR is suffered by the accumulative errors with the time going on. In addition, the existence of turns possibly causes the deterioration of positioning accuracy. Therefore, to guarantee the effectiveness of fingerprint database, we construct it by using the data collected on the short-distance straight lines.

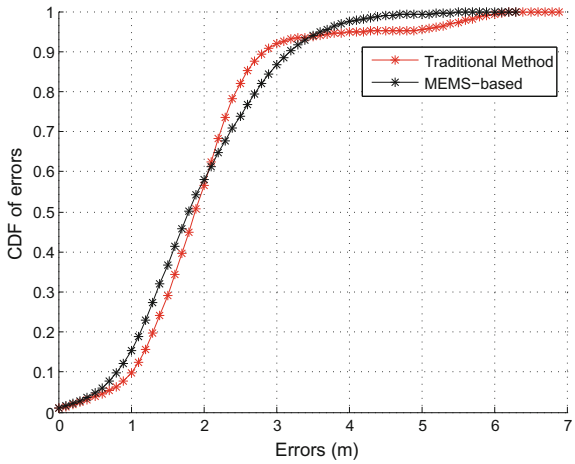



**Fig. 8** CDFs of errors

### *4.3 Comparison of Different Methods for Fingerprint Construction*

Finally, we compare two different methods used for fingerprint database construction, i.e., the conventional location fingerprinting method and the proposed method with the help of MEMS. The conventional method requires measuring the location coordinate of each RPs, which is time-consuming and labor intensive, while in the proposed one, we only need to measure the location coordinates of the starting locations of 24 paths in target environment. After the fingerprint database is constructed, the user installing the *WifiSensors* acquires a newly RSS measurements of a place. The RSS vectors are matched against the fingerprint database, and then the fingerprints corresponding to the highest similarities are selected as the neighbors for Weighted K-nearest Neighbor (WKNN) positioning. The Cumulative Distribution Functions (CDFs) of errors by using these two different methods for fingerprint database construction is shown in Fig. 8. We can find that the proposed method is able to achieve the similar positioning performance to the conventional one, but is featured with much low cost for site survey.

## 5 Conclusion

In this paper, we have proposed a new method of constructing the fingerprint database quickly with the help of multi-sensor. This method enables the users to construct the fingerprint database and real-time updates it. In addition, to enhance the accuracy of the PDR, we employ the EKF to estimate the heading and velocity of the pedestrian. Finally, a comparative experiment is designed to verify the effectiveness and

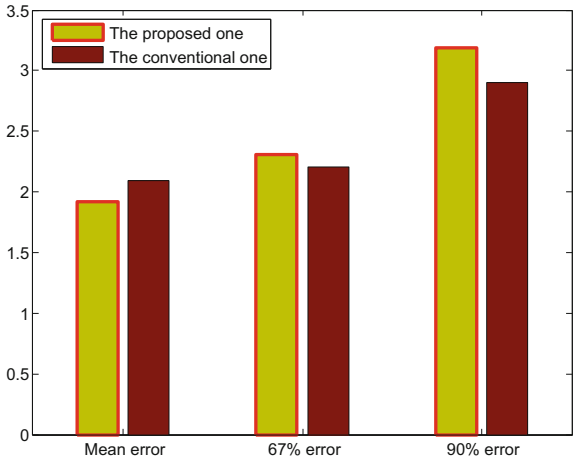



**Fig. 9** Statistical results of localization errors

efficiency of the proposed method used for fingerprint database construction, which demonstrates that the proposed method is much suitable for the large-scale indoor environment. In future, the improvement of positioning accuracy by integrating the indoor geographical information forms an interesting topic Fig. 9.

**Acknowledgements** The authors wish to thank the reviewers for the careful review and valuable suggestions. This work was supported in part by the Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), National Natural Science Foundation of China (61301126), Special Fund of Chongqing Key Laboratory (CSTC), and Fundamental and Frontier Research Project of Chongqing (cstc2015jcyjBX0065).

## References

1. G. Lui, T. Gallagher, B. Li, A.G. Dempster, C. Rizos, Differences in RSSI readings made by different Wi-Fi chipsets: a limitation of WLAN localization, in *2011 International Conference on Localization and GNSS (ICL-GNSS)* (2011), pp. 53–57
2. He. Suining, S.H. Gary Chan, Wi-Fi fingerprint-based indoor positioning: recent advances and comparisons. IEEE Commun. Surv. Tutor. **18**(1), 466–490 (2016)
3. W. Chen, R. Chen, Y. Chen, H. Kuusniemi, J. Wang, An effective pedestrian dead reckoning algorithm using a unified heading error model, in *Position Location and Navigation Symposium (PLANS), 2010 IEEE/ION)* (2010), pp. 340–347
4. Z. Tian, X. Fang, M. Zhou, L. Li, Smartphone-based indoor integrated WiFi/MEMS positioning algorithm in a multi-floor environment. Micromachines **6**(3), 347–363 (2015)
5. V. Renaudin, M. Susi, G. Lachapelle, step length estimation using handheld inertial sensors. Sensors **12**(7), 8507–8525 (2012)
6. X. Yuan, S. Yu, S. Zhang, Quaternion-based unscented Kalman filter for accurate indoor heading estimation using wearable multi-sensor system. Sensors **15**(5), 10872–10890 (2015)

# Carrier Phase Based Attitude Determination Using Tightly Coupled BDS/INS

**Zengshan Tian, Yuezhong Zhang, Mu Zhou and Zipeng Wu**

**Abstract** In order to realize the attitude determination in high accuracy and stability of BeiDou Navigation Satellite System (BDS) receiver carrier, a BDS/INS tightly coupled attitude determination algorithm was proposed. First, the error model of BDS system was put forward. Then, an extended Kalman Filter System with the double differences carrier phase as the main observation and the error state equation of INS as the system state equation which can ensure the attitude in a high precision were designed. Last, a platform was set up for testing the effectiveness of the algorithm with a single-frequency BDS receiver and an inertial sensor. The results show that the algorithm can effectively improve the measurement accuracy and output frequency of the attitude.

**Keywords** Beidou navigation satellite system (BDS) · Inertial navigation system (INS) · Integration attitude determination · Carrier phase · Extended Kalman filter

## 1 Introduction

With the rapid development of Global Navigation Satellite System, satellite based attitude determination has obtained more and more attention in recent years. And it is of great importance to obtain accurate ambiguity in the determination progress. Traditional algorithms such as Least-squares Ambiguity Decorrelation Adjustment (LAMBDA) [1], Constrained LAMBDA (CLAMBDA) [2], and Modified LAMBDA (MLAMBDA) [3] can provide accurate ambiguity only with the reasonable observation from receiver. Moreover, due to multipath effect, the satellite signal is often blocked which hinders the ambiguity accuracy [4].

Z. Tian · Y. Zhang (✉) · M. Zhou · Z. Wu (✉)
Chongqing Key Lab of Mobile Communication Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: hpuzhang@163.com

Z. Wu
e-mail: 846333478@qq.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_4

The ideas of using inertial sensors to improve GPS ambiguity accuracy have been proposed recently [5]. These methods adopt inertial sensors in Micro Electro Mechanical Systems (MEMS) to meet different ambiguity resolution requirement. Specifically, the MEMS attitude is utilized to reduce the search region for integer ambiguity, thus improve the instantaneity and reliability of ambiguity resolution [6]. However, the accuracy will be deteriorated when the satellite signal is blocked.

In this paper, a novel approach is proposed to improve the accuracy of attitude by the data fusion of BDS and INS. First, we use the ambiguity dilution of precision (ADOP), Geometric Dilution of Precision (GDOP), as well as the Signal Noise Ratio (SNR) from BDS measurement to estimate the quality of ambiguity. Secondly, system measurement equation based the carrier phase and pseudo-rate is designed. Finally, the BDS/INS integrated filter is adopted to estimate and correct the errors of attitude and the measurement from sensor.

This paper is organized as follows. Section 2 describes the process in which makes use of ADOP, GDOP, and SNR to estimate the quality of BDS attitude. Section 3 describes the carrier phase and pseudo-rate based system measurement equation. The experimental results are illustrated in Sect. 4. And Sect. 5 concludes this paper.

## 2 The Error Model of BDS Attitude Determination

The double differences (DD) observation of BDS is as follows:

$$\begin{cases} P_{r_{12}}^{S_{12}} = \rho_{r_{12}}^{S_{12}} + I_{r_{12}}^{S_{12}} + T_{r_{12}}^{S_{12}} + \varepsilon_{P,r_{12}}^{S_{12}} \\ \boldsymbol{\Phi}_{r_{12}}^{S_{12}} = \rho_{r_{12}}^{S_{12}} + IS_{12_{r_{12}}} + I_{r_{12}}^{S_{12}} + T_{r_{12}}^{S_{12}} + \lambda \cdot z_{r_{12}}^{S_{12}} + \varepsilon_{\boldsymbol{\Phi},r_{12}}^{S_{12}} \end{cases} \tag{1}$$

where $P_{r_{12}}^{S_{12}}$ and $\boldsymbol{\Phi}_{r_{12}}^{S_{12}}$ are the vector of DD pseudo and DD carrier phase respectively; $\rho_{r_{12}}^{S_{12}}$ is the vector of DD the real distance from receiver to satellite; $I_{r_{12}}^{S_{12}}$ and $T_{r_{12}}^{S_{12}}$ are the vector of DD error of ionosphere and troposphere respectively; $z_{r_{12}}^{S_{12}}$ is the DD Ambiguity; $\varepsilon_{\boldsymbol{\Phi},r_{12}}^{S_{12}}, \varepsilon_{P,r_{12}}^{S_{12}}$ are the vector of DD unknown error.

The linearized single-baseline double differences model is described as:

$$y = Aa + Bb + e \tag{2}$$

where a is the of DD integer ambiguity; b is the vector of baseline and e is the unknown error.

Then we can get the fixed solution and the covariance of baseline through least square estimation and CLAMBDA, as follows:

$$\begin{cases} \hat{b}(a) = \hat{b} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1} (\hat{a} - a) \\ Q_{\hat{b}(a)} = Q_{\hat{(b)}} - Q_{\hat{b}\hat{a}} Q_{\hat{a}}^{-1} Q_{\hat{a}\hat{b}} \end{cases} \tag{3}$$

An evaluation factor that captures the ambiguity precision as:

$$ADOP = |Q_{\hat{a}}|^{1/2n}[cyc] \tag{4}$$

where *ADOP* stands for ambiguity dilution of precision which properly describes the precision and correlation of the ambiguities. Since the *ADOP* is determined by the value of the $Q_{\hat{a}}$, and the number of the available satellite. And we can decompose the $Q_{\hat{a}}$ as follows:

$$|Q_{\hat{a}}| = (\frac{n+1}{2^n})(\frac{\sigma_{\Phi}}{\lambda_f})|Q|^n(1+\frac{\sigma_P^2}{\sigma_{\Phi}^2})^{3n} \tag{5}$$

where $\sigma_P$ and $\sigma_{\Phi}$ stand for the standard deviation of pseudo and carrier phase respectively. Furthermore, a gain factor can be defined as:

$$\zeta = \frac{ADOP(Q_{\hat{Z}\hat{Z}})}{ADOP(Q_{\hat{Z}_R\hat{Z}_R})} = 1 + \frac{\sigma_P^2}{\sigma_{\Phi}^2} \tag{6}$$

where $Q_{\hat{Z}\hat{Z}}$ is the optimal covariance matrix of the solution using LAMBDA. And $Q_{\hat{Z}_R\hat{Z}_R}$ stands for the suboptimal solution. Then the precision factor of attitude can be calculated by:

$$\xi = ADOP \times GDOP \times e^{\zeta} \tag{7}$$

## 3 BDS/INS Integrated Filter

### *3.1 State Equation of the Integrated Filter*

The state vector of integrated filter is given as:

$$X(t) = [\delta V, \delta\boldsymbol{\Phi}, \delta p, \delta\varepsilon, \bigtriangledown_a]^T \tag{8}$$

where $\delta V = [\delta v_x, \delta v_y, \delta v_z]$ is the velocity errors vector of east, north, and up; $\delta\boldsymbol{\Phi} = [\delta\theta, \delta\gamma, \delta\psi]$ is the vector of pitch, roll, and yaw errors; $\delta p = [\delta L, \delta\lambda, \delta h]$ is the vector of longitude, latitude and height errors respectively; $\delta\varepsilon$ and $\nabla_a$ are gyro and accelerometer error vectors. And the system state equation is described by:

$$\dot{X}(t) = F(t)X(t) + G(t)W(t) \tag{9}$$

where $F(t)$ is the state transition matrix, W(t) is the system noise and the G(t) is the noise coefficient matrix.

## 3.2 Measurement Equations of the Integrated Filter

The system measurement equations include carrier phase and pseudo-rate measurement equations.

### 3.2.1 Carrier Phase Measurement Equation

The DD carrier phase observation is described in Eqs. (1), and its linearized model is described as

$$(\boldsymbol{\Phi}_{r_{12}}^{S_{12}} + Z_{r_{12}}^{S_{12}})\lambda = (l^{S_{12}}, m^{S_{12}}, n^{S_{12}})(x - x_0, y - y_0, z - z_0)^T + \varepsilon \tag{10}$$

where the position of base receiver is set as $(x_0, y_0, z_0)$, $\varepsilon$ is the unknown error. If the INS and the rover receiver are in the same position, we can describe the DD equation between INS and base receiver as:

$$P_I^{S_{12}} = (l_I^{S_{12}}, m_I^{S_{12}}, n_I^{S_{12}})(x - x_0, y - y_0, z - z_0)^T + \varepsilon_I \tag{11}$$

The carrier phase measurement equation is the difference between Eqs. (10) and (11), which can be described as follows:

$$(\varphi^{S_{12}} - P_I^{S_{12}}) = (l^{S_{12}} - l_I^{S_{12}}, m^{S_{12}} - m_I^{S_{12}}, n^{S_{12}} - n_I^{S_{12}}) \begin{pmatrix} x - x_I \\ y - y_I \\ z - z_I \end{pmatrix} + (\varepsilon - \varepsilon_I) \tag{12}$$

where $\varphi^{S_{12}} = (\boldsymbol{\Phi}_{r_{12}}^{S_{12}} + Z_{r_{12}}^{S_{12}})\lambda$ is the DD range calculated by the DD carrier phase, $(x_I, y_I, z_I)$ is the position calculated by INS. Then we describe the measurement as follows:

$$Z_\varphi = H_\varphi X + V_\varepsilon \tag{13}$$

### 3.2.2 Pseudo-Rate Measurement Equation

The pseudo-rate of rover receiver can described as:

$$\dot{\rho}_{Gi} = \dot{r}_i - \delta t_{ru} - \nu_{\rho i} \tag{14}$$

where $\dot{r}_i$ can be calculated by:

$$\dot{r}_i = e_{ix}(\dot{x} - \dot{x_{si}}) + e_{iy}(\dot{y} - \dot{y_{si}}) + e_{iz}(\dot{z} - \dot{z_{si}}) \tag{15}$$

And the distance between satellite and rover receiver can be described as:

$$\dot{\rho}_{Ii} = \dot{r}_i + e_{ix}\delta\dot{x} + e_{iy}\delta\dot{y} + e_{iz}\delta\dot{z} + \frac{1}{r_i}[(\dot{x} - \dot{x_{si}} - \dot{r}_i e_{ix})\delta x + (\dot{y} - \dot{y_{si}} - \dot{r}_i e_{iy})\delta y + (\dot{z} - \dot{z_{si}} - \dot{r}_i e_{iz})\delta z] \tag{16}$$

where $(\delta x, \delta y, \delta z)$ and $(\delta\hat{x}, \delta\hat{y}, \delta\hat{z})$ are the position error and velocity error of receiver in CGCS2000 coordinate system respectively. We can calculate the difference of the pseudo-rate between BDS and INS by Eqs. (14) and (16):

$$\delta\dot{\rho} = D_v \begin{bmatrix} \delta v_x \\ \delta v_y \\ \delta v_z \end{bmatrix} + D_p \begin{bmatrix} \delta L \\ \delta\lambda \\ \delta h \end{bmatrix} + V_\rho \tag{17}$$

where $D_v = eC_n^e$, $C_n^e$ is the transition matrix from n-coordination to e-coordination, $D_p = eD_E + ED_a$. Then we can describe the Eq. (18) as:

$$Z_{\dot{\rho}}(t) = H_{\dot{\rho}}(t)X(t) + V_{\dot{\rho}}(t) \tag{18}$$

## 4 Experimental Results

Experimental tests in the real environment were conducted to verify the performance of the proposed attitude determination system. The testbed is shown in Fig. 1 which consists of INS, the BDS attitude determination system, turntable, and the display unit. The BDS attitude determination system includes two receivers, an interface unit, and a navigation processing unit. The output frequency of the BDS/INS integrated system is around 20–100 Hz.

In order to test the overall performance of the proposed approach, we conducted the experiment in three conditions as static, rotation, and signal obscured. In order to illustrate the experimental results more clearly, we use the yaw to represent the attitude in the following.

### *4.1 Verification in Static Condition*

The static experiment was conducted in the bad observation condition when the BDS fluctuates a lot, which is in order to verifying the performance of the integrated system.

It can be seen from Fig. 2 that the precision of BDS system is lower than the integrated system. Since the bad weather can deteriorate the precision of observation and the ambiguity was wrong and fluctuates a lot which caused the attitude disable. Meanwhile, the performance of proposed integrated system remains accurate, instant, and stable. The result illustrates that the integrated system can improve the precision and output frequency of attitude even the BDS attitude is inaccurate Fig. 3.

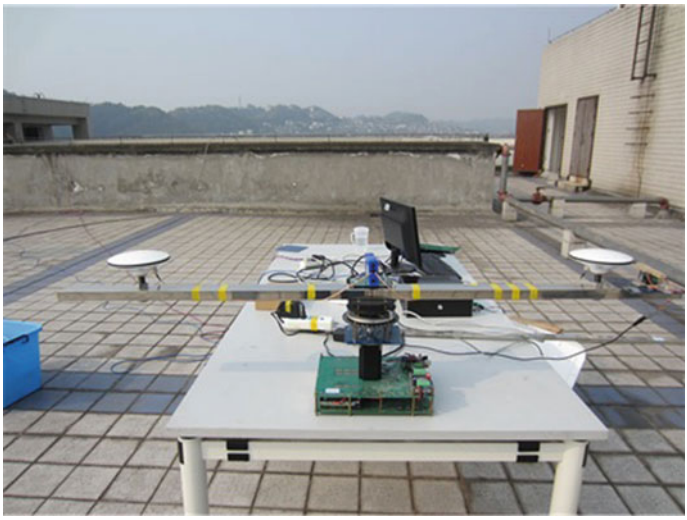


**Fig. 1** The BDS/INS attitude determination system

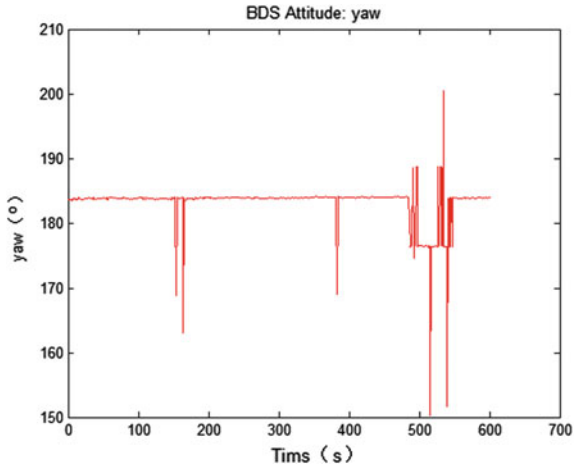



**Fig. 2** The BDS system output attitude of static condition

## *4.2 Verification in Rotation Condition*

When the rotation experiment was conducted, the turnable uniformly rotated a round. And the results of the experiment are shown as follows.

The Figs. 4 and 5 show the attitude and the attitude error of BDS system and the integrated system respectively. It can be seen from Fig. 5 that there is a minor fluctuation in BDS attitude during the receiver's rotation. While the attitude calculated by the BDS/INS integrated system is nearly no fluctuation. This indicates that the sys-

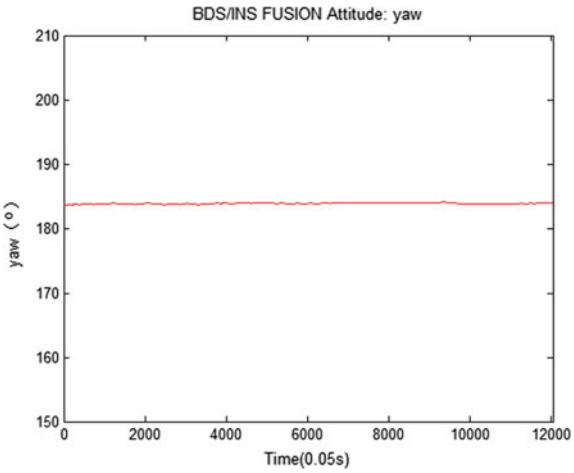



**Fig. 3** The integrated system output attitude of static condition

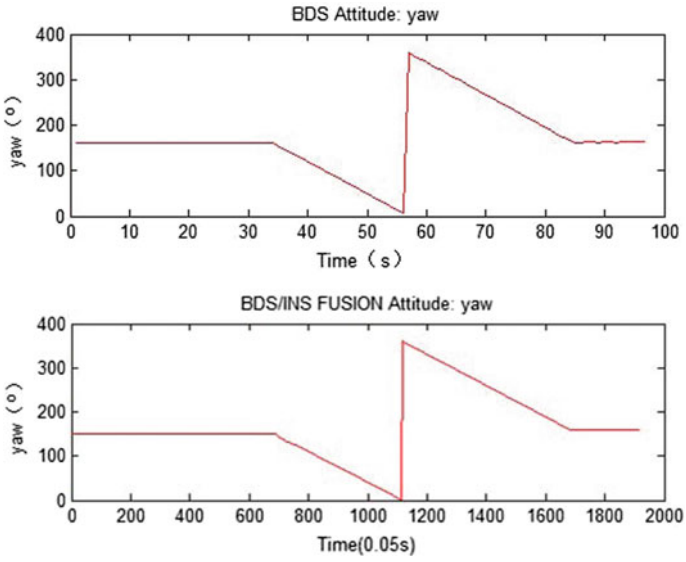



**Fig. 4** The attitude of BDS system and integrated system under rotation

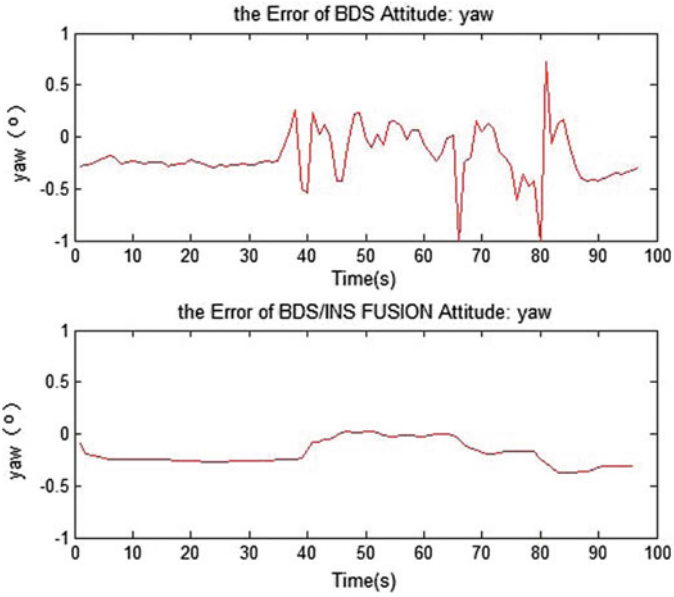



**Fig. 5** The attitude error of BDS system and integrated system under rotation

tem improves the accuracy, stability, and the output frequency of the attitude during the motion.

## *4.3 Verification in Obscured Signal Condition*

Finally, we tested the performance of static obscured experiment as scheme. In this experiment, the satellite signal was shield by hand-held double printed circuit boards (25 cm * 30 cm) above the main antenna.

The number of satellites observed in the case of occlusion is described in Fig. 6, in which the red and blue lines stand for the number of the satellite observed by main receiver and rover receiver respectively. And the green line stands for the number of the available satellite of the BDS.

It can be seen from Fig. 7 that the BDS attitude is accurate and stable, when the number of the observed satellites is 5 or more. But between 700 and 800 epochs in which the number descends to 4 or less, the attitude calculated by the BDS was terrible and unavailable. Meanwhile the fusion system corrects the attitude by the extended Kalman filter. And the fusion system attitude is accurate and almost no fluctuation.

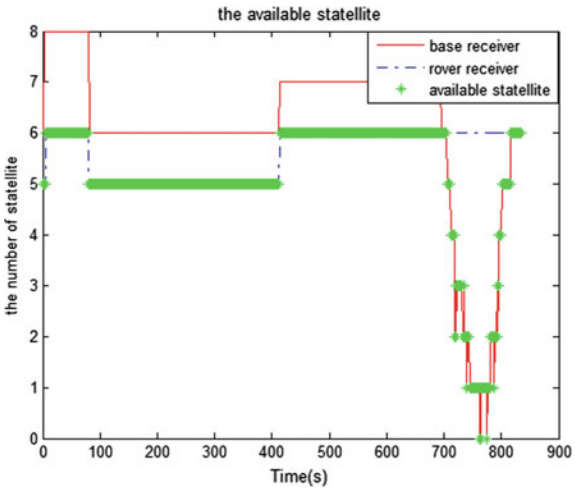



**Fig. 6** The number of the available satellite

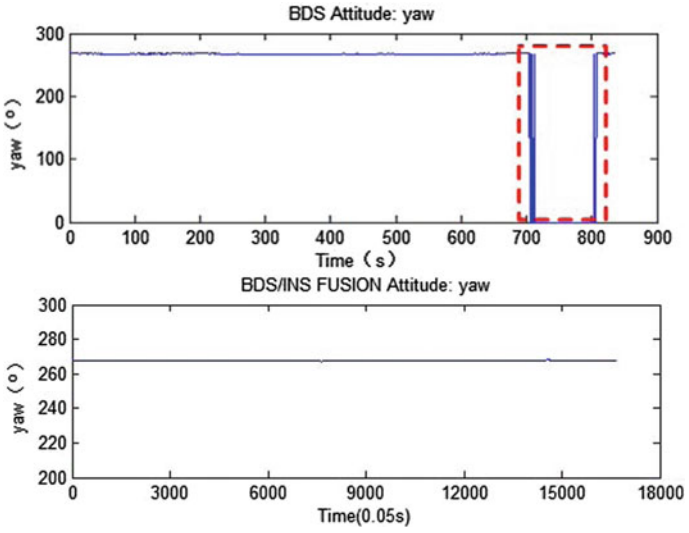



**Fig. 7** The comparison of the two systems attitude under obscured condition

## *4.4 Result Analysis*

At last, we counted several experimental tests results and calculated the standard deviation. The comparison of the standard deviation between BDS system and the fusion system is shown in Table 1.

**Table 1** Comparison of the attitude standard deviation between BDS system and integrated system

| | Stationary | | | Motion | | |
|---|---|---|---|---|---|---|
| Groups | 1 | 2 | 3 | 4 | 5 | 6 |
| The number of the epoch | 580 | 900 | 850 | 140 | 135 | 125 |
| The standard deviation of BDS attitude | 0.121 | 0.015 | 0.013 | 0.231 | 0.283 | 0.266 |
| The standard deviation of fusion attitude | 0.009 | 0.012 | 0.010 | 0.122 | 0.125 | 0.123 |

From the above results, it can be seen that the standard deviation calculated by the BDS/INS integrated system is smaller than the BDS systems whether the vehicle is stationary or in motion. So the integrated system can improve the accuracy of the attitude, even though the BDS system cannot provide accurate attitude. Besides, the integrated system can provide the output frequency from 20 to 100 Hz.

## 5 Conclusions

In order to improve the performance of the BDS/INS attitude determination system, a novel integrated attitude determination approach is presented. We calculate the attitude covariance of BDS which is effected by the corresponding error model. And the proposed algorithm is based on carrier phase which is measured accurately. The experimental tests prove that the proposed scheme can improve the accuracy, instantaneity of the BDS attitude determination.

## References

1. P.J.G. Teunissen, The least-squares ambiguity decorrelation adjustment: a method for fast GPS integer ambiguity estimation [J]. J. Geodesy **70**(70), 65–82 (1995)
2. P.J. Buist, The baseline constrained LAMBDA method for single epoch, single frequency attitude determination applications [J], in *Proceedings of ION GPS-2007*, vol. 3 (2007), pp. 2962–2973
3. X.W. Chang, X. Yang, T. Zhou, MLAMBDA: a modified LAMBDA method for integer least-squares estimation [J]. J. Geodesy **79**(9), 552–565 (2005)
4. L. Cong, E. Li, H. Qin et al., A performance improvement method for low-cost land vehicle GPS/MEMS-INS attitude determination. Sensors **15**(3), 5722–5746 (2015)
5. C. Wang, G. Lachapelle, M.E. Cannon, Development of an integrated low-cost GPS/rate gyro system for attitude determination. J. Navig. **57**(01), 85–101 (2004)
6. C. Eling, P. Zeimetz, H. Kuhlmann, Development of an instantaneous GNSS/MEMS attitude determination system. GPS Solut. **17**(1), 129–138 (2013)

# A Deficit-Round-Robin-Based Variable-Length Packets Scheduling Scheme for Satellite Onboard Switches

**Le Yang, Qinghua Chen, Lufeng Qiao, Pengze Lv and Qian Chen**

**Abstract** Due to the limitation of hardware resources in satellite onboard switches, this paper proposes a three-level deficit-round-robin-based variable-length scheduling algorithm, which could improve the throughput as well as saving the hardware resources. The circuits occupy 6607 slice registers and 8092 slice lookup tables while used in a 16 × 16 switch fabric, which can meet the requirements of triple modular Redundancy. The scheme can meet the requirements of 160 Gbps switch fabric with 16, 10 Gbps ports, and the design is implemented in a Xilinx xc6vlx240t FPGA. Typical simulated results are presented to show the availability of the scheme.

**Keywords** Satellite onboard switches · Deficit-round-robin · Scheduling scheme · FPGA

## 1 Introduction

The rapid development of information network has proposed an urgent requirement of improving the performance of satellite onboard switches, which play an important role in the modern networks [1].

An efficient packet scheduling algorithm is of great importance to the performance of switches. Packet scheduling is the strategy of selecting among all the flows backlogged according to different scheduling granularity.

The current packet scheduling algorithms are divided into two categories: work-conserving and non-work-conserving [2]. Round Robin [3] was proposed originally in order to get various traffic flows treated equally. The router discriminates flows on the basis of their ID or service classes and put them into corresponding output queues, then the nonempty queues would be polled in cyclic order. The flaw of this

L. Yang · Q. Chen (✉) · L. Qiao · P. Lv · Q. Chen
Institute of Communication Engineering, PLA University Science and Technology, Nanjing, China
e-mail: 3601237@qq.com

L. Yang
e-mail: 1601654705@qq.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_5

solution is the ignorance of packet length. The idea of Fair Queuing [4] is proposed under the circumstance, Demers proposed an ideal algorithm, in which each output link is split infinitesimally and each flow is serviced instantaneously. The model is difficult to realize, several approximate algorithms are proposed such as WFQ [4], WF2Q [5], SFQ [6] and so on. These algorithms perform the virtual time function which makes computation of departure time and sorts the packet based on the departure time. This method leads to high implementation cost with the algorithm complexity of O(log N) at least. By comparison, DRR can remove the flaw with the complexity of O(1) [7], which is suitable for hardware implementation.

## 2 Switch Fabric Architecture

The design consists of ingress/egress interface, 32-bit standard interface, local scheduler, central scheduler, and the on-chip SRAM for data storage. The local scheduler and central scheduler are connected by serial links. On account of the three-level scheduling employed by the switch fabric, the realization of the scheduling scheme needs the cooperation of local schedulers and central scheduler. A local scheduler lies in each VOQ, as each ingress port maintains N VOQs, there are $N^2$ local schedulers in total in N × N switch fabric.

## 3 The Optimization of Switch Fabric

The design adopts a three-level variable-length scheduling combined-input-output-queuing (CIOQ) structure, which is suitable when there are large amounts of ports with high-speed. Unlike the output-buffered or the input-buffered mechanism, the CIOQ mechanism is the proper way to guarantee the scalability and alleviate the output contention. Although the output-buffered mechanism has excellent performance on traffic throughput and delay control, the switch size is limited by the memory access bandwidth [8]. When the speed of each port reaches multi-gigabits, simple output-buffered switch fabrics cannot meet the requirements. Then, the CIOQ mechanism can be used. It has been shown that CIOQ switches with speedup of two can achieve 100% throughput with any maximal scheduling algorithm [9, 10] Figs. 1, 2, and 3.

The design adopts the VOQ mechanism to solve the problem of HOL blocking [7], which means establishing a dedicated queue associated with each egress port at the buffer of each ingress port, this method improves the overall throughput of input-buffered switch fabric, both the theoretical research and the simulation practice manifest that a VOQ router employing the maximum weight match scheduling algorithm can achieve 100% throughput [11].

Variable-length scheduling can be employed in the CIOQ switches to reduce the hardware consumption in satellite onboard switches. Compared with fixed-length

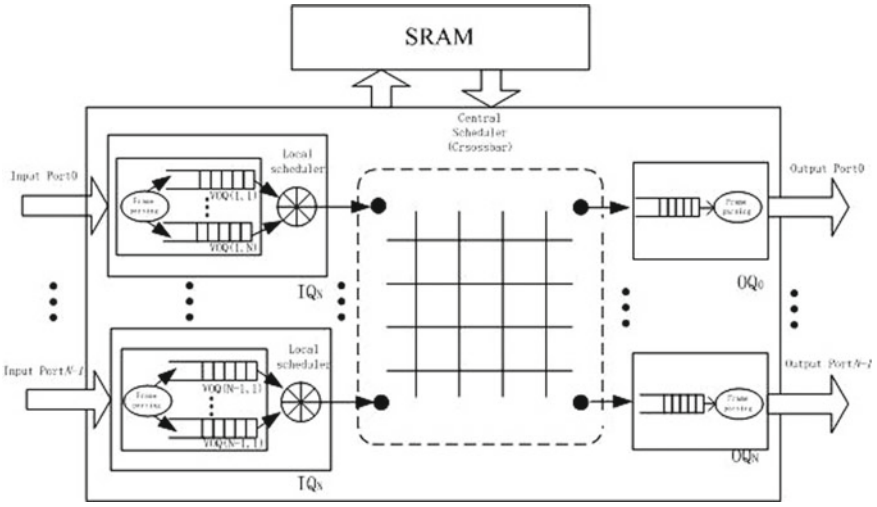



**Fig. 1** Structure of switch fabric

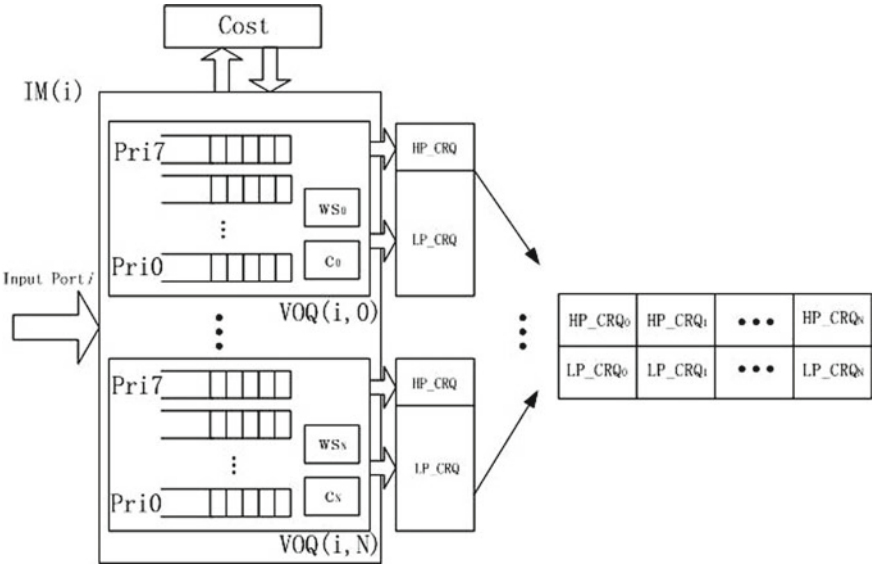



**Fig. 2** First-level scheduling example

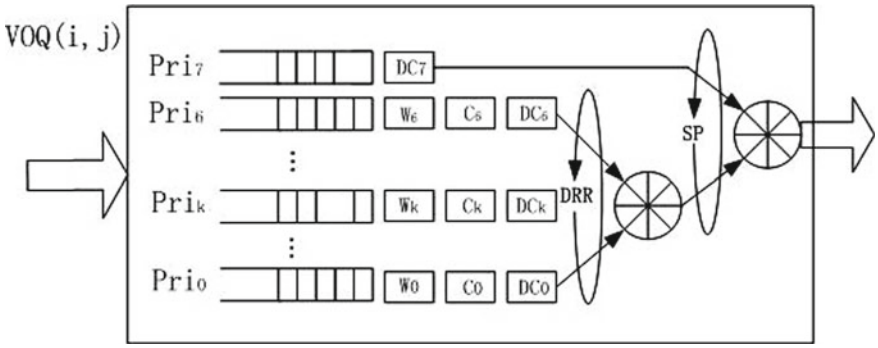



**Fig. 3** Third-level scheduling example

scheduling, variable-length scheduling has two main advantages: Firstly, the latency would be reduced significantly because the time expanded on reassembly of cells could be saved using the variable-length scheduling [12]. Secondly, hardware implementation cost would also be lowered.

## 4 The Scheduling Algorithm

In this section, the implementation of the algorithm will be introduced, the realization of three-level scheduling process involved the local and central schedulers, the algorithm employed by the local schedulers during the first- and third-level scheduling is presented in this section.

In this CIOQ structure, the input buffer is split into N VOQs. There are one high-priority queue and seven low-priority queues in each VOQ. VOQ($i, j$) denotes that the stored cells coming from ingress port $i$ are aiming at egress port $j$. IM($i$) presents ingress port $i$.

Each priority queue in a VOQ corresponds to a variable $weight_{i,j,k}$ and $credit_{i,j,k}$, $k = 0, 1, 2 \ldots 6$. Each VOQ corresponds to a variable $weight\_sum_{i,j}$ and $credit_{i,j}$, let the B($t$) denote the all the priority queues holding data to transmit at time $t$, the value of the $weight\_sum_{i,j}$ can be obtained by calculating the following equation:

$$weight\_sum_{i,j}(t) = \sum_{k \in B(t)} weight_{i,j,k}(t)(4-1)$$

During the first-level scheduling process, the $credit_{i,j}$ is used and its variation is based on the $weight_{i,j}$, the cost is used as the judgement criterion of whether the VOQ could send the request. During the third-level scheduling process, the $credit_{i,j,k}$ is used and its variation is based on the $weight_{i,j,k}$, the cost is used as the judgement criterion of whether the priority in the certain VOQ could be responded. In both scheduling process, the cost is a dynamic variable and its value changes dynamically based on the current traffic condition.

### *4.1 First-Level Scheduling Algorithm: CRQ Generation*

The input buffer is split into N VOQs, so the scheduler needs to deal with at most $8N^2$ connection request, in order to reduce the complexity of iterative operation, the design set certain threshold for the transmission of each connection request, the VOQs holding data frames cannot transmit the connection request unless its corresponding credit exceeds the predefined threshold and all cells in a frame has entered the switch fabric. The specific process is as follows:

Step 1: The credits of all VOQs in IM($i$) are set to 0. The cost will be first set to the sum of all weights of low-priority classes holding frames waiting to be transmitted, then the maximum cost across all the VOQs is used as the global cost.

Step 2: If $Pri_7$ logic queue in VOQ($i,j$) is not empty, the corresponding $HP\_CRQ_n$ is asserted.

Step 3: The credits of all VOQs in IM($i$) would be added by their weight sum, let the $t_n(n = 2, 3, \ldots)$ be time at which the *nth* scheduling process occurs. The $credit_{i,j}$ evolves as follows,

$$credit_{i,j}(t_n) = credit_{i,j}(t_{n-1}) + weight\_sum_{i,j}(t_n), j = 0, 1, \ldots N(4-2)$$

Then there are two cases as follows:

(1) If the current total weight of VOQ($i,j$) is less than the cost and the accumulated credits of VOQ($i,j$) exceed the cost, the value of the register $LP\_CRQ_n$ which indicates the corresponding low-priority connection request is asserted, then the current credit will be decreased by the cost, that is,

$$credit_{i,j}(t_n) = credit_{i,j}(t_{n-1}) - cost(t_n)(4-3)$$

(2) If the current weight sum of VOQ($i,j$) is equivalent to the cost, the credit is not used and the CRQ is sent without modifying the credit value, that is,

$$credit_{i,j}(t_n) = credit_{i,j}(t_{n-1})(4-4)$$

Step 4: The high-priority connection requests (*HP_CRQs*) and low priority connection requests (*LP_CRQs*) from N VOQs constitute the header before the payload of the frame which is sent in acknowledged to the last scheduling period. The unicast and multicast frames are processed identically. For the unicast and multicast frame with the same priority aiming at exactly the same destination output ports, whichever meets the demand can generate the connection request accordingly.

### 4.2 Third-level Scheduling Algorithm: Queue Selection

Traditional DRR algorithm serves all the queues backlogged in a round robin manner, although the manner guarantees the fairness of scheduling process in certain degree while accomplishing variable-length scheduling, it ignores the real situation of the existence of different priorities in various traffic, which results in that the traffic with high-priority cannot be served first.

In order to allocate the bandwidth according to the demands of transmission service, the design implements the traditional DRR algorithm on the basis of introducing the concept of credit. Each priority queue is still allocated a *DC* variable and the *DC* is added by a certain quantity during each scheduling period. The specific process is as follows:

Step 1: If VOQ$(i,j)$ gets granted, all logical queues are accessed in descending order of priority, saying, from $Pri_7$ to $Pri_0$. The eight logical queues are classified into two levels of priorities. The $Pri_7$ belongs to high priority and the other seven logic queues belong to low priorities. They are served by the rule of strict priority. If connection request exists in $Pri_7$, it could be served immediately. The low-priority queue cannot be serviced until the high-priority queue is empty.

Step 2: If there is no frame in the high-priority queue, the pointer $P_j$ goes down to the classes with lower priorities, assume there is at least one packet in the logic queue with $Pri_k(0 \le k < 7)$, the corresponding weight of the *kth* queue is set to the connection threshold of current scheduling cycle, the credit of the *kth* queue is set to 0.

Step 3: The priority queues with $Pri_{k-1} \sim Pri_0$ are checked continuously, once the accumulated credits of the $pth(0 \le p < k)$ queue exceed the cost, the corresponding priority queue will be responded and the current credit of the queue is decreased by the cost. Otherwise, under the circumstance that all the accumulated credits of all the k priority queues are not large enough to meet the request, the priority queue with $Pri_k$ will be responded without modifying the credit. The $credit_{i,j,k}$ evolves as follows:

$$credit_{i,j,k}(t_n) = 0(4-5)$$

$$credit_{i,j,p}(t_n) = credit_{i,j,p}(t_{n-1}) - cost(t_n), 0 \le p < k(4-6)$$

$$credit_{i,j,p}(t_n) = credit_{i,j,p}(t_{n-1}) + weight_{i,j,q}(t_n), 0 \le q < p(4-7)$$

Step 4: The sequence number of selected priority queue is fed back to the queue manager, it is noteworthy that the HOL frame will be scheduled only if its length is shorter than the value maintained in deficit counter, the remainder is left in the deficit counter preparing for the next scheduling. Afterwards, the local scheduler sends back acknowledgement signal to the queue manager for the withdrawal of the request signal.

### *4.3 The Bandwidth Allocation of Unicast and Multicast Traffic*

As the unicast and multicast scheduler work independently, the problem of bandwidth allocation needs to be considered. We use a 32-bit WRR priority configuration register, 0 or 1 lies in the horizontal direction according to the configuration set in advance. "0" represents polling unicast frame first, "1" represents polling multicast frame first. The scheduler has a pointer *r* in the horizontal direction which polls each bit of the register in cyclic order.

When the process of queue selection terminates within unicast and multicast queues, the configuration register would be queried to determine a unicast connection request or a multicast connection request can be sent. When the bit indicated by the pointer *r* is "0" and no unicast frame waits to be transmitted currently, the multicast frame would be transmitted and vice versa. The quantitative distribution of 0 and 0 is in direct proportion to the predefined distribution ratio of unicast and multicast traffic. If there are no frames currently, an idle frame would be transmitted. If the amount of 0 in the priority configuration register is K, then the proportion which the unicast traffic occupies in the total bandwidth of egress link is K/32, the proportion that the multicast traffic occupies is (32−k)/32.

## 5 Simulated Results and Analysis

The design is implemented in a Xilinx xc6vlx240t FPGA and the development environment is ISE Design Suite 14.7. The scheduling algorithm is realized with Verilog HDL and simulated with Modelsim SE 10.2c.

### *5.1 Simulation of CRQ Generation Scheduling Algorithm*

Figure 4 shows the variation of the accumulated credits of VOQ1, VOQ3, VOQ5 and VOQ10 during different scheduling periods. As presented in Fig. 4, once the accumulated credits of a certain VOQ exceed the cost or its current weight sum is equivalent to the cost, it will acquire the output permission.

### *5.2 Simulation of Queue Selection Scheduling Algorithm*

Figure 5 shows the simulated results of the third-level scheduling process, the DRR scheduler starts to operate after VOQ3 is granted. In the position of ①, where is the beginning of the scheduling process, the connection requests exist in $Pri_2$, $Pri_3$ and

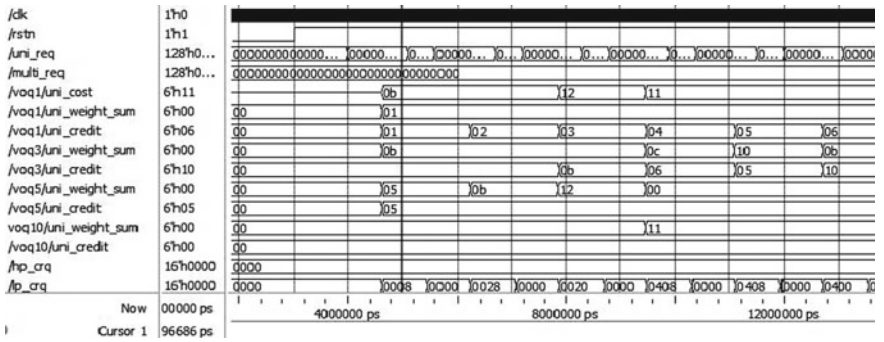



**Fig. 4** Simulated waveform of first-level scheduling process

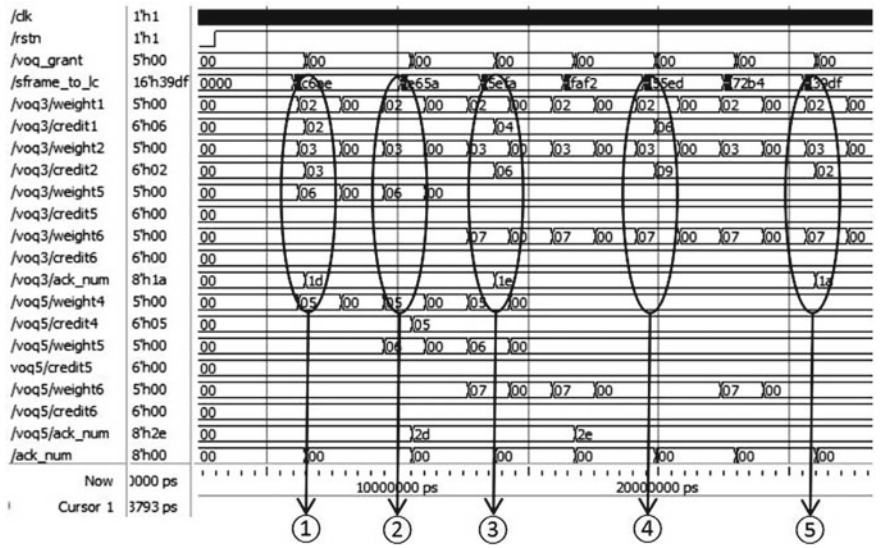



**Fig. 5** Simulated waveform of third-level scheduling process

$Pri_6$ with the weight of 2, 3 and 6 respectively. The weight of $Pri_6$ becomes the cost and credit of $Pri_6$ is set to 0, then its queue number is sent to the queue manager through *ack_num* with the value of 8'h1d. The VOQ5 is granted in the following scheduling period, $Pri_6$ get the output permission at the position of ②. At the position of ③, the VOQ3 is granted again, the credits of $Pri_2$ and $Pri_3$ accumulate further, but is still insufficient to output a frame, and the traffic of $Pri_6$ sends out the frame in the head, the scheduler then withdraws the corresponding connection request. Meanwhile, the connection request for $Pri_7$ is generated which leads to variation of cost from 6 to 7. As the accumulated credits of $Pri_2$ and $Pri_3$ are still lower than the cost, *ack_num* is set to be 8'd1e. The $Pri_7$ in the VOQ5 get the output permission at the

position of ④. At the position of ⑤, The credit of $Pri_3$ turns to be 9 which exceeds the cost, the logic queue with $Pri_3$ gets the output permission and its credit is decreased by the cost, which turns out to be 2 [13].

## 6 Conclusion

A three-level deficit-round-robin-based variable-length scheduling scheme is presented in this paper. The whole design is implemented in a Xilinx xc6vlx240t FPGA. For a $16 \times 16$ switches, the hardware resource consumption is reduced significantly to make it more suitable for satellite onboard switches. The simulated results of the scheduling module show that it works correctly.

## References

1. D. Whitefield, R. Gopal, S. Arnold, Spaceway now and in the future: on-board IP packet switching satellite communication network, in *Military Communications Conference* (2006), pp. 1–7
2. H. Wang, G. Guanqun, The research of packet scheduling algorithms within integrated services network. Chinese J. Comput. **22**(10), 1091–1099 (1999)
3. J. Nagle, On packet switches with infinite storage, IEEE Trans. Commum. **COM-35**(4), (1987)
4. A. Demers, S. Keshav, S. Shenke, Analysis and simulation of a fair queueing algorithm, *ACM SIGCOMM89* **19**(4), 3–12 (1989)
5. J. Bennett, Z. Hui, WF2Q: worst-case fair weighted fair queueing. *IEEE INFOCOM96* (CA, San Francisco, 1996), pp. 120–128
6. P. Goyal, H.M. Vin, H. Chen, Start-time fair queueing: a scheduling algorithm for integrated services. *ACM SIGCOMM96* (CA, Stanford, 1996), pp. 157–168
7. M. Shreedhar, G. Varghese, Efficient fair queuing using deficit round-robin. IEEE/ACM Trans. Netw. **4**(3), 375–385 (1996)
8. H.J. Chao, B. Liu, *High Performance Switches and Routers* (Wiley, 2007), pp. 179–181
9. J.G. Dai, B. Prabhakar, The throughput of data switches with and without speedup. *IEEE INFOCOM00*, vol. 2 (Tel Aviv, Israel, 2000), pp. 556–564
10. D. Pan, Y. Yang, Pipelined two step iterative matching algorithm for CIOQ crossbar switches, in *Proceedings of ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS)*, Princeton, NJ (2005)
11. N. McKeown, A. Mekkittikul, V. Anantharam, J. Walrand, Achieving 100 percent throughput in an input-queued switch. IEEE Trans. Comm. **47**(8), 1260–1267 (1999)
12. J. Lou, X. Shen, Frame-based packet-mode scheduling for input-queued switches. IEEE Trans. Comput. **58**(7), 956–969 (2009)
13. Z.M. Shen, L.F. Qiao, Q.H. Chen, S.L. Shao, Design of switch fabric in satellite onboard IP switch based on a multi-priority variable-length packets scheduling. Acta Electronica Sinica **42**(10), 2045–2049 (2014)

# Research and Implementation of the DDR2-Based Shared Memory Switch Fabric for Onboard Switches

**Qian Chen, Lufeng Qiao, Qinghua Chen, Huansheng Shen, Pengze Lv and Le Yang**

**Abstract** A shared memory output queuing structure used in satellite onboard switches is presented in this paper, where DDR2 memory is chosen as the main storage resource. The DDR2 memory can improve the storage capacity and the ability to resist flow fluctuation. The whole shared memory switch fabric is realized with Verilog HDL, simulated with ModelSim SE 10.0b and occupied 8921 LUTs and 77 Block RAMs in a Xilinx xc4vsx55 FPGA, which indicates that the key resources consumption can meet the requirements of triple modular redundancy. When the data is 64 bit wide and the system clock is 100 MHz, the peak throughput of the switch fabric can reach 6.4 Gbps.

**Keywords** Onboard switches · Shared memory switch fabric · DDR2 · Throughput

## 1 Introduction

Satellite Onboard IP Switches is the key node equipment of satellites-based communication networks [1]. The application environment of onboard switches have a notable difference with terrestrial equipments. The power consumption, performance of available components and hardware scale of onboard switches are all quite limited, which lies higher demands of algorithm complexity, hardware resource occupation and working frequency on onboard switches [2, 3].

Considering the strictly hardware resources limitation, the Shared Memory(SM) switch fabric, which has higher utilization of storage resources, lower packet loss rate and higher throughput compared to other switch fabrics, is widely used in the design of onboard switches [4].

Q. Chen · L. Qiao (✉) · Q. Chen · H. Shen · P. Lv · L. Yang
Institute of Communication Engineering, PLA University Science and Technology, Nanjing, China
e-mail: 13357837783@189.cn

Q. Chen
e-mail: njice-qlf@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_6

The shared memory switch fabric of traditional onboard switches mainly choose SRAM as the shared memory which is simple to realize. Using the on-chip SRAM as shared memory can greatly improve the access speed. However, the resources of on-chip SRAM are limited, which lead to poor ability to resist flow fluctuation. As the increasing switch capacity of onboard switches, on-chip SRAM can't meet the design requirements obviously. Off-chip SRAM has greater storage capacity than on-chip SRAM and is commonly used as shared memory in current onboard switches. In this case, off-chip SRAM don't use the hardware resources of FPGA. However, the off-chip SRAMs access speed is slow, which will decrease the throughput of the shared memory switch fabric. To meet the requirements of high speed onboard IP switches, memories with larger storage capacity and higher access speed are needed.

At present, DDR2 SDRAM can be used in satellite onboard switches. The DDR2 SDRAM has higher storage density, double rate of reading and writing, higher operating frequency, lower power consumption and so on [5]. It can be used to improve the key performance of onboard switches.

## 2 The Function and Architecture of Switch Circuit

In this paper, Microns DDR2 MT4HTF3264HY-667D3 is used as the shared storage medium. The storage capacity of the DIMM module is 256 MB with a data bus width of 64.

Figure 1 shows the circuit structure of the $4 \times 4$ DDR2-based shared memory switch fabric. The input interface circuit is used to multiplex the four ports' input data and store the data into local memory. The input interface module sends request to the write preprocessing module to transmit the local stored cells to the corresponding output queues. Then write preprocessing module request a free pointer, which points to a certain memory space in the DDR2, from the free pointer management circuit. The free pointer management circuit acquires the free pointers from DDR through the multi-user interface manager and sends them to the write preprocessing module. After that, write preprocessing module reads cells from input interface circuit and sends write request to multi-user interface manager. After the request is responded, the cells are written into DDR2 and associated pointers are added to corresponding output logic queues.

The Queue Controllers (QCs) send the read requests to multi-user interface manager through queue interface manager. The multi-user interface manager issues a response to one of requests, according to the rules of polling. The logic queue who acquire the response sends the associated pointer to multi-user interface manager. The multi-user interface manager reads out the cell from DDR2 and sends it to queue interface manager. At last, the queue interface manager sends the cell to output interface.

In this structure, the Queue Controller (QC) is used to manage corresponding output port's pointer logic queues in the form of link list. The latest arrival pointer is added to the tail of the link list. If a multicast cell arrives, its corresponding pointer has to be copied and written into the multiple output ports' QC link lists. The corre-

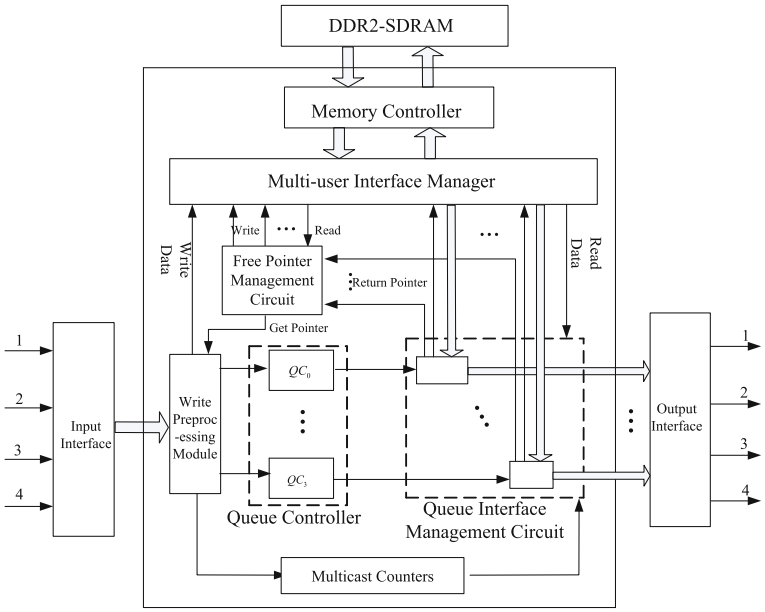



**Fig. 1** Structure of the DDR2-based SM switch circuit

sponding amount of output ports is written into the Multicast Counters (MCs), which is related to the port map of the pointer. If one cell is transmitted through the output interface, the corresponding MC value decreases by one. The pointer cannot be returned to the free pointer manager, unless the corresponding MC value becomes zero.

## *2.1 Design of QC Module*

If the QC logic queues were stored in the DDR2, it was necessary to maintain every user's link list in the way of burst write and burst read. In this way, large amounts of read/write bandwidth and memory space of DDR2 was occupied and wasted, which may influence the performance of the switch fabric. Therefore, in this paper, the pointers' logic queues are stored in the on-chip SRAM.

Figure 2 shows the structure of the QCs, one QC for a port. Every QC includes eight priority-based logic queues and these logic queues are stored in one SRAM. Each logic queue has a head register and a tail register. The head register is used to store the head pointer of the logic queue. The tail register is used to store the tail pointer of the logic queue. When a new pointer arrives, it will be the new tail pointer and will be written into the place where previous tail pointer pointed to. When the previous head pointer is read out, the next pointer of the logic queue will be the new head pointer.

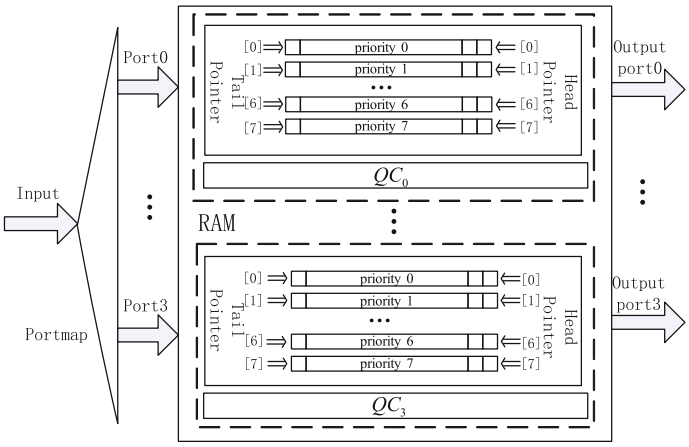



**Fig. 2** Structure of QCs

## *2.2 Design of Free Pointer Management Circuit*

In this switch fabric, all the free pointers are stored in the DDR2 SDRAM. This paper designed the free pointer management circuit, aimed at improving the management efficiency of free pointers. This circuit is used both for the initialization of the free pointers queue and acquiring/returning free pointers during the switching process. Figure 3 shows the structure of free pointer management circuit.

The circuit starts to initialize the queue of free pointers after the DDR2 completed the initialization operation. The DDR data memory are divided into 64-byte blocks, and these pointers are written into the free pointer queue through initialization operation. In addition, the free pointers buffer for each QC is built by the circuit. Every free pointers buffer includes the returned pointer buffer and the acquired pointer buffer. Each buffer is large enough to accommodate two DDR2 burst write or burst read.

The threshold is set to prevent the overflow of the return pointer buffer. When the number of free pointers that stored in this buffer is greater than threshold, a write request is sent to multi-user interface manager. On receiving the response, a certain number of pointers are returned to the queue of free pointers in a DDR2 burst write operation. In addition, a threshold is set to prevent the read error caused by the empty of acquired pointer buffer. When the number of free pointers that stored into this buffer is less than the threshold, a read request is sent to multi-user interface manager. On receiving the response, a certain number of pointers are read out from the DDR2 in the form of DDR2 burst read.

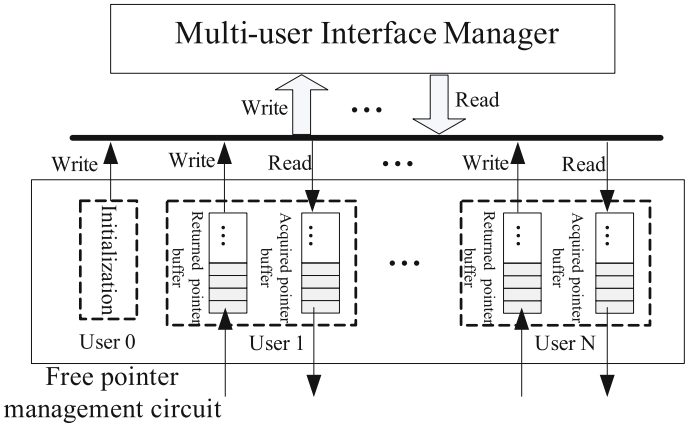



**Fig. 3** Structure of free pointer management circuit

## *2.3 Design of the DDR2-Based Multi-user Interface Manager*

In this switch fabric we designed a multi-user storage access interface manager to process write or read requests from different circuits. In order to manage the multi-user read/write requests fairly, a round robin polling scheme is used.

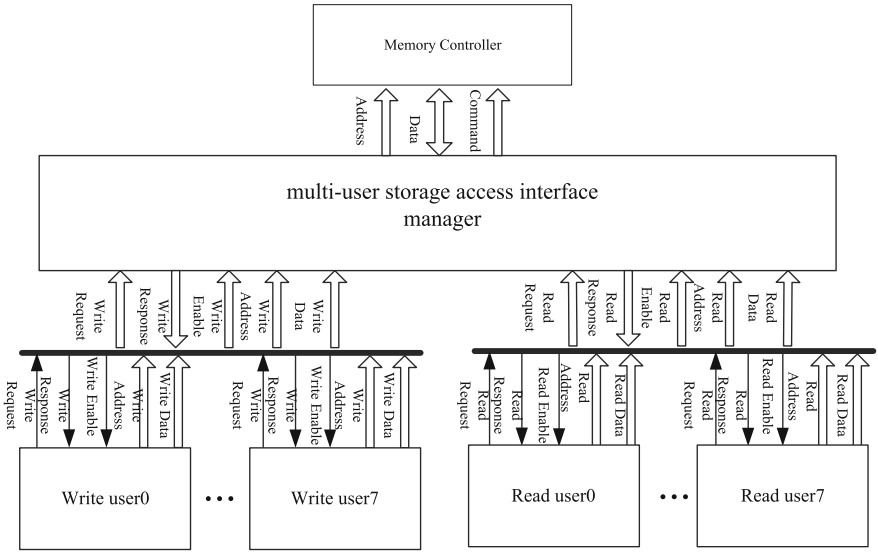



**Fig. 4** Structure of DDR2-based multi-user interface manager

In addition, free pointers are stored in the DDR. The initialization and read/write operations of free pointer queue in the DDR2 need to be completed by the multi-user storage access interface manager. Figure 4 shows the structure of DDR2-based multi-user interface manager.

In this figure, eight read users and eight write users are showed. The basic functions of the circuit include dealing with the write/read requests according to a certain polling rule, writing user data or reading corresponding data through burst write or burst read operations. In addition, the memory controller is used to manage the DDR2 [6].

## 3 The Storage Resources Allocation of SM Switch Fabric

During the design of SM switch fabric, the allocation of memory resources is just as important as the improvement of throughput. In the case of the high flow fluctuation and uneven distribution of the data flow, some users may gain more data buffer than others. Therefore, in order to guarantee the QOS of every user, the design and research of the resources allocation mechanism of SM switch fabric are important.

### *3.1 Allocation Mechanism of Private and Shared Memory Resources*

The free pointer queue consists of several private free pointer queues and one shared free pointer queue. The number of private free pointer queues is the same as the QCs.

In this paper, a part of free pointers are reserved for each QC, which constitute the private free pointer queues. The rest of free pointers are stored in the shared free pointer queue. These free pointer queues are worked in an first in first out manner. Figure 5 shows the allocation mechanism of the storage resources.

In Fig. 5, four private pointer queues corresponding to four QCs are shown. If the private pointers for a QC are all occupied, it can use the shared pointers from the shared free pointer queue. For instance, the total quantity of the free pointers is 16 k, each QC gains 1 k private free pointers and the rest of the free pointers are shared. In this mechanism, multicast cells can only use the shared free pointers.

Base on this allocation mechanism, each port can gain enough storage resources that guarantee the fairness of the switch fabric and ensure the ability to resist flow fluctuation.

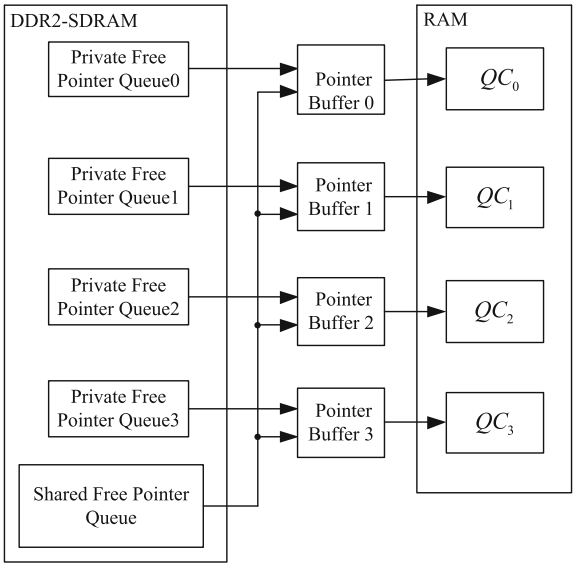



**Fig. 5** Allocation mechanism of the storage resources

## 4 Simulationed Results and Analysis

The design is realized with Verilog HDL and implemented in Xilinx xc4vsx55 FPGA. The integrated development environment of ISE Design Suite 14.7 is used and the whole design is simulated with Modelsim SE 10.0a.

### *4.1 Simulated Result of Multi-user Write Process*

The DDR2-based multi-user interface manager uses a 16 bits called sel-local to control the scheduling of the write or read operations. The lower eight bits represent the write requests of eight users. The higher eight bits represent the read requests of eight users. Figure 6 shows the simulated result of the multi-user write process.

The wr-req is a 8 bits signal of write requests. When one of eight users sends the write request to the multi-user interface manager, the corresponding bit of the wr-req will be set. As shown in Fig. 6, the multi-user interface manager sends the wr-ack to the corresponding user, acquires the user data wr-rd-din and receives the address information presented on wr-addr. In this simulation, user1, user2, user4, user5 and user7 carry out the write operation.

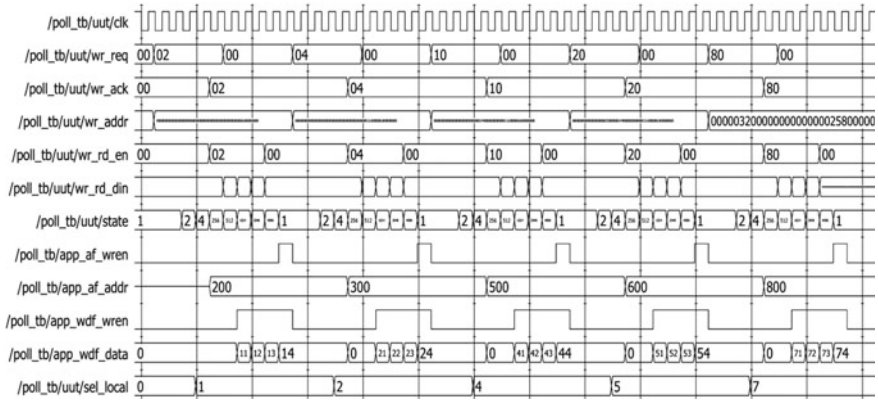



**Fig. 6** Simulated waveform of the multi-user write process

| Device Utilization Summary (estimated values) | | | |
|---|---|---|---|
| **Logic Utilization** | **Used** | **Available** | **Utilization** |
| Number of Slices | 5765 | 24576 | 23% |
| Number of Slice Flip Flops | 7744 | 49152 | 15% |
| Number of 4 input LUTs | 8921 | 49152 | 18% |
| Number of bonded IOBs | 394 | 640 | 61% |
| Number of FIFO16/RAMB16s | 77 | 320 | 24% |
| Number of GCLKs | 6 | 32 | 18% |
| Number of BUFIOs | 8 | 44 | 18% |

**Fig. 7** The comprehensive report of system

## *4.2 Analysis of the Synthesized Results*

Figure 7 shows the synthesis report of the design. The whole switch fabric uses 8921 LUTs that occupies 18% of total resource and 77 Block RAMs that occupies 24% of total resource in a Xilinx xc4vsx55 FPGA, which indicates that the key resources consumption can meet the requirements of triple modular redundancy [7].

With the 64 bit data width and 100 MHz system clock, the throughput of SM switch fabric, being proposed in literature [8], is 1.6 Gbps. This paper choose DDR2 as the shared memory, under the same premise, the peak throughput of switch fabric mentioned in this paper is 6.4 Gbps, which is three times higher than that using the off-chip SRAM. Apparently, in this way, the throughput of the switch fabric can be improved immensely.

## 5 Conclusion

A DDR2-based SM switch fabric is presented in this paper, which is proposed for satellite onboard switches. Then the function of each key module is introduced. The structure of QC, which is used to manage the local link list to build logical output queues, is given in detail. The multi-user interface manager can treat the write and read request from different users in a round robin mode. To save the hardware resource used in FPGA and allocate the DDR2 storage resource reasonably, a private buffer and shared buffer allocation scheme is used. The whole design is realized with Verilog HDL and simulated with ModelSim SE 10.0b, the switch fabric is implemented in a Xilinx xc4vsx55 FPGA, 8921 LUTs and 77 Block RAMs are used, which can meet the requirements of triple modular redundancy, a common method used in satellite onboard equipments to improve reliability. With a 64 bit data width and 100 MHz clock, the peak throughput of the switch fabric can reach 6.4 Gbps.

## References

1. A. Iera, A. Molinaro, S. Marano, Call admission control and resource management issues for real-time VBR traffic in ATM-satellite networks. IEEE J. Sel. Areas Commun. **18**(11), 2393–2403 (2000)
2. K.-C. Yi, L. Yi, S. Chen-hua, N. Chun-guo, Recent development and its prospect of satellite communication. J. Commun. **36**(6), 1–16 (2015)
3. X. Peng, Z. Tao, L. Feng, Integrated onboard route switching system based on wireless access. Comput. Eng. **34**(12), 277–279 (2008)
4. A.G. Wassal, M.A. Hasan, A VLSI switch architecture for broadband satellite networks. Circuits Syst. (1998)
5. L. Guan-nan, O. Ming-shuang, S. He-juan, The design of DDR2 SDRAM controller and implementation in FPGA. Chin. lntegr. Circuit (131), 42–45, 71 (2010)
6. S.Y. Chen, D.H. Wang, S. Rui, et al., An innovative design of the DDR/DDR2 SDRAM compatible controller, in *IEEE. ASICON International Conference* (2005), pp. 62–66
7. M.-F. Yang, S.-L. Guo, Z.-Q. Sun, Onboard computer techniques for spacecraft control. Areosp. Control **23**(2), 47–53 (2005)
8. S. Ze-min, Q. Lu-feng, C. Qing-hua, S. Shi-lei, Design of switch fabric in satellite onboard IP switch based on a multi-priority variable-length packets scheduling. Acta Electron. Sinica. **42**(10), 2045–2049 (2014)

# Design and Implementation of Credit-Based Dynamic WRR Scheduler For Satellite Onboard Switches

**Pengze Lv, Lufeng Qiao, Qinghua Chen, Qian Chen and Le Yang**

**Abstract** Considering the hardware resource limitation in satellite onboard switches, a Credit-based Dynamic WRR (CDWRR) Scheduler is presented in this paper. The scheduler can improve the throughput under nonuniform traffics and afford good support for multicast traffics. It can also guarantee the service priority and ensure the fairness concurrently. With the 128 bit data width and 100 MHz system clock, the peak throughput of switch fabric mentioned in this paper is 12.8 Gbps, which can meet the requirement of the 10 Gbps input port for the satellite onboard switches. The CDWRR scheduler is used in a Combined Input Output Queued (CIOQ) switch structure and realized with Verilog HDL. The scheduler occupies 2251 LUTs and 3 Block RAMs in a Xilinx xc6vlx130t FPGA, which indicates that the key resources consumption can meet the requirements of triple modular redundancy. Typical simulated results show that the scheduler works correctly.

**Keywords** Satellite onboard switch · Scheduling algorithm · Credit · FPGA

## 1 Introduction

As the key node of the satellite communication networks, the satellite onboard IP switches are mainly used to forwarding IP packets between different ports [1]. The bandwidth of terrestrial network keeps growing greatly to meet the requirements of multi-medium services. Meanwhile, the bandwidth of satellite networks grows correspondingly which makes it important to realize high-speed satellite onboard IP switches [2, 3].

An effective switch fabric is crucial to improve its throughput. The performance of a switch fabric depends on the buffering strategy and scheduling algorithm.

P. Lv · L. Qiao (✉) · Q. Chen · Q. Chen · L. Yang
Institute of Communication Engineering, PLA University Science and Technology, Nanjing, China
e-mail: 13357837783@189.cn

P. Lv
e-mail: njice-qlf@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_7

Buffering strategies are divided into three categories depend on its location: Output Queued (OQ), Input Queued (IQ) and Combined Input Output Queued (CIOQ). The output queuing structure benefits from better QoS control ability, as the destining cells are always available at the output ports to be scheduled based on their priority levels, allocated bandwidth, and QoS requirements. However, this structure suffers from memory access bandwidth constraint and thus switch size limitation. This is because there can be up to N cells from different input ports destined for the same output port at the same cell time. Due to the switch size limitation of the OQ structure, the IQ structure that works at link speed has attracted more attention [4]. The problem of IQ is the so called head-of-line (HOL) blocking. HOL blocking can be solved through the mechanism of Virtual Output Queuing (VOQ). Another way to alleviate HOL blocking is to increase the speedup factor S, whose value is between 1 and N. It has been proved that with a speedup factor of 2, the CIOQ structure can reach the performance of OQ structure, no matter the number of the ports and traffic distribution [5].

The traffic arriving at CIOQ switch needs to be buffered at the input ports and output ports after forwarding by the switch fabric. The input queue structure is more complex than the output queue structure in the CIOQ scheme. So the selection of input scheduling algorithm is important to meet the QoS and throughput requirements of the CIOQ structure, which is most concerned in this paper.

The most commonly used scheduling algorithm for input queuing structure is Maximal Size Matching (MSM) and Maximum Weight Matching (MWM). MSM has lower complexity and has been widely used. There are some typical MSM algorithms such as PIM, iSLIP, FIRM, SRR and so on [6]. Presently, iSLIP is widely used, and simulated results show that it can achieve 100% throughput under uniform traffics. As for nonuniform traffics, the throughput of iSLIP will decrease greatly.

MWM scheduling algorithm has better performance than MSM algorithm when processing nonuniform traffics [7]. Longest queue first (LQF) is a typical algorithm of MWM and the length of a queue is also used as the weight of the queue. LQF can achieve 100% throughput under both uniform and nonuniform traffics. But some input port may be starved in some case. If a queue has only one cell and there are no other cells after that, and the other queues have at least two cells and new cells arrive continuously, then the first queue will be starved.

Based on the analysis mentioned above, this paper presents the CDWRR algorithm to solve the problem involved in LQF algorithm. It allocates weights based on the priority of the queue and accumulates credits for service. The phenomenon of starvation can be avoided. Just like LQF algorithm, CDWRR also can achieve 100% throughput under nonuniform traffic. Moreover, the CDWRR algorithm can afford good support for multicast traffic.

## 2 The Structure and Scheduling Schemes of the Scheduler

The CDWRR Scheduler implements three levels of scheduling. Figure 1 shows the circuit structure of the 16 * 16 CIOQ switch. The switch includes the input queue managers, the switch fabric and the output queue managers. One input queue manager for an input port. Similarly, there is one output queue manager for an output port. The input queue manager includes the input queues and the local scheduler. The local scheduler is used to implement the first level scheduling and the third level scheduling. The central scheduler located in the switch fabric is used to implement the second level scheduling.

### *2.1 First Level Scheduling: CRQ Generation*

Figure 2 shows the structure of the first level scheduling. As shown in Fig. 2, there are 16 VOQs and one VOQ for an output port and can solve the problem of HOL blocking. In addition, each VOQ consists of a unicast VOQ and a multicast VOQ.

In this level, the local scheduler conducts the scheduling based on the CDWRR algorithm. For each scheduling cycle, the scheduler can generate connection requests (CRQs) from the input queues to the switch fabric for frames that wait to access.

The unicast traffic is divided into two categories, high priority (HP) and low priority (LP). Class7 is HP, while Class0 to Class6 are LP. The schedulers always schedule HP traffic before LP traffic, so the HP class is well suited for traffic with very stringent latency and jitter requirements.

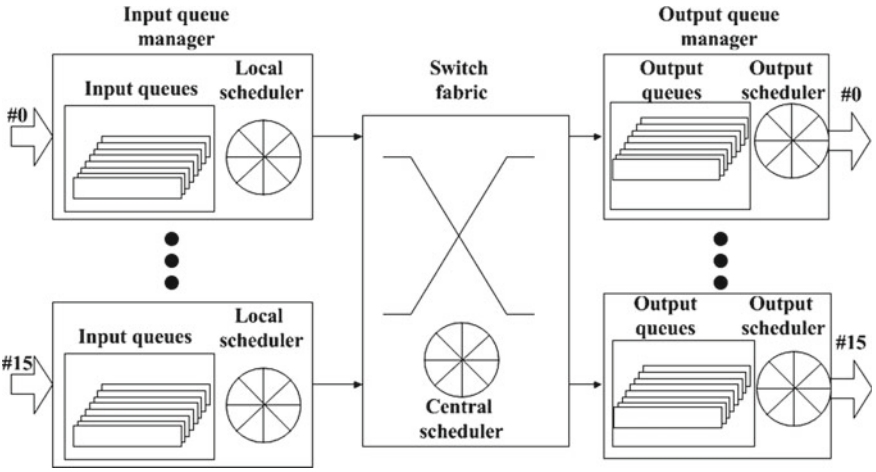



**Fig. 1** Structure of CIOQ switch

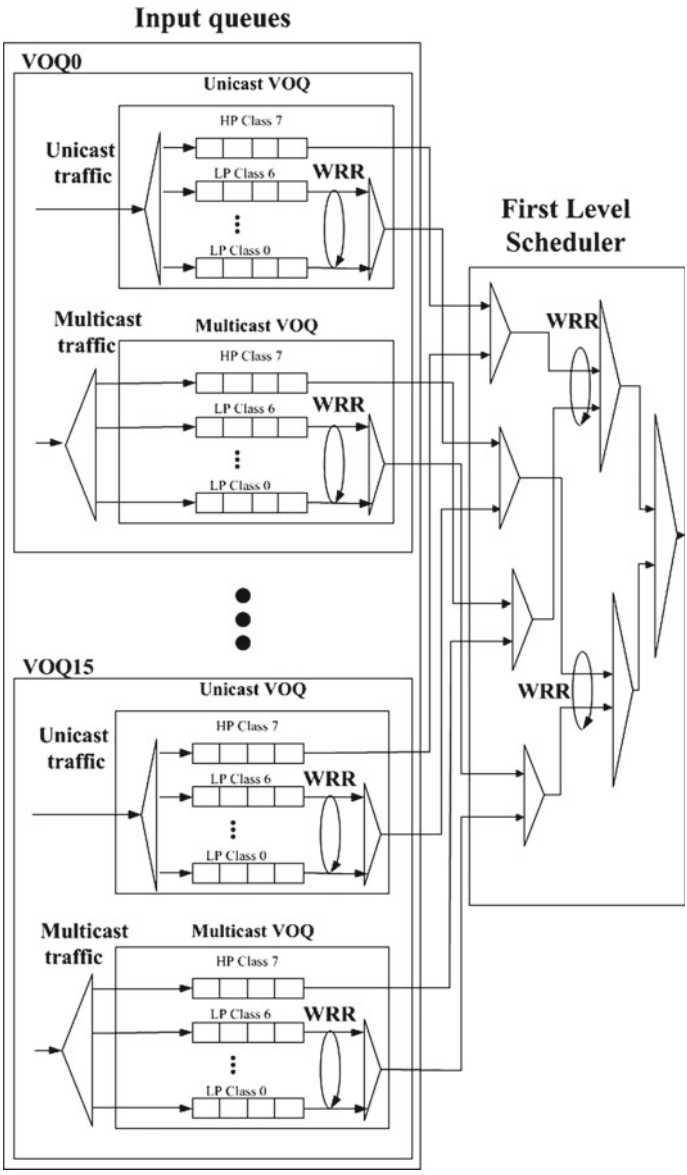



**Fig. 2** Structure of the first level scheduling

The queue structure for the multicast traffics is identical to the unicast one. Like the unicast VOQs, two categories are defined for multicast VOQs, and Class7 is HP, while class0 to Class6 are LP. The scheduling of multicast VOQs is the same as the unicast VOQs.

For both HP and LP traffics, multicast traffics share the bandwidth with unicast traffics according to a user-programmable weight, which can be set between 0 and 31. Level 0 configures unicast traffics with a strictly higher priority over multicast traffics. Level 31 configures the device so that the multicast traffics have a strictly higher priority than the unicast traffics.

When there is a HP traffic in a VOQ, the corresponding request bit will be asserted. If there is a LP traffic in a VOQ, the local scheduler will generate connect request based on the CDWRR algorithm. As shown in Fig. 2, there are sixteen VOQs totally, so there will be 16 LP-CRQs and 16 HP-CRQs when a schedule completed. The key variables of CDWRR scheduling algorithm are as follows.

Cost:
A global variable which represents a threshold. When the credit of a queue reaches or exceeds the cost, it can be served. The cost is configured dynamically at every cell slot. The cost of each VOQ is equal to the maximum weight of the nonempty queues. Then the global cost is set to the maximum cost of all the VOQs.

Credit:
Each VOQ has its own dynamic credit. The credit will accumulate at every cell slot when it cannoot reach the global cost. The accumulated credit of each slot is the maximum weight of the queues which have data in the corresponding VOQ. When the credit exceeds the global cost, a connection request will be sent and the credit will be reduced by the amount of the global cost.

Weight:
It is the static value configured to LP queues and is used to allocate bandwidth among different queues. Every LP traffic in a VOQ has a programmable weight and this weight is suitable for the unicast and the multicast traffics. The queue with higher priority will be configured with a bigger weight. These weights are used to decide the cost of VOQs and the credits added after a scheduling cycle.

## *2.2 Second Level Scheduling: ACK Generation*

The second level scheduling generates a request acknowledgement (ACK) from the switch fabric after a matching iteration. Each ACK gives an associated local scheduler the permission to send a cell to one specific switch output.

In this level, the central schedulers check all the CRQs coming from all the local schedulers and goes through an iterative matching process to send grants in an equitable and timely manner. The iterative matching process will continue until the next

cell comes. After four iterations for HP-CRQs, the rest of iterations are left for LP-CRQs. This mechanism of iteration will allocate available bandwidth for HP traffics prior to LP traffics.

As for the central scheduler, the iterative process of matching inputs to outputs determines the throughput of the whole switch, and different iterative matching algorithm has been well-studied. Among all the algorithms, the maximum achievable throughput depends mainly on the number of iterations during the matching process.

## 2.3 *Third Level Scheduling: Queue Selection*

Figure 3 shows the structure of the third level scheduling. When the local scheduler receives an ACK from the central scheduler, it must find the most suitable frame to send from its corresponding VOQ. The third level scheduling selects the cell to be sent using an algorithm similar to the algorithm of the first level.

In this level, the HP traffic will be served first. If the VOQ has no HP traffics, the scheduler will find out the most suitable cell based on the CDWRR algorithm from the LP traffics. The selected cell will be sent to the switch fabric.

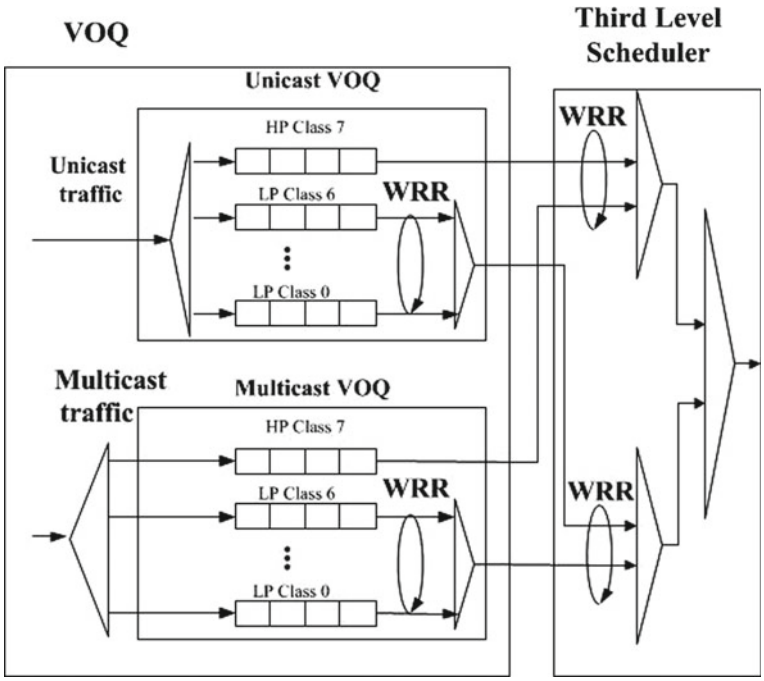



**Fig. 3** Structure of the third level scheduling

The scheduler will send an idle cell carrying current CRQ information if the cells in the VOQ granted have been transmitted during the previous time slots or no VOQ receives an ACK from the central scheduler.

A delay of a certain number of cell slots will be generated from the local scheduler generates connection requests to receive the ACK. During these periods, the credit of each queue may be accumulated or decreased.

The third level scheduling is a little different from the first level, and the output cell is selected among the queues with different classes in the granted VOQ. Among this VOQ, the cost is the maximum weight of the nonempty queues. The credit will be adjusted for the queue served. If there are no queues have enough credits, the nonempty queue with the biggest weight will be served. Concurrently, its credit will be cleared.

## 3 Simulated Results and Analysis

The design of CDWRR scheduler is realized with Verilog HDL and implemented in Xilinx xc6vlx130t FPGA. The integrated development environment of ISE Design Suite 14.7 is used and the whole design is simulated with ModelSim SE 10.2c.The simulated result of some key circuit are shown as follows.

### *3.1 Simulated Result of the First Level Scheduler*

Figure 4 shows the simulated waveform of first level scheduling for unicast traffics, which is the same for multicast traffics. In Fig. 4, VOQ0, VOQ5 and VOQ15 are nonempty. The signal of i-req0 is 128 bit wide and its used to represent the requests of the unicast queues from 16 VOQs with 8 classes. The h-crq signal represents the scheduled results of the high priority queues and l-crq signal represents the scheduled results of the low priority queues. The initial value of credits is 0. The credit value of each VOQ is accumulated and the global cost changes dynamically during the scheduling process. At last, the credit of VOQ15 meets the requirements and acquires the chance to send a request. Meanwhile, its credit value is decreased. As shown in Fig. 4, the scheduling process occupies three clocks.

### *3.2 The Simulated Result of the Third Level Scheduler*

Figure 5 shows the simulated waveform of the third level scheduling. The third level scheduling is used to select a cell to send among the granted VOQ. The voq-num signal represents the granted VOQ and the voq-req signal represents the requests from queues of different classes. As shown in Fig. 5, queues of Class0, Class1, Class2,

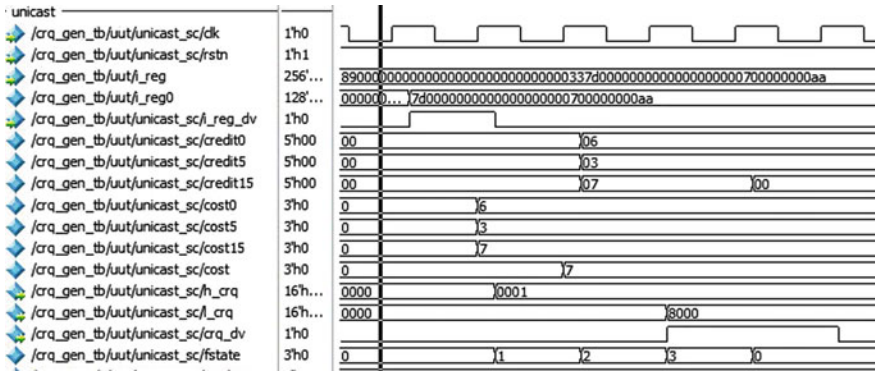



**Fig. 4** Simulated waveform of first level scheduling for unicast

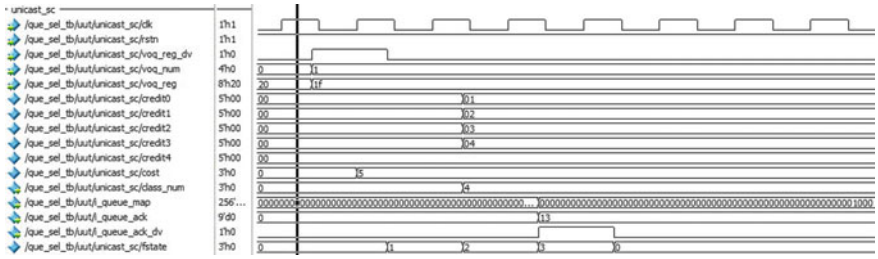



**Fig. 5** Simulated waveform of third level scheduling

Class3 and Class4 have cells to send. The initial credits of these queues are all zero and the queue of Class4 acquires the chance to send a cell finally. When the scheduling completed, the queue number and the map information of the queue are all returned by the i-queue-ack signal and the i-queue-map signal. Figure 5 shows the simulated results of the third level scheduling, which occupies three clocks.

## *3.3 Performance Comparison*

Table 1 shows the synthesized results of the design. The whole scheduler occupies 2167 Slices, which is 12% of the total slice resources in a Xilinx xc6vlx130t FPGA, and 3 Block RAMs are used, which is only 1% of total block RAM resources.

With the 128 bit data width and 100 MHz system clock, the peak throughput of switch fabric mentioned in this paper is 12.8 Gbps, which can meet the requirement of the 10 Gbps input port for the satellite onboard switches. For a 10 Gbps port, the key resources consumption are relatively low and apt to meet the requirements of triple modular redundancy.

**Table 1** Hardware resource requirements

| | iCDWRR Scheduler |
|---|---|
| Logic unit | Used/Utilization (%) |
| Slices | 2167 /12 |
| LUTs | 2251 /2 |
| Slices FFs | 1228 /21 |
| IOBs | 231/38 |
| Block RAM/FIFO | 3/1 |

## 4 Conclusion

A Credit-based Dynamic WRR (CDWRR) scheduler is presented in this paper, which is proposed for satellite onboard switches. While ensuring the fairness of scheduling process, the scheduler improves the throughput under nonuniform traffics with this algorithm implemented in 16 * 16 switches. As the structure of the design is simple and scalable, the consumption of the hardware resource is reduced dramatically to satisfy the demand of outer space environment.

## References

1. A. Iera, A. Molinaro, S. Marano, Call admission control and resource management issues for real-time VBR traffic in ATM-satellite networks. IEEE J. Sel. Areas Commun. **18**(11), 2393–2403 (2000)
2. X. Peng, Z. Tao, L. Feng, Integrated onboard route switching system based on wireless access. Comput. Eng. **34**(12), 277–279 (2008)
3. Y. Ke-chu, L. Yi, S. Chen-hua, N. Chun-guo, Recent development and its prospect of satellite communication. J. Commun. **36**(6), 1–16 (2015)
4. X. Qing-xu, Research on packet scheduling in input-queued switches. J. Commun. **26**(06), 118–129 (2005)
5. T. Ren, J.L. Lan, H.C. Hu, Practical parallel packet switch architecture based on CIOQ, **32**(05), 14–21 (2011)
6. B. Pand, S. He, W. GAO, A survey on input-queued scheduling algorithms in high-speed IP routers. J. Softw. **14**(05), 1011–1022 (2003)
7. X. Zhang, M. Hu, G. Shi, The analyse of scheduling algorithms for high-performance switch. Inf. Technol. **10**, 34–38 (2002)

# OpenFlow-Based Load Balancing in WLAN: Throughput Analysis

**Syed Mushhad M. Gilani, Heng Meng Heang, Tang Hong, Guofeng Zhao and Chuan Xu**

**Abstract** Software-Defined Wireless Network (SDWN) aims to build a flexible wireless network infrastructure that can support future Internet services. In this paper, we present WLAN architecture to take advantage of OpenFlow that provides the global view of the entire network including wireless network configuration, resource allocation, and flow control policies to make the load balanced network environment. We build simulation environment through Mininet-WiFi to analysis throughput and jitter values of associated stations. The results demonstrate that proposed architecture can divide the load between APs that increase the average throughput of associated stations.

**Keywords** OpenFlow · WLAN · Load balancing · Software-defined network

## 1 Introduction

IEEE802.11-based wireless network considers to provide high-speed Internet with robust throughput for extensive sort of devices. The capacity of WLAN users rapidly increased although the performance is reducing. The wireless network infrastructure is becoming gradually more complex, dispute and deficient with predefined existing standards, rules, and technologies. The most dominating issues of enterprise WLAN includes load-imbalance [1] between APs that produce transmission delay, minimized throughput, lack of resources, and low responsiveness.

Load balancing in IEEE802.11 wireless network initiated when different access points coverage area is overlapped. In this kind of scenario, at least, two access points available for association with the user. Thus, to make a fair selection of access point apply the load balancing technique that calculates the network load and distributes the load equally using predefined rules or algorithms.

S.M.M. Gilani (✉) · H.M. Heang · T. Hong · G. Zhao · C. Xu
Research Center of Future Internet (RFI), Chongqing University of Posts and Telecommunications, Chongqing 400065, China
e-mail: mushhad@gmail.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_8

Software-Defined Wireless Network (SDWN) [2] technology has been obtaining a significant interest for separation of control plane and data plane. It has been applied to various domains of a wireless network with tremendous flexibility, dynamic scheduling, fine-grained packet control and global view of the entire network.

We propose a system model for WLAN within load balancing environment, based on SDN paradigm. The main objective of this research work is to construct OpenFlow-based simulation topology for testing and analysis of TCP throughput and UDP throughput in WLAN. We focus on improving the performance of the wireless network regarding transmission throughput which affected due to load-imbalance between APs.

## 2 Related Work

As mentioned earlier, SDWN spreads into depth level of the wireless networks such as on-demand resource allocation with resource optimization, simultaneous support for heterogeneous wireless networks, and execution of open APIs. Moreover, SDWN is helpful for multiple entities of a wireless network including ISP providers, network operators, and consumers [2]. However, SDWN is not bounded in only wireless network dimensions. The other dimension of SDWN is mobile network covering radio technology that targets the base station [3]. The multidimensional nature of SDWN makes it more complex then SDN-wired networking approach. OpenRoads [4] is the pilot project about SDWN that enables user mobility between WiFi and WiMax. NOX [5] controller deploy to control network devices, and FlowVisor [6] makes virtualization with the isolation of traffic paths. However, OpenRoads limited to mobility in WLAN environment without awareness of load balancing.

Recently, researchers have been introduced some load balancing schemes in the wired SDN environment. Such as [7], in which authors proposed an architecture for mobility management with an extensive feature of load balancing between switches to reduce the packet delay. Another [8] load balancing technique introduced for OpenFlow-enabled switches which installed in data center networks. In comparison to schemes mentioned above our proposed scheme specifically, targets the load balancing between APs.

## 3 System Model

The proposed load balancing system model consists of a centralized controller installed with load balancing applications, OpenFlow-enabled switches, and APs. OpenFlow-based controller POX [9] is to provide a global view of entire wireless access network regarding packet control, traffic flow management, end-terminal

association, and load balancing. OpenFlow/SDN enables rapid development and deployment of innovative applications. We deploy three applications on the top of the controller as mentioned in Fig. 1.

Wired Ethernet switches configured with OpenFlow table to keep records of packet flow statistics includes packet forwarding rules and flow entries with switch port numbers. OpenFlow-enabled AP (OFAP) received instructions from controller regarding packet forwarding, packet discarding, packet re-transmitting, and packet broadcasting policies. The proposed system consists OFAP with Load calculation agent (LCA) as depicted in Fig. 2. The centralized controller forwards the station probe request to LCA that creates load calculation virtual AP (LCVAP) to provide a dedicated virtual connection for an individual station.

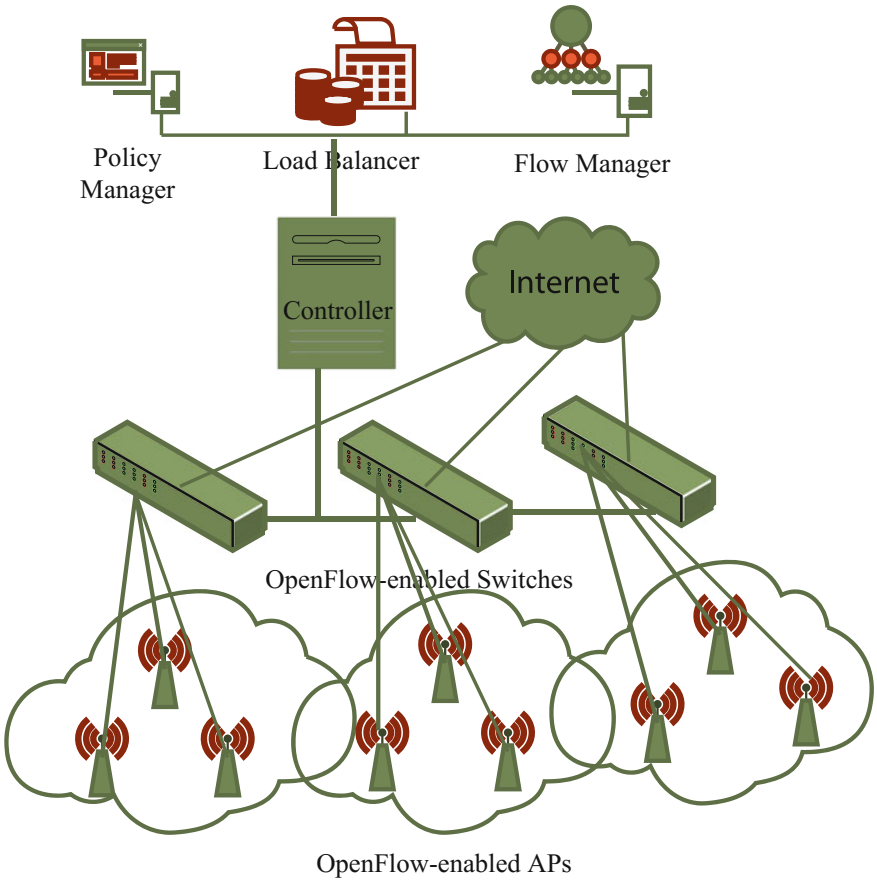



**Fig. 1** Overall system model

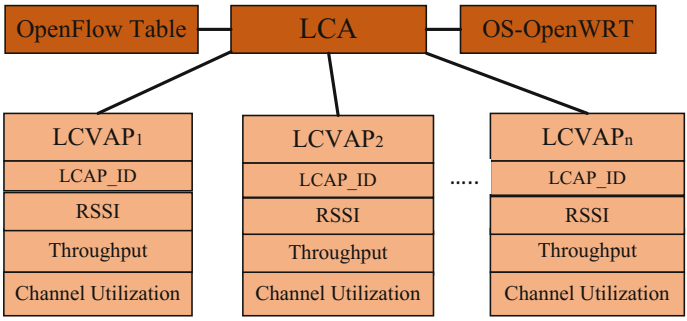



**Fig. 2** OFAP structure

## 4 Simulation Test

We evaluate the SDN-based load balancing mechanism within WLAN through Mininet-WiFi [10] simulation environment as illustrated in Fig. 3. We consider ten stations which are associated with three different APs in this simulation topology. In this experiment, we deployed APs in overlapping area to analyze the traffic load.

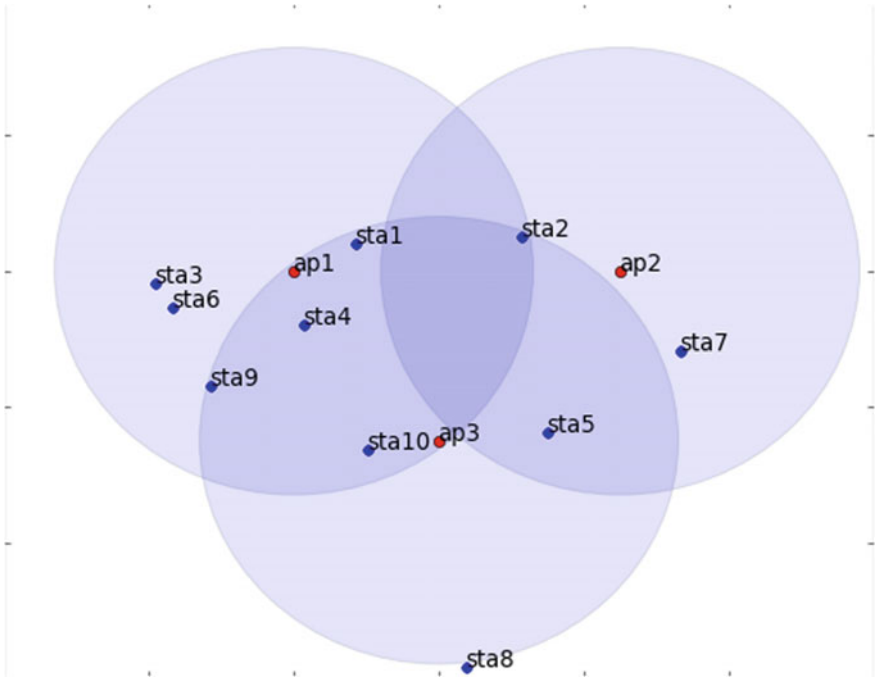



**Fig. 3** Simulation environment to analyze transmission throughput

**Table 1** Simulation configuration parameters

| Controller | Wireless standard | Channel | Data rate | Data range | Tx-power | Transmission type |
|---|---|---|---|---|---|---|
| POX | 802.11 g | Channel 6 (20 MHz) | 54.0 Mbit/s | 100 m | 20 dBm | TCP and UDP |

After creating a custom topology, two stations are taken as HTTP server and the remaining stations are performing as HTTP client. According to the traffic load, the server schedules the clients. The centralized controller which configures with load balancer defined the policy and rewrites the destination address of incoming packets for forwarding towards less loaded APs. Ipref uses to measure the throughput between stations and server. Moreover, it also permits to perform various tests that enabled insight view of the current network performance with packet loss ratio, delay, and jitter (Table 1).

In our simulation topology, we collected the distance value of each station with associated AP and related-received signal strength indicator (RSSI).

## *4.1 TCP Throughput*

Initially, TCP Server configured on Sta2 with port 5566 at default TCP window size 58.3 Kbyte. TCP Clients on Sta1, Sta3, Sta4… Sta10, associated with sta2 for sending TCP traffic at different transfer rate through various APs. The association interval time is set to 15 s with various transfer rates as mentioned in Table 3. The first test performed using the traditional approach in which each station association is based only on RSSI that creating an imbalance between APs and effect the bandwidth. Figure 4a represents two stations sta1 and sta3 that associated with

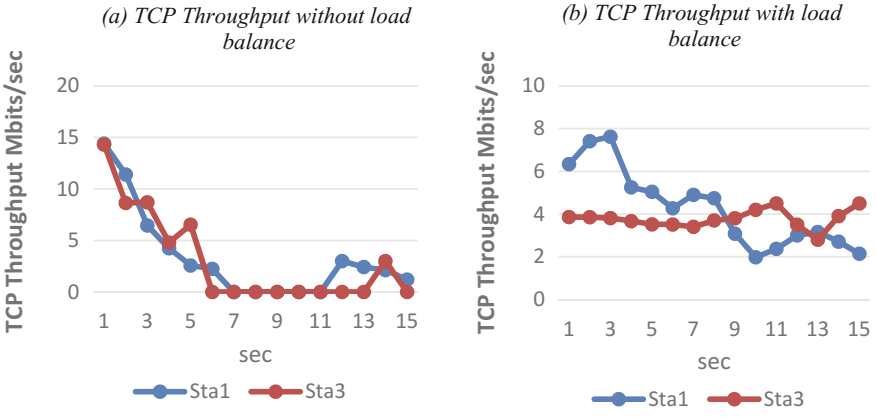



**Fig. 4** TCP throughput analysis between sta1 and sta3

**Table 2** Comparison of TCP throughput during access to server sta2

| | | Traditional approach | | OF-based approach | |
|---|---|---|---|---|---|
| Station | Time interval (s) | Transfer (Mb) | Bandwidth (Mb/s) | Transfer (Mb) | Bandwidth (Mb/s) |
| Sta1 => Sta2 | 0–15 | 2.38 | 1.23 | 8.00 | 4.32 |
| Sta3 => Sta2 | 0–15 | 2.10 | 1.11 | 8.25 | 4.51 |
| Sta4 => Sta2 | 0–15 | 1.82 | 0.8 | 8.30 | 4.6 |
| Sta5 => Sta2 | 0–15 | 1.09 | 0.7 | 8.00 | 4.33 |
| Sta6 => Sta2 | 0–15 | 3.02 | 1.85 | 7.92 | 4.1 |
| Sta7 => Sta2 | 0–15 | 2.08 | 1.02 | 8.25 | 4.51 |
| Sta8 => Sta2 | 0–15 | 3.00 | 2.01 | 9.00 | 5.2 |
| Sta9 => Sta2 | 0–15 | 0.79 | 0.52 | 8.30 | 4.6 |
| Sta10 =>Sta2 | 0–15 | 0.98 | 0.56 | 7.92 | 4.1 |

overloaded AP and in results throughput degradation started and reached to 0 Mbits/s.

The same test performed with OF-based approach in which stations association based on the load of individual OFAP. After adopting the proposed approach the throughput of sta1 and sta3 enhanced as illustrated in Fig. 4b. Meanwhile, others associated stations transfer rate and bandwidth also improved. Table 2 provides comparison results between the traditional approach and OF-based approach.

## *4.2 UDP Throughput*

The second experiment performed on UDP Server on Sta10 with port 5566 and monitored the results after every second. Start the UDP Clients on Sta1, Sta3… Sta10. We select two stations Sta7 and Sta9 to analyze their performance in a traditional environment and OF-based load-balancing environment. We measured the traffic using Ipref and collect samples of traffic throughput.

Figure 5a shows the results of sta7 and sta9 which are connected with server sta10 at 1.5 Mbits/s. Meanwhile, the associated AP received others stations association request due to strong RSSI and connected with them that make overloaded the AP. Due to an unbalanced network, the throughput of sta7 and sta9 gradually decreased and till reached to 0 Mbits/s.

The same experiment conducted in OF-based load-balancing approach that provides association on the basis of LCA that makes the overall balanced WLAN. Figure 5b shows the sta7 and sta9 performance which maintains better average throughput in comparison to traditionally based approach.

In this experiment, we also compare the traditional approach with OF based approach regarding congestion level of each station that reveals the OF-based WLAN is less congested. The results obtained from traditional and OF-based approach regarding jitter values are shown in Table 3.

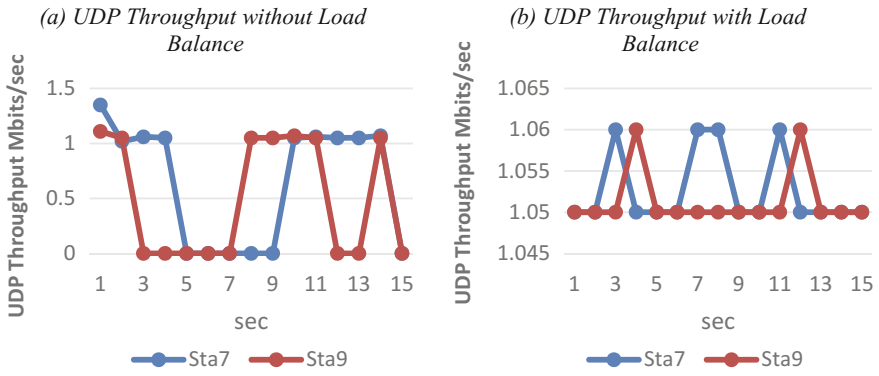



**Fig. 5** UDP throughput analysis between sta1 and sta3

**Table 3** Result of jitter values when all stations accessed to server Sta10

| Station | Interval (s) | Jitter (Traditional) (ms) | Jitter (OF-based approach) (ms) |
|---|---|---|---|
| Sta1 => Sta10 | 0–15 | 1.196 | 0.230 |
| Sta2 => Sta10 | 0–15 | 2.613 | 0.182 |
| Sta3 => Sta10 | 0–15 | 2.087 | 0.375 |
| Sta4 => Sta10 | 0–15 | 1.653 | 0.128 |
| Sta5 => Sta10 | 0–15 | 1.879 | 0.448 |
| Sta6 => Sta10 | 0–15 | 2.275 | 0.275 |
| Sta7 => Sta10 | 0–15 | 3.012 | 0.248 |
| Sta8 => Sta10 | 0–15 | 0.897 | 0.190 |
| Sta9 => Sta10 | 0–15 | 1.673 | 0.162 |

We collect statistics of each station and make an overall comparison in traditionally based approach and OF-based approach. The results demonstrated in Fig. 6 exposes that OF-based load balancing approach enhanced the individual station throughput.

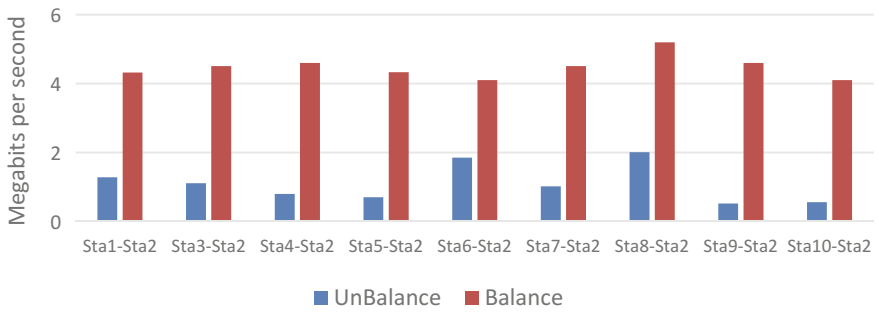



**Fig. 6** Comparative bandwidth analysis of associated stations

## 5 Conclusion

The objective of this research paper is to evaluate the throughput of associated mobile stations and provide a comparison between traditional and proposed environment. We conducted experiments through Mininet-WiFi to the analysis of stations through in load balance and load-imbalance environment. In future work, we shall deploy the proposed architecture with different applications to evaluate the performance of OFAP.

**Acknowledgements** This work was supported by Prospective Research Project on National Nature Science Foundation (61402065, 61501075), Future Networks (BY2013095-2-03) fund by Jiangsu Future Networks Innovation Institute and Project Foundation of Chongqing Municipal Education Committee (No. KJ1500429).

## References

1. A.J. Nicholson, Y. Chawathe, M.Y. Chen, B.D. Noble, D. Wetherall, Improved access point selection, in *Proceedings of 4th International Conference on Mobile Systems Applications and Services* (ACM, 2006), pp. 233–245
2. M. Yang, Y. Li, D. Jin, L. Zeng, X. Wu, A.V. Vasilakos, Software-defined and virtualized future mobile and wireless networks: a survey. Mob. Netw. Appl. (2014). doi:10.1007/s11036-014-0533-8
3. A. Gudipati, D. Perry, L.E. Li, S. Katti, SoftRAN: software defined radio access network (ACM, 2013), pp. 25–30
4. K.-K. Yap, M. Kobayashi, R. Sherwood, T.-Y. Huang, M. Chan, N. Handigol, N. McKeown, OpenRoads: empowering research in mobile networks. ACM SIGCOMM Comput. Commun. Rev. **40**, 125–126 (2010)
5. http://www.noxrepo.org/
6. R. Sherwood, G. Gibb, K.-K. Yap, G. Appenzeller, M. Casado, N. McKeown, G. Parulkar, *Flowvisor: A Network Virtualization Layer* (OpenFlow Switch Consort, Tech Rep, 2009)
7. A. Bradai, A. Benslimane, K.D. Singh, Dynamic anchor points selection for mobility management in software defined networks. J. Netw. Comput. Appl. **57**, 1–11 (2015). doi:10.1016/j.jnca.2015.06.018
8. R. Wang, et al. Openflow-based load balancing gone wild. Proc. Hotice (2011)
9. NOX OpenFlow controller, http://www.noxrepo.org/. Accessed March 2016
10. R.R. Fontes, S. Afzal, S.H. Brito, M.A. Santos, C.E. Rothenberg, Mininet-WiFi: emulating software-defined wireless networks, in *Network and Service Management CNSM 2015 11th International Conference On* (IEEE, 2015), pp. 384–389

# A Packet Dispatching Scheme with Load Balancing Based on iSLIP for Satellite Onboard CIOQ Switches

**Li-Chun Mei, Lu-Feng Qiao, Qing-Hua Chen, Le Yang and Jian Yang**

**Abstract** Under the circumstance of high reliability demand in satellite onboard switches, an Iterative Round-Robin with SLIP (iSLIP) matching scheduling algorithm with load balancing suitable for Combined Input and Output Queuing (CIOQ) switch is presented in this paper. The implementation of load balancer improves the system reliability and the ability of recovery from failure. The iSLIP algorithm with the function of load balancing is used in a $16 \times 16$ CIOQ switch, and the whole switch fabric is implemented in a Xilinx xc7vx690t FPGA. Typical simulated results are given and analysized.

**Keywords** Satellite onboard switches · CIOQ switch · Load balancing scheduling scheme · iSLIP matching scheduling algorithm

## 1 Introduction

The satellite communication network plays an important role in modern communication networks, which has many advantages, such as wide coverage, convenient deployment, larger bandwidth, and less influenced by geographical environments. Considering the harsh space environment which involves the outer space radiation, Hardware complexity of the satellite onboard switches must be controlled to satisfy the limitation of weight, volume, power consumption, and design redundancy [1].

With the development of high-performance FPGAs embedded with high-speed Serializer and Deserializer (SerDes) that could be used in satellite, complex switch architectures, such as Combined Input and Output Queuing (CIOQ), can be designed. CIOQ structure, which can meet the requirements of high throughput, quality of

L.-C. Mei · L.-F. Qiao (✉) · Q.-H. Chen · L. Yang · J. Yang
Institute of Communication Engineering, PLA University of Science and Technology, Nanjing 210000, China
e-mail: 13357837783@189.cn

L.-C. Mei
e-mail: mlchun@qq.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_9

service guarantee, high reliability, has been widely used in high throughput switches and routers in territorial networks.

The crossbar takes an important part in CIOQ switch and its reliability and throughput are affected by the link state and matching scheduling algorithm directly. If a link of the crossbar is broken down, the performance of CIOQ switch would be lowered. And the unsuitable matching scheduling algorithm will lower the throughput of crossbar and make it become the bottleneck of the switch fabric.

To solve these problems, load balancing technology and proper matching scheduling algorithm are necessary to be applied in the crossbar of CIOQ switch. Load balancing technology can recover the broken links rapidly and realize load sharing [2]. In this paper, the iSLIP algorithm with load balancing is implemented in a 16 × 16 CIOQ switch.

## 2 The Structure of CIOQ Switch Architecture

How to achieve high throughput and low latency is one of the main challenges in high-speed switch fabric [3]. CIOQ is a queuing scheme that combines input and output queuing [4], which can achieve a good balance between performance and scalability. The CIOQ switch architecture is proposed by Chuang et al. in 1999 [5].

In this section, an N × N CIOQ switch architecture is presented, as shown in Fig. 1. This switch architecture consists of Input Modules (IMs), Output Modules (OMs), and Switch Fabric [6]. Each IM comprises *n* Virtual Output Queues (VOQs), and

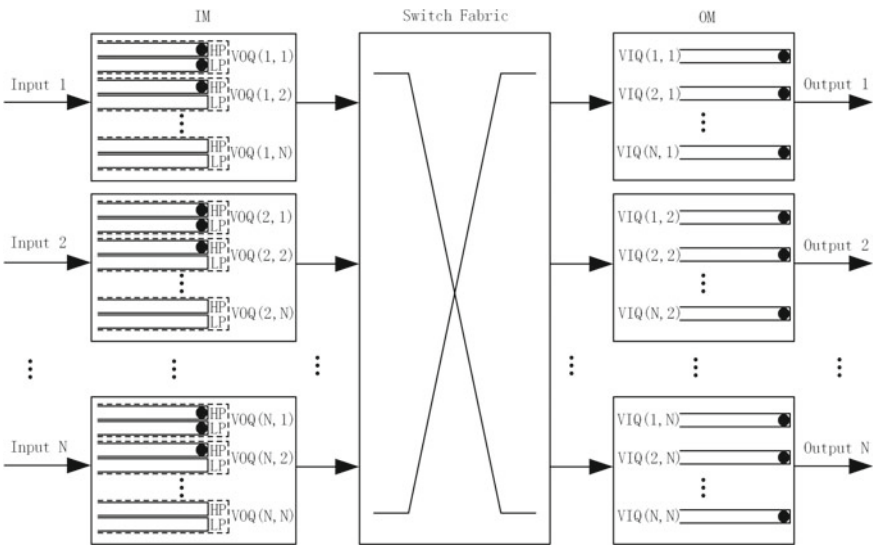



**Fig. 1** The structure of CIOQ switch architecture

each VOQ comprises High Priority (HP) traffic and Low Priority (LP) traffics. Every OM is made up of $n$ Virtual Input Queues (VIQs). The Switch Fabric provides interconnection between IMs and OMs. There are $n$ VOQs in an IM to eliminate the HOL. A VOQ, notated as VOQ($i$, $j$), buffers the cells from input port $i$ to output port $j$. In an OM, there are $n$ VIQs, each of which stores the cells from the relevant input port.

## 3 Load Balancing Scheme

Load balancing is a technique to increase the total throughput of the switch by distributing traffics from each IM to two or more links to different crossbars and then to the associated OMs [2].

As shown in Fig. 2, there are several crossbars between IMs and OMs and can provide multiple links between an IM and an OM. To reduce the probability of contention in the crossbar and increase the link capacity, load balancing should be used. Load balancing can also provide a method for the user to influence the traffic distribution between crossbars. With load balancing, the traffics from one IM can go through crossbars to the desired OMs in a uniform distribution manner, and can avoid to send cells through a link or even a crossbar with error, this will increase the reliability of the whole design which is crucial for satellite onboard switches.

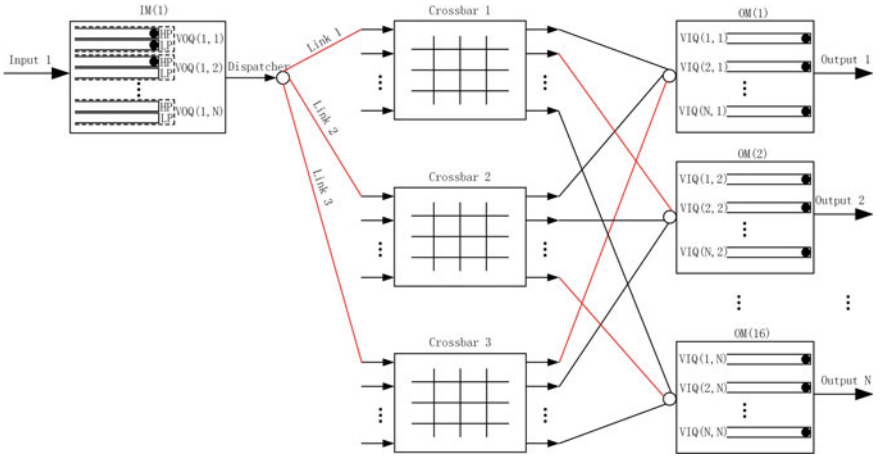



**Fig. 2** The structure of load balancing

## 4 Cell Scheduling Algorithms

The requirements for algorithms in high-performance CIOQ switches are listed as follows [7]:

*Simple to Implement*: For high-throughput CIOQ switches, the scheduling algorithms must be simplified and easy to implementation with hardware logics.

*High Throughput*: Each scheduler works independently in different crossbars and so contentions may occur, the scheduler in each crossbar can desynchronize from each other to get higher throughput.

*Qos Guarantee and Starvation Free*: Traffics with high priority should be served prior to traffics with lower priority, but the traffics with lower priority should not be starved under heavy load.

In this paper, the iSLIP matching scheduling algorithm is presented to meet these strict requirements. The main character of iSLIP algorithm is that the implementation is relatively simple and can be operated at high speed in hardware. The three steps for the iSLIP scheduling are as follows [8]:

*Step 1*: Request. Each unmatched input sends a request to every output for which it has a queued cell [8].

*Step 2*: Grant. If an unmatched output receives multiple requests, it chooses the one that appears next in a fixed, round-robin schedule starting from the highest priority element. The Grant Arbiters pointer $g_i$ is incremented (module N) to one location beyond the granted input if and only if the grant is accepted in step 3 of the first iteration [8].

*Step 3*: Accept. If an input receives multiple grants, it accepts the one that appears next in a fixed, round-robin schedule starting from the highest priority element. The pointer $a_j$ is incremented (module N) to one location beyond the accepted output. The Accept Arbiters pointer $a_i$ is only updated in the first iteration [8].

## 5 Design and Implementation of Each Module

In this section, the design of a $16 \times 16$ CIOQ switch fabric with fixed-length cell scheduling is presented. The whole design contains a load balancing module and an iSLIP matching module. Load balancing module is used to achieve load sharing, link backups and failure recovery. The function of iSLIP matching module is to realize iterative matching between different ports.

### *5.1 The Structures of CIOQ and Scheduling Process*

As shown in Fig. 3, the $16 \times 16$ CIOQ switch consists of 16 input ports, 16 output ports, and 3 crossbars. Three crossbars are used to provide interconnection between IMs and OMs, which can afford speedup, load balancing and data switching.

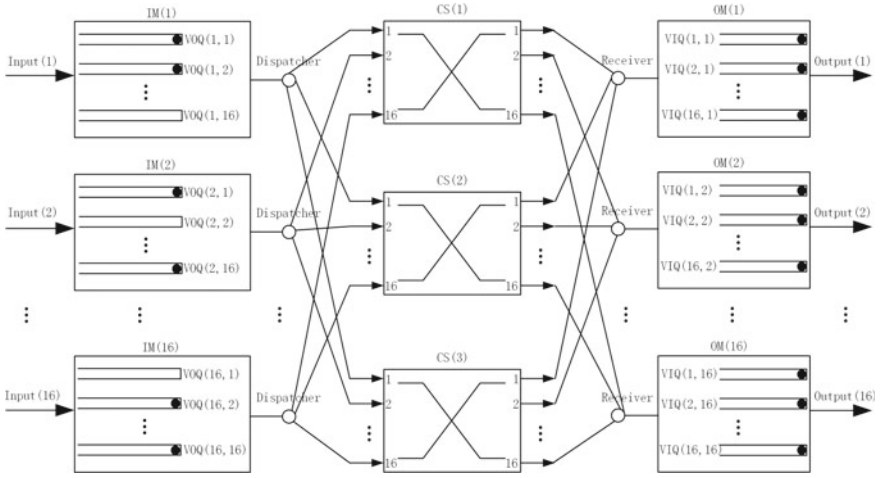



**Fig. 3** The structure of CIOQ

When a packet arrives at the IM, it is divided into fixed-length local cells logically, these cells are put into the HP queue or one of the LP queues of a VOQ according to the output port information in its local header. When a packet is reassembled in the OM successfully, it will be sent out from the corresponding output port. When two or more VOQs in an IM are nonempty, the scheduler will select cells in different VOQs, then transfer them to different OMs through different Crossbar simultaneously.

## *5.2 Implementation of Load Balancing Module*

As illustrated in Fig. 3, three Crossbar Schedulers (CSs) are employed here to provide load balancing and interconnection and works as follows:

*Sending Req signal from IMs to CSs* The port scheduler checks which VOQs have cells waiting to be transmitted and which links are available. Assume that three cells in three VOQs are waiting to be transmitted, if three links are available, these links would be used to send the Req signal. If link 2 breaks down, the Req signal will be transmitted only through link 1 or link3.

*Receiving Ack signal from CSs* Acks would be sent out from CSs after matching and scheduling. The schedulers in IMs check that which CSs have send Ack to it. If the scheduler received an Ack from one CS, the relevant cells would be selected and wait to be sent to this CS.

*Sending cells from IMs to CSs* The IMs would send one or more cells to these CSs which have sent Acks to it.

*Receiving cell frame from CS* The OMs receive cells from CSs and reassemble these cells to variable length frames depend on the following information in the cell local header: input port number, sequence number of the cell and priority.

### 5.3 Implementation of iSLIP Algorithm Module

The iSLIP algorithm module employs request-grant-accept handshake operation in one iteration.

*Updating Request signals* When a cell enters the CS, the request information in its local header will be extracted and used by the scheduler. When a request from one IM has been granted during a matching iteration, all the requests from that IM will be deasserted. More than one iterations are needed to reach higher matching ratio.

*Generating Grant signals* The requests from all 16 IMs to each OM are combined (G_Req1 to G_Req16 in Fig. 4) and sent to the associated output arbiter in each CS. Each arbiter checks the G_Req according to its pointer. If there is no request and G_Req is zero, the grant from the arbiter is zero, otherwise the arbiter will check

| | G–Req16: | G–Req15: | G–Req14: | G–Req13: | G–Req12: | G–Req11: | G–Req10: | G–Req9: | G–Req8: | G–Req7: | G–Req6: | G–Req5: | G–Req4: | G–Req3: | G–Req2: | G–Req1: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Req16: | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| Req15: | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 0 | 0 | 1 |
| Req14: | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| Req13: | 0 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 |
| Req12: | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| Req11: | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 |
| Req10: | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| Req9: | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| Req8: | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| Req7: | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |
| Req6: | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 1 | 0 | 0 |
| Req5: | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| Req4: | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 0 |
| Req3: | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 1 |
| Req2: | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 |
| Req1: | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 1 |

**Fig. 4** The value of G-Req to destination arbiters

each bit from the position it initially pointed to find out the first none zero bit and return the grant signal accordingly.

*Generating Accept signals* It receives the grant from each destination arbiter and generate the accept signals to the destination arbiters who grant its request with same algorithm used by the destination arbiters and then refresh its pointer and pointed to the position next the accepted one. The destination arbiter received an accept will refresh the it pointer accordingly.

## 6 Simulated Results and Analysis

The design is realized with Verilog HDL language and implemented in Xilinx xc7vx690t FPGA. The integrated development environment of Vivado 2015.4 is used and the whole design is simulated with Questa Sim-64 10.2c.

### *6.1 Simulation of Load Balancing*

As presented in Fig. 5, the value of *iframe_num* is '2', which means that there are two cells in a VOQ waiting to be scheduled, and *HSSL_state* indicates the link status between the IM and all the CSs. When the falling edge of the signal *simul-temp* appears, the request signal 'ea35' will be transmitted in the header of the local cell. Since *HSSL_state* of link 3 is '0', which means link 3 has broken down as shown in Fig. 5, and so accordingly 'ea35' doesn't appear in *HSSL3*.

### *6.2 Simulation of iSLIP in CSs*

Figures 6 and 7 illustrate the detailed scheduling process of iSLIP. As shown in Fig. 6, since IM(1) matches successfully after one iteration, the value of *Port00_request* changes from '1111' to '0000', meanwhile, *Port01_requst* ~ *Port15_request* is

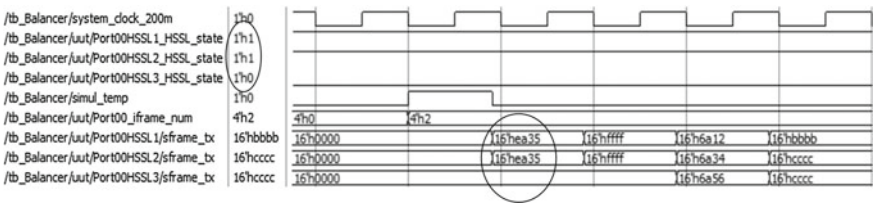



**Fig. 5** Simulated result of load balancing

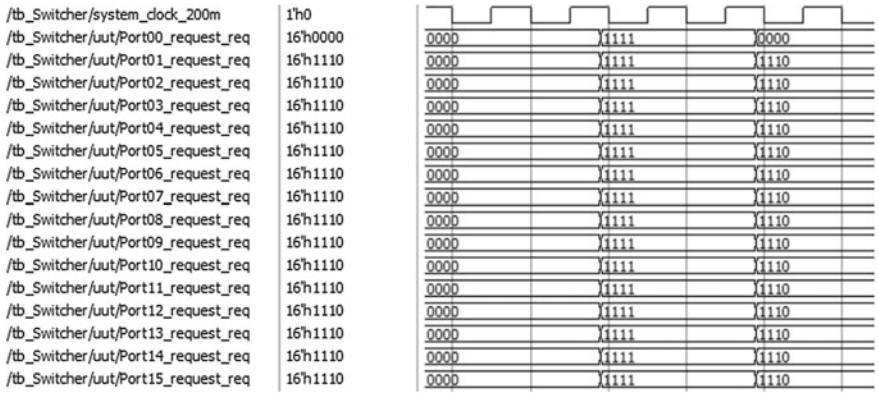



**Fig. 6** Updating request result of iSLIP algorithm module

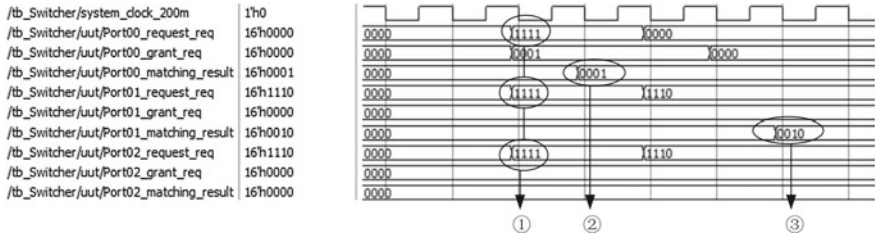



**Fig. 7** Scheduling result of iSLIP algorithm module

updated from '1111' to '1110'. As shown in Fig. 7, in the position of ①, all the output arbiters in the CS receive the requests from all IMs. The value of *matching_result* is '0001' in the position of ②, which means that IM(1) has matched with OM(1). In the position of ③, the value of *matching_result* is '0010', which means that IM(2) has matched with OM(5).

## 7 Conclusion

The Iterative Round-Robin with SLIP (iSLIP) matching scheduling algorithm with load balancing is presented in this paper. The load balancing technology is proposed for satellite onboard CIOQ switches and is implemented in a 16 × 16 CIOQ network, it can realize load sharing, failure recovery and reliable link connection. An iSLIP algorithm is presented and analyzed in the crossbar of CIOQ switch. The iSLIP algorithm with load balancing is simple and reliable, and can satisfy the demands of harsh space environment and provide higher throughput and reliability.

## References

1. F.M. Chiussi, J.G. Kneuer, V.P. Kumar, Low-cost scalable switching solutions for broadband networking: the ATLANTA architecture and chipset. IEEE Commun. Mag. **35**(12), 44–53 (1997)
2. K. Hashiguchi, Y. Kon, M. Hasegawa, et al., Traffic allocation control using support vector machine in heterogeneous wireless load balancing, in *Consumer Communications and Networking Conference (CCNC)* (IEEE, 2011), pp. 653–657
3. X. Yang, C. Kachris, M. Katevenis, Efficient implementation of CIOQ switches with sequential iterative matching algorithms, in *International Conference on Field-Programmable Technology (FPT)* (IEEE, 2010), pp. 433–436
4. C.R.D. Santos, S.A Motoyama, QoS Provisioned CIOQ ATM Switch with m Internal Links, in *International Conference on Telecommunications* (Springer, 2004), pp. 698–703
5. S.T. Chuang, A. Goel, N. Mckeown et al., Matching output queueing with a combined input/output-queued switch. IEEE J. Selected Areas Commun. **17**(6), 1030–1039 (1999)
6. A. Awan, R. Venkatesan, Design and implementation of enhanced crossbar CIOQ switch architecture. Electr. Comput. Eng. 1045–1048 (2004). IEEE
7. N. Mckeown, The iSLIP scheduling algorithm for input-queued switches. IEEE/ACM Trans. Netw. **7**(2), 188–201 (1999)
8. H. Chao, B. Liu, in *High Performance Switches and Routers* (Wiley, 2007), pp. 225–467

# A Constrained Conjugate Cyclic Adaptive Beamforming Algorithm with Symmetric Uniform Linear Array

**Yue Cui and Junfeng Wang**

**Abstract** A conjugate cyclic adaptive beamforming algorithm is proposed with the conjugate symmetric constraint. By using the symmetric structure of uniform linear array, it can be first proved that the conjugate cyclic correlation matrix is centro-Hermitian matrix, and the weight vector of the linearly constrained conjugate cyclic adaptive beamformer has a conjugate symmetric structure which is particular to conjugate cyclostationary signal. Then, the conjugate symmetric constraint for weight vector is added to the original algorithm, and the weight vector is derived from the proposed algorithm by recursive least squares. Compared to the traditional algorithms, the proposed method can achieve a higher steady state output SINR and a faster convergence speed. Moreover, in our method, the number of variables in the update equation are reduced effectively by half, which leads to significantly improve the overall performance. Simulation results demonstrate the effectiveness of the proposed method.

## 1 Introduction

Over the past decades, many blind adaptive beamforming algorithms [1] have been proposed without the a priori knowledge of direction-of-arrival (DOA) of the desired signal, autocorrelation matrix of interference and noise, training sequence, array calibration, and so on. Generally, they are based on constant modulus

Y. Cui (✉)
College of Computer and Information Engineering,
Tianjin Normal University, Tianjin, China
e-mail: cuiyue_enya@126.com

J. Wang (✉)
School of Computer and Communication Engineering,
Tianjin University of Technology, Tianjin, China
e-mail: great_seal@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_10

property [2, 3], high-order cumulants [4, 5] and cyclostationary property [6–10], respectively. The blind adaptive beamforming algorithm utilizing the cyclostationary property of communication signals has more advantages than the others. It is easy to find the desired signal and differentiate from the other interferences and noise, because the most communication signals have different cycle frequencies of their own which depend on the features of signal such as the baud rate, the carrier frequency offset, the pilot tone, etc. The SCORE algorithm based on cyclostationary property is used to increase the capacity of the cellular radio systems [6]. Nevertheless, its convergence speed is relatively slow, which could result in a low output signal-to-interference-plus-noise ratio (SINR) [7]. Furthermore, its computation complexity is high. In cyclic adaptive beamforming (CAB) algorithm [8], its convergence speed is relatively fast at the high signal-to-noise ratio (SNR). However, it does not consider the suppression of interferences and noise. Thus, in case of strong interferences or low SNR, its output SINR would degrade sharply. To overcome the above shortcoming, the linearly constrained CAB (LCCAB) algorithm [9] is proposed, which can form nulling and suppress strong interferences by using the linear-constrained minimum variance (LCMV) criterion. Nevertheless, it only has a little higher output SINR than CAB algorithm when interferences exist, and its convergence speed is relatively slow.

In this paper, a conjugate cyclic adaptive beamforming algorithm is proposed with the conjugate symmetric constraint. In theory, it can be first proved that the conjugate cyclic correlation (CCC) matrix is centro-Hermitian matrix, and the weight vector of the linearly constrained CAB (LC-CCAB) beamformer has a conjugate symmetric structure based on the symmetric uniform linear array, which are particular to conjugate cyclostationary signal. In order to improve the anti-jamming performance, the constraint for weight vector by using the conjugate symmetric property is then added. Finally, the weight vector is derived from the proposed algorithm by recursive least squares (RLS). Compared to the traditional algorithms, a much higher output SINR can be achieved when interferences and noise coexist, and less number of snapshots is required to converge to the steady state by the proposed method. Moreover, the number of variables in the update equation is reduced effectively by half in our method, which leads to significantly improve the overall performance even at the changed situation or small number of snapshots.

## 2 Problem Statement and Preliminaries

Consider a uniform linear array (ULA) with $M = 2N + 1$ isotropic sensors and take the middle one as the reference. Without loss of generality, assume that $K$ narrowband and cyclostationary signals come from distinct directions $\theta_k$, $k = 1, 2, \ldots, K$ and are uncorrelated impinging on the ULA. The received signal, including signals, interferences and noise, can be written as

$$\begin{aligned}\mathbf{x}(t) &= [x_{-N}(t), \cdots, x_{-1}(t), x_0(t), x_1(t), \cdots, x_N(t)]^T \\ &= \sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k(t) + \sum_{l=1}^{L} \mathbf{b}(\theta_l) J_l(t) + \mathbf{N}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{J}(t) + \mathbf{N}(t)\end{aligned} \tag{1}$$

where $s_k(t)$ and $\mathbf{a}(\theta_k)$ are respectively the $k$th signal and its steering vector, $J_l(t)$ and $\mathbf{b}(\theta_l)$ are the $l$th interference and its steering vector, and $\mathbf{N}(t)$ is the additive white Gaussian noise vector. For simplicity, we denote $\mathbf{s}(t) = [s_1(t), \ldots, s_K(t)]^T$, $\mathbf{A} = [a(\theta_1), \ldots, a(\theta_K)]$ and $\mathbf{J}(t)$ as the signal vector, array manifold and interference vector, respectively. The steering vector $\mathbf{a}(\theta_k)$, $k = 1, \ldots, K$ is

$$\begin{aligned}\mathbf{a}(\theta_k) &= \left[ e^{j2\pi f_0 N d \sin\theta_k / c} \quad \cdots \quad e^{j2\pi f_0 d \sin\theta_k / c} \quad 1 \quad e^{-j2\pi f_0 d \sin\theta_k / c} \quad \cdots \quad e^{-j2\pi f_0 N d \sin\theta_k / c} \right]^T \\ &= \mathbf{J}\mathbf{a}^*(\theta_k)\end{aligned} \tag{2}$$

where $d$ denotes the inter-element spacing of the sensors and equals half of the wavelength, $c$ is the light velocity, $f_0$ is the carrier frequency of signals, and $\mathbf{J}$ is the exchange matrix.

For the received signal, the conjugate cyclic correlation matrix is defined as

$$\mathbf{R}_{\mathbf{x}\mathbf{x}^*}^{\alpha}(\tau) = \langle \mathbf{x}(t)\mathbf{x}^T(t+\tau)e^{-j2\pi\alpha t} \rangle_P = \lim_{P\to\infty} \frac{1}{P} \sum_{t=1}^{P} \mathbf{x}(t)\mathbf{x}^T(t+\tau)e^{-j2\pi\alpha t} \tag{3}$$

where $\langle \cdot \rangle_P$ denotes the infinite-time average, and $\alpha$ is referred to as cycle frequency. As analysis in [10], $\mathbf{R}_{\mathbf{J}\mathbf{J}^*}^{\alpha}(\tau) = \mathbf{R}_{\mathbf{w}\mathbf{w}^*}^{\alpha}(\tau) = 0$ and the conjugate cyclic cross-correlations between the signals and interferences (or noise) are zeros since they are cyclically uncorrelated. According to the above properties, we get

$$\mathbf{R}_{\mathbf{x}\mathbf{x}^*}^{\alpha}(\tau) = \mathbf{A}\mathbf{R}_{\mathbf{s}\mathbf{s}^*}^{\alpha}(\tau)\mathbf{A}^H = \begin{bmatrix} R(-2N) & \cdots & R(-1) & R(0) \\ R(2N-1) & \cdots & R(0) & R(1) \\ \vdots & \ddots & \vdots & \vdots \\ R(0) & \cdots & R(2N-1) & R(2N) \end{bmatrix} \tag{4}$$

where $\mathbf{R}_{\mathbf{s}\mathbf{s}^*}^{\alpha}(\tau) = diag\left\{ R_{s_1 s_1 *}^{\alpha}(\tau), \ldots, R_{s_K s_K *}^{\alpha}(\tau) \right\}$ is the conjugate cyclic correlation matrix of $\boldsymbol{s}(t)$, and $\mathbf{R}_{\mathbf{x}\mathbf{x}^*}^{\alpha}(\tau)$ is a Hankel matrix, i.e.,$R_{\mathbf{x}_p \mathbf{x}_q *}^{\alpha}(\tau) \approx \sum_{n=1}^{K} R_{\mathbf{s}_n \mathbf{s}_n *}^{\alpha}(\tau) e^{-j\pi(2f_0+\alpha)[p+q-(M+1)]\tau_n} = R(p+q-M-1)$. Owing to the property of Hankel matrix, we can get

$$\left[\mathbf{R}_{\mathbf{x}\mathbf{x}^*}^{\alpha}(\tau)\right]^T = \mathbf{R}_{\mathbf{x}\mathbf{x}^*}^{\alpha}(\tau), \quad \left[\mathbf{R}_{\mathbf{x}\mathbf{x}^*}^{\alpha}(\tau)\right]^H = \left[\mathbf{R}_{\mathbf{x}\mathbf{x}^*}^{\alpha}(\tau)\right]^*. \tag{5}$$

Meanwhile, for the symmetric configuration of ULA, $\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)$ has the following property:

$$\mathbf{J}\left[\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\right]^H\mathbf{J}=\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\ . \tag{6}$$

Substituting (5) into (6), it can be proved $\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)$ is centro-Hermitian matrix, i.e.,

$$\mathbf{J}\left[\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\right]^*\mathbf{J}=\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\ . \tag{7}$$

In fact, when the array has a special structure, such as a symmetric ULA or more general symmetrically distributed array, we can also obtain $\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)$ is centro-Hermitian matrix.

## 3 The Proposed Beamforming Algorithm

### *3.1 The Weight Vectors for Conjugate CAB (CCAB) Algorithm and Linearly Constrained Conjugate CAB (LC-CCAB) Algorithm*

By using the cyclostationary property of the communication signals, the CAB algorithm [8] aims at forming beam toward the direction of the desired signal. The criterion of CAB algorithm can be written as

$$\max_{\mathbf{w},\mathbf{c}}\left|\mathbf{w}^H\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\mathbf{c}\right|^2 \quad \text{subject to} \quad \mathbf{w}^H\mathbf{w}=\mathbf{c}^H\mathbf{c}=1 \tag{8}$$

where CCC matrix $\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)$ is used in this paper. Hereafter, (8) is referred to as CCAB algorithm. In theory, we can prove that the weight vector of CCAB algorithm has the conjugate symmetric structure

$$\mathbf{W}_C=\mathbf{J}\mathbf{W}_C^* \tag{9}$$

**Proof** Let $\mathbf{U}=\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\left[\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\right]^H$, so the optimum weight vector $\mathbf{W}_C$ for the CCAB algorithm satisfies

$$\begin{aligned}\mathbf{U}\mathbf{W}_C&=\lambda\mathbf{W}_C\\ \mathbf{J}\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\mathbf{J}\cdot\mathbf{J}\left[\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\right]^H\mathbf{J}\cdot\mathbf{J}\mathbf{W}_C&=\lambda\mathbf{J}\mathbf{W}_C\\ \left[\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\right]^*\cdot\left[\mathbf{R}_{\mathbf{xx}^*}^{\alpha}(\tau)\right]^T\cdot\mathbf{J}\mathbf{W}_C&=\lambda\mathbf{J}\mathbf{W}_C\\ \mathbf{U}^*\cdot\mathbf{J}\mathbf{W}_C&=\lambda\mathbf{J}\mathbf{W}_C\end{aligned} \tag{10}$$

The above equation indicates $\mathbf{W}_C$ and $\mathbf{JW}_C$ are eigenvectors of $\mathbf{U}$ and $\mathbf{U}^*$, respectively. Meanwhile, $\mathbf{W}_C^*$ is eigenvectors corresponding to $\mathbf{U}^*$, so $\mathbf{W}_C = \mathbf{JW}_C^*$. Under ideal conditions, $\mathbf{W}_C$ has the conjugate symmetric structure.

In order to suppress interferences, the LCMV criterion is applied in CCAB algorithm, and the LC-CCAB algorithm is proposed in [9]. The cost function is

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w} \quad \text{subject to} \quad \mathbf{W}_C^H \mathbf{w} = 1 \tag{11}$$

where $\mathbf{R}_{\mathbf{xx}} = E\{\mathbf{x}(t)\mathbf{x}^H(t)\} = \langle \mathbf{x}(t)\mathbf{x}^H(t) \rangle_P$ is autocorrelation matrix of the received signals. The optimum weight vector is $\mathbf{w}_{opt} = \mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{W}_C$. According to (6), (7), and (9), we can demonstrate that $\mathbf{w}_{opt}$ also has the following conjugate symmetric structure

$$\mathbf{w}_{opt} = \mathbf{J}\mathbf{w}_{opt}^* \tag{12}$$

**Proof** $\mathbf{R}_{\mathbf{xx}}$ is a Hermitian Toeplitz matrix, so $\mathbf{R}_{\mathbf{xx}}$ is also a centro-Hermitian matrix, i.e., $\mathbf{R}_{\mathbf{xx}} = \mathbf{JR}_{\mathbf{xx}}^T\mathbf{J} = \mathbf{JR}_{\mathbf{xx}}^*\mathbf{J}$. Substituting into (13), we can get

$$\mathbf{w}_{opt} = \left(\mathbf{JR}_{\mathbf{xx}}^*\mathbf{J}\right)^{-1} \cdot \mathbf{JW}_C^* = \mathbf{J}\left(\mathbf{R}_{\mathbf{xx}}^{-1}\right)^*\mathbf{J} \cdot \mathbf{JW}_C^* = \mathbf{J}\left(\mathbf{R}_{\mathbf{xx}}^{-1}\mathbf{W}_C\right)^* = \mathbf{Jw}_{opt}^* \tag{13}$$

## 3.2 *The Recursive Least Squares for Constrained Conjugate Cyclic Adaptive Beamforming Algorithm*

In the general sense, the more information is added to the algorithm, the higher SINR of algorithm tends to be and less number of snapshots is required to arrive at its steady state. Using the conjugate symmetric structure of $\mathbf{w}_{opt}$ as demonstrated in (14), it can be combined with (11) in the following way

$$\min_{\mathbf{w}} \mathbf{w}^H \mathbf{R}_{\mathbf{xx}} \mathbf{w} \quad \text{subject to} \quad \mathbf{W}_C^H \mathbf{w} = 1 \quad \text{and} \quad \mathbf{w} = \mathbf{Jw}^* \tag{14}$$

When searching the optimal weight vector, the least mean squares (LMS) criterion only takes advantage of the information at the current moment, which is suitable for stationary signals. However, for the nonstationary or cyclostationary signals, its adaptation performance deteriorates significantly and convergence rate is reduced dramatically. In that case, the least squares (LS) technique, which utilizes all information from beginning until now, has the potential ability to achieve good performance independently of the spread of the eigenvalues of autocorrelation matrix. By employing RLS method, the following set of update equation can be got

$$\begin{aligned}\tilde{\mathbf{w}}(n+1) &= \mathbf{R}_{\mathbf{xx}}^{-1}(n+1)\mathbf{W}_C(n+1)\\ \mathbf{w}(n+1) &= \frac{1}{2}\left[\tilde{\mathbf{w}}(n+1) + \mathbf{J}\tilde{\mathbf{w}}^*(n+1)\right]\end{aligned} \tag{15}$$

where

$$\begin{aligned}\mathbf{g}(n+1) &= \frac{\mathbf{R}_{\mathbf{xx}}^{-1}(n)\mathbf{x}(n+1)}{\lambda + \mathbf{x}^H(n+1)\mathbf{R}_{\mathbf{xx}}^{-1}(n)\mathbf{x}(n+1)}\\ \mathbf{R}_{\mathbf{xx}}^{-1}(n+1) &= \frac{1}{\lambda}\left\{\mathbf{R}_{\mathbf{xx}}^{-1}(n) - \mathbf{g}(n+1)\mathbf{x}^H(n+1)\mathbf{R}_{\mathbf{xx}}^{-1}(n)\right\}\\ \mathbf{W}_C(n+1) &= \frac{n}{(n+1)}\mathbf{W}_C(n) + \frac{1}{(n+1)}\left(\sum_{i=-N}^{N} x_i(n+1+\tau)e^{-j2\pi\alpha t}\right)\mathbf{x}(n+1)\end{aligned} \tag{16}$$

In the iterative process, the initial values are $\mathbf{W}_C(0) = [1,\ 0,\ \ldots,\ 0]^T$ and $\mathbf{R}_{\mathbf{xx}}^{-1}(0) = \delta\mathbf{I}_M$. It should be noted that for an even $M$, if we know $w_1(n+1), w_2(n+1), \ldots, w_{M/2}(n+1)$ in weight vector $\mathbf{w}(n+1)$, another half of the coefficients would have been decided based on the conjugate symmetric property $\mathbf{w} = \mathbf{J}\mathbf{w}^*$. For an odd $M$, half of the information about the coefficient at the middle $w_{(M+1)/2}(n+1)$ would be determined according to (12), and the same conclusion as in the case of an even $M$ can be got. In a word, the number of variables in the update equation has been reduced effectively by half, and we would expect a faster convergence speed and a higher output SINR because the extra constraint guarantees $\mathbf{w}(n+1)$ in a form corresponding to $\mathbf{w}_{opt}$.

## 4 Simulation Results

The performance of the proposed method is illustrated in this section. For each level, we run 4000 Monte Carlo trials. Hereafter, the proposed method is compared with the fast implementation of CCAB (FCCAB) algorithm [7], LC-CCAB algorithm [8] and the fast implementation of LC-CCAB (FLC-CCAB) algorithm [9]. Note that the weight vector of LC-CCAB algorithm corresponds to the closed-form solution, where the singular value decomposition and sample matrix inversion would lead to a very high computational complexity, although a good performance has been achieved.

The first simulation assumes the desired signal and interference impinging on the ULA come from $\theta_1 = 40°$ with SNR = 10 dB and $\theta_l = -30°$ with JNR = 7 dB, respectively. The number of snapshots varies from 1 to 800. The output SINR of the

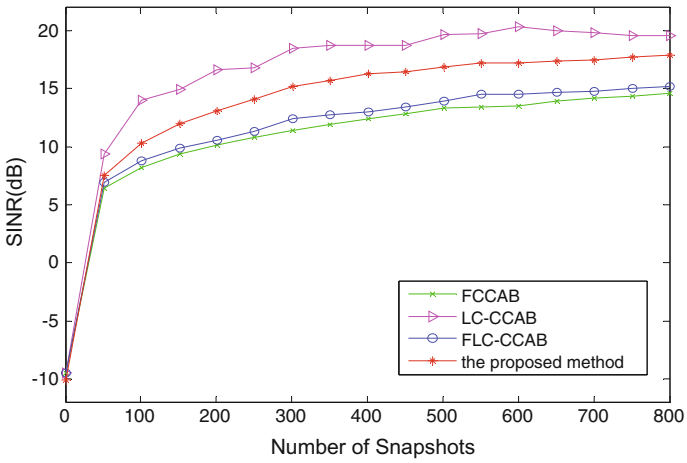



**Fig. 1** The output SINR versus the number of snapshots

proposed method versus the number of snapshots is investigated in Fig. 1. The result illustrates that the output SINR of the proposed method improves as the number of snapshots increases until the steady state is reached. Moreover, the output SINR of the proposed method is higher than that of FCCAB and FLC-CCAB, and approaches the closed-form solution of LC-CCAB.

In what follows, the normalized weight vector error $\left\|\mathbf{w}-\mathbf{w}_{opt}\right\|^2/\left\|\mathbf{w}_{opt}\right\|^2$ is assessed as criterion in detail. As can be seen in Fig. 2, the weight vector of the proposed algorithm converges to the steady state in a slightly faster rate than the original one. Theoretically speaking, this is because the weight vector has a

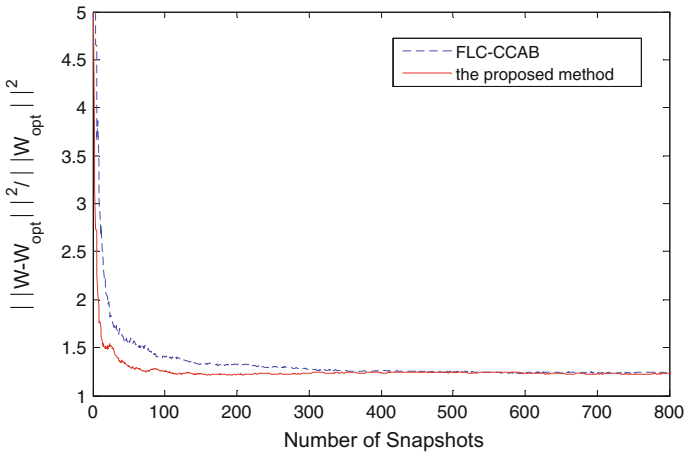



**Fig. 2** The normalized weight vector error versus the number of snapshots

conjugate symmetric structure proved in (13), and the constraint based on this property is employed to the proposed algorithm. Hence, the recursive process of weight vector with the constraint can converge faster and becomes closer to the optimum solution. Moreover, the number of variables in the update equation has been reduced effectively by half, which leads to a much improved overall performance, even at the changed situation or small number of snapshots.

## 5 Conclusions

In this paper, it can be proved that the CCC matrix possesses centro-Hermitian property and the weight vector for LC-CCAB beamformer has a conjugate symmetric structure based on the symmetric ULA, which are particular to conjugate cyclostationary signal. Then, we present a conjugate cyclic adaptive beamforming algorithm with the conjugate symmetric constraint. Different from the traditional algorithms, the proposed method can achieve a higher steady state output SINR and a faster convergence speed. Moreover, using conjugate symmetric structure of weight vector, the number of variables in the update equation has been reduced effectively by half. Simulation results validate the effectiveness of the proposed method and illustrate that our method outperforms the existing methods.

**Acknowledgements** This work is supported in part by Tianjin Research Program of Application Foundation and Advanced Technology under Grant 16JCQNJC01100 and 15JCQNJC01800, in part by Doctoral Foundation of Tianjin Normal University under Grant 52XB1201, in part by High School Science and Technology Developing Foundation of Tianjin under Grant 20140706.

## References

1. B. Allen, Adaptive array systems, fundamentals and applications (2005). Chichester, U.K
2. Y. Chen, T. Le-Ngoc, B. Champagne, Recursive least squares constant modulus algorithm for blind adaptive array. IEEE Trans. Signal Process. **52**(5), 1452–1456 (2004)
3. L. Zhang, W. Lei, R. Langley, A class of constant modulus algorithms for uniform linear arrays with a conjugate symmetric constraint. Signal Process. **90**(9), 2760–2765 (2010)
4. M.C. Dogan, J.M. Mendel, Cumulant-based blind optimum beamforming. IEEE Trans. Aerosp. Electron. Syst. **30**(3), 722–740 (1994)
5. K. Yang, T. Ohira, Y. Zhang, Super-exponential blind adaptive beamforming. IEEE Trans. Signal Process. **52**(6), 1549–1563 (2004)
6. R. Ho, O. Wu, K.M. Wong, Implementation of cyclic beamforming techniques on mobile communication systems, in *Proceedings of 2nd workshop cyclostationary signals* (1994), pp. 16.1–16.9
7. Q. Wu, K.M. Wong, R. Ho, Fast algorithm for adaptive beamforming of cyclic signals. IEE Proc. Radar Sonar Navig. **141**(6), 312–318 (1994)
8. Q. Wu, K.M. Wong, Adaptive beamforming of cyclic signal and fast implementation, in *Proceedings of IEEE international conference on acoustics, speech, and signal processing* (1994), pp. 157–160

9. Q. Wu, K.M. Wong, Blind adaptive beamforming for cycloststionary signals. IEEE Trans. Signal Process. **44**(11), 2757–2767 (1996)
10. W.J. Zeng, X.L. Li, X.D. Zhang, An improved signal-selective direction finding algorithm using second-order cyclic statistics, in *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing* (2009), pp. 2141–2144

# PAPR Reduction for Cognitive AIS Using Transforming Sequence of Frank-Heimiller and Artificial Bee Colony Algorithm

**Junfeng Wang, Yue Cui, Shexiang Ma, Lanjun Liu and Jianfu Teng**

**Abstract** A cognitive automatic identification system (CAIS) employing some promising technologies, such as spectrum sensing and OFDM, has been investigated by us in recent 4 years. In the CAIS, the normal location messages and security video information will be loaded by employing the OFDM. However, OFDM signals have a high peak-to-average power ratio (PAPR), which causes signal distortion. Lots of the PAPR reduction techniques have been presented in the literature, among which, a technique of dynamically selecting sequences has been taken considerable suggestion, but its high computational complexity and bandwidth expansion impedes practical implementation. In this paper, transforming sequence of Frank-Heimiller (TSFH) is proposed for the first time, which is with the ideal correlation properties; then we propose a dynamic spreading code allocation (DSCA) based on the set of TSFH and artificial bee colony algorithm (DSCA-TSFH and ABC) scheme to obtain low PAPR. Simulation results show that the proposed DSCA-TSFH and ABC algorithm is an efficient one to achieve significant PAPR reduction, with a low computational complexity.

J. Wang (✉) · S. Ma
School of Computer and Communication Engineering, Tianjin University of Technology, Tianjin, China
e-mail: great_seal@163.com

Y. Cui (✉)
College of Computer and Information Engineering, Tianjin Normal University, Tianjin, China
e-mail: cuiyue_enya@126.com

L. Liu
College of Engineering, Ocean University of China, Qingdao, China

J. Teng
School of Electronic Information Engineering, Tianjin University, Tianjin, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_11

## 1 Introduction

In more recent years, automatic identification system (AIS) that can minimize maritime traffic accidents and improve maritime traffic efficiency is becoming an urgent need all around the world. However, with the increments of shipborne AIS users and the emerging requirements for some new applications, many drawbacks of the AIS, such as slot collision, incommunicable security video information and limited coverage, are more and more prevalent. To overcome those, a cognitive AIS (CAIS) employing some promising technologies, such as spectrum sensing and OFDM, is proposed by us in recent 4 years [1–4]. The CAIS has many virtues, such as the transmissions of the security video information which is favourable to dealing with the emergency, and the satellite-based CAIS receivers which enlarge the usage range of shipborne AIS information. Moreover, in the CAIS, system capacity will be increased by utilizing spectrum sensing technique; coverage range will be enlarged by using satellite-based CAIS equipments in which users are identified via transforming sequence of Frank-Heimiller (TSFH) possessing approximately ideal correlation and random phase; and the normal location messages and security video information will be loaded by employing OFDM which has many advantages such as high bit-rate, robustness against channel fading, and high frequency efficiency. However, OFDM signals have a high peak-to-average power ratio (PAPR), which causes signal distortion because of the use of a high-power amplifier in the transmitter. In fact, as a new communication system, many research challenges and standardization work have to be addressed before the CAIS is widely applied. This paper will concentrate on one of the most important and fundamental challenges: how to reduce the PAPR of the CAIS signals?

Many PAPR reduction techniques have been proposed in the literature [5–10], including clipping, coding, tone injection, constellation extension, partial transmit sequences (PTS), the PTS based on optimization method, dynamically selecting sequences, and so on. Among which, the recent PTS algorithm is a distortionless technique which is based on combining signal subblocks, whose phase is shifted by constant phase factors. This technique can obtain sufficient PAPR reduction and side information need to be sent at the same time, but the exhaustive search complexity of the optimal phase combination increases exponentially with the number of subblocks. Thus, suboptimal PTS methods employing particle swarm optimization (PSO) have been suggested. Additionally, a technique of dynamically selecting sequences has been taken considerable investigation, but its PAPR is also high for the single user and its high computational complexity and bandwidth expansion impedes practical implementation.

In this paper, we propose a dynamic spreading code allocation (DSCA) based on the set of TSFH and artificial bee colony algorithm (DSCA-TSFH & ABC) scheme to obtain low PAPR. The presented DSCA-TSFH & ABC method can efficiently reduce the PAPR of OFDM signals, which employs ABC algorithm to search the

TSFH with lower PAPR. Moreover, the suggested TSFH is a new code method with approximately ideal correlation and random phase. In addition, simulation results show that the proposed DSCA-TSFH & ABC scheme is an efficient one to achieve the significant PAPR reduction, with a low computational complexity.

## 2 System Model and PAPR Expressions

In the CAIS, the normal location messages and security video information are loaded by employing OFDM when the useful spectrum is obtained via the spectrum sensing method. It is well known that the high PAPR of the transmitted signal is the one of the major drawbacks. In this section, the system model of the CAIS and PAPR expressions of the OFDM signals will be given.

### *2.1 System Model of the Cognitive AIS*

Users are identified via the TSFH in the CAIS, and it is assumed that the system has $U$ active users. For the $u$ th user, $\mathbf{d}^{(u)} = \begin{bmatrix} d_1^{(u)} & d_2^{(u)} & \cdots & d_M^{(u)} \end{bmatrix}$ denotes $M$ number of QPSK or QAM modulated data symbols. Each symbol of the user is multiplexed by a TSFH spreading code $\mathbf{c}^{(u)} = \begin{bmatrix} c_1^{(u)} & c_2^{(u)} & \cdots & c_N^{(u)} \end{bmatrix}$, where $N$ is the spreading factor or spreading length. Assuming there are $N$ subcarriers, the transmission of the $m$ th symbol of the arbitrary user corresponds to

$$s^{(m)}(t) = d_m \sqrt{E_d/N} \sum_{n=0}^{N-1} c_n e^{j2\pi n \Delta f t} \tag{1}$$

where $E_d$ is the energy of the symbols, and $n\Delta f$ is the frequency position of the $n$ th carrier component. Note that selection of $\Delta f = 1/T_s$ should ensure subcarrier orthogonality over symbol duration $T_s$.

The transmitted signal[1] at one OFDM symbol at carrier frequency $f_c$ is

$$x(t) = \sqrt{E_d/N} \sum_{m=0}^{M} d_m \sum_{n=0}^{N-1} c_n e^{j2\pi (f_c + n\Delta f)t} \tag{2}$$

[1]In the ship-to-ship or ship-to-coast communications of the CAIS, the transmitted signal is considered as uplink one. Meanwhile, the satellite or coast-based signal can be considered as downlink one. In this paper, for simplicity, we only consider the case one. In fact, the extension of the proposed method is straightforward and the interested readers can further investigate.

### 2.2 PAPR Expressions of the OFDM Signals

Based on the above analyses, the PAPR of $x(t)$ is defined as the ratio of the maximum instantaneous power to the average power, that is

$$PAPR = \|x\|_{\infty}^{2} / \|x\|_{2}^{2} = \max_{t} |x(t)|^{2} / E\left\{ |x(t)|^{2} \right\} \tag{3}$$

where $E\{\cdot\}$ denotes the statistical average operator.

The complementary cumulative distribution function (CCDF), represented the probability that the PAPR of an OFDM symbol exceeds the given threshold $\mathrm{PAPR}_0$ and always utilized as performance measures for the PAPR reduction, is denoted as

$$\mathrm{CCDF} = \Pr\{PAPR > PAPR_0\} \tag{4}$$

## 3 PAPR Reduction Algorithm for the Cognitive AIS

In this section, we will use the TSFH with lower PAPR which can be selected by the ABC algorithm as the set of the spreading code of the transmitted signal. First, sequence of Frank-Heimiller and its transforming one will be given. Then, ABC algorithm will be introduced briefly. Finally, we propose a DSCA-TSFH & ABC scheme to obtain low PAPR efficiently.

### 3.1 Transforming Sequence of Frank-Heimiller

In multi-user communication systems, it is often require spread spectrum sequences with desirable correlation properties. Low correlations can distinguish different user with different sequences or/and different shifts of a single sequence. Thus, lots of spread spectrum sequences were suggested. Among them, Frank-Heimiller sequence of period $N^2$ is with $N$ phases, which can be constructed by writing the $N$ by $N$ discrete Fourier transform matrix by rows, i.e. matrix is given first

$$\mathrm{F}_N = \left[ \omega^{ik \bmod N} \right], \quad i, k = 0, 1, 2, \ldots, N-1 \tag{5}$$

where $\omega = \exp(j2M\pi/N)$. However, the correlation properties of the truncate Frank-Heimiller sequence usually employed in the wireless systems are unsatisfactory. Fortunately, we found that when the truncate Frank-Heimiller $\mathrm{F}_N^{'}$ is processed by IDFT or IFFT, i.e. the TSFH is $\mathbf{c} = \mathrm{FT}\left\{ \mathrm{F}_N^{'} \right\}$, here $\mathrm{FT}\{\cdot\}$ denotes IDFT

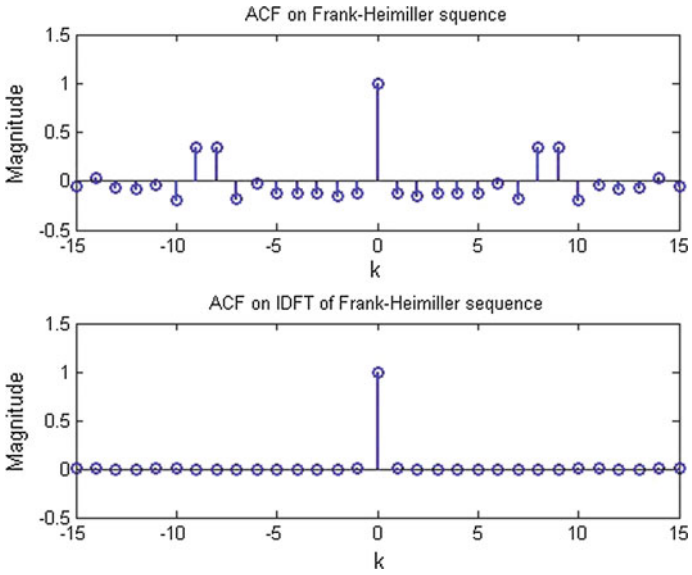



**Fig. 1** The autocorrelation function of original FH and TSFH

or IFFT, its properties become desirable, especially for autocorrelation sidelobes. In other words, the periodic autocorrelation sidelobes for such sequence are all zero, which is verified by computer simulation as shown in Fig. 1.

## *3.2 Artificial Bee Colony Algorithm*

The ABC algorithm proposed by Karaboga introduces a bee swarm algorithm for numerical optimization problems. In this algorithm, three groups of bees are contained: employed bees, onlookers, and scouts. Moreover, there are three steps in each cycle of the search: (1) placing the employed bees onto the food sources and then calculating their nectar amounts; (2) selecting the food sources by the onlookers after sharing the information of employed bees and determining the nectar amount of the foods; (3) determining the scout bees and placing them onto the randomly determined food sources. In the ABC algorithm, a food source position represents a possible solution to the problem to be optimized and the nectar amount of a food source corresponds to the quality (fitness) of the associated solution. Since the length of article is limited, for a complete understanding of the ABC method, the reader can also refer to [11–13].

### 3.3 PAPR Reduction for Cognitive AIS Based on DSCA-TSFH & ABC Algorithm

In this subsection, we will employ the DSCA-TSFH & ABC scheme to reduce the PAPR of the OFDM signals efficiently. In order to apply ABC algorithm to search the better spreading code set of the TSFH, the original ABC algorithm is modified. In the DSCA-TSFH & ABC algorithm, a food source position represents a TSFH vector $\mathbf{c}^{(u)} = \left[ c_1^{(u)} \quad c_2^{(u)} \quad \cdots \quad c_N^{(u)} \right]$, $u = 1, 2, \ldots, K$, where $K$ represents the size of a randomly distributed initial population. The fitness value of a solution $\mathbf{c}^{(u)}$ in the population is determined by

$$fitness\left(\mathbf{c}^{(u)}\right) = 1/\left[1 + f\left(\mathbf{c}^{(u)}\right)\right] \tag{6}$$

where $f\left(\mathbf{c}^{(u)}\right)$ is the PAPR value of the solution.

If the fitness value of the new TSFH vector is higher than the previous TSFH vector, the bee memorizes the new TSFH vector. After calculating the fitness value of a TSFH vector, an employed bee chooses the new TSFH vector by

$$\mathbf{c}^{(u)'} = \mathbf{c}^{(u)} + \phi^{(u)}(\mathbf{c}^{(u)} - \mathbf{c}^{(k)}) \tag{7}$$

where $\mathbf{c}^{(k)}$ is a solution within the neighbourhood of the $\mathbf{c}^{(u)}$ and $\phi^{(u)} \in [-1, +1]$. After all the employed bees complete the tasks, they share the fitness values with the onlooker bees.

For each onlooker bee, they watch the dances of the employed bees and select a food source depending on the knowledge taken from the employed bees, through the following formula

$$p_i = fitness(\mathbf{c}^{(u)}) / \sum_{u=1}^{K} fitness(\mathbf{c}^{(u)}) \tag{8}$$

In ABC algorithm, if a source cannot be improved when the number of cycles is greater than a predetermined limit value, the source is considered to be exhausted. The employed bee associated with the exhausted source becomes a scout and makes a random search in problem domain by

$$\mathbf{c}^{(u)} = \min\left(\mathbf{c}^{(u)}\right) + rand(0, 1) \times \left[\max\left(\mathbf{c}^{(u)}\right) - \min\left(\mathbf{c}^{(u)}\right)\right] \tag{9}$$

where $rand(0, 1)$ is a random number with a uniform distribution and $\min\left(\mathbf{c}^{(u)}\right)$ and $\max\left(\mathbf{c}^{(u)}\right)$ are the lower and upper bounds of the TSFH vector, respectively.

To be here, all procedures of our DSCA-TSFH & ABC scheme are performed, whose performance will be simulated and evaluated in the next section.

## 4 Performance Evaluation

In order to evaluate and compare the performance of the DSCA-TSFH & ABC algorithm for the PAPR reduction of the OFDM signals, numerous simulations have been conducted in this section. In our simulation, the spreading factor length $N = 64$, 16-QAM modulation is used, $10^3$ random OFDM symbols are generated to get CCDF.

First, we present the simulation results with $K = 16$ and $K = 32$ as shown in Fig. 2a. From the comparison of the PAPR reduction, it can be seen that 2 dB at $K = 16$ is less than at $K = 32$ when CCDF $= 10^{-3}$, which delivers better performance via the better optimization and allocation.

Under the same scenarios, the comparisons of the PAPR reduction between the DSCA-TSFH & PSO and DSCA-TSFH & ABC algorithm are given in Fig. 2b. From this figure, it is evident that the performance of the proposed algorithm outperforms that of the DSCA-TSFH & PSO.

Finally, the PAPR convergence performance is verified with maximum iteration number $\text{iter}_{\max} = 10^3$, the limit value limit $= 10^2$, and the initialization iteration number $\text{iter}_{\text{init}} = 0$. Figure 3 shows the simulation results on the mean of the best

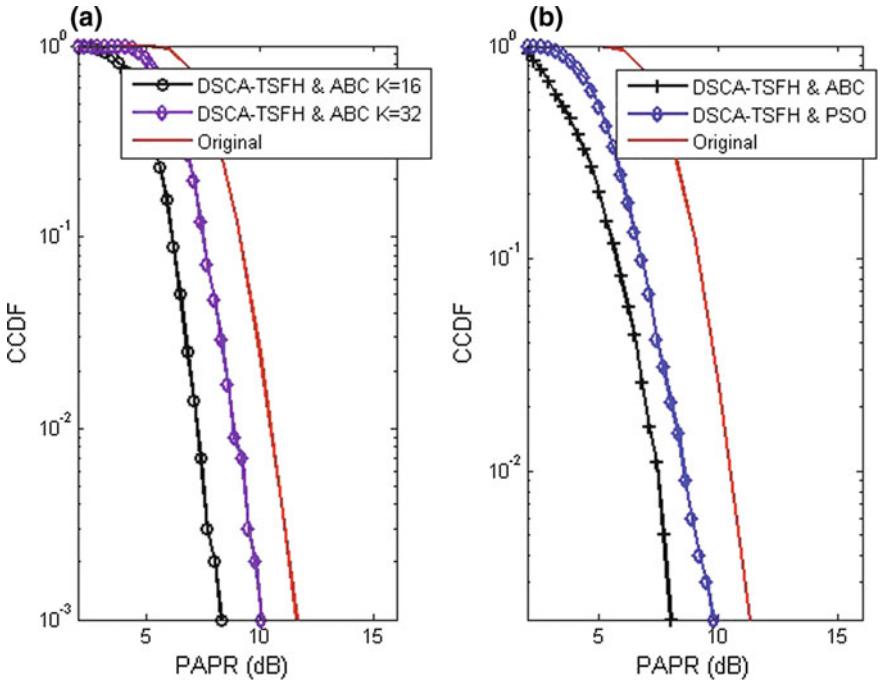



**Fig. 2** The comparison of the PAPR reduction for different $K$ (**a**) and different algorithms (**b**)

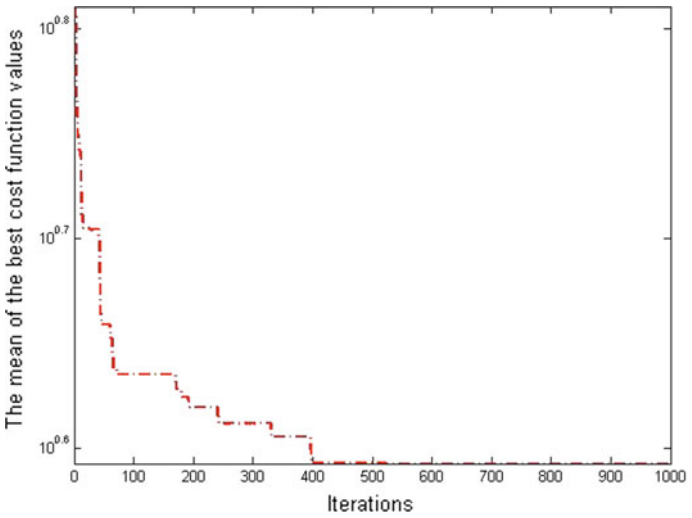



**Fig. 3** The mean of the best cost function values versus the iteration number

cost function values. As can be from this figure, although the performance is better and better as iteration number increases, the mean of PAPR getting by the iteration number iter = 390 is only $10^{0.59}$ which is a suitable choice for applications.

## 5 Conclusion

OFDM technique has many advantages such as high bit-rate, robustness against channel fading, and high frequency efficiency, which has been employed in the CAIS investigated by us in recent 4 years. However, the major drawback of the OFDM signals is the high PAPR of the transmitted signal. In order to overcome this deficiency, a DSCA-TSFH & ABC scheme is proposed for PAPR reduction in this paper. In addition, we give the TSFH with the desirable correlation properties for the first time, which can be considered as partially theoretical guidance for the spread spectrum wireless communication systems in the future. Finally, computer simulations show that our presented DSCA-TSFH & ABC algorithm can achieve significant PAPR reduction with a low computational complexity.

**Acknowledgements** This work is supported in part by Tianjin Research Program of Application Foundation and Advanced Technology under Grant 15JCQNJC01800 and 16JCQNJC01100, in part by High School Science and Technology Developing Foundation of Tianjin under Grant 20140706, in part by Doctoral Foundation of Tianjin Normal University under Grant 52XB1201, and in part by the National Natural Science Foundation of China under Grant 61371108, and 61431005.

## References


1. J. Wang, Y. Cui, S. Ma et al., A scheme designed for cognitive automatic identification system, P.R. China Patent 201510995727.2 (2015)
2. J. Wang, Y. Cui, S. Ma et al., A non-isotropic simulation model for ship-to-ship fading channels, in *Proceedings of IET International Conference Wireless Mobile Multimedia Network*, pp. 115–119 (2015)
3. J. Wang, Y. Cui, S. Ma et al., On the spectrum sensing for cognitive automatic identification system, in *Proceedings of IEEE International Conference Signal Process*, in press (2016)
4. J. Wang, Y. Cui, Y.R. Zheng et al., Statistical modeling and simulation of ship-to-ship fading channels. unpublished (2016)
5. X. Li, L.J. Cimini Jr., Effect of clipping and filtering on the performance of OFDM. IEEE Commun. Lett. **2**(5), 131–133 (1998)
6. E. Jones, A.T. Wilkinson, K.S. Barton, Block coding scheme for reduction of peak to mean envelope power ratio of multicarrier transmission scheme. Electron. Lett. **30**(25), 2098–2099 (1994)
7. Y.J. Kou, W.S. Lu, A. Antoniou, A new peak-to-average power-ratio reduction algorithms for OFDM systems via constellation extension. IEEE Trans. Wireless Commun. **6**(5), 1823–1832 (2007)
8. L.J. Cimini Jr., N.R. Sollenberger, Peak-to-average power ratio reduction of an OFDM signal using partial transmit sequences. IEEE Commun. Lett. **4**(3), 86–88 (2000)
9. H.L. Hung, Y. F. Huang, C.M. Yeh et al., Performance of paticle swarm optimization techniques on PAPR reduction for OFDM systems, in *Proceedings of IEEE International Conference on Systems, Man, Cybernetics*, pp. 2390–2395 (2008)
10. T. Ginige, N. Rajatheva, K.M. Ahmed, Dynaic spreading code selection method for PAPR reduction in OFDM-CDMA systems with 4-QAM Modulation. IEEE Commun. Lett. **5**(10), 408–410 (2001)
11. D. Karaboga, An idea based on honey bee swarm for numerical optimization. Technical Report-TR 06, Erciyes University (2005)
12. D. Karaboga, B. Basturk, A powerful and efficient algorithm for numeric function optimization: artificial bee colony (ABC) algorithm. J. Global Optimization **39**, 459–471 (2007)
13. D. Karaboga, B. Basturk, On the performance of artificial bee colony (ABC) algorithm. Appl. Soft Comput. **8**(1), 687–697 (2008)

# Beamforming of Sparse Cylindrical Arrays

**Na Wu and Qilian Liang**

**Abstract** As sparse arrays cost fewer elements, they are less expensive than dense arrays and could be applied widely in antenna and sonar deployment, two sparse cylindrical arrays are proposed in this paper. According to the characteristic of cylindrical array, it can be seen as a linear array whose elements are the identical circular arrays. Therefore the co-prime linear array and nested linear array could be combined with circular arrays. Based on the beam pattern of uniform cylindrical array, 1D and 2D beam patterns of co-prime cylindrical array and nested cylindrical array are derived respectively. In addition, when more than one sources are coming from arriving directions, the performance of sparse arrays are analyzed and compared. The simulation results show that the new proposed sparse cylindrical arrays not only reduce the number of elements, but also improves the resolution in comparison with an equal length uniform cylindrical array.

**Keywords** Beam pattern · Sparse array · Cylindrical array

## 1 Introduction

Conformal arrays are valuable in many applications, such as underwater sonar arrays and antenna arrays. Conforming the array to the surface not only saves the space, but also makes the elements less visually intrusive by integrating it into existing objects. In this paper, we focus on the cylindrical array, which is one type of conformal arrays. A cylindrical array contains of elements in three directions and this provides wide cover in both the azimuth plane and the elevation plane. As a result, cylindrical arrays are widely used in wireless communication and underwater environment [1, 4].

N. Wu (✉) · Q. Liang
Department of Electrical Engineering, University of Texas at Arlington,
Arlington, TX 76019, USA
e-mail: na.wu@mavs.uta.edu

Q. Liang
e-mail: liang@uta.edu



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_12

One big difference between conformal array and traditional array is that the beam pattern cannot be obtained by the product of array factor and element pattern. However, since the cylindrical array consists of a series of identical circular arrays, the whole array can be seen as a linear array whose elements are these identical circular arrays. In this paper, we propose the nested cylindrical array (NCA) and co-prime cylindrical array (CCA) which combined the nested linear array [5] and co-prime linear array [6] with uniform circular array. Some studies about nested arrays and co-prime arrays have been published, such as two-dimensional nested arrays in [7] and moving co-prime sensor arrays [8]. Adhikari in [9] analyzed beamforming with extended co-prime sensor arrays and Na in [10] presented the underwater direction of arrival estimation based on 1D and 2D nested arrays. Besides, there are some work related with the sparse cylindrical arrays in [11, 12]. Based on the previous work, we derive the new beam pattern of the proposed CCA and NCA, and also compare it with the uniform cylindrical array (UCA).

The remainder of this paper is outlined as follows. In Sect. 2, we give a brief overview of the co-prime linear array, nested linear array, and the cylindrical array beam pattern synthesis. Then the theoretic analysis of the proposed CCA and NCA is introduced in Sect. 3. After that, in Sect. 4, simulation results of beam patterns are provided and in Sect. 5, conclusions would be presented.

## 2 Preliminaries

In this section, we provide a brief summary of sparse linear arrays and beamforming of uniform cylindrical array, respectively.

### *2.1 Co-Prime Linear Array and Nested Linear Array*

Co-prime linear array interleaves two uniform subarrays which are sampled by two prime integers $C_1$, $C_2$, and the locations of the elements are in set $\mathbf{K} = \{\mathbf{K}_1, \mathbf{K}_2\}$, where

$$\begin{aligned} \mathbf{K}_1 &= \left\{0, C_2, 2C_2, 3C_2, \ldots, (C_1 - 1) \times C_2\right\} \\ \mathbf{K}_2 &= \left\{0, C_1, 2C_1, 3C_1, \ldots, (C_2 - 1) \times C_1\right\} \end{aligned} \tag{1}$$

Nested linear array is a little different from co-prime linear array, where the elements are placed with two-level density, and level 1 has $N_1$ elements, level 2 has $N_2$ elements,

$$\begin{aligned} &1 \le l \le N_1 \\ &(N_1 + 1)m, \quad 1 \le m \le N_2 \end{aligned} \tag{2}$$

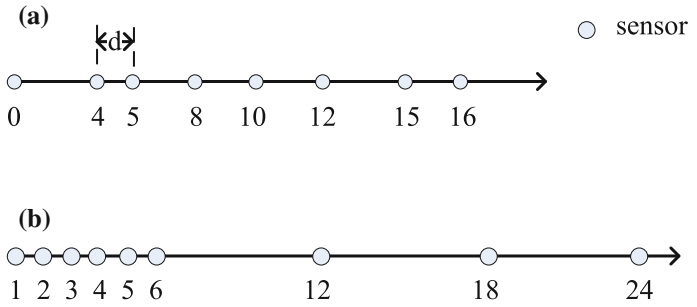



**Fig. 1** Example of co-prime linear array and nested linear array, **a** co-prime pair $(4, 5)$, **b** nested pair $(5, 4)$

In Fig. 1a, it plots an example of co-prime linear array with co-prime integers $(4, 5)$, and in Fig. 1b, it shows the example of nested linear array with pair $(5, 4)$.

## *2.2 Cylindrical Array Beam Pattern Synthesis*

Figure 2 shows the cylindrical array geometry and the coordinate system. The elevation is the direction parallel to the z-axis and the azimuth plane is the x-axis and y-axis. Assuming all the elements are isotropic and are placed on the surface of a staggered grid. If there are $N$ circular arrays, $M$ elements in each circular array and

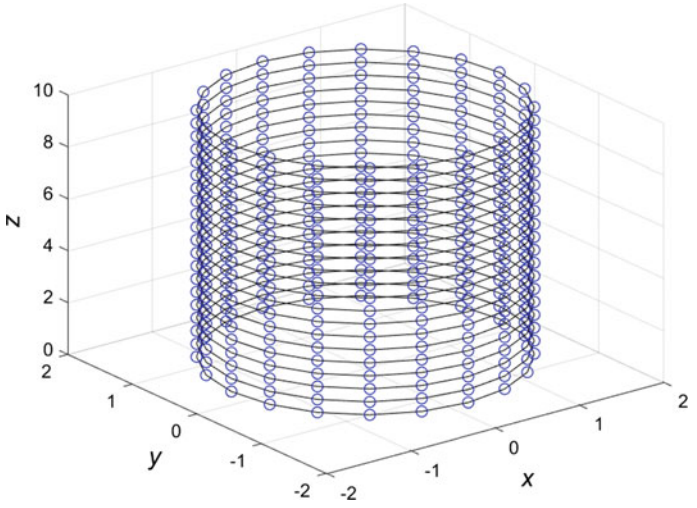



**Fig. 2** Geometry of uniform cylindrical array

the radius of circular arrays is $2\pi R = 10\lambda$ ($\lambda$ is the wavelength), the far-field beam pattern in the generic direction $(\theta, \phi)$ of the cylindrical array could be represented as [13]

$$B(\theta,\phi) = \sum_{n=0}^{N-1}\sum_{m=1}^{M} w_{nm}^{*} e^{jk_0[Rsin\theta cos(\phi-\phi_1)+z_n cos(\theta)]} \tag{3}$$

here $k_0 = |k| = 2\pi/\lambda$.

If we further expand (3) into (4), it is easy to see that the second term in the braces is the beam pattern of the $n$th circular array.

$$B(\theta,\phi) = \sum_{n=0}^{N-1} e^{jk_0 z_n cos(\theta)} \left\{ \sum_{m=1}^{M} w_{nm}^{*} e^{jk_0 Rsin\theta cos(\phi-\phi_1)} \right\} \tag{4}$$

Replace the second term with $B_{cir,n}(\theta,\phi)$, then (4) is simplified as

$$B(\theta,\phi) = \sum_{n=0}^{N-1} e^{jk_0 z_n cos(\theta)} B_{cir,n}(\theta,\phi) \tag{5}$$

Thus if $w_{nm}^{*}$ is separable,

$$w_{nm}^{*} = w_{n}^{*} w_{m}^{*} \tag{6}$$

(5) could reduce to

$$\begin{aligned} B(\theta,\phi) &= \sum_{n=0}^{N-1} w_{n}^{*} e^{jk_0 z_n cos(\theta)} B_{cir}(\theta,\phi) \\ &= B_{lin}(\theta,\phi) B_{cir}(\theta,\phi) \end{aligned} \tag{7}$$

## 3 Numerical Analysis of Co-Prime Cylindrical Array and Nested Cylindrical Array

First, the geometry of co-prime cylindrical array and nested cylindrical array is shown in Fig. 3 and Fig. 4 respectively. In the z-axis direction, the elements are placed as co-prime linear array or nested linear array, which are the same as plotted in Fig. 1.

If we consider the sensors in the z-direction to be elements of a linear array and each column has uniform weighting, then the total array factor is the product of the two array factors as expressed in (7). For the $B_{lin}$ part, if it is the uniform linear array with length $N$, then the beam pattern is

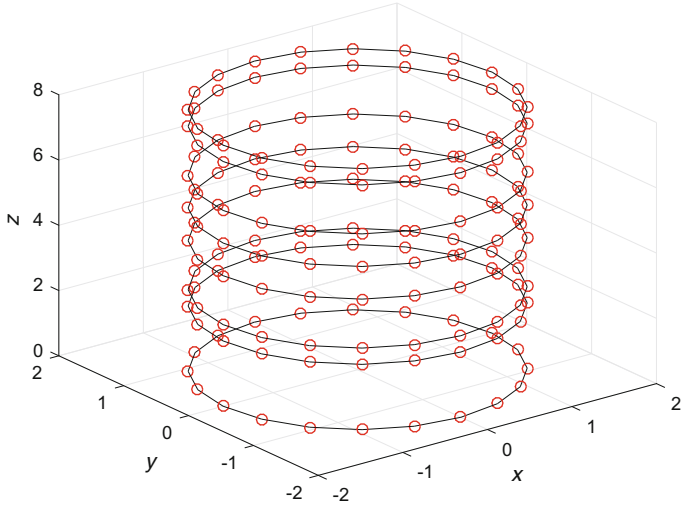



**Fig. 3** Geometry of co-prime cylindrical array

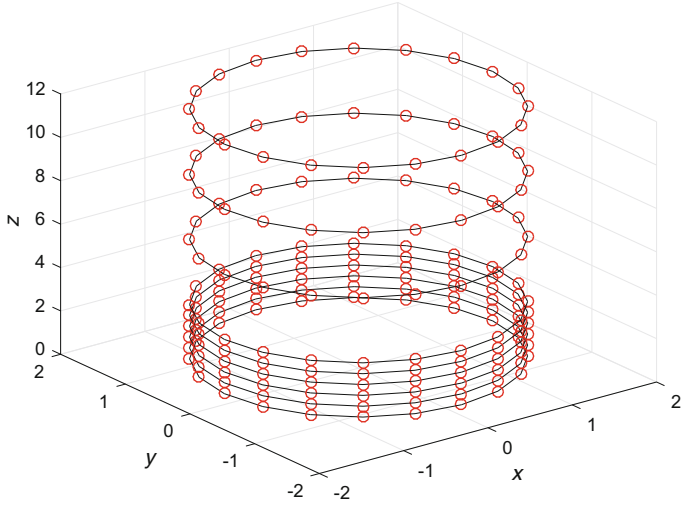



**Fig. 4** Geometry of nested cylindrical array

$$B_{lin} = \frac{1}{N} \cdot \frac{1 - e^{jN\pi u}}{1 - e^{j\pi u}} \tag{8}$$

When the elements in z-direction are placed as co-prime linear array with pair $(C_1, C_2)$, the beam pattern of the co-prime array is

$$B_{lin,C_1,C_2} = \frac{1}{C_1 C_2} \cdot \frac{(1 - e^{jC_1C_2\pi u})^2}{(1 - e^{jC_1\pi u})(1 - e^{jC_2\pi u})} \tag{9}$$

where $u = cos(\theta)$.

$B_{lin,C_1,C_2}$ has been derived in [14], but the authors did not provide the beam pattern of nested array. For the nested array with pair $(N_1, N_2)$, we also apply uniform weighting and the beam pattern of the nested array is

$$B_{lin,N_1,N_2} = \frac{e^{j(N_1+2)\pi u}}{(N_1+1)N_2} \cdot \frac{1 - e^{j(N_1+1)N_2\pi u}}{1 - e^{j\pi u}} \tag{10}$$

To be noticed is that, in the first subarray, we choose one more elements to compute the beam pattern. For example, the locations of the nested array in Fig. 1b is $[1, 2, 3, 4, 5, 6, 12, 18, 24]$. The beam pattern of first subarray is based on elements at $[1, 2, 3, 4, 5, 6]$ and the second subarray is calculated by $[6, 12, 18, 24]$. As a consequence, there is an overlapping element in the two subarrays.

The absolute value of (10) is

$$\left|B_{lin,N_1,N_2}\right| = \left|\frac{1}{(N_1+1)N_2} \cdot \frac{1 - e^{j(N_1+1)N_2\pi u}}{1 - e^{j\pi u}}\right| \tag{11}$$

Comparing to (8), it is straightforward to see that (11) equals to the absolute value of (8) if $N = (N_1 + 1)N_2$ sensors. As a matter of fact, nested linear array could achieve exactly the same beam pattern of a much longer ULA. This property helps us to further analyze the performance improvement of the beam pattern when comparing NCA, CCA, and UCA.

## 4 Simulation Results

In order to test the resolution of CCA and NCA, we assume there are two sources coming from the direction of 40° and 80°. The elements in the linear array are located as following:

$$\begin{aligned} UCA &= [0, 1, 2, \ldots, 29] \\ CCA &= [0, 5, 6, 10, 12, 15, 18, 20, 24, 25] \\ NCA &= [1, 2, 3, 4, 5, 6, 12, 18, 24, 30, 36] \end{aligned}$$

where both the co-prime pair and nested pair is (5, 6). The length of UCA equals to the length of CCA and NCA including their virtual elements, which is almost triple of the real elements in CCA or NCA.

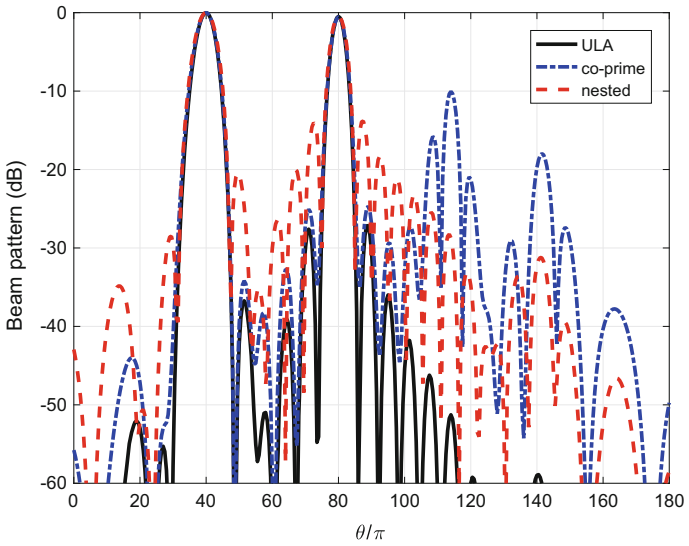



**Fig. 5** 1D beam pattern of two far sources for UCA, CCA, and NCA

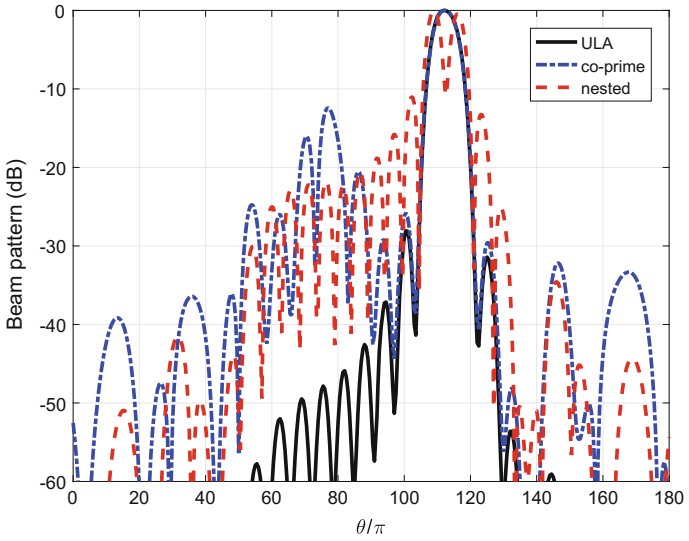



**Fig. 6** 1D beam pattern of two close sources for UCA, CCA, and NCA

Figure 5 shows the beam pattern over 100 snapshots. The two sources are successfully detected by UCA, CCA, and NCA. But if the angles of the two sources are close to each other, for example, one is 100° and the other 105°, the result of beam pattern is illustrated in Fig. 6. Both UCA and CCA fail to distinguish these two sources while the NCA detects the two close sources. As a result, when the sources are far away,

co-prime cylindrical array and nested cylindrical array could be applied to detection, however, if the sources are close to each other, we would better choose nested cylindrical array.

## 5 Conclusion

Motivated by the recent studies about co-prime linear array and nested linear array, two new sparse cylindrical arrays—co-prime cylindrical array and nested cylindrical array are proposed in this paper. To the best of our knowledge, this is the first time that co-prime array and nested array are combined with cylindrical array. A detailed analysis of the beam pattern for CCA and NCA is provided. Since beam pattern is related with the ability of resolution, we compare the performance of beam pattern among uniform cylindrical array and the two sparse cylindrical arrays. Simulation results prove our theoretical analysis, which is the two sparse cylindrical arrays could achieve better performance with few elements comparing with uniform cylindrical array. In the future study, we will focus on some other properties of the CCA and NCA, such as applying different weighting functions.

**Acknowledgements** This work was supported in part by U.S. Office of Naval Research under Grant N00014-13-1-0043 and U.S. National Science Foundation under Grants CNS-1247848, CNS-1116749, CNS-0964713.

## References

1. S. Stergiopoulos, A.C. Dhanantwari, J.J. Grodski, Implementation of adaptive processing in integrated active-passive sonars deploying cylindrical arrays, in *IEEE Proceedings of International Symposium in Underwater Technology*, Apr 1998, pp. 289–294
2. N. Herscovici, Z. Sipus, P.-S. Kildal, The cylindrical omnidirectional patch antenna. IEEE Trans. Antennas Propagat. **49**, 1746–1753 (2001)
3. M.G. Hussain, Theory and analysis of adaptive cylindrical array antenna for ultrawideband wireless communications. IEEE Trans. Wireless Comm. **4**(6), 3075–3078 (2005)
4. R. Salamon, J. Marszal, W. Lesniak, Broadband sonar with a cylindrical antenna. Acta Acustica United Acustica **92**(1), 153–155 (2006)
5. P. Pal, P.P. Vaidyanathan, Nested arrays: a novel approach to array processing with enhanced degrees of freedom. IEEE Trans. Signal Process **58**(8), 4167–4181 (2010)
6. P.P. Vaidyanathan, P. Pal, Sparse sensing with co-prime samplers and arrays. IEEE Trans. Signal Process **59**(2), 573–586 (2011). Feb
7. P. Pal, P.P. Vaidyanathan, Nested arrays in two dimensions, Part II: application in two dimensional array processing. IEEE Trans. Signal Process **60**(9), 4706–4718 (2012). Sept
8. N. Wu, Q. Liang, Underwater mobile co-prime sensor array for angle of arrival estimation based on space-domain sensor synthesis, in *IEEE INFOCOM Workshop*, Apr 2016
9. K. Adhikari, J.R. Buck, K.E. Wage, Extending coprime sensor arrays to achieve the peak side lobe height of a full uniform linear array. EURASIP J. Adv. Signal Process. **2014**(1), 3592–3608 (2014). Sep

10. N. Wu, Q. Liang, Underwater DoA estimation based on nested array, in *IEEE MILCOM*, pp. 216-221, Oct 2015
11. V. Voipio, P. Vainikainen, Narrowbeam cylindrical antenna array with sparse antenna spacing. IEEE Veh. Technol. Conf. **48**(1), 465–469 (1998). May
12. J.E. Kirkebo, A. Austeng, Sparse cylindrical sonar arrays. IEEE J. Ocean. Eng. **33**(2), 224–231 (2008). Apr.
13. H.L. Van Trees, Detection, estimation, and modulation theory, in *Optimum Array Processing* (Wiley, 2004)
14. K. Adhikari, J.R. Buck, K.E. Wage, Beamforming with extended co-prime sensor arrays, in *Proceedings of IEEE International Conference Acoust, Speech, Signal Process*, May 2013, pp. 4183–4186

# ACO-GA Combined Algorithm for Solving Spectrum Allocation Problem in D2D Communications

**Chenguang He, Tingting Liang, Shouming Wei and Weixiao Meng**

**Abstract** D2D communication is considered to be one of the key technologies in LTE-A network even in 5G communication system. Spectrum allocation problem is an important part in the study of D2D communication. Considering the spectrum allocation problem under D2D communication scenario, this paper proposes a system model with graph theory and adopt the concept of interference weights. A novel spectrum allocation algorithm ACO-GA combined algorithm is proposed. This algorithm combines ant colony algorithm and modified genetic algorithm based on the theory of graph coloring. Simulation results show that ACO-GA combined algorithm performs superior than traditional ant colony algorithm and genetic algorithm on spectrum efficiency and interference cost.

**Keywords** D2D communication · Spectrum allocation · Ant colony algorithm · Genetic algorithm

## 1 Introduction

Device-to-Device (D2D) communication technology [1, 2] allows the neighboring mobile terminals use spectrum resource for point-to-point data transmission directly, without the need for additional infrastructure. D2D communication can be classified as inband and outband D2D, which use licensed and unlicensed spectrums, respec-

C. He · T. Liang (✉) · S. Wei · W. Meng
Communication Research Center, Harbin Institute of Technology, Harbin, China
e-mail: wsltt1992@sina.com

C. He
e-mail: hechenguang@hit.edu.cn

S. Wei
e-mail: weishouming@hit.edu.cn

W. Meng
e-mail: wxmeng@hit.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_13

tively. Depending on whether transmission links reuse spectrum with cellular links, there are two D2D types: underlay, which means they share the same spectrum, and overlay, meaning D2D links use dedicated resource. D2D communication has many advantages, for instance, the short distance between D2D terminals makes it possible to get higher data transmission rate with smaller transmission power, and to further improve spectrum efficiency by reusing spectrum resource with other users. In addition, D2D communication can be employed to emergency situations when infrastructure cannot work normally, so as to improve the robustness of cellular networks.

Based on the aforementioned potential benefits, D2D communication is widely considered to be one of the significant technologies in LTE-A network even in 5G communication system. However, D2D communication using cellular network spectrum brings more complex electromagnetic interference to the existing network environment, and interference is the key factor affecting the system capacity and Quality-of-Service (QoS). Therefore, appropriate spectrum allocation scheme and effective interference control mechanism has been the focus in the study of the D2D communication.

Spectrum allocation model based on graph coloring theory [3, 4] is a relatively mature model in mobile cellular networks, which has been widely adopted to analyze the spectrum planning and assignment of cellular cells. The principle of graph coloring theory to solve spectrum allocation is as follows: the network topology is abstracted to a graph; vertexes in the graph can either represent users in the network or represent transmission link requests, and edges means two relevant vertexes can communicate directly or there is interference between these two vertexes, etc.; and colors represent spectrum resources. To find proper coloring scheme for graph vertexes to make sure neighbor nodes have different colors, so as to avoid the co-channel interference and improve spectrum efficiency.

Most intelligent algorithms have made certain progress in solving graph coloring problem, mainly including the exhaustive search method, backtracking method, greedy algorithm, ant colony optimization algorithm (ACO), genetic algorithm (GA), tabu search algorithm, and simulated annealing algorithm. The exhaustive search method and the backtracking method belong to precise algorithms [5, 6] and they are superior to other optimization algorithms in the case you can get solutions, but the search space of precise algorithms is very complex, so they only apply to solving small-scale graph coloring problem; Greedy algorithm, ant colony algorithm, genetic algorithm, tabu search algorithm, and simulated annealing algorithm are approximation algorithms [7–12]. These algorithms cannot find the optimal solution, but they often find a better solution, especially when dealing with large-scale graph coloring problem, approximation algorithms are more efficient than precise algorithms. But these algorithms have some shortcomings inevitably, like that ant colony algorithm easily gets suboptimal solution because of its local convergence feature, and need large iterations to find the optimal solution; genetic algorithm has cumbersome crossover and mutation operation; however, in addressing the graph coloring problem, crossover operation can produce instability and may destroy the former excellent results, which complicate the search process.

This paper considers inband overlay D2D communication scenario, which means D2D users use dedicated spectrum resources in licensed spectrum, so only interferences between D2D links are taken into account without considering interference between D2D and cellular users. Meanwhile a weighting factor [13] is presented to show the interference size between D2D communication links. Combining ant colony algorithm and genetic algorithm, ACO-GA combined algorithm, a new spectrum allocation algorithm based on graph coloring theory is proposed. Establish initial solution using the local convergence feature of ACO algorithm, and then using modified GA to optimize the solution, finally we can get the best spectrum allocation scheme balancing spectrum efficiency and interference. And then through the simulation analysis, we compare the performance of ACO-GA combined algorithm with traditional ACO algorithm and GA algorithm.

The remainder of this paper is organized as follows. Section 2 shows the system model of inband overlay D2D communication scenario. In Sect. 3 the principle of the proposed ACO-GA combined algorithm is demonstrated separately by modified ACO and GA algorithms. Section 4 contains simulating results while finally, Sect. 5 concludes this paper.

## 2 System Model

Considering the inband overlay D2D communication scenario in LTE-A network, there only exists interference between D2D communication links and there is no interference between D2D users and cellular users. All D2D users use small and fixed transmission power to ensure only neighboring terminals, which have short distance with each other, and can communicate directly through D2D mode. So it is possible to make more D2D users in the network reuse the same spectrum resource and do not cause large co-channel interference. In order to minimize the number of spectrum used in the D2D communication, the maximum number of D2D communication links that can reuse the same resource need to be found, and at the same time to make sure that the interference among these D2D links, is within the acceptable range.

The network is abstracted to a graph $G=(V,E)$, vertex set $V=\{v_1, v_2, \ldots, v_n\}$ represents the number of D2D communication requests, and edge set $E=\{(v_1,v_2),(v_1,v_3),\ldots,(v_1,v_n),(v_2,v_3),\ldots,(v_{n-1},v_n)\}$ represents any possible interference between two D2D communication requests, a weighting factor is used to describe the value of interference, which depends on the actual channel conditions between these two links. In this model, value of the interference weight set ranges from 1 to 15 and it is calculated as formula (1):

$$w_{ij} = round(14 \times (1 - \frac{d_{ij}}{\sqrt{a^2+b^2}})) + 1. \quad (1)$$

In formula (1), $w_{ij}$ means the interference weight between D2D communication request $i$ and request $j$, $d_{ij}$ represents the distance between request $i$ and request $j$, the network range is $a \times b$, $round(x)$ is the function to calculate the nearest integer to $x$, and then we can get $w_{ij} \in [1, 15]$. The greater the distance between these two communication requests, the less the interference weight, which means the interference between the two requests is smaller, therefore the more likely it is to reuse the same spectrum resource. Colors represent spectrum resources and two vertexes will coat with the same color if they reuse the same resource.

## 3 ACO-GA Combined Algorithm

### *3.1 Establish Initial Solution with ACO Algorithm*

In the graph $G = (V, E)$ edge set $E$ represents interference weight. We set a threshold $T$, when $w_{ij} > T$, D2D communication requests $i$ and $j$ are regarded as neighbor nodes, which means the interferences between requests $i$ and $j$ are too heavy to coat with the same color; otherwise, requests $i$ and $j$ can coat with one color. Therefore, an adjacent matrix $A = \{a_{ij}\}_{n \times n}$ can be described as formula (2):

$$a_{ij} = a_{ji} = \begin{cases} 1, (w_{ij} > T) \\ 0, (w_{i,j} \leq T) \end{cases}. \tag{2}$$

After acquiring the adjacent matrix A, the degree of each vertex is $degree(i) = \sum_{k=1}^{n} a_{ik}$, which shows the number of neighbor nodes that a certain node has, so the degree of the graph is $\Delta G = \max_{1 \leq i \leq n} (degree(i))$. Due to the systematic characteristics of ACO algorithm, and its local convergence feature that are adaptive for solving graph coloring problem with $k$ colors, in the meantime the fact that the number of colors used for graph coloring problem is less than $p = \Delta G + 1$, so first, ACO algorithm is used to solve graph coloring problem of $p$ colors, then the solution of $p$ colors is adopted as initial population of an improved genetic algorithm to find the minimum number of colors that color the graph.

Pheromone updating rules of ACO algorithm are as follows:

$$\Delta\tau_{ij} = \begin{cases} 1/C_k, (i,j \in visited_k) \\ 0, (else) \end{cases} \tag{3}$$

$$\tau_{ij}(t) = (1 - \rho) \bullet \tau(t-1) + \sum_{k=1}^{p} \Delta\tau_{ij}^{k}. \tag{4}$$

In formulas (3) and (4), $\Delta\tau_{ij}$ is pheromone increment between request $i$ and $j$, $visited_k$ represents the set of vertexes that ant $k$ visited. Every ant represents one color so value of ant $k$ is from 1 to $p$. $C_k$ means the total interference cost among vertexes that ant $k$ visited and $C_k = \sum_{i,j \in visited_k} w_{ij}$. $\rho$ is the pheromone evaporation factor.

Ant chooses the next vertex to visit according to the probability, which is calculated as follows:

$$P_{ij}^{k} = \frac{\tau_{ij}^{\alpha} \bullet \eta_{ij}^{\beta}}{\sum_{nodes\_available} \tau_{ij}^{\alpha} \bullet \eta_{ij}^{\beta}}. \tag{5}$$

In formula (5), $\eta = \frac{1}{M}$ is the heuristic factor and M is the number of colors that have been used. *nodes_available* represents the vertex set that ant $k$ can access; $\alpha, \beta$ show the importance of pheromone and heuristic factor respectively.

## 3.2 Find Optimization Solution with Modified GA Algorithm

Traditional genetic algorithm adopt crossover and mutation operations, but in the graph coloring problem, using crossover operation may make the neighbor nodes coat with the same color, which destroys the former good group and makes the solution becomes poorer. So a modified genetic algorithm is proposed, using transposition operator and shifting operator as well as cut operator and splice operator to replace crossover and mutation.

Transposition operator is a kind of random transposition operation to any characters on the gene string according to the given rules or probability. For example, for a binary coding chromosome A:1100100110, implementing transposition operation to the characters of position number 3 and 8 we can get a new chromosome B:1110100010.

Shifting operator is to move a substring of certain length one position backward in turn and at the same time move the last character to the first position of the string according to the given rules or probability. For example, for a binary coding chromosome A:1100100110, implementing shifting operator to substring of position number 3, 4 bits length, we get a new chromosome B:1100010110.

Cut operator is to cut off the gene string at a particular location to make the original chromosome string into two separate individuals according to the preset probability or rules. For instance, for a binary coding chromosome A: 1100100110, cut it off at the position of 6 to make it become two substrings 11001 and 00110 and both length is 5 bits.

Splice operator is to make two gene strings link together to become a single individual according to the preset probability or rules. For example, for two binary coding chromosomes A:11001 and B:00110, implementing splice operator to them we get a new longer chromosome 1100100110.

Modified genetic algorithm works as follows. First, the initial solution of ACO algorithm is adopted as the initial population of GA and vertexes are divided into a number of substrings depending on the color they use. Then, genetic variation is accomplished through verifying the different gene substrings. The concrete implementation process of genetic variation is as follows: choose substring 1; verify each vertex of substring 1 with all of the vertexes in other substrings, such as test the first vertex in substring 1, marked as $x$, with all the vertices in substring 2; if $x$ is not neighbor of any vertex in substring 2, then delete $x$ from substring 1 and move it to substring 2; otherwise $x$ remain at its place. Then do the same operation to the other vertices of substring 1 as well as vertexes in other substrings. In the above process, cut operator and splice operator are also used to change the length of gene strings. After iterations, the number of substrings of new chromosome will become smaller, namely the number of colors used, and finally it will converge to the optimal solution that has minimum number of colors.

## 4 Simulation Results

The simulation is realized in MATLAB; the scope of the network is $1000 \times 1000\,\mathrm{m}^2$, and D2D communication request nodes randomly distributed in the network. Under different numbers of D2D requests, ACO-GA combined algorithm together with the traditional ant colony algorithm, genetic algorithm, and ant colony algorithm considering interference weight proposed in reference [13] are simulated and compared on the performance of spectrum reuse efficiency and total interference cost. These four algorithms are simply marked as “combine”, “ACO”, “GA”, and “interfere” in sequence in the figures.

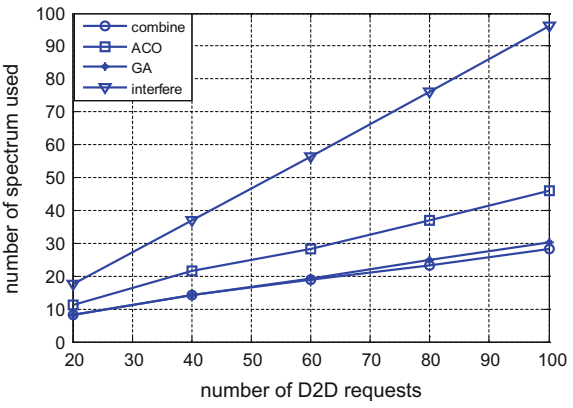



**Fig. 1** The spectrum reuse curves under small number of D2D requests

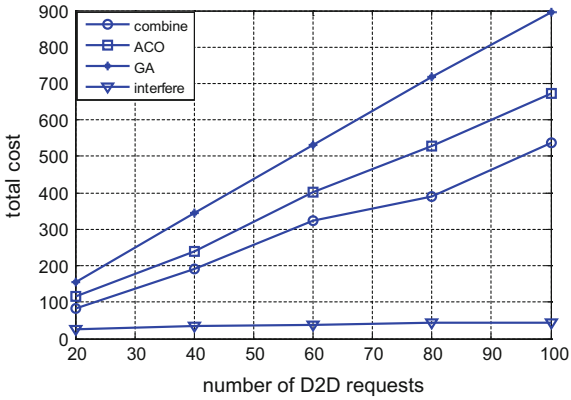



**Fig. 2** The total cost curves under small number of D2D requests

Figures 1 and 2 show spectrum reuse curves and total cost curves of four algorithms respectively when D2D requests number range [20,100]. As shown in Fig. 1, with the number of requests increasing, spectrum reuse efficiency of ACO-GA combined algorithm and traditional genetic algorithm are very close and are better than the other two algorithms. For total cost in Fig. 2, "interfere" curve has remained at very low cost, but this advantage is the product of sacrificing spectrum resource, and "combine" has less total cost compared with "ACO" and "GA".

As the number of D2D requests growing, from Figs. 3 and 4 we can see that "combine" curve displays its advantage in spectrum reuse performance, while in the same way, "interfere" curve shows its superiority in total interference cost performance. But considering the balance of spectrum reuse efficiency and interference cost performance, apparently, ACO-GA combined algorithm exceeds the other three algorithms.

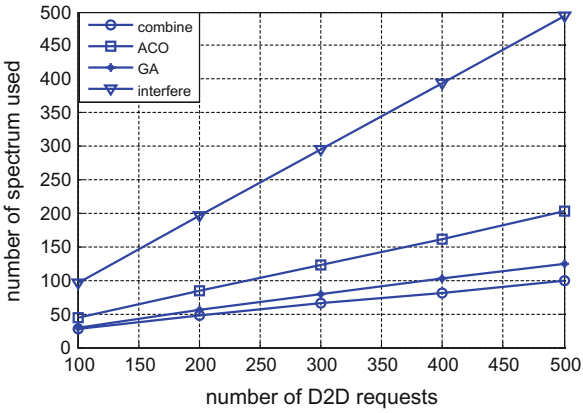



**Fig. 3** The spectrum reuse curves under large number of D2D requests

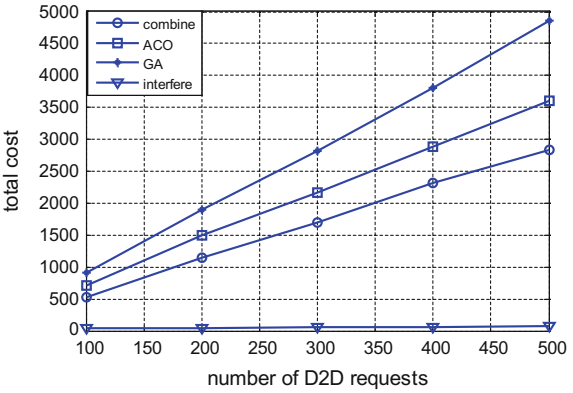



**Fig. 4** The total cost curves under large number of D2D requests

## 5 Conclusion

In this paper, we consider the spectrum allocation problem under the inband overlay D2D communication scenario. With interference only exists among D2D users, we adopt the concept of interference weights and establish the system model. Then ACO-GA combined algorithm is proposed which combine ant colony optimization algorithm and modified genetic algorithm based on the theory of graph coloring. After simulation analysis, ACO-GA combined algorithm performs better on spectrum reuse efficiency and total interference cost than traditional ant colony algorithm and genetic algorithm as well as the ant colony algorithm using interference weights that are proposed in reference [13]. Our work provides a novel idea for spectrum allocation problem in D2D communications.

**Acknowledgements** This work was supported by the National Science and Technology Major Specific Projects of China (Grant No.2015ZX03004002-004) and the Fundamental Research Funds for the Central Universities (Grant No. HIT. NSRIF. 201616).

## References

1. A. Asadi, Q. Wang, V. Mancuso, *A Survey on Device-to-Device Communication in Cellular Networks* (IEEE Commun. Surv, Tutorials, 2014)
2. J. Liu, N. Kato, J. Ma, et al., Device-to-device communication in LTE-Advanced networks: a survey. IEEE Commun. Surv. Tuts. PP(99):1–19 (2014)
3. P. Janis, V. Koivunen, C. Ribeiro, J. Korhonen et al., Interference-aware resource allocation for device-to-device radio underlaying cellular networks, in *69th IEEE Vehicular Technology Conference (VTC Spring)*, Barcelona, Apr 2009, pp. 1–5
4. R. Zhang, X. Cheng, L. Yang, B. Jiao, Interference graph based resource allocation (InGRA) for D2D communications underlaying cellular networks. IEEE Trans. Veh. Technol. (2014)
5. R. Dondi, G. Fertin, S. Vialette, Complexity issues in vertex-colored graph pattern matching. J. Discrete Algorithms **9**(1), 82–99 (2011)

6. S.K. Pal, S.S. Sarma, Graph coloring approach for hiding of information. Procedia Technol. **4**, 272–277 (2012)
7. L. Ouerfelli, H. Bouziri, Greedy algorithms for dynamic graph coloring. International Conference on Communications. Comput. Control Appl. (2011)
8. N. Zhao, Z. Wu, Y. Zhao et al., Ant colony optimization algorithm with mutation mechanism and its applications. Expert Syst. Appl. **37**(7), 4805–4810 (2010)
9. L.E. Agusti'n-Bias, S. Salcedo-Sanz, S. Jimenez- Fernandez, L. Carro-Calvo, J. DelSer, J.A. Portilla- Figueras. A new grouping genetic algorithm for clustering problems. Expert Syst. Appl. **39**(10), 9695– 9703 (2012)
10. B.B. Mabrouk, H. Hasni, Z. Mahjoub, On a parallel genetic-tabu search based algorithm for solving the graph coloring problem. Eur. J. Oper. Res. **197**(3), 1192–1201 (2009)
11. T. Zheng, Y. Qin, D. Gao, H. Zhang, A novel channel assignment method using tabu search based on graph theory in wireless sensor networks, in *3rd IEEE International Conference on Broadband Network and Multimedia Technology* (2010)
12. Z.H. Che, Clustering and selecting suppliers based on simulated annealing algorithms. Comput. Math Appl. **63**(1), 228–238 (2012)
13. E. Liotou, D. Tsolkas, N. Passas, L. Merakos, Ant colony optimization for resource sharing among D2D communications, in *International Workshop on Computer Aided Modeling & Design of Communication Links & Networks*, (2014), pp. 360–364

# The Requirement for Mobile Relay Nodes Under Highway Scenarios

**Chen-Guang He, Kai-Yu Zhang, Yu-Long Gao and Wei-Xiao Meng**

**Abstract** As we access data widely in our daily life, the present situation is that there is a lot of demand for data access in the process of moving while there used to access data in the static context. Under this situation, using the mobile relay nodes (MRN) in conventional cellular communication to support auxiliary transmission communication can effectively improve the ability of data transmission in the process of mobile performance. In this paper, we discuss that under the driving on the highway scenario, with the using of mobile relay nodes, we chose the whole vehicle as the communication transfer object and the middle vehicle carry as mobile relay node (MRN). Under the condition of considering the link capacity, this paper analyzes the traffic flow with the demand of the mobile relay nodes (MRNs) and conveys a digital expression. Simulation and probability analyses indicate that for three different traffic flows, it is expounded to the change of traffic to the mobile relay nodes (MRNs) demand changes.

**Keywords** Highway communication · Vehicle-to-vehicle · Relay node · Assist transmission

C.-G. He · K.-Y. Zhang (✉) · Y.-L. Gao · W.-X. Meng
Communications Research Center, Harbin Institute of Technology, Harbin, China
e-mail: zhangkaiyuwoniu@163.com

C.-G. He
e-mail: hechenguang@hit.edu.cn

Y.-L. Gao
e-mail: ylgao@hit.edu.cn

W.-X. Meng
e-mail: wxmeng@hit.edu.cn

C.-G. He · W.-X. Meng
Key Laboratory of Police Wireless Digital Communication,
Ministry of Public Security, Beijing, People's Republic of China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_14

## 1 Introduction

Now in our daily life, while the dependent of communication network is gradually increasing, the results then lead to the daily data traffic increasing gradually. For these extra data accesses, we need to rebuild more communication facilities, such as building more base station (BS) to improve the utilization rate of data transmission. But because urban planning pattern has fixed in our society, it is hard to improve the transfer ability of communication through rebuild the construction of macro-base station. To solve these facing problems, the concept about relay nodes (RNs) to support auxiliary transmission is presented, while it is also known as the single-hop transmission. The function of relay nodes (RNs) is to enhance the ability of communication in the base station coverage [1]. Follow to different application environments, the mature small cells include femtocell for indoor, microcell for outdoor, and picocell for both. Those small cells act as the extension to 3G/4G macro-cellular to provide customer better wireless internet and voice data service at a lower overhead. Relay nodes (RNs) are divided into two kinds, mobile relay nodes (MRNs) and fixed relay nodes (FRNs). Similar to their name, they are respectively applied to improve transmission communication ability in the process of moving and stationary.

The remainder of the paper is organized as follows. Section 2 reviews related work in the literature. Section 3 derives the system model and analyzes the probability of the MRN requirement. Simulation and numerical results are provided in Sect. 5. Section 6 proposes the conclusion.

## 2 Related Work

Recent years, the fixed communication can be compensated through wire transmission; however, the requirement of mobile communications capacity is growing. In addition to the normal utilization, using the vehicles is also increasing, for example, car and subway train. The utilization of communication transmission system in these media still exists some problems and application prospects. First, the car adopts wireless transmission channel in the process of high-speed moving, while the channel model is the multi-path random channel. When the speed of the mobile station is too high, the receiving signal carrier occurs Doppler shift which will produce distortion and inter-symbol interference. Though the working relay node still have a Doppler shift in the process of the auxiliary transmission, relative to the each other users affected by Doppler shift, the application of the relay nodes is obvious to improve this problem [2]. Second, as the users are surrounded in the vehicular, the signal accomplishes the direct transmission process through the wireless penetrate of the vehicles, resulting in vehicular penetration loss (VPL). To reduce the energy consuming, the technology of vehicular crowd cell exploits MRN which derive from 3GPP R12 [3]. These two points mean it is need more capacity

to complete transmission, so it is essential to utilize relay to compensate the loss of transmission [4]. Third, the on-board network is also a trending topic under the situation of the highway. According to the IEEE- and ASTM-adopted Dedicated Short Range Communication (DSRC) standard, a lot of reference about highway cooperative collision avoidance is showing up, which is available to enhance the security in the daily life. Thus, to improve the vehicular transmission performance, we discuss the highway scenario in this paper, with using vehicular-to-vehicular (V2 V) model as the communication model, to analyze the quantity of MRNs.

## 3 System Model

In this section, there is a simple highway model, where we only consider the single-track vehicles, and all the vehicles have a constant speed v m/s.

In this model, as Fig. 1 shows, there are n exit/entrance in the highway, and in each exit/entrance, the vehicles new into the entrance obey the Poisson distribution, where the distribution of average is $\lambda_n$ (n is the nth exit/entrance, $1 \leq n$ and n as an integer). In the (n + 1)th section connecting, we assume that the vehicles from the nth to the (n + 1)th account for $\alpha_n$ of all the nth vehicles, and deduce the probability of the vehicles pulling away $1 - \alpha_n$. Similarly, we can deduce all the vehicles driving on the nth which are also object to a Poisson distribution [5], where the distribution of average is $C_{n+1}$:

$$C_{n+1} = \lambda_{n+1} + \sum_{n=1}^{n} \lambda_n \times \alpha_n. \tag{1}$$

Also, all the vehicles $T_n$ in each highway denote:

$$T_n = \frac{C_n}{v} \times d_n \tag{2}$$

Molisch [6] in this paper, we choose D = 100 m as the working coverage of the MRN to calculate the transmit power. Because of the $d_{break}$ existing, the transmit power is

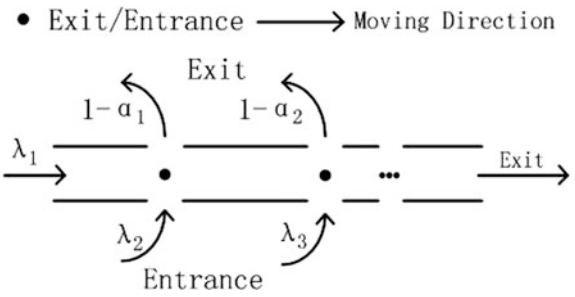



**Fig. 1** The model of the highway with many exit/entrances

$$\mathrm{P}_{ri} = P_{ti} \times cL_i^{-2}, \tag{3}$$

where $\mathrm{P_r}$ is the receive transmit power (considering the $\mathrm{P_r}$ of the each user is constant); $\mathrm{P_t}$ is the transmit power of MRN; i denotes the ith vehicle in the working coverage. Also, $P_{ti} = \frac{\mathrm{P}_{ri} \times L_i^2}{c}$.

Similarly, the working coverage (D) of MRN is 100 m, so we can deduce all the vehicles moving in the working coverage (D) which is also object to a Poisson distribution, where the distribution of average is

$$A_n = \frac{C_n}{v} \times 100. \tag{4}$$

Though we consider that the quantities of the vehicles obey Poisson distribution in working coverage of the mobile relay nodes (100 m), but the position of these vehicles obeys uniform distribution, which means the location of the vehicles is interval distribution in the coverage. So the probability distribution for $\mathrm{P_r}$ is expressed as

$$\Pr(x; m) = \frac{1}{m} \tag{5}$$

(m denote the quantity of the vehicles in the scope of the mobile relay nodes).

With the consideration of the different situations, we can assume three conditions about the deployment of MRNs:

(1) There is no need for the MRN in the working coverage;
(2) There needs only 1 MRN in the working coverage;
(3) There needs at least 2 MRNs in the working coverage.

## 4 Probability Analysis

### *4.1 No MRN Existing*

Because there is no MRN for auxiliary transmission, which means there is no vehicle existing in the scope of this 100 m section, so the probability of no vehicle expression is

$$\Pr\{k = 0\} = \frac{A_n^0}{0!} e^{-A_n} = e^{-A_n}. \tag{6}$$

## 4.2 Only One MRN Existing

When we consider the only one mobile relay node (MRN) in the model, it can be continued to maintain the utilization of the MRN until the critical stability limit in the working coverage. When we assume the critical stability limit is m, we can calculate the probability as follows:

$$\Pr\{0 < k \le m\} = \frac{A_n^1}{1!}e^{-A_n} + \frac{A_n^2}{2!}e^{-A_n} + \frac{A_n^3}{3!}e^{-A_n} + \cdots = \sum_{k=1}^{m}\frac{A_n^k}{k!}e^{-A_n}. \quad (7)$$

After the probability formula is confirmed, there are two kinds of circumstances about the value of the m as follows: $P_r$ denotes the receive power of the vehicle and $P_{tsum} = \varepsilon$ is the peak efficiency rating, namely $\varepsilon$.

(1) m is the odd number which is marked is $m_1$, namely $m_1 = 2k_1 + 1$. So the MRN is on the (k + 1)th vehicle, and the distance of the each vehicle is $L_i = \frac{D}{m-1} \times |k_1 + 1 - i|$.

So the probability of $P_{tsum}$ is

$$P_{tSUM} = \sum_{i=1}^{m} P_{ti} = \sum_{i=1}^{m}\frac{P_r \times L_i^2}{c} = \frac{P_r}{c}\sum_{i=1}^{m}(k_1 + 1 - i)^2 = \frac{D^2 \times P_r \times k_1(k_1+1)(2k_1+1)}{12ck_1^2} = \varepsilon \quad (8)$$

$$2k_1 + \frac{1}{k_1} = \frac{12c \times \varepsilon}{D^2 \times P_r} - 3 \quad (9)$$

subject to $k_1$ is a integer a number

(2) m is the even number which is marked as $m_2$, namely $m_2 = 2k_1 + 1$. So the MRN is on the kth vehicle, and the distance of the each vehicle is $L_i = \frac{100}{m-1} \times |k - i|$

$$P_{tSUM} = \sum_{i=1}^{m} P_{ti} = \sum_{i=1}^{m}\frac{P_r \times L_i^2}{c} = \frac{P_r}{c}\sum_{i=1}^{m}(k_2 - i)^2 = \frac{D^2 \times P_r \times (2k_2^3 + k_2)}{12ck_2^2} = \varepsilon \quad (10)$$

$$2k_2 + \frac{1}{k_2} = \frac{12c \times \varepsilon}{D^2 \times P_r} \quad (11)$$

subject to $k_2$ is a integer.

As shown in (1) and (2), we can calculate the $k_1$ and $k_2$ values. Under the condition of such values, we can obtain the final result $m_{max}$, $m_{max} = Max\{m_1, m_2\}$.

So

$$\Pr\{0 < k \le m_{\max}\} = \frac{A_n^1}{1!}e^{-A_n} + \frac{A_n^2}{2!}e^{-A_n} + \frac{A_n^3}{3!}e^{-A_n} + \cdots = \sum_{k=1}^{m_{\max}} \frac{A_n^k}{k!}e^{-A_n}. \quad (12)$$

### 4.3 At Least Two MRN Existing

From Sect. 2, we could obtain the $m_{max}$, which can maintain the maximum in the mobile relay node. So, when parts of the vehicle are more than maximum, there is need to add additional mobile relay nodes to guarantee the performance of the transmission.

First, we only can consider at least one extra relay probability is expressed as

$$\Pr\{m_{\max} < k\} = 1 - (\frac{A_n^0}{0!}e^{-A_n} + \frac{A_n^1}{1!}e^{-A_n} + \frac{A_n^2}{2!}e^{-A_n} + \cdots) = 1 - \sum_{k=0}^{m_{\max}} \frac{A_n^k}{k!}e^{-A_n}. \quad (13)$$

So we can analyze the next step:

Step 1 Two mobile relay nodes, namely $D_1 = \frac{D}{2}$ instead of D in Sect. 2, so the existing probability is

$$\begin{aligned}\Pr\{m_{1\max} < k \le m_{2\max}\} &= \frac{A_n^{m_{1\max}+1}}{(m_{1\max}+1)!}e^{-A_n} + \frac{A_n^{m_{1\max}+2}}{(m_{1\max}+2)!}e^{-A_n} + \cdots + \frac{A_n^{m_{2\max}}}{m_{2\max}!}e^{-A_n} \\ &= \sum_{k=m_{1\max}+1}^{m_{2\max}} \frac{A_n^k}{k!}e^{-A_n}.\end{aligned} \quad (14)$$

Step 2 Three mobile relay nodes, namely $D_3 = \frac{D}{3}$ instead of D in Sect. 2, so the existing probability is

$$\begin{aligned}\Pr\{m_{2\max} < k \le m_{3\max}\} &= \frac{A_n^{m_{2\max}+1}}{(m_{2\max}+1)!}e^{-A_n} + \frac{A_n^{m_{2\max}+2}}{(m_{2\max}+2)!}e^{-A_n} + \cdots + \frac{A_n^{m_{3\max}}}{m_{3\max}!}e^{-A_n} \\ &= \sum_{k=m_{2\max}+1}^{m_{3\max}} \frac{A_n^k}{k!}e^{-A_n}.\end{aligned} \quad (15)$$

By that analogy,we can obtain the following:

Step 3 There is Z mobile relay nodes, namely $D_z = \frac{D}{Z}$ instead of D in Sect. 2, so the existing probability is

$$\begin{aligned}\Pr\{m_{(z-1)\max} < k \le m_{z\max}\} &= \frac{A_n^{m_{(z-1)\max}+1}}{(m_{(z-1)\max}+1)!}e^{-A_n} + \frac{A_n^{m_{(z-1)\max}+2}}{(m_{(z-1)\max}+2)!}e^{-A_n} + \cdots + \frac{A_n^{m_{z\max}}}{m_{z\max}!}e^{-A_n} \\ &= \sum_{k=(z-1)\max+1}^{m_{z\max}} \frac{A_n^k}{k!}e^{-A_n}.\end{aligned} \tag{16}$$

We think the upper limit is z MRNs, so the else probability is

$$\Pr\{m_{z\max} < k\} = 1 - \sum_{k=0}^{m_{z\max}} \frac{A_n^k}{k!}e^{-A_n} = t. \tag{17}$$

If the probability over than t = $10^{-6}$, we do not consider rebuild another MRN.

## 5 Simulation Results

Considering the third entrance (n = 3), we assume and analyze the demand of MRN under the following situation as shown in Table 1. We can get the result $m_{1\max} = 40$; $m_{2\max} = 166$; $m_{3\max} = 376$, and assume the three cases as shown in Table 2.

**Table 1** The main parameters

| Parameter | Values | Parameter | Values |
|---|---|---|---|
| $P_r$ | 50 dBm | c | $3.5 \times 10^{-4}$ |
| v | 20 m/s | $\alpha_1$ | 0.6 |
| $\varepsilon$ | 1 w | $\alpha_2$ | 0.6 |

**Table 2** The three case parameters

| | $\lambda_1$ | $\lambda_2$ | $\lambda_3$ |
|---|---|---|---|
| Case 1 | 0.1 | 0.15 | 0.2 |
| Case 2 | 1.5 | 1.5 | 1.5 |
| Case 3 | 2.8 | 3 | 3.2 |

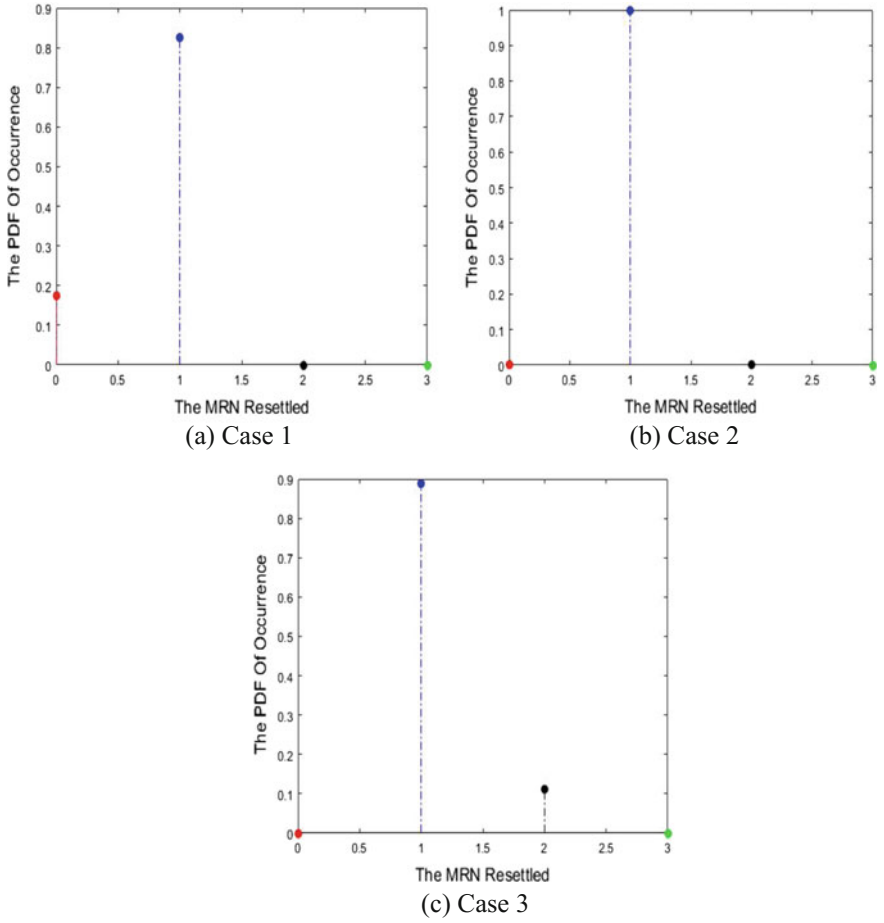



**Fig. 2** The PDF of occurrence in three case

Case 1: In the light traffic time, less vehicle volume pass. As shown in Fig. 2a, in most cases (82%), there all need a MRN which can complete the transmission, but in a few cases (18%), there are no vehicles through the continuous range.

Case 2: In the normal traffic time, normal vehicle volume pass. As shown in Fig. 2b, in nearly all cases (99%), there all need a MRN which can complete the transmission

Case 3: In the heavy traffic time, more vehicle volumes pass. As shown in Fig. 2c, in most cases (89%), there all need a MRN which can complete the transmission, but in a few cases (11%), there need two vehicles to support the transmission.

Comparing the three figure, in all case, one MRN is enough to support the transmit mission while each PDF of using one MRN is the highest (82, 99, 89%).

## 6 Conclusion

In this paper, we discussed transmission models of vehicle carrier user in highway and analyzed the changes of vehicular flow to the distribution of MRN. Especially, we know that vehicular traffic flow is related to the position the MRN. The simulation and numerical result shows the one MRN is nearly efficient to complete the transmission process under three different situations as we discussed above. With the change of traffic flow, the demand of MRN will change slightly. Therefore, how to find out the traffic flow at different times in the highway in order to enhance the coverage of the transmission is the focus of the future research.

**Acknowledgements** This work was supported by the National Science and Technology Major Specific Projects of China (Grant No. 2015ZX03004002-004) and "the Fundamental Research Funds for the Central Universities" (Grant No. HIT. NSRIF. 201616).

## References

1. S. Chen, J. Zhao, The requirements, challenges and technologies for 5G of terrestrial mobile telecommunication. IEEE Commun. Mag. **52**(5), 36–43 (2014)
2. Y. Sui et al., The energy efficiency potential of moving and fixed relays for vehicular users, in *38th IEEE Vehicular Technology Conference*, *1988* (2013) pp. 1–7
3. 3GPP TR 36.836, Technical specification group radio access network; mobile relay for evolved universal terrestrial radio Access (E-UTRA). Technical Report, http://www.3gpp.org/. Accessed 20 Nov 2012
4. S. Biswas, R. Tatchikou, F. Dion, Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety. IEEE Commun. Mag. **44**(1), 74–82 (2006)
5. Y. Wang, J. Zheng, A connectivity analytical model for a highway with an entrance/exit in vehicular ad hoc networks, in *IEEE International Conference on Communications* (IEEE, 2016)
6. A. F. Molisch, *Wireless Communications*, 2nd edn. (Wiley, West Sussex, UK, 2010)

# Research of Improved ALOHA Anti-collision Algorithm in RFID System

**Ye Tian and Hui Kang**

**Abstract** Dynamic time slot frame ALOHA algorithm is currently the most widely used anti-collision technology in radio frequency identification (RFID) system. Based on the traditional ALOHA algorithm analysis, we propose an improved algorithm. The algorithm is based on a packet adaptive ALOHA anti-collision algorithm (PA-ALOHA). First, the reader scans and counts the time slots which are randomly selected by tags, and sends it to each tag. The tags, then, adjust accordingly the time slot that enable the reader to skip idle slots and collision slots, adaptively allocate effective slots, and then identify tags rapidly. The algorithm employs packet and dynamically adjusts the frame size and other strategies, in order to reduce the time of processing slots. Simulation results show that PA-ALOHA algorithm improves the efficiency and stability of the system and reduce the transmission overhead. Especially, when the number of tags is over 1000, the algorithm throughput is still above 70% that has been greatly improved system efficiency than conventional ALOHA algorithm.

**Keywords** RFID · ALOHA algorithm · Packet adaptive ALOHA · Throughput

## 1 Introduction

RFID (Radio Frequency Identification, RFID) is a noncontact two-way data-transmission technique which gets the information of identified object using electromagnetic wave propagation mode between the reader and the tags [1]. RFID system that is widely used in supply chain management, traffic control, and other

Y. Tian (✉)
Zhonghuan Information College Tianjin University of Technology,
Tianjin, China
e-mail: aveling_84@163.com

H. Kang
Deviser Instruments Co., Ltd., Tianjin, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_15

fields usually consists of three major components: electronic tags (tag), reader (reader), and back-end database (database), in order to efficiently and accurately as possible to identify a large number of items [2]. In many applications, it is inevitable to happen two or more tags simultaneously communicate with a reader, and then the signals will interfere with each other, resulting in a collision. The identification efficiency of the RFID system will be reduced when a collision occurs, and therefore, in a dense RFID tag system, often need to adopt an effective anti-collision algorithm to identify multiple tags.

Domestic and international scholars have done a lot of research of collision problem in multi-tag communicate with a reader at the same time [3]. These methods are generally divided into four categories: space division multiplexing method, frequency division multiplexing method, code division multiplexing method, and the time-division multiplexing method [4]. Due to the low power consumption, low storage capacity, and limited computing power and other characteristics of tags, time-division multiplexing method anti-collision methods are mainly used. In time-division multiplexing method, the most commonly used multi-tag anti-collision algorithm is divided into two categories, one is based on binary tree algorithm; its main representative algorithms are binary search algorithm [5], dynamic binary search algorithm [6], leaping dynamic tree algorithm [7], and the query tree algorithm [8]. This kind of algorithm selects a tag to communicate by the reader based on the unique tag ID, so the performance of the search algorithm will sharply deteriorate with ID-digit increase. Another is based on the ALOHA algorithm. The main algorithms include slotted ALOHA(SA) algorithm [9], the frame-slotted ALOHA(FSA) algorithm [10], the dynamic frame-slotted ALOHA (DFSA) algorithm [11], in addition to the EPC Class-1 Generation-2 (EPC C1G2) standard [12] proposed by EPC global, which uses a random frame-slotted algorithm based on a Q value [13] and the collision prediction-ALOHA (CP-ALOAH) algorithm [14] proposed on recent literature, etc. The anti-collision algorithms are relatively uncomplicated; it also does not need to design complex internal circuitry of tag, so the cost of the tag is lower. However, as the number of tags increases, or even thousands, the probability of collision also increases with a sharp decline in the performance of system identification. For the problem of large number of tags to identify, researchers have also proposed the concept of grouping, including enhanced dynamic frame-slotted algorithm, grouped dynamic frame-slotted ALOHA (GDFSA) algorithm [15], etc. Compared with the previous algorithm, throughput performance of packet-type anti-collision method is relatively stable, but the throughput is less than 50% as the same as other algorithms.

In this paper, we propose an improved adaptive ALOHA algorithm and packet adaptive ALOHA (PA-ALOHA) algorithm. The simulation results show that the new algorithm reduces the transmission overhead and improves the recognition efficiency and stability of the system. When the number of tags is greater than 1000, recognition efficiency is still up to 70%. Compared with the EPC standard protocol, transmission overhead declined by about 48%, and with the number of tags increases, the superiority of the new algorithm is more obvious.

## 2 Algorithm Description

### 2.1 The Basic Definition of the Algorithm

Set frame length L. When n tags are identified, each tag will randomly select a time slot to send information of its own identifier. According to the binomial theorem, the probability of any frame time slot occupied by a tag is p = 1/L, so the probability of m tags in the same time slot can be expressed as

$$P(L,n,m) = C_n^m \times \left(\frac{1}{L}\right)^m \times \left(1-\frac{1}{L}\right)^{n-m}, \tag{1}$$

When m = 0, that is to say in the time slot no tags request to be identified, the time slot is called idle time slot; its probability is

$$P_i = P(L,\ n,\ 0) = \left(1-\frac{1}{L}\right)^n. \tag{2}$$

When m = 1, that is to say in the time slot two or more tags request to be identified, the time slot is called collision time slot, and its probability is

$$P_e = P(L,n,1) = \frac{n}{L} \times \left(1-\frac{1}{L}\right)^{n-1}. \tag{3}$$

When $m \geq 2$, that is to say in the time slot only one tag requests to be identified, the time slot is called effective time slot, and its probability is

$$P_c = 1 - P_i - P_e. \tag{4}$$

Then after a period of identification, the expectation of the number of idle time slots is $\varepsilon_i^{L,n}$, the expectation of the number of effective time slots is $\varepsilon_e^{L,n}$, and the expectation of the number of collision time slots is $\varepsilon_c^{L,n}$. They can be expressed as

$$\varepsilon_i^{L,n} = L \times P_i = L(1-\frac{1}{L})^n \tag{5}$$

$$\varepsilon_e^{L,n} = L \times P_e = n(1-\frac{1}{L})^{n-1} \tag{6}$$

$$\varepsilon_c^{L,n} = L \times P_c = L - \varepsilon_i^{L,n} - \varepsilon_e^{L,n}. \tag{7}$$

Within identifying a time frame length, the proportion of the number of tags which the reader transfers effective information is called the throughput of the RFID system. The throughput is denoted as T:

$$T = \frac{\varepsilon_e^{L,n}}{L} = P_e. \tag{8}$$

## 2.2 Optimization Analysis of RFID System Frame Length

Before the start, PA-ALOHA algorithm requires the reader to estimate the remaining the number of tags. Vogt algorithm is used in this paper because of the advantages of this algorithm, which is low error and stable. After its estimate, reader will adjust the frame length number of tags dynamically according to the number of tags, within the readable range. If the frame length is too long, a lot of idle time slot will emerge, and on the contrary will lead to a sharp rise in the collision slot. Both situations will result in reduced efficiency of identification system eventually, so in order to achieve a higher throughput efficiency, the corresponding relationship between the frame length and the number of tags must be found that determine the optimal frame length. Demand Eq. (8) the derivative:

$$\frac{dT}{dn} = \frac{1}{L}\left(1 - \frac{1}{L}\right)^{n-1} + \frac{n}{L}\left(1 - \frac{1}{L}\right)^{n-1} \ln\left(1 - \frac{1}{L}\right). \tag{9}$$

Order of Eq. (8) is equal to 0, and frame length L and the number of labels should be met:

$$n = -\frac{1}{\ln\left(1 - \frac{1}{L}\right)}. \tag{10}$$

The Taylor series expansion of Eq. (10) is

$$L \approx \frac{1 + \frac{1}{n}}{1 + \frac{1}{n} - 1} = n + 1. \tag{11}$$

Equation (11) shows that, when the number of tags n is much greater than 1, and n is close to the frame length L, the system throughput is maximized. Figure 1 shows simulation results of the throughput when fixed values of L are different:

According to Fig. 1, the intersection of the throughput curves of adjacent frame length is the critical point which is to adjust the frame length.

## 2.3 Tag Packet Principle

Due to limited costs, so that the number of frame slots cannot unlimited increase with the number of tags. So in the case for a large number of tags, response number

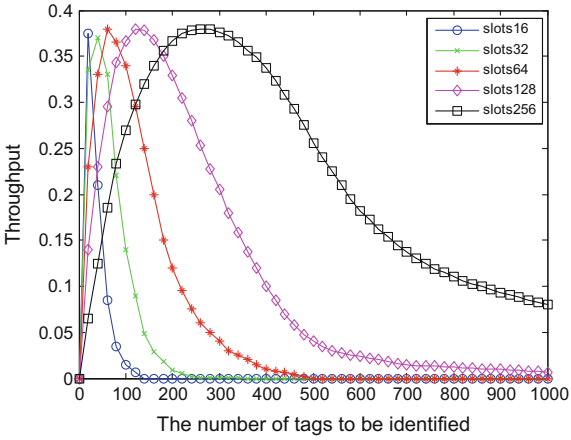



**Fig. 1** Throughput of different L

of tags should be limited, in order to keep the system relatively high throughput. According to Eq. (8), select the threshold of grouping tags. If the length of time slot frames L is 256,

$$\frac{n}{256\times j}\times\left(1-\frac{1}{256}\right)^{n/j-1}=\frac{n}{256\times k}\times\left(1-\frac{1}{256}\right)^{n/k-1}. \tag{12}$$

Wherein, j and k represent the number of neighboring groups, for example, when j = 1 and k = 2, by Eq. (12), we can obtain $n\approx 355$, that is, in order to maintain high efficiency of throughput, the number n of identification tags cannot be greater than 355, when n is greater than 355, the unrecognized tags need to be divided into packets. However, with the increase in the number of packets, slot consumption will also increase and slot time adjustment will be longer. As shown in Fig. 2, when

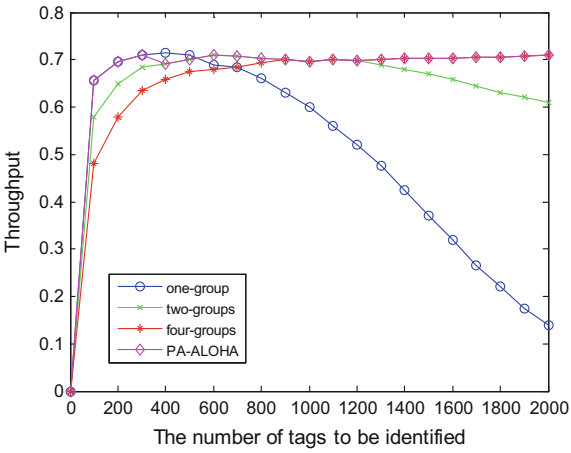



**Fig. 2** Throughput of different packets

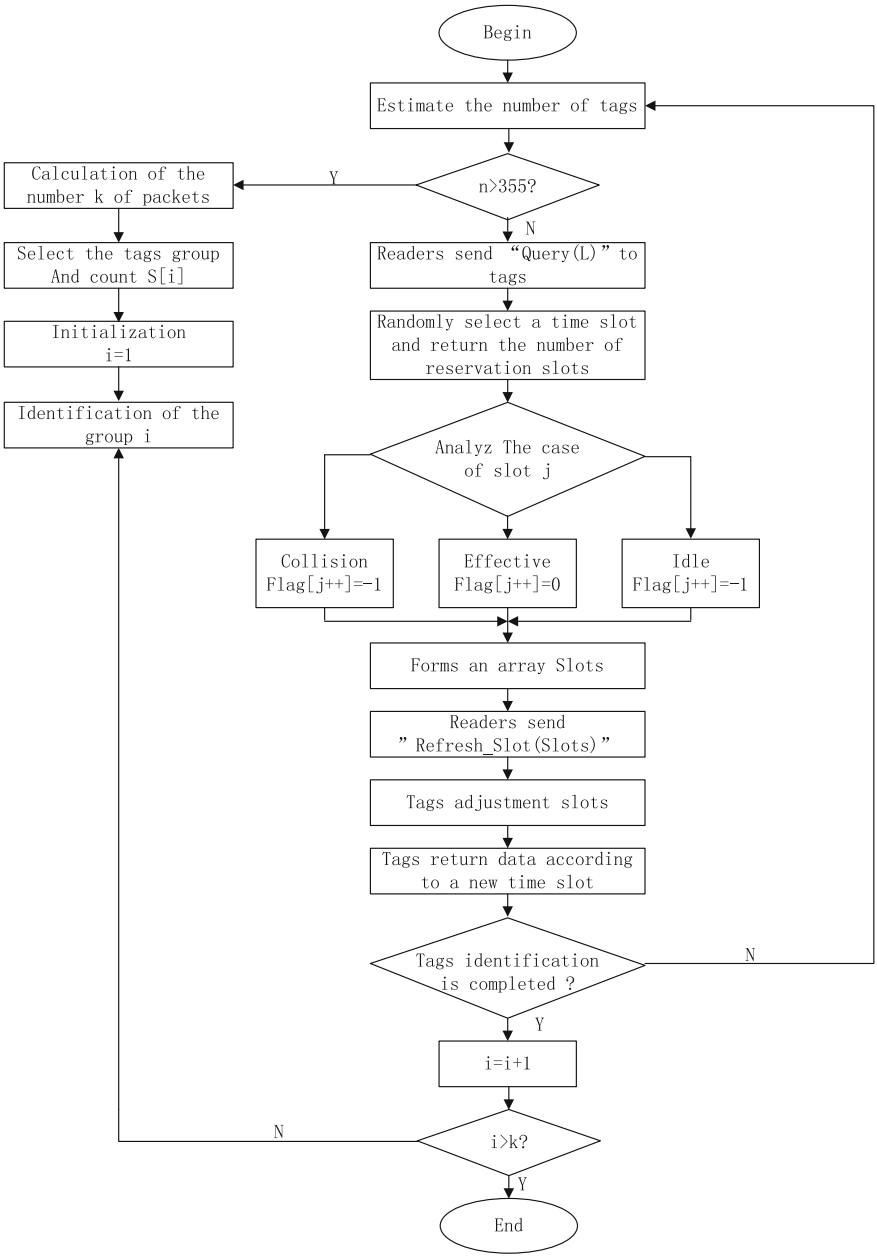



**Fig. 3** Flow chart of PA-ALOHA algorithm

**Table 1** Command and parameter length of PA-ALOHA

| Commands and parameters | Features | Length/bits |
|---|---|---|
| Query | Adjust the frame length | 8 |
| Refresh_slot | Adjust the number of slots | 32 |
| Count | Slot count | 8 |
| Collection | Confirm communication | 18 |
| Sleep | Into a quiet state | 2 |

the number of tags between 0 and 600, with divided into four groups, it reduces the system performance. Therefore, PA-ALOHA algorithm takes adaptive packet; the tags to be identified are grouped according to the number. When the number of tags is large, this policy can solve the problem of low recognition rate, but also make the system performance remained relatively stable, and reduce the time to identify the tag.

### *2.4 Anti-collision Protocol of PA-ALOHA Algorithm*

Before the tag is identified, PA-ALOHA algorithm scans time slots and records the case of the tags reservation first, and then let the reader to skip the collision slots and idle slots, and allocate effective time slot directly. So the tag which has reserved time slot successfully can be identified directly to improve the utilization of slots. The flow chart of PA-ALOHA algorithm is shown in Fig. 3.

The command sequences which appear in the flow chart are shown in Table 1.

## 3 Experimental Results

In order to confirm the validity of PA-ALOHA algorithm, the experiments have been carried out. FSA-256 algorithm, DFSA algorithm, EPC standard algorithm, GDFSA algorithm, and PA-ALOHA algorithm are simulated respectively using MATLAB, in Windows7 operating system, 2G memory environment. Assume tags are well distributed and the number of tags is more than 1500. In order to make the results convincing, the simulation results are taken from the average of 100 experiments in the same conditions.

The first experiment was used to compare reader overhead of various algorithms. Set the number of tags in the interval [0, 2000], the number of bits that is transmitted by reader is shown in Fig. 4. With the increase of the number of tags, reader overhead also increases. When the number of the tags is less than 1300, the reader overheads of adaptive ALOHA algorithm and PA-ALOHA algorithm are almost the same. When the number exceeds 1300, the advantages of PA-ALOHA algorithm began to stand out. When the number of tags is 2000, the reader overhead of

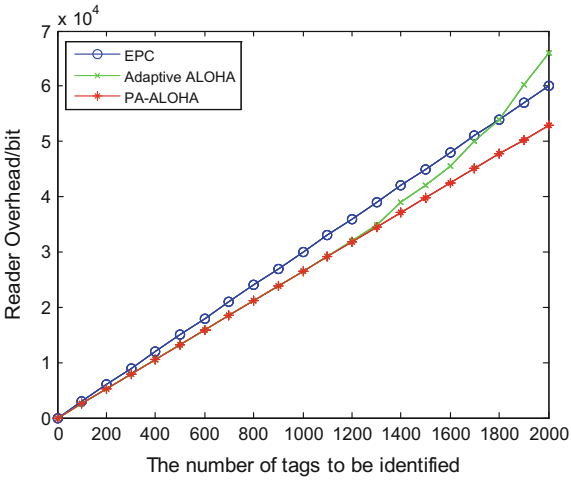



**Fig. 4** Reader overhead

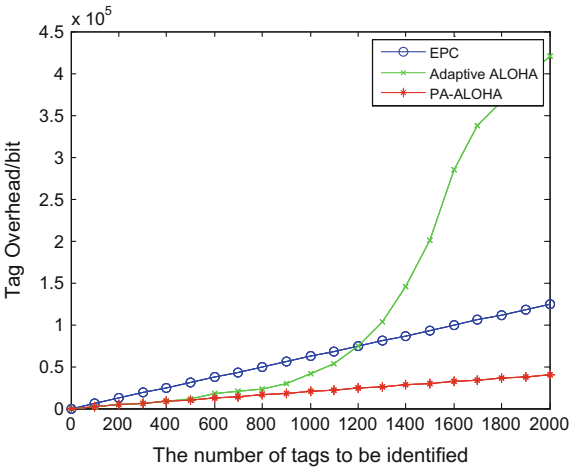



**Fig. 5** Tag overhead

adaptive ALOHA algorithm is 66790, EPC standard algorithm is 60527 and PA-ALOHA algorithm is 54245. Compared with adaptive ALOHA algorithm and ECP standard algorithm, the reader overhead of PA-ALOHA algorithm decreased by 18.8% and by 10.4%, respectively.

The second experiment was used to compare tag overhead of various algorithms. Figure 5 shows tag overhead of three algorithms. In adaptive ALOHA algorithm the number of bits which is transmitted by tags is nearly exponential growth. In ECP standard algorithm and PA-ALOHA algorithm the number of bits which is transmitted by tags is nearly linear growth. However, tag overhead increases the most slowly in PA-ALOHA algorithm. When the number of tags is 2000, the tag overhead of adaptive ALOHA algorithm is 422135, EPC standard algorithm is

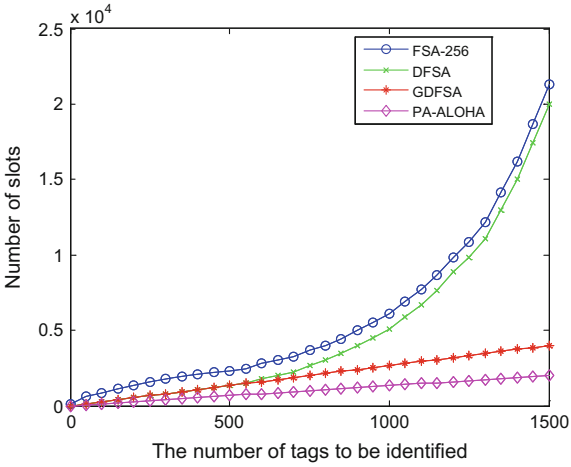



**Fig. 6** Total number of slots

121545 and PA-ALOHA algorithm is 39601. Compare with adaptive ALOHA algorithm and ECP standard algorithm, the tag overhead of PA-ALOHA algorithm decreased by 96.1% and by 67.4%, respectively.

Total time slot is a key factor in determining the efficiency of the system, the less the total number of slots, the better the performance of the system. The simulation results are shown in Fig. 6. Set the number of tags in the interval [0, 1500]. The total number of the slots of FSA-256 algorithm and DFSA algorithm is nearly exponential growth. The total number of the slots of GDFSA algorithm and PA-ALOHA algorithm is nearly linear growth. The total number of slots increases the most slowly in PA-ALOHA algorithm. Especially when the number of tags is large, the advantages of PA-ALOHA algorithm are more obvious. When the number of tags is 1000, PA-ALOHA algorithm only needs about 1400 slots, about 4165 less than FSA-256, 3727 less than DFSA, and 1366 less than GDFSA.

System throughput is also an important measure of system performance. As shown in Fig. 7, when the number of tags is less than 355, the throughputs of GDFSA and DFSA are the same, and FSA-256 algorithm is the lowest, only about 0.2. PA-ALOHA algorithm is the highest, up to 0.7 or more. GDFSA, DFSA, and PA-ALOHA algorithms can allocate effective time slots adaptively base on the actual number of tags, and the FSA-256 algorithm uses a fixed frame length 256. When the tag number is greater than 355, FSA-256, DFSA algorithm decreases rapidly, and GDFSA and PA-ALOHA algorithm will divide tags into multiple, by dynamically adjusting the frame length to identify tag for each group, so that the throughput is stable within a certain range. System throughput of PA-ALOHA algorithm is obviously much larger than the other three algorithms. The throughput of FSA-256 algorithm is between 0.1 and 0.25. The DFSA algorithm is between 0.2 and 0.36, and GDFSA algorithm only maintains at about 0.36. When the number of tags reaches 1000, compare with the FSA-256 and GDFSA, system efficiency of PA-ALOHA algorithm increases by 300 and 97.2%.

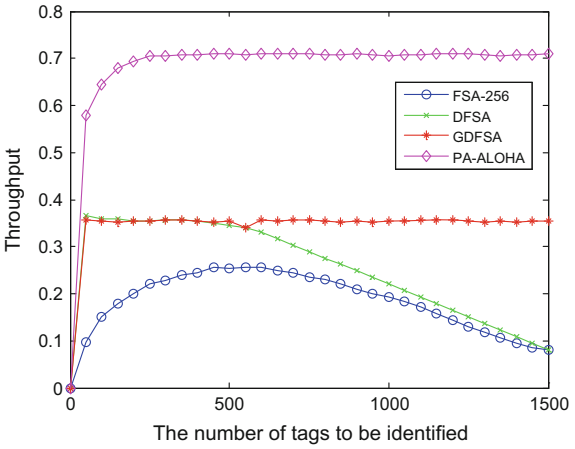



**Fig. 7** Throughput comparison

## 4 Conclusions

This paper presents an adaptive packet ALOHA anti-collision (PA-ALOHA) algorithm. PA-ALOHA can identify tags rapidly by estimating the number of tags and grouping, the slot reservation and other adaptive strategies. The simulation results show that with the increasing number of tags, especially when the number of tags exceeds 1000, throughput of PA-ALOHA algorithm maintains at 0.71 or more. It is substantial increase to compare with the traditional algorithm-based ALOHA. The total number of time slots and transmission overhead remain almost linear increase. The efficiency of RFID systems can be effectively improved. The stability of the system throughput increases and the cost of tags reduces. When the number of tags is large, the advantage of PA-ALOHA algorithm is particularly significant, and has broad application prospects.

## References

1. R. Want, An introduction to RFID technology, in *Proceedings of IEEE Pervasive Computing* (2006), pp. 25–33
2. V.D. Hunt, A. Puglia, M. Puglia, *RFID: A Guide to Radio Frequency Identification* (Wiley, NJ, 2007)
3. H. Wu, Bayesian tag estimate and optimal frame length for anti-collision ALOHA RFID system. IEEE Trans. Autom. Sci. Eng. 963– 969 (2010)
4. W. Ya-qi, J. Guo-ping, Anti-collisionalgorithm based on grouping mechanism and jumping dynamic binary. Acta Automatica Sinica 1390–1400 (2010)
5. C.N. Yang, J.Y. He, An effective 16-bit random number aided query tree algorithm for RFID tag anti-collision. IEEE Commun. Lett. 539– 541 (2011)
6. G. Liva, Graph-based analysis and optimization of contention resolution diversity slotted ALOHA. IEEE Trans. Commun. 477–487 (2011)

7. Wu H, "Efficient framed slotted ALOHA protocol for RFID tag anti-collision. IEEE Trans. Autom. Sci. Eng. 581–588 (2011)
8. H.W. Tsao, Optimaldynamic framed slotted ALOHA based anti-collision algorithm for RFID systems. Wireless Pers. Commun. **59**, 109–122 (2011)
9. Z. Xiao-hong, Research on RFIDanti-collision algorithm of slot responding in real-time and co-processing. Acta Electronica Sinica **42**, 1139–1146 (2013)
10. C.Y. Wang, C.C. Lee, M.C. Lee, An enhanced dynamic framed slotted ALOHA anti-collision method for mobile RFID tag identification. J. Convergence Inf. Technol. **6**, 340–351 (2011)
11. C.F. Lin, Efficient estimation for dynamic frame-slotted ALOHA in RFID networks. IEEE Trans. Autom. Sci. **7**, 840–848 (2010)
12. C.H. Chen, Mutual authentication protocol for RFID conforming to EPC class-1 generation-2 standards. Comput. Stand. Interfaces **29**, 254–259 (2007)
13. P. Yu, Reducing tag collision in radio frequency identification systems by using a grouped dynamic frame slotted ALOHA algorithm. Acta. Phys. Sin. 484–492 (2013)
14. Wu Hai-feng, Yu. Zeng, Tagestimate and fame length for dynamic frame slotted ALOHA anti-collision RFID system. Acta Automatica Sinica **36**(7), 620–624 (2010)
15. W.T. Chen, An accurate tag estimate method for improving the performance of an RFID anti-collision algorithm based on dynamic frame length ALOHA. IEEE Trans. Autom. Sci. Eng. **6**, 9–15 (2009)

# Performance Analysis for Multiple Tags Inventory in RFID

**Li Wang, Wenyuan Tao, Weibo Hu and Jiwei Song**

**Abstract** This paper analyzed and contrasted multiple tags inventory performances of the two international air Interface standards and Chinese national standard in RFID. Given the innovation of technology and design ideas of the Chinese national standard, the problems of inventory were analyzed in the three standards. Afterwards, the performance was simulated by software and hardware platforms, and the performance of Chinese national standard [1] is better than EPC (EPCglobal Class-1 Generation-2 UHF RFID Protocol) [2] or ISO 18000–6B (Information Technology-Radio Frequency Identification for Item Management-Part 6: Parameters for Air Interface Communications at 860–930 MHz) [3].

**Keywords** RFID · Anti-collision algorithm · Chinese national standard

## 1 Introduction

The RFID system consists of Radio Frequency tags, reader, and middleware, an application system. The reader communicated with tags through the air interface protocol. The air interface protocol is the crucial technology to the RFID system [4].

The air interface protocol is separated into a physical layer and a MAC layer. The physical layer defines the physical parameters of the interface between reader and tag, while the MAC layer specifies the operation process and order between

L. Wang (✉) · W. Tao (✉)
School of Computer Software, Tianjin University, Tianjin, China
e-mail: wangli@tju.edu.cn

W. Tao
e-mail: taowenyuan@tju.edu.cn

W. Hu
Tianjin Electronic Information College, Tianjin, China

J. Song
China Electronics Standardization Institute, Beijing, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_16

reader and tag. The radio frequency identification has many tags in the same reader communication field in application scenes. In order to inventory tags, multiple access technology is needed in the anti-collision algorithm. Section 2 of this paper introduces the anti-collision algorithm in the Chinese nation standard, which is designed by the team that is led by the author. Afterwards, Sect. 3 presents simulation results and the experiment results.

## 2 Anti-Collision Algorithm [5, 6]

- Binary tree algorithm in 6B

The anti-collision algorithm in 6B is a binary tree-splitting algorithm. The core idea is to split the collision tags into two subsets of 0 and 1. Then the "0" subset is queried first. If the slot does not contain a collision, then the reader identified the tags correctly. If there is a collision, then the tags of "0" subset are separated into "00" and "01" subsets. After identifying all tags in subset "0," then query subset "1" in the same method.

- ALOHAL algorithm in EPC [7, 8]

The EPC protocol is an improved ALOHA algorithm. In EPC, the tags are divided into several discrete time slots. The tags randomly select one of these slots to send data. The range that tag selected is dynamically changing in the anti-collision process by adjusting the Q value.

- GB algorithm

The anti-collision algorithm in GB is DDS-BT. There are some improvement ideas such as dispersing, shrinkage, and re-splitting.

The anti-collision algorithm in GB is improved by using the following ideas:

- Pre-splitting: In the binary tree algorithm, when there is a collision, only the 0 tags are separated into 0 and 1. At the beginning, the slot counters all tags that are 0. There is a large number of collisions when there are many tags. Pre-splitting is designed to solve this question. If several consecutive slots are designated for collision, then a large amount of tags are in the same slot. Then the reader sends the disperse command to divide all tags at the same time. Then the number of collision slots is reduced.
- Shrinking: Pre-splitting is helpful for concentrated tags. However, it may establish a large number of idle slots. Then shrinking is introduced to reduce the number of idle time slots by simultaneously reduce idle slots.
- Re-splitting: In the binary tree algorithm, when two tags both counter 0, then a collision occurs. The reader sends a Fail command to divide the tags. Then the two tags generate random numbers. There is 25% probability that two random numbers are both 1, and if so, then the next slot is idle and the reader will send a Success command. If these two tags become counter 0 and collision again, then

this process the Success command is actually invalid. If there is no response in the next slot after collision, it means that all colliding tags are counter 1. Then it should directly split tags with counter 1.

- Answering data structure: In the binary tree algorithm, each answering datum is the whole tag ID. As a result, there is too much interaction information. In EPC protocol, each answering datum is RN16. That can reduce the interaction information to improve the system speed, but another problem is that the answering data is not checked. Transmission errors need a few processes to solve so the GB is designed to improve the answering data structure. The data is 16bit with the CRC check data, and the data structure is in RN11 + CRC5 mode.
- Ending algorithm: It gives a ending algorithm in GB. The reader tracks the largest number of tags countered when counting tags. When the biggest number of tags counter is zero, all tags are inventoried in the communication field.

Using these algorithms, the GB is better than EPC and 6B when inventory tags which is analyzed in Sect. 3.

## 3 Experimental Classification Results and Analysis

The evaluation for anti-collision is inventory speed. This is the number of inventory tags in the unit time. Inventory speed is related to the anti-collision algorithm, but is also related with the link rate, tag answering data structure, and state machine. The evaluation for anti-collision algorithm without other factors is inventory throughput. Inventory throughput is the rate between valid slots and total slots.

### *3.1 Evaluation Method*

There are two methods of evaluation. One is software [9] and the other is hardware.

Software simulation platform: The Matlab on the PC simulates the air interface between reader and tag and obtains the throughput and speed. The number of tags is from 10 to 1000 with interval of 10, and the software is performed 1000 times. The inventory speed is related with link rate and other communication parameters. The simulation parameters obtained are listed below, and the minimum pulse width of forward coding is 12.5 us and the backward data rate is 128 kbps.

(A) **EPC:**

| | | |
|---|---|---|
| Pulse width: 12.5 (us) | RTcal = 2.500 | T1 = 78.100 (us) |
| Tari = 25 (us) | RTcal = 62.500 (us) | T2 = 78.100 (us) |
| Data-1:1.5 Tari | Forward clock: 31.25 (us) | T3 = 0.000 (us) |
| Forward data rate = 32.0 (kbps) | TRcal = 166.6 (us) | T4 = 125.00 (us) |
| Backward data rate:128 (kbps) | DR = 64/3 | Q init = 4 |
| TRext = 0 | ΔQ = 0.300 | |

(B) **6B:**

| Pulse width: 12.5 us | Forward data rate: 40 kbps | Backward data rate: 128 kbps |
|---|---|---|

*Note*: The backward data rate is assumed to be 128 kbps in order to be consistent with the other protocol.

(C) **GB**

| Tc: 12.5 (us) | Backward data rate: 128.00 (kbps) | T1 = 78.100 (us) |
|---|---|---|
| Forward data rate: 45.71 (kbps) | Tpri: 7.81 (us) | T2 = 78.100 (us) |
| Forward clock: 21.875 (us) | Backward coding: FM0 | T3 = 0.000 (us) |
| CIN = 4 | TRext = 0 | T4 = 125.00 (us) |
| CCN = 3 | | |

Hardware platform: The reader and tags are designed to communicate in three protocols that focus on the inventory. The experiment is done in an indoor environment. The inventory time and all information are recorded. The inventory throughput and speed are calculated later. There are 100 tags for each protocol. The number of tag in experiment ranges from 10 to 100 with interval of 10, and the experiment is performed 1000 times.

The reader is shown as follows (Fig. 1).

The tags are pasted on the glass wall as follows (Fig. 2).

**Fig. 1** The antenna of the reader


**Fig. 2** The tag

## 3.2 *Inventory Throughput*

Inventory throughput is the rate between valid slots and total slots. It is calculated with the following formula:

$$\text{Inventory efficient} = \frac{\text{number of valid slots}}{\text{number of total slots}}$$

Figure 3 shows the graph of the result of inventory throughput by software simulation. The abscissa is the number of tags, which ranges from 10 onto 1000. In addition, the ordinate is the inventory throughput. As shown in Fig. 3, the throughput of GB is the highest, the 6B is second and the EPC is the worst.

When the hardware is tested, the inventory throughput is calculated by analyzing all the commands sent by the reader. The result in hardware experiment is similar with the result in software simulation and is shown in Fig. 4.

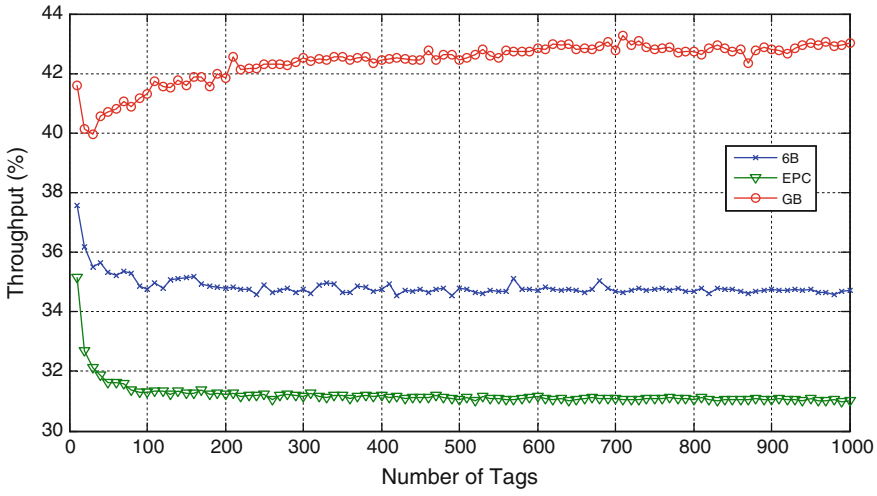



**Fig. 3** Comparison in inventory efficiency of three protocols by software simulation

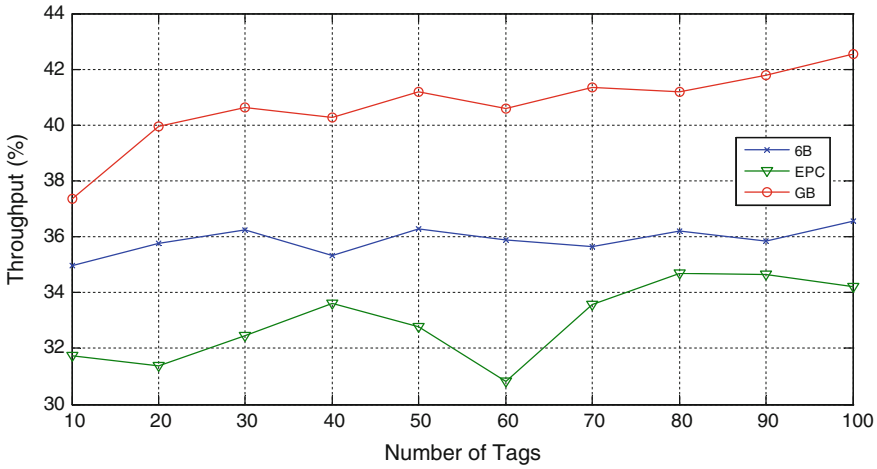



**Fig. 4** Comparison in inventory throughput of three protocols by hardware experiment

## *3.3 Inventory Speed*

The inventory speed refers to the number of tags that are in inventory in a unit time. It is calculated according to the following formula:

$$\text{Inventory Speed} = \frac{\text{Number of tags}}{\text{Inventory time}}$$

Figure 5 shows a graph depicting inventory speed by software simulation. As shown in Fig. 5, the speed of GB is the highest, and the EPC is second, the 6B is the slowest.

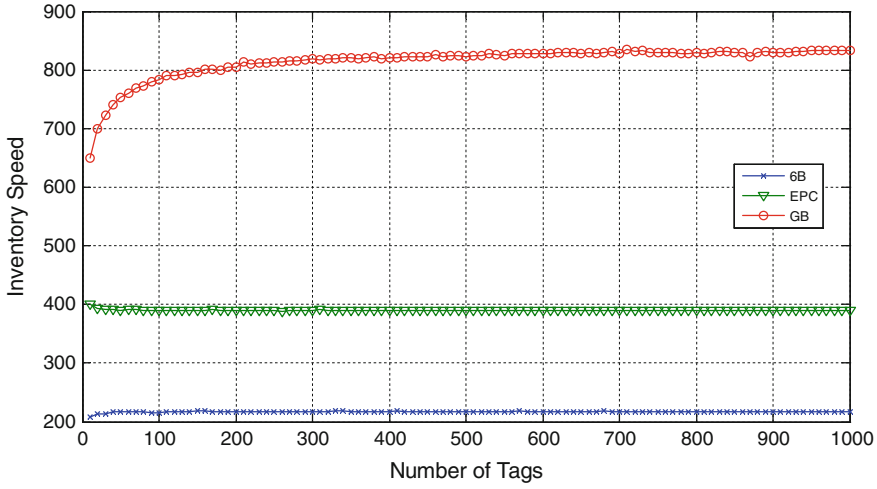



**Fig. 5** Comparison in inventory speed of three protocols by software simulation

After analyzing the results, the main factors of the inventory speed are inventory forward and backward data rate. With the same forward and backward data rate, the results are shown in Fig. 6:

The result of hardware experiment is shown below. The inventory speed of GB is highest, speed of EPC is second, and speed of 6B is the slowest. The experimental results are consistent with simulation (Fig. 7).

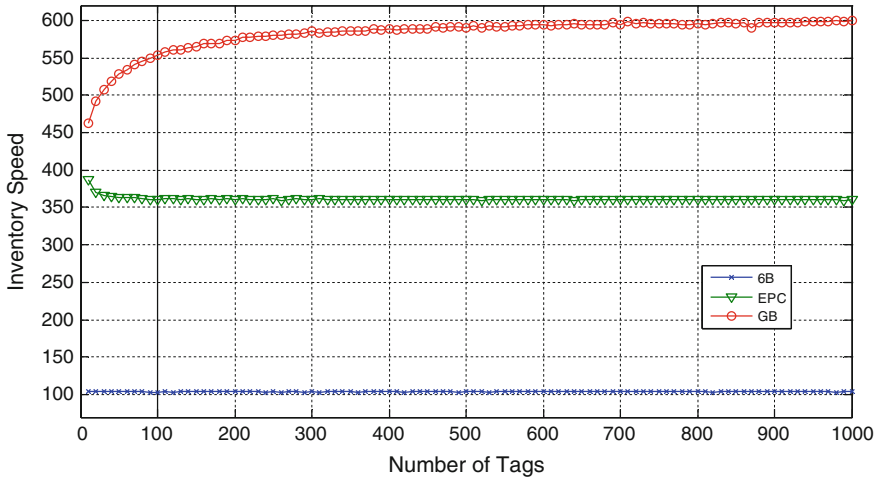



**Fig. 6** Comparison in inventory speed of three protocols by software simulation with same data rate

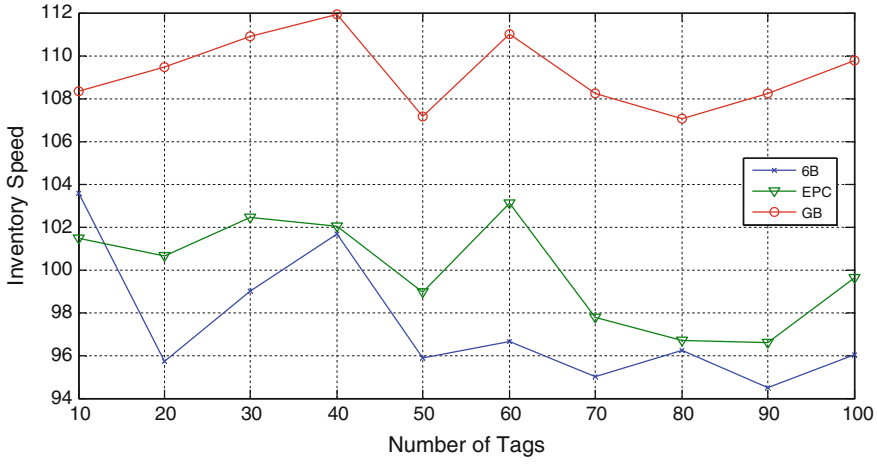



**Fig. 7** Comparison in inventory speed of three protocols by hardware experiment

**Table 1** Complexity of the EPC and GB

| Algorithm | Code line number of simulation | Logic gate of actual chip | Logic gate of actual reader |
|---|---|---|---|
| EPC | 319 | 8000–9000 | 9883 |
| GB | 384 | 8000–9000 | 11,176 |

### *3.4 Complexity*

Unfortunately, there is very little application of 6B. As a result, complexities between EPC algorithm and GB algorithm are compared (Table 1).

## 4 Conclusion

The anti-collision algorithm of GB/T 29768 is more efficient compared with EPC and ISO18000–6B not only by software simulation but also by hardware experiment.

## References

1. GB/T 29768-2013, Information Technology—Radio Frequency Identification Air Interface Protocol at 800/900 MHz
2. EPC Radio-Frequency Identity Protocols Class-1 Generation-2 UHF RFID Protocol for Communications at 860 MHz–960 MHz Version 1.2.0
3. ISO 18000-62 Parameters for air interface communications at 860 MHz to 960 MHz Type B
4. S. Meng, D.K.W. Chiu, E. Kafeza et al., Automated management of assets based on RFID triggered alarm messages. Inf. Syst. Front. **12**(5), 563–578 (2010)
5. K.-H. Yeh, N.-W. Lo, K.-Y. Tsai, A novel RFID tag identification protocol: adaptive n-resolution and k-collision arbitration. Wirel. Pers. Commun. **77**(3), 1775–1800 (2014)
6. Z. Li, C. He, J. Li, Adaptive hierarchical artificial immune system and its application in RFID reader collision avoidance. Appl. Soft Comput. **21**, 119–138 (2014)
7. P. Solic, J. Radic, N. Rozic, Energy efficient tag estimation method for ALOHA-based RFID systems. IEEE Sens. J. **14**(10), 3637–3647 (2014)
8. W.-T. Chen, A fast anticollision algorithm for the EPCglobal UHF Class-1 Generation-2 RFID standard. IEEE Commun. Lett. **18**(9), 1519–1522 (2014)
9. Q. Peng, C. Zhang, Z. Wang, A multi-tag emulator for the UHF RFID system. IEEE Trans. Instrum. Meas. **63**(6), 1461–1469 (2014)

# Ergodic Capacity Upper Bound for Multi-hop Full-Duplex Decode-and-Forward Relaying

**Liang Han**

**Abstract** Full-duplex relaying (FDR) can receive and transmit simultaneously over the same frequency band, thus enabling a significant increase of spectral efficiency and has attracted much research interests. In this paper, we investigate the ergodic capacity of multi-hop decode-and-forward (DF) FDR systems, in which the relay nodes suffer not only from self-interference but also from inter-relay interference. We consider two cases for the inter-relay interference, i.e. one relay node knows perfect channel state information (CSI) from all other relay nodes or only knows the CSI from the previous relay node. The ergodic capacity upper bounds are derived for each case, respectively. Finally, we present numerical results to compare the ergodic capacity of multi-hop FDR with multi-hop half-duplex relaying (HDR).

**Keywords** Multi-hop relaying · Full-duplex · Decode-and-forward · Ergodic capacity upper bound

## 1 Introduction

In multi-hop relaying system, a source communicates with a destination via multiple intermediate relays [1]. Compared with direct transmission, multi-hop relaying can improve system capacity, save transmitter power and extend network coverage [2–4].

So far, most studies on traditional wireless relaying have focused on the half-duplex mode, where each relay receives and retransmits the signal over orthogonal channels. Half-duplex relaying (HDR) can make the system design and implementation simpler, but it will result in significant loss of spectrum efficiency. In recent years, encouraged by the development of self-interference cancellation (SIC) technologies [5–8], full-duplex relaying (FDR), where the relay node receives

L. Han (✉)
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
e-mail: hanliang_tjnu@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_17

and transmits at the same time and over the same frequency band, has drawn significant research interests.

Early works on performance analysis of FDR assumed that the transmit antenna and receive antenna of a relay are perfect isolated and thus the self-interference can be neglected [9]. However, the self-interference cannot be completely cancelled in practice. So far, several works have taken the effect of residual self-interference (RSI) into account. In [10], the authors investigated the outage probability of an infrastructure-based FDR. In [11], a hybrid relaying scheme which could switch between full-duplex and half-duplex mode was proposed. In [12], the authors investigated the outage performance of a selective decode-and-forward (DF) FDR that consider both RSI and direct link. We note that most of the works on FDR systems focused on a single relay case (i.e. dual-hop relaying). Since the relay nodes suffer not only from RSI but also from inter-relay interference in multi-hop FDR, the performance analysis and design of multi-hop FDR is more difficult. In [13], the authors considered all the inter-relay signals except the signal from previous relay as interference and investigated the outage performance of multi-hop HDR.

In this paper, we investigate the ergodic capacity of multi-hop DF FDR systems. Two cases are considered for the inter-relay interference, i.e. one relay node knows perfect channel state information (CSI) from all other relay nodes or only knows the CSI from the previous relay node. We assume all the channels including the RSI are Rayleigh fading. The upper bound of the ergodic capacity is derived considering the RSI and inter-relay interference.

## 2 System Model

The multi-hop FDR system is illustrated in Fig. 1, where a source node $S$ communicates with a destination node $D$ through $N$ relay nodes $R_1, R_2, \ldots, R_N$. For notational convenience, we denote $S$ and $D$ as $R_0$ and $R_{N+1}$, respectively. The distance between $R_i$ and $R_j$ is denoted by $d_{i,j}$, and the transmit power of $R_m$ is denoted by $P_m$. We employ DF relaying in this paper, thus each relay node will decode the message based on the received signal and then forward the message to the respective successor node if it is correctly decoded. We assume that $S$ and $D$ are equipped with a single antenna, and $R_m$ is equipped with two antennas (one for receiving and one for transmitting). All the antennas are assumed to be omnidirectional. The received signal at $R_m (m = 1, 2, \ldots, N+1)$ can be written as

$$y_m^{FD}(t) = \sqrt{P_{m-1}} h_{m-1,\,m} x_{m-1}(t) + \sum_{\substack{i=0,\\ i \neq m-1}}^{N} \sqrt{P_i} h_{i,\,m} x_i(t) + n_m(t), \qquad (1)$$

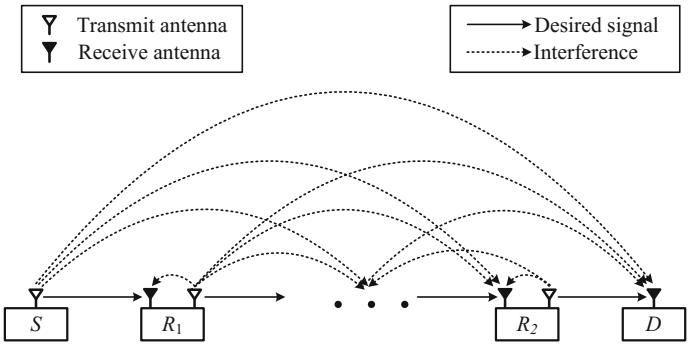



**Fig. 1** System model of multi-hop FDR

where $x_i(t)$ denotes the unit-power signal transmitted from $R_i$ at time instant $t$, and $h_{i,\,m}$ denotes the channel from $R_i$ to $R_m$. Specifically, $h_{m,\,m}$ denotes the RSI channel when $i=m$, and the interference channel caused by other relays when $i\neq m$ and $m-1$.

For simplicity, we assume the processing delay at each relay is 1 time unit, then the transmit signal $x_i(t)$ for $i>m$ at time instant $t$ can be given by $x_i(t)=x_m(t-i+m)$. Moreover, we assume all the channels including the RSI channel follow Rayleigh fading; then the channel coefficient $h_{i,\,m}$ is a zero-mean complex Gaussian random variable. Using the distance-dependent path loss model, the variance of $h_{i,\,m}$ is $d_{i,\,m}^{-\alpha}$ when $i\neq m$. When $i=m$, the variance of $h_{m,\,m}$ is assumed to be equal to $\tau$ for all the relay nodes.

Each relay decodes the message based on the received signal, so we consider two cases in this paper:

**Case I**: $R_m$ only knows channel $h_{m-1,\,m}$. As such, all other signals are considered to be interference.

**Case II**: $R_m$ knows perfect CSI from all other relays. In this case, since $R_m$ knows the transmit signal $x_i(t)$ for $i>m$, the interference from $R_{m+1}$ to $R_N$ can be subtracted from the received signals (1), and the equivalent received signals can be given as

$$\begin{aligned}\tilde{y}_m^{FD}(t)&=y_m^{FD}(t)-\sum_{i\,=\,m+1}^{N}\sqrt{P_i}h_{i,\,m}x_m(t-i+m)\\&=\sqrt{P_{m-1}}h_{m-1,\,m}x_{m-1}(t)+\sum_{\substack{i\,=\,0,\\ i\,\neq\,m-1}}^{m}\sqrt{P_i}h_{i,\,m}x_i(t)+n_m(t).\end{aligned}\tag{2}$$

## 3 Ergodic Capacity Analysis

For case I of multi-hop FDR system, we denote the received signal-to-interference-plus-noise ratio (SINR) at $R_m$ by $\gamma_m$, which is given as

$$\gamma_m = \frac{P_{m-1}|h_{m-1,m}|^2}{\sum\limits_{\substack{i=0,\\ i\neq m-1}}^{N} P_i|h_{i,m}|^2 + N_0} = \frac{\gamma_{m-1,m}}{\sum\limits_{\substack{i=0,\\ i\neq m-1}}^{N} \gamma_{i,m} + 1}, \tag{3}$$

where $\gamma_{i,m} = P_i|h_{i,m}|^2/N_0$. Since the interference channel also follows Rayleigh distribution, the PDF of $\gamma_{i,m}$ is given by $f_{\gamma_{i,m}}(x) = \frac{1}{\bar{\gamma}_{i,m}}\exp\left(-\frac{x}{\bar{\gamma}_{i,m}}\right)$, where $\bar{\gamma}_{i,m} = P_i d_{i,m}^{-\alpha}/N_0$. The equivalent SINR for case I of multi-hop FDR is given by $\gamma_{FD} = \min(\gamma_1, \gamma_2, \ldots, \gamma_{N+1})$. The CDF of $\gamma_{FD}$ can be written as

$$F_{\gamma_{FD}}(x) = \Pr(\min(\gamma_1, \gamma_2, \ldots, \gamma_{N+1}) \le x) = 1 - \prod_{m=1}^{N+1}\left(1 - F_{\gamma_m}(x)\right), \tag{4}$$

where $F_{\gamma_m}(x)$ denotes the CDF of $\gamma_m$. Using the result in [14], we have

$$F_{\gamma_m}(x) = 1 - \exp\left(-\frac{x}{\bar{\gamma}_{m-1,m}}\right) \prod_{\substack{i=0,\\ i\neq m-1}}^{N} \frac{1}{1 + x\bar{\gamma}_{i,m}/\bar{\gamma}_{m-1,m}}. \tag{5}$$

By substituting (5) into (4), we get

$$F_{\gamma_{FD}}(x) = 1 - \exp\left(-\sum_{m=1}^{N+1}\frac{x}{\bar{\gamma}_{m-1,m}}\right) \times \prod_{m=1}^{N+1} \prod_{\substack{i=0,\\ i\neq m-1}}^{N} \frac{1}{1 + x\bar{\gamma}_{i,m}/\bar{\gamma}_{m-1,m}}. \tag{6}$$

Using the equivalent SINR, the instantaneous capacity for case I of multi-hop FDR is given by $C_{FD} = \log_2(1+\gamma_{FD})$, and the ergodic capacity is given by $\bar{C}_{FD} = E[\log_2(1+\gamma_{FD})]$. The upper bound of the ergodic capacity is obtained as $\bar{C}_{FD} \le \log_2(1+E[\gamma_{FD}])$, where $E[\gamma_{FD}]$ can be given by

$$\begin{aligned} E[\gamma_{FD}] &= \int_0^{+\infty}\left(1 - F_{\gamma_{FD}}(x)\right)dx \\ &= \int_0^{+\infty}\exp\left(-\sum_{m=1}^{N+1}\frac{x}{\bar{\gamma}_{m-1,m}}\right)\prod_{m=1}^{N+1}\prod_{i=0,\, i\neq m-1}^{N}\frac{1}{1 + x\bar{\gamma}_{i,m}/\bar{\gamma}_{m-1,m}}dx. \end{aligned} \tag{7}$$

To make (7) more mathematically tractable, we use partial fraction expansion

$$\prod_{m=1}^{N+1}\prod_{i=0,\, i\neq m-1}^{N}\frac{1}{1+x\bar{\gamma}_{i,\,m}/\bar{\gamma}_{m-1,\,m}}=\sum_{m=1}^{N+1}\sum_{i=0,\, i\neq m-1}^{N}\frac{\lambda_{i,\,m}}{1+x\bar{\gamma}_{i,\,m}/\bar{\gamma}_{m-1,\,m}}, \tag{8}$$

where $\lambda_{i,m}$ can be obtained by using method of undetermined coefficients. Note that in (8), we assume the coefficients $\bar{\gamma}_{i,m}/\bar{\gamma}_{m-1,m}$, $m=1,2,\ldots,N+1$, $i=1,2,\ldots,N$, $i\neq m-1$ are unequal. When some of the coefficients are equal, the expression should be modified accordingly.

Substituting (8) into (7), we have

$$E[\gamma_{FD}]=\sum_{m=1}^{N+1}\sum_{i=0,\, i\neq m-1}^{N}\lambda_{i,\,m}\frac{\bar{\gamma}_{m-1,\,m}}{\bar{\gamma}_{i,\,m}}\exp\left(\frac{\bar{\gamma}_{m-1,\,m}}{\bar{\gamma}_{i,\,m}}\Phi\right)\mathrm{Ei}\left(1,\frac{\bar{\gamma}_{m-1,\,m}}{\bar{\gamma}_{i,\,m}}\Phi\right), \tag{9}$$

where $\Phi=\sum_{m=1}^{N+1}\frac{1}{\bar{\gamma}_{m-1,\,m}}$, and $\mathrm{Ei}(n,x)=\int_1^{\infty}\frac{\exp(-xt)}{t^n}dt$ is the well-known exponential integrals function. Then the upper bound of the ergodic capacity can be obtained.

For case II of multi-hop FDR system, the SINR at $R_m$ can be written as

$$\tilde{\gamma}_m=\frac{P_{m-1}|h_{m-1,m}|^2}{\sum\limits_{\substack{i=0,\\ i\neq m-1}}^{m}P_i|h_{i,m}|^2+N_0}=\frac{\gamma_{m-1,m}}{\sum\limits_{\substack{i=0,\\ i\neq m-1}}^{m}\gamma_{i,m}+1}. \tag{10}$$

Note that the interference from $R_{m+1}$ to $R_N$ has been cancelled. Similar to case I of multi-hop FDR system, the equivalent SINR for case II is given by $\tilde{\gamma}_{FD}=\min(\tilde{\gamma}_1,\tilde{\gamma}_2,\ldots,\tilde{\gamma}_{N+1})$. The instantaneous capacity and ergodic capacity can be obtained by replacing $\gamma_{FD}$ with $\tilde{\gamma}_{FD}$. Suppose the partial fraction expansion for this case is given by

$$\prod_{m=1}^{N+1}\prod_{i=0,\, i\neq m-1}^{m}\frac{1}{1+x\bar{\gamma}_{i,\,m}/\bar{\gamma}_{m-1,\,m}}=\sum_{m=1}^{N+1}\sum_{i=0,\, i\neq m-1}^{m}\frac{\mu_{i,\,m}}{1+x\bar{\gamma}_{i,\,m}/\bar{\gamma}_{m-1,\,m}}. \tag{11}$$

We can obtain the expected value of the equivalent SINR as

$$E[\tilde{\gamma}_{FD}]=\sum_{m=1}^{N+1}\sum_{i=0,\, i\neq m-1}^{m}\mu_{i,\,m}\frac{\bar{\gamma}_{m-1,\,m}}{\bar{\gamma}_{i,\,m}}\exp\left(\frac{\bar{\gamma}_{m-1,\,m}}{\bar{\gamma}_{i,\,m}}\Phi\right)\mathrm{Ei}\left(1,\frac{\bar{\gamma}_{m-1,\,m}}{\bar{\gamma}_{i,\,m}}\Phi\right). \tag{12}$$

Then the upper bound of the ergodic capacity for this case can be obtained.

## 4 Numerical Results

In this section, we present numerical results to compare the ergodic capacity of multi-hop FDR with multi-hop HDR. For simplicity, we normalize the distance between the source and the destination to unity, i.e. $d=1$. All the relays are assumed to be located on the line connecting the source and the destination, and the path loss exponent $\alpha$ is set to 4.

Figures 2 and 3 illustrate the ergodic capacity of multi-hop HDR and FDR for $N=2$ and $N=3$, respectively. We assume all the nodes are equidistant, i.e. $d_i=1/(N+1),\ i=0,1,\ldots,N$. The transmit power of the source and all the relays are equal to $P_0$ and the SNR is defined as $(N+1)P_0/N_0$. We can see that the ergodic capacity of multi-hop HDR is higher than that of multi-hop FDR case I in the high SNR region but lower than that of multi-hop HDR case II in the low SNR region. This is because all the signals except the desired signal are treated as interference in multi-hop FDR case I while the interference from $R_{m+1}$ to $R_N$ can be subtracted from the received signal of $R_m$ in multi-hop FDR case II. Therefore, the interference of multi-hop FDR case II is much smaller than that of multi-hop FDR case I. From Figs. 2 and 3, we can conclude that (i) if a relay cannot handle the interference from other relays, it is not beneficial to use FDR and (ii) multi-hop FDR has performance limits because it is an interference-limited system.

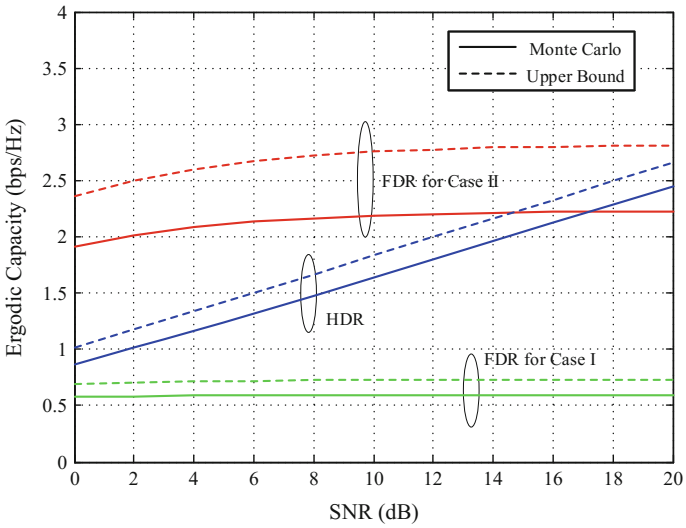



**Fig. 2** Ergodic capacity of multi-hop HDR and FDR for $N = 2$

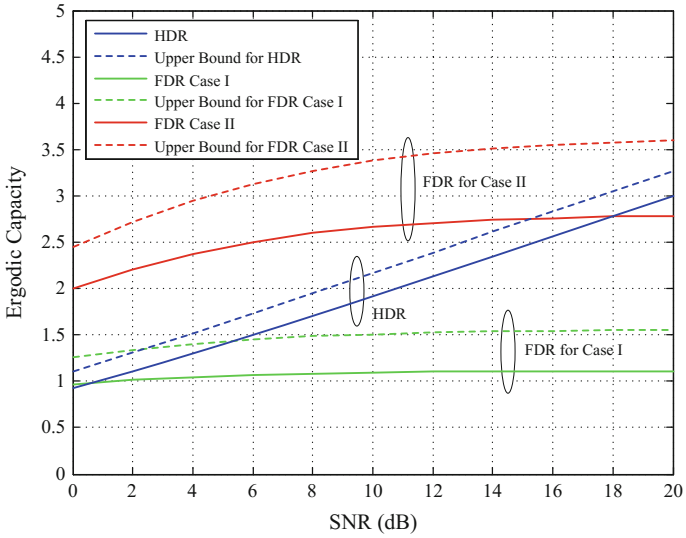



**Fig. 3** Ergodic capacity of multi-hop HDR and FDR for $N = 3$

## 5 Conclusion

We investigated the ergodic capacity of multi-hop DF FDR systems, whose main characteristic is that the relay nodes suffer from both self-interference and inter-relay interference. Two cases were considered for the interference. In case I we considered all the signals except the desired signal as interference while in case II we subtracted the interference from $R_{m+1}$ to $R_N$. We derived the upper bound of the ergodic capacity. It was shown by the numerical results that the multi-hop FDR case II presents a better performance than multi-hop HDR.

**Acknowledgements** This work was supported by the China Scholarship Council (201508120012) and the Doctoral Foundation of Tianjin Normal University (5RL135).

## References

1. M. Conti, S. Giordano, Multihop ad hoc networking: the reality. IEEE Commun. Mag. **45**(4), 88–95 (2007)
2. J. Boyer, D.D. Falconer, H. Yanikomeroglu, Multihop diversity in wireless relaying channels. IEEE Trans. Commun. **52**(10), 1820–1830 (2004)
3. M.O. Hasna, M.S. Alouini, Outage probability of multihop transmission over Nakagami fading channels. IEEE Commun. Lett. **7**(5), 216–218 (2003)
4. G. Farhadi, N.C. Beaulieu, On the ergodic capacity of multi-hop wireless relaying systems. IEEE Trans. Wirel. Commun. **8**(5), 2286–2291 (2009)

5. J.I. Choi, M. Jain, K. Srinivasan, P. Levis, S. Katti, Achieving single channel, full duplex wireless communications, in *Proceedings of ACM Mobicom*, Chicago, IL, USA, Sept 2010, pp. 1–12
6. D. Bharadia, E. McMilin, S. Katti, Full duplex radios, in *Proceedings of ACM SIGCOMM*, Hong Kong, China, Aug 2013, pp. 375–386
7. M. Heino, D. Korpi, T. Huusari et al., Recent advances in antenna design and interference cancellation algorithms for in-band full duplex relays. IEEE Commun. Mag. **53**(5), 91–101 (2015)
8. S. Hong, J. Brand, C. Jung, M. Jain, J. Mehlman, S. Katti, P. Levis, Applications of self-interference cancellation in 5G and beyond, IEEE Commun. Mag. **52**(2) (2014)
9. G. Kramer, M. Gastpar, P. Gupta, Cooperative strategies and capacity theorems for relay networks. IEEE Trans. Inf. Theory **51**(9), 3037–3063 (2005)
10. T. Riihonen, S. Werner, R. Wichman, J. Hamalainen, Outage probabilities in infrastructure-based single-frequency relay links, in *Proceedings of the IEEE WCNC* (2009), pp. 1–6
11. T. Riihonen, S. Werner, R. Wichman, Hybrid full-duplex/half-duplex relaying with transmit power adaptation. IEEE Trans. Wirel. Commun. **10**(9), 3074–3085 (2011)
12. M. Khafagy, A. Ismail, M.S. Alouini, S. Aissa, On the outage performance of full-duplex selective decode-and-forward relaying. IEEE Commun. Lett. **17**(6), 1180–1183 (2013)
13. T.K. Baranwal, D.S. Michalopoulos, R. Schober, Outage analysis of multihop full duplex relaying. IEEE Commun. Lett. **17**(1), 63–66 (2013)
14. S. Kandukuri, S. Boyd, Optimal power control in interference-limited fading wireless channels with outage-probability specifications. IEEE Trans. Wirel. Commun. **1**(1), 46–55 (2002)

# Transmit Multi-beamforming for Colocated Uniform Linear Array Using MISL Beamformers

**Haisheng Xu, Jian Wang, Wenyun Gao and Zhihui Yuan**

**Abstract** This paper considers transmit multi-beamforming for colocated uniform linear array (ULA). First, the colocated ULA emitting sum of independent orthonormal baseband waveforms by each element is briefly introduced, and its transmit beampattern expressed by the sum of sub-beampatterns of all the waveforms is formulated. Then the beamforming algorithms using the criterion of minimum integrated sidelobe level (MISL) for different types of steering angle-space are proposed. The algorithms can be summarized by optimizing the weighting coefficients of each waveform, which controls one sub-beampattern, to form a beam based on parallel MISL beamformers. And each beamformer is then converted to a Rayleigh quotient (RQ) minimization in which the optimal closed-form solutions can be obtained. Finally, numerical comparisons and computational complexity analyses are provided to validate the MISL beamformers.

**Keywords** Transmit multi-beamforming · Uniform linear array (ULA) · Optimization · Minimum integrated sidelobe level (MISL) · Floating point operations (FLOPs)

H. Xu (✉)
Department of Automation, Tsinghua University, Beijing 100084, China
e-mail: hsxu@tsinghua.edu.cn

J. Wang · W. Gao
Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
e-mail: jian-wang@tsinghua.edu.cn

W. Gao
e-mail: gaowenyun.thu@hotmail.com

Z. Yuan
School of Electrical and Information Engineering, Changsha University of Science and Technology, Changsha 410114, China
e-mail: yuanzh.2003@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_18

## 1 Introduction

Colocated uniform linear array (ULA) has attracted lots of attention for it can take the advantage of the good properties of waveform diversity and antenna coherence. These two properties guarantee that the colocated ULA can have extra degrees of freedom in transmit beamforming with significant coherent gains [1–4].

In transmit antenna radiation theory, sidelobe level of the transmit beampattern is one of the most important performance indexes [5, 6]. And in array applications, targets may not just present at a single angle-direction or angle-space but may come from several scattered directions or areas, which makes us need to carefully inspect these several discrete spaces simultaneously. Thus transmit multi-beamforming with lower sidelobe levels is also an important demand in reality. As we know, a phased-array transmitting a single waveform can obtain narrowly focused beampatterns [1]. However, the mode of transmitting the same waveform for all elements limits its applications in multi-beamforming. Since transmitting multiple waveforms can provide extra degrees of freedom in waveform design to realize flexible transmit beampattern formation, it makes sense to further exploit the potentials of colocated ULA transmitting multiple waveforms in multi-beamforming.

Therefore, this paper considers transmit multi-beamforming for the colocated ULA using minimum integrated sidelobe level (MISL) [7] beamformers. First, the colocated ULA emitting sum of independent orthonormal baseband waveforms for each element is briefly introduced, and its transmit beampattern expressed by the sum of sub-beampatterns of all the waveforms is formulated. Then the beamforming algorithms using the criterion of MISL for different types of steering angle-space are proposed. The algorithms can be summarized by optimizing the weighting coefficients of each waveform, which controls one sub-beampattern, to form a beam based on parallel MISL beamformers. And each beamformer is then converted to a Rayleigh quotient (RQ) minimization in which the optimal closed-form solutions can be obtained. Finally, numerical comparisons and computational complexity analyses among different beamformers are provided to show the good performance and high-efficiency of the our beamformers, respectively.

## 2 Transmit Multi-beamforming by MISL

Consider a colocated ULA equipped with $M$ elements, which are separated by a half wavelength. Assume each element emits a weighted sum of $Q$ independent orthonormal baseband waveforms represented by $\boldsymbol{\phi}(t) = \left[\phi_1(t)\ \phi_2(t)\ \cdots\ \phi_Q(t)\right]^T$, where superscript $T$ denotes transpose operation. The weightings are described by a coefficient matrix $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \cdots\ \mathbf{c}_Q]$, where $\mathbf{C} \in \mathbb{C}^{M\times Q}$ and $\mathbf{c}_q$ is used to control the amount of the $q$th waveform added into a sum at each element [8]. The total transmit power is fixed at $E_t$ by setting $\sum_{q=1}^{Q} \left\|\mathbf{c}_q\right\|^2 = E_t$. Then the well-known transmit beampattern

[1, 9] is given by

$$P(\theta) = \mathbf{a}^H(\theta)\mathbf{C}\mathbf{C}^H\mathbf{a}(\theta) \tag{1}$$

where $\theta \in \Theta = [\frac{-\pi}{2}, \frac{\pi}{2}]$, subscript $H$ denotes complex conjugate-transpose (or Hermitian) operation and $\mathbf{a}(\theta)$ is the transmit steering vector [10] formulated as $\mathbf{a}(\theta) = [1\ e^{j\pi\sin(\theta)}\ \cdots\ e^{j\pi(M-1)\sin(\theta)}]^T$. Then $P(\theta)$ can be reformulated as

$$\begin{aligned} P(\theta) &= \mathbf{a}^H(\theta)\left(\sum_{q=1}^{Q}\mathbf{c}_q\mathbf{c}_q^H\right)\mathbf{a}(\theta) \\ &= \sum_{q=1}^{Q}\mathbf{c}_q^H\mathbf{a}(\theta)\mathbf{a}^H(\theta)\mathbf{c}_q \\ &= \sum_{q=1}^{Q}\mathbf{c}_q^H\mathbf{A}_M(\theta)\mathbf{c}_q \end{aligned} \tag{2}$$

where $\mathbf{A}_M(\theta) \triangleq \mathbf{a}(\theta)\mathbf{a}^H(\theta)$.

From (2), it can be seen that the transmit beampattern of the ULA can be formulated by a sum of the sub-beampatterns of the waveforms and each sub-beampattern is controlled by the weighting coefficients of one waveform. In the following, we present two transmit multi-beamforming cases by using MISL beamformers under the strategy that one sub-beampattern forms one beam according to the different types of steering angle-space.

## 2.1 *Transmit Multi-beamforming Toward Multiple Wide Angle Areas*

Consider forming $Q$ beams each toward a continuous interested wide area $\Theta_q$, where $\Theta_q \subset \Theta$ and $q = 1, 2, \ldots, Q$. For the $q$th beam, to maximize the power of the beam concentrated in $\Theta_q$ while minimize it outside $\Theta_q$, we can formulate the following MISL [7, 9] optimization

$$\min_{\|\mathbf{c}_q\|=\sqrt{\frac{E_t}{Q}}} \frac{\mathbf{c}_q^H\mathbf{\Gamma}_u^{(q)}\mathbf{c}_q}{\mathbf{c}_q^H\mathbf{\Gamma}_d^{(q)}\mathbf{c}_q},\ q = 1, 2, ..., Q \tag{3}$$

where $\mathbf{\Gamma}_d^{(q)}$ and $\mathbf{\Gamma}_u^{(q)}$ ($\in \mathbb{C}^{M\times M}$) are calculated by two integral operations over $\Theta_q$ and $\Theta \setminus \Theta_q$, respectively, and given by

$$\begin{aligned} \mathbf{\Gamma}_d^{(q)} &\triangleq \int_{\Theta_q} \mathbf{A}_M(\theta)\mathrm{d}\theta \\ \mathbf{\Gamma}_u^{(q)} &\triangleq \int_{\Theta\backslash\Theta_q} \mathbf{A}_M(\theta)\mathrm{d}\theta \end{aligned}. \tag{4}$$

Note that except the constraint of equal norm for all waveform weighting vectors, we can also use some other constraints, for example, setting equal maximal-value for all sub-beampatterns, for the $Q$ optimizations formulated in (3). As $\Theta_q$ is a wide angle area, we can reckon the rank of $\mathbf{\Gamma}_d^{(q)}$ is greater than one. Although we cannot guarantee that $\mathbf{\Gamma}_d^{(q)}$ is positive definite, we can prove that $\Gamma_d^{(q)} + \mathbf{\Gamma}_u^{(q)}$ for $q = 1, 2, \ldots, Q$ are always positive definite since $\mathbf{x}^H \left(\Gamma_d^{(q)} + \mathbf{\Gamma}_u^{(q)}\right) \mathbf{x} = \mathbf{x}^H \int_\Theta \mathbf{A}_M(\theta)\mathrm{d}\theta\mathbf{x} = 2\left\|\mathbf{x}\right\|^2 > 0$ for any nonzero vector $\mathbf{x}$. Then we can use case (1) or case (3) of the following theorem to solve (3).

**Theorem 1** *[11, ERQM Theorem]: Define Hermitian matrices* $\mathbf{A}$, $\mathbf{B} \in \mathbb{C}^{n\times n}$, *and assume* $\mathbf{A} + \mathbf{B}$ *is positive definite while* $\mathbf{A}$ *is nonnegative definite with* $\mathrm{rank}(\mathbf{A}) = \varsigma$ $(0 < \varsigma \leq n)$. *Focusing on the optimization problem formulated as*

$$\min_{\mathbf{r}} \ \frac{\mathbf{r}^H\mathbf{B}\mathbf{r}}{\mathbf{r}^H\mathbf{A}\mathbf{r}}, \tag{5}$$

*then*

*(1) If* $2 \leq \varsigma < n$, *the optimal solution to (5) is given by*

$$\mathbf{r}_{opt} = \mathbf{U}\begin{pmatrix} \left(\mathbf{\Lambda}_1^{-\frac{1}{2}}\right)^H \\ -\widetilde{\mathbf{B}}_1^{-1}\widetilde{\mathbf{B}}_2^H\left(\mathbf{\Lambda}_1^{-\frac{1}{2}}\right)^H \end{pmatrix}\mathbf{x}_\bigstar, \tag{6}$$

*where* $\mathbf{r}_{opt}$ *is the minimal solution when* $\mathbf{x}_\bigstar$ $(\in \mathbb{C}^{\varsigma\times 1})$ *is the eigenvector corresponding to the minimum eigenvalue of* $\mathbf{\Lambda}_1^{-\frac{1}{2}}\widetilde{\mathbf{B}}\left(\mathbf{\Lambda}_1^{-\frac{1}{2}}\right)^H$; $\widetilde{\mathbf{B}} \in \mathbb{C}^{\varsigma\times\varsigma}$ *such that* $\widetilde{\mathbf{B}} \triangleq \widetilde{\mathbf{B}}_0 - \widetilde{\mathbf{B}}_2\widetilde{\mathbf{B}}_1^{-1}\widetilde{\mathbf{B}}_2^H$; $\mathbf{\Lambda}_1 \in \mathbb{R}^{\varsigma\times\varsigma}$ *is an invertible diagonal matrix obtained from the eigen decomposition of* $\mathbf{A}$ *by* $\mathbf{A} = \mathbf{U}\begin{pmatrix} \mathbf{\Lambda}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix}\mathbf{U}^H$, *and* $\widetilde{\mathbf{B}}_0 \in \mathbb{C}^{\varsigma\times\varsigma}$, $\widetilde{\mathbf{B}}_1 \in \mathbb{C}^{(n-\varsigma)\times(n-\varsigma)}$ *and* $\widetilde{\mathbf{B}}_2 \in \mathbb{C}^{\varsigma\times(n-\varsigma)}$ *are the submatrices of* $\mathbf{U}^H\mathbf{B}\mathbf{U}$ *given by*

$$\mathbf{U}^H\mathbf{B}\mathbf{U} = \begin{pmatrix} \widetilde{\mathbf{B}}_0 & \widetilde{\mathbf{B}}_2 \\ \widetilde{\mathbf{B}}_2^H & \widetilde{\mathbf{B}}_1 \end{pmatrix}. \tag{7}$$

*(2) If* $\varsigma = 1$, *let* $\mathbf{A} = \mathbf{u}\mathbf{u}^H$, *then the optimal solution to (5) is given by*

$$\mathbf{r}_{opt} = \mathbf{u} - \mathbf{U}_2 \left(\mathbf{U}_2^H \mathbf{B} \mathbf{U}_2\right)^{-1} \mathbf{U}_2^H \mathbf{B}\mathbf{u}, \tag{8}$$

*where* $\mathbf{U}_2$ *is a submatrix of the unitary matrix* $\mathbf{U}$ *obtained by eigen-decomposing* $\mathbf{A}$ *and formulated as* $\mathbf{U} = \left(\frac{\mathbf{u}}{\|\mathbf{u}\|}\ \mathbf{U}_2\right)$.

(3) *If* $\varsigma = n$, *the optimal solution of (5) is given by*

$$\mathbf{r}_{opt} = \left(\mathbf{A}^{-1/2}\right)^H \mathbf{x}_{\bigstar}, \tag{9}$$

*where* $\mathbf{r}_{opt}$ *is the minimal solution when* $\mathbf{x}_{\bigstar}$ $(\in \mathbb{C}^{n\times 1})$ *is the eigenvector corresponding to the minimum eigenvalue of* $\mathbf{A}^{-1/2}\mathbf{B}(\mathbf{A}^{-1/2})^H$.

## 2.2 Transmit Multi-beamforming Toward Discrete Angle Points

The problem in (3) seeks multi-beamforming toward wide angle areas. However, in some applications beamforming toward several discrete angle points is also desired, requiring multiple beams steered to several interesting directions. Suppose the directions are $\theta_q$, $q = 1, 2, \ldots, Q$, i.e., $\Theta_q = \{\theta_q\}$, $q = 1, 2, \ldots, Q$, then we use Riemann sum [12] instead of continuous integral to modify $\mathbf{\Gamma}_d^{(q)}$ and $\mathbf{\Gamma}_u^{(q)}$ as $\mathbf{\Gamma}_d^{(q)} = \mathbf{A}_M(\theta_q)\Delta\theta$ and $\mathbf{\Gamma}_u^{(q)} = \sum_{\Theta\setminus\{\theta_q\}} \mathbf{A}_M(\theta)\Delta\theta$, respectively, where $\Delta\theta$ represents the infinitesimal (but nonzero) differential element, and obtain the altered optimization by

$$\min_{\|\mathbf{c}_q\|=\sqrt{\frac{E_t}{Q}}} \frac{\mathbf{c}_q^H \left(\sum\limits_{\Theta\setminus\{\theta_q\}} \mathbf{A}_M(\theta)\right) \mathbf{c}_q}{\mathbf{c}_q^H \mathbf{A}_M(\theta_q)\mathbf{c}_q},\ q = 1, 2, ..., Q, \tag{10}$$

where we still use the equal-norm constraint to formulate our optimization.

The above problem is a quadratic ratio minimization with a rank-one matrix in the denominator. As we have that $\mathbf{\Gamma}_d^{(q)} + \mathbf{\Gamma}_u^{(q)}$ for any $q$ is always positive definite, we can obtain that $\sum_{\Theta\setminus\{\theta_q\}} \mathbf{A}_M(\theta) + \mathbf{A}_M(\theta_q) = \sum_{\Theta} \mathbf{a}(\theta)\mathbf{a}^H(\theta)$ is also positive definite, hence, we can use case (2) of Theorem 1 to directly obtain the optimal closed-form solutions to (10).

## 3 Numerical Results

In this section, we present numerical results to demonstrate the effectiveness of the MISL beamformers for the two multi-beamforming problems.

### *3.1 Performance Comparisons on Different Beamformers*

For the two cases of multi-beamforming problems introduced in Sect. 2, we present two numerical examples. To show the better precision of the MISL beamformers using Theorem 1 (called MISL for short), the pseudoinverse-based solution (PBS) of [13, 14] is compared. Meanwhile, to show the better performance of the MISL, the beamformers obtained by solving the discrete prolate spheroidal sequences (DPSS) [15], which employed the typical RQ optimizations, are also compared. Set the number of array elements $M = 20$, and the beamforming results are shown in Fig. 1, where we normalize all the sub-beampatterns of each case to a same value before merging them together. From the two subfigures of Fig. 1, it can be seen that the MISL can form radiation patterns with lower sidelobe levels than that of the DPSS beamformers, and especially, the DPSS beamformers may split the mainlobes of the patterns when beamforming toward a wide range of space (see Fig. 1a). Additionally, although the PBS uses the same beamforming criterion as the MISL to form radiation patterns, it employs pseudoinverses to deal with singular matrixes, which causes approximation errors in transmit beamforming and results in the worst radiation patterns among the three algorithms. And these approximation errors are bigger as shown in Fig. 1b, where the mutli-direction beam based on the PBS cannot form a shaped radiation pattern any more for which applies the pseudoinverse to a rank-one matrix.

### *3.2 Complexity Analyses on Different Beamformers*

In this subsection, we compare the computational complexity of the MISL beamformer with the PBS and DPSS. To perform the analyses, floating point operations (FLOPs) [16] of the PBS, DPSS, and MISL are counted. The matrix operations involved in the three algorithms are mainly eigen decomposition, matrix inversion, matrix–matrix product, matrix–vector product, matrix summation/subtraction, and square-root operation. Then given the matrix dimension and rank of $\mathbf{A}$ defined in (5) are $n$ and $\varsigma$ ($1 \leq \varsigma < n$), respectively, the FLOPs of the three algorithms are calculated and presented in Table 1. Observe the table, we can coarsely know that the DPSS has the lowest complexity and when $2 \leq \varsigma < n$ the complexity of MISL lies between the PBS and DPSS while when $\varsigma = 1$ the complexity of MISL approaches to the PBS. To validate our conclusion, we draw the complexity curves of the three

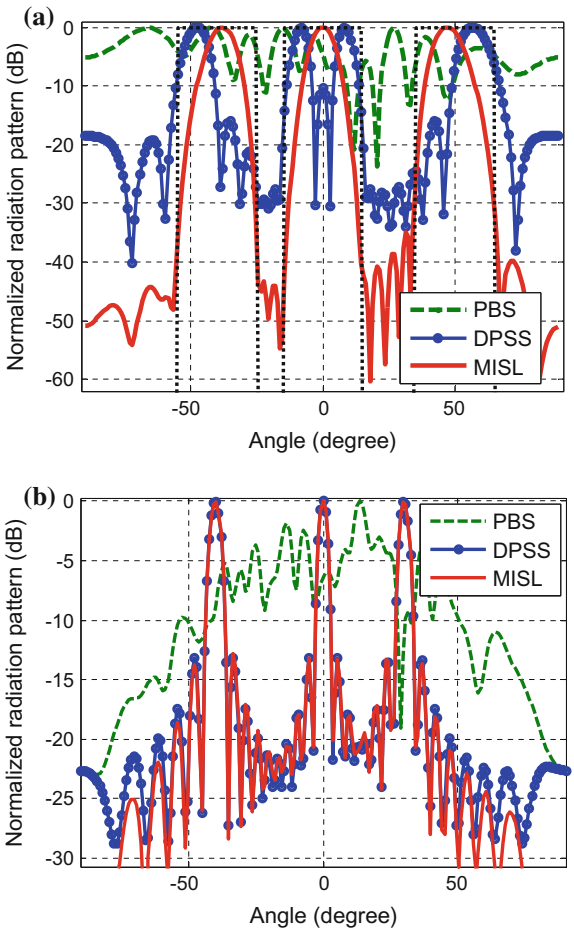



**Fig. 1** Radiation patterns of multi-beamforming toward **a** three wide areas: $\Theta_1 = [-55°, -25°]$, $\Theta_2 = [-15°, +15°]$ and $\Theta_3 = [+35°, +65°]$ and **b** three angles: $\Theta_1 = \{-40°\}$, $\Theta_2 = \{0°\}$, and $\Theta_3 = \{+30°\}$

**Table 1** FLOPs of the PBS, DPSS and MISL

| | PBS | DPSS | MISL |
|---|---|---|---|
| $2 < \varsigma < n$ | $\frac{32}{3}n^3 - n^2 - \frac{17}{3}n + 2 + \varsigma$ | $\frac{1}{3}n^3 + n^2 - \frac{7}{3}n + 1$ | $\frac{16}{3}n^3 + \frac{10}{3}\varsigma^3 + 3n\varsigma^2 + n^2\varsigma + \varsigma^2 - \frac{7}{3}(n + \varsigma) - 3n\varsigma + 2$ |
| $\varsigma = 1$ | | | $\frac{34}{3}n^3 - 16n^2 + \frac{20}{3}n$ |

algorithms based on Table 1 and show them in Fig. 2, where Fig. 2a shows that the curves of MISL, i.e., the solid red lines, for different $\varsigma$ ($2 \leq \varsigma < n$) basically all lies between the PBS and DPSS, while Fig. 2b and its partial enlarged details show that the curve of MISL for $\varsigma = 1$ nearly overlaps with that of the PBS.

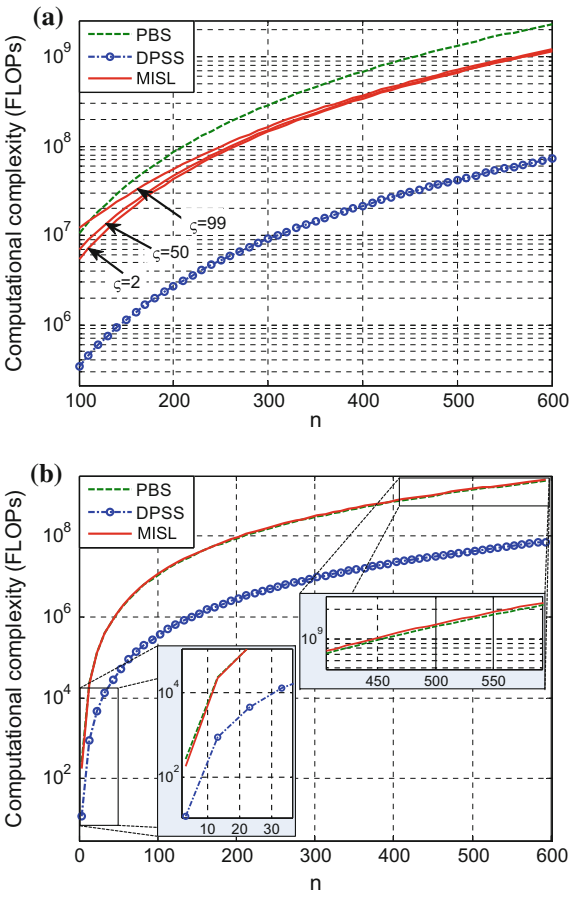



**Fig. 2** Computational complexity comparisons among the PBS, DPSS, and MISL for **a** case $2 \leq \varsigma < n$ and **b** case $\varsigma = 1$

The above analysis results are objective and consistent with our expectations. When we pursue better shaped radiation patterns than that of the DPSS, we need to take some other measures at the cost of complexity or other things. Thus it is not strange that the computational complexity of the MISL are higher than that of the DPSS. But then again, the complexity discrepancy of the two kinds of beamformers can be ignored for which are actually at the same complexity order, i.e., $\mathcal{O}(n^3)$. On the other hand, under the same beamforming criterion, MISL not only can have better performance but also can have lower computational complexity than that of using the PBS. Therefore, in summary, we can still conclude that our MISL beamformers can have high-efficiency.

## 4 Conclusion

This paper considered transmit multi-beamforming for ULA. First, the colocated ULA emitting sum of independent orthonormal baseband waveforms for each element was briefly introduced, and its transmit beampattern expressed by the sum of sub-beampatterns of all the waveforms was formulated. Then the beamforming algorithms using the criterion of minimum integrated sidelobe level (MISL) for different types of steering angle-space were proposed. The algorithms could be summarized by optimizing the weighting coefficients of one waveform for one beam based on parallel MISL beamformers. And each beamformer was then converted to a RQ minimization in which the optimal closed-form solutions could be obtained. Finally, numerical comparisons and computational complexity analyses were provided to validate the MISL beamformers.

**Acknowledgements** This work was supported by China Postdoctoral Science Foundation Funded Project under Grant No. 2015M581095.

## References

1. D. Fuhrmann, G. San Antonio, Transmit beamforming for MIMO radar systems using partial signal correlation, in *Proceedings of the 38th Asilomar Conference on Signals, Systems, and Compututer*, vol. 1 (Pacific Grove, CA, Nov 2004), pp. 295–299
2. J. Li, P. Stoica, MIMO radar—diversity means superiority, in *Proceedings of the 14th Annual Workshop Adaptive Sensor Array Process* (MIT Lincoln Laboratory, Lexington, MA, June 2009), pp. 1–6
3. A.A. Gorji, R. Tharmarasa, T. Kirubarajan, Widely-separated MIMO vs. multistatic radar for target localization, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Process. (ICASSP)* (Kyoto, Mar 2012), pp. 2461–2464
4. G. Cui, H. Li, M. Rangaswamy, MIMO radar waveform design with constant modulus and similarity constraints. IEEE Trans. Signal Process. **62**(2), 343–353 (2014)
5. K. Gerlach, Adaptive array transient sidelobe levels and remedies. IEEE Trans. Aerosp. Electron. Syst. **26**(3), 560–568 (1990)
6. C.A. Balanis, *Antenna Theory: Analysis and Design*, 3rd edn. (Wiley, Hoboken, NJ, USA, 2005)
7. A. Martinezl, J. Marchand, SAR image quality assessment. Revista de Teledeteccion **2**, 12–18 (1993)
8. H. Xu, J. Wang, J. Yuan, X. Shan, Colocated MIMO radar transmit beamspace design for randomly present target detection. IEEE Signal Process. Lett. **22**(7), 828–832 (2015)
9. H. Xu, R. Blum, J. Wang, J. Yuan, Colocated MIMO radar waveform design for transmit beampattern formation. IEEE Trans. Aerosp. Electron. Syst. **51**(2), 1558–1568 (2015)
10. M.I. Skolnik, *Radar Handbook*, 3rd edn. (McGraw-Hill, New York, NY, USA, 1990)
11. H. Xu, J. Wang, J. Yuan, C. Jiang, C. Zhang, Generalized RQ minimization with applications in array transmit beamforming, *IEEE Antennas Wireless Propagation Letters*, vol. PP, no. 99, pp. 1–1, 2016
12. K.E. Atkinson, *An Introduction to Numerical Analysis*, 2nd edn. (Wiley, Hoboken, NJ, USA, 1988)
13. J. Ye, R. Janardan, C.H. Park, H. Park, An optimization criterion for generalized discriminant analysis on undersampled problems. IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 982–994 (2004)

14. T. Bogale, L. Vandendorpe, Max-Min SNR signal energy based spectrum sensing algorithms for cognitive radio networks with noise variance uncertainty. IEEE Trans. Wirel. Commun. **13**(1), 280–290 (2014)
15. H.L. van Trees, *Optimum Array Processing: Part IV of Detection, Estimation and Modulation Theory* (Wiley, New York, NY, 2002), pp. 90–230. Ch. 3
16. R. Hunger, *Floating point operations in matrix-vector calculus* (Munich University of Technology, Institute for Circuit Theory and Signal Processing Munich, 2005)

# Part II
# Radar Techniques

# A Multi-antenna Receiving Technique for Terahertz Radar CSAR Imaging

**Jubo Hao, Jin Li, Jie Zou and Diqiu Bai**

**Abstract** Circular synthesis aperture radar (CSAR) is an efficient way to achieve high-resolution imaging and a common way to realize 3D imaging. In this paper, a multi-antenna receiving (MAR) technique for terahertz radar is presented. In the vertical direction, one antenna transmits radar signal while multi-antenna receives echoes scattered from targets. Data fusion based on threshold selection method is used for different antenna imaging results to get a better image. Imaging quality of the horizontal plane by the proposed technique is remarkably increased. The simulations verify the effectiveness of the proposed technique.

**Keywords** Circular synthetic aperture radar (CSAR) · Multi-antenna receiving (MAR) · Terahertz · 3D imaging

## 1 Introduction

Circular synthetic aperture radar (CSAR) has caught great attention owing to its capability of obtaining high-resolution image and three-dimensional scatterers' distributions [1–4]. Compared with traditional linear trajectory SAR, CSAR is able to observe the scene from all directions, which leads to a better understanding of the scattering properties of the targets. Besides, the circular track makes the aspect angle wide enough to reconstruct a 3-D target image.

J. Hao · J. Li (✉)
School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu, China
e-mail: lijin@uestc.edu.cn

J. Zou
The Second Research Institute of Civil Aviation Administration of China, Chengdu, China

D. Bai
Glasgow College, University of Electronic Science and Technology of China, Chengdu, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_19

As far, several CSAR systems have been employed. AFRL operated a CSAR system in X-band with the center frequency of 9.6 MHz and the bandwidth of 640 MHz [1]. Meanwhile, MITL obtained the first China's CSAR data using a P-band, fully polarimetric SAR system in Sichuan [3]. In order to acquire higher precision image, improved experiments were conducted. In [5], a multibaseline fully polarimetric circular SAR experiment using DLR's F-SAR system at L-band is accomplished. The experiment achieved higher resolution in the direction perpendicular to the line-of-sight (LOS) compared to the single-pass ones.

Moore and Potter pointed out that the resolution of CSAR has a direct relevance with the wavelength of propagating wave and its bandwidth [6]. The shorter the wavelength and the wider the bandwidth, the higher imaging resolution could be got. Terahertz wave has a frequency of more than 100 GHz, and can achieve a more than 10 GHz bandwidth [7]. Thus, terahertz radar has advantages over traditional microwave radar systems in SAR imaging applications [8]. In this paper, we use the terahertz radar to achieve the CSAR imaging.

With the extending application of SAR images, higher resolution and better quality images are needed. In this paper, a multi-antenna receiving (MAR) technique is introduced to realize a better performance of SAR imaging quality.

## 2 Theory and Methodology

### *2.1 Signal Model*

The geometry of the MAR CSAR system is shown in Fig. 1. The radar platform moves along a circular trajectory whose radius is $R_g$ and height is $H$. The imaging area is centered about the coordinate $(0, 0)$ with radius $R_0$. The multiple receiving antennas distribute in the direction perpendicular to the ground plane with an interval of $d$.

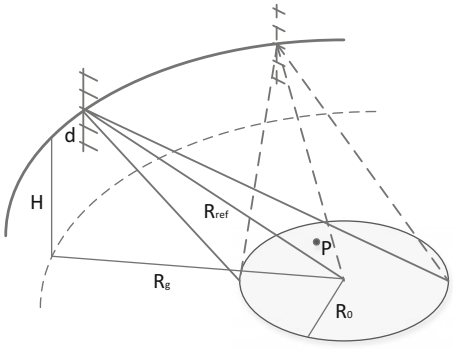



**Fig. 1** Geometry of MAR CSAR

The linear frequency modulated (LFM) signal is transmitted, which is shown as follows:

$$s(t) = A\,rect\left(\frac{t}{T_p}\right)\exp\left(j2\pi f_c t + j\pi\gamma t^2\right), \tag{1}$$

where $A$ is the signal amplitude, $f_c$ is the carrier frequency, $\gamma$ is the modulated rate, and $T_p$ is the pulse duration time. Suppose that there are a total of $I$ receiving antennas, the received echo of the $i$th receiving antenna is

$$s(t,\theta,i) = rect\left(\frac{t-t_d(i)}{T_p}\right)\exp[j2\pi f_c(t-t_d(i))]\cdot\ \exp\left[j\pi\gamma(t-t_d(i))^2\right], \tag{2}$$

where $\theta$ is the azimuth angle, $t_d(i) = R_d(i)/c$ is the time delay of the $i$th antenna. $R_d(i)$ is the propagating distance of electromagnetic wave, which is presented as $R_d(i) = R_t + R_r(i)$, where $R_t$ is the distance from the transmitting antenna to the target and $R_r(i)$ is the distance from the target to the $i$th receiving antenna.

Assuming that a point target is located at $(x_p, y_p, z_p)$, then

$$R_t = \sqrt{\left(R_0\cos(\theta)-x_p\right)^2 + \left(R_0\sin(\theta)-y_p\right)^2 + \left(H-z_p\right)^2} \tag{3}$$

$$R_r(i) = \sqrt{\left(R_0\cos(\theta)-x_p\right)^2 + \left(R_0\sin(\theta)-y_p\right)^2 + \left(H+i\cdot d-z_p\right)^2}. \tag{4}$$

From Eqs. (3) and (4), we know that the propagating distances are relevant with the slow time and the location of the point.

For terahertz wave, a broadband of more than 10 GHz can be achieved to dramatically increase the range resolution. Generally, matching-filtering method is implemented for pulse compression; however, it would result in large data volume for such a broadband. We use the dechirping method to process the broadband LFM signal.

It is assumed that the distance from the transmitting antenna to the center of imaging area is $R_{ref}$, and the reference function is presented as Eq. (5):

$$s_{ref}(t) = rect\left(\frac{t-2R_{ref}/c}{T_p}\right)\exp\left[j2\pi f_c\left(t-2R_{ref}/c\right)\right]\cdot\ \exp\left[j\pi\gamma\left(t-2R_{ref}/c\right)^2\right]. \tag{5}$$

The dechirped signal is the conjugate production of (2) and (5), which is shown as

$$s_{if}(t,\theta,i)=s(t,\theta,i)\cdot s_{ref}^{*}(t)=A\,rect\left(\frac{t-t_d(i)}{T_p}\right)\exp\left(-\frac{4\pi\gamma}{c}\left(t-\frac{2R_{ref}}{c}\right)\Delta R\right)\cdot\exp\left(-j\frac{4\pi\gamma f_c}{c}\Delta R\right)\cdot\exp\left(j\frac{4\pi}{c^2}(\Delta R)^2\right) \tag{6}$$

where $\Delta R=R_d(i)-R_{ref}$. Transform the range direction from time domain to frequency domain with Fourier transform and neglect the amplitude, we get

$$S_{if}(f_r,\theta,i)=FFT_t\left\{s_{if}(t,\theta,i)\right\}=\sin c\left[T_p\left(f_r+\frac{\gamma}{c}\Delta R\right)\right]\exp\left(-j\frac{4\pi f_c}{c}\Delta R\right)\cdot\exp\left(-\frac{4\pi f_r}{c}\Delta R\right)\cdot\exp\left(-j\frac{4\pi\gamma}{c^2}(\Delta R)^2\right) \tag{7}$$

In (7), the second phase exponent is envelope term and the third phase exponent is the residual video phase. Both are removed before the following process.

## 2.2 The MAR CSAR Imaging Algorithm

Generally, in discrete signal processing area, a target is seen as a single point target or the summation of several point targets with corresponding scattering coefficients. Assuming that the scattering coefficient of a point target is $\sigma_n(n=1,\ldots,N)$, then the echo of the whole scene is expressed as

$$S_a=\sum_n \sigma_n S_{if}(f_r,\theta,i). \tag{8}$$

For circular trajectory SAR, it has the ability to reconstruct three-dimensional image. Therefore, multiple 3D image could be reconstructed with multiple receiving antennas, which means that the data received from a single antenna can be processed for a 3D image. Thus, we will process the data from different antennas separately, while the processing is non-interfering and thus the imaging can be done parallelly.

The *Back-Projection* (BP) algorithm is employed to each data set from corresponding antenna. The principle of BP algorithm is coherently accumulating the pulses from different azimuth directions after phase compensation [9]. The expression is shown as

$$\widehat{\sigma}_n=\sum_a S_a\exp\left(j\frac{2\pi f_c}{c}R\right). \tag{9}$$

In (9), $R$ is propagating distance of electromagnetic wave, which is similar to $R_d(i)$. The main step of BP algorithm is as follows: (1) Divide the region of interest (ROI) into cells. Here, we divide the imaging area into three-dimensional gridding. (2) Get the first pulse and compensate the phase error for each cell. (3) Accumulate pulse with interpolation. (4) Repeat (2) and (3) until all the pulses are processed.

After the imaging process, we get a total of $I$ 3D images $G_i(i = 1, \ldots, I)$ and the data fusion method is used to form the final image. The following principle is established: first, the mean value of the $I$ images is calculated, which is

$$G_m = \frac{1}{I}\sum_i G_i. \tag{10}$$

After that, we compare each pixel in $G_m$ with a given threshold $Th$, if $G_m(k) > Th$, $G = \max\{G_i, i = 1, \ldots, I\}$, else, $G = \min\{G_i, i = 1, \ldots, I\}$.

The flowchart of multi-antenna receiving technique for Terahertz radar CSAR imaging is shown in Fig. 2.

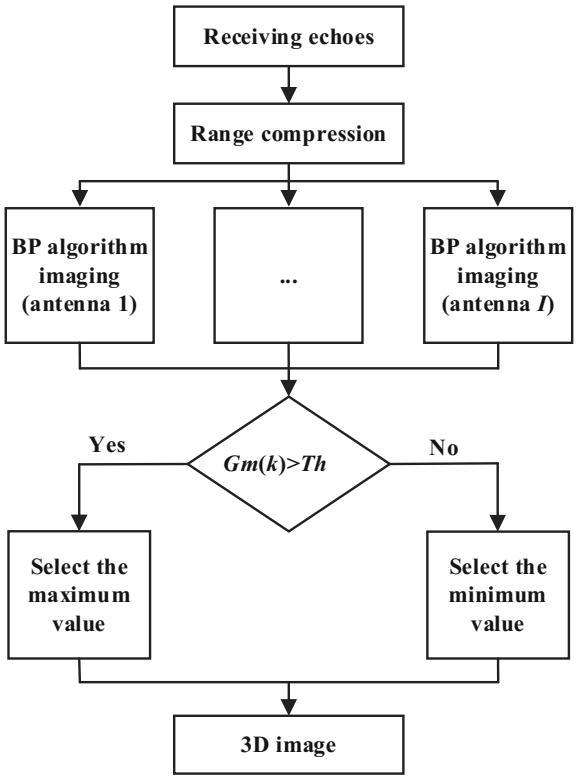



**Fig. 2** Flowchart of MAR CSAR imaging

## 3 Simulation

The numerical simulations have been performed to illustrate the effectiveness of the proposed MAR technique. The main system parameters are as follows: the carrier frequency $f_c$ is 0.3 THz, bandwidth $B$ is 10 GHz, and PRF is 1 MHz. The height of the transmitter $H$ is 2 m, and the moving radius of radar $R_g$ is 0.4 m. The radar moves a whole circle, which means that the azimuth angle is from 0 to $2\pi$. Eight point targets with different altitudes are selected for the imaging. The scatterers' coordinates are $[0.05\,0.05\,0.05]$ $[-0.05\,0.05\,0.05]$, $[-0.05\ -0.05\,0.05]$, $[0.05\ -0.05\,0.05]$, $[0.05\,0\ -0.05]$, $[0\,0.05\ -0.05]$, and $[-0.05\,0-0.05]$, respectively. It is assumed that all the simulation scatters are persistently illuminated by radar with constant reflection coefficient. White Gaussian noise with SNR which equals to 10 dB is added to the echoes.

In the simulation, three receiving antennas are used; the imaging results are shown in Fig. 3. Figure 3a and b shows the images obtained via single antenna receiving technique, while Fig. 3c and d shows the images obtained via multiple antenna receiving techniques. Comparing (a) and (b) with (c) and (d) in Fig. 3, it

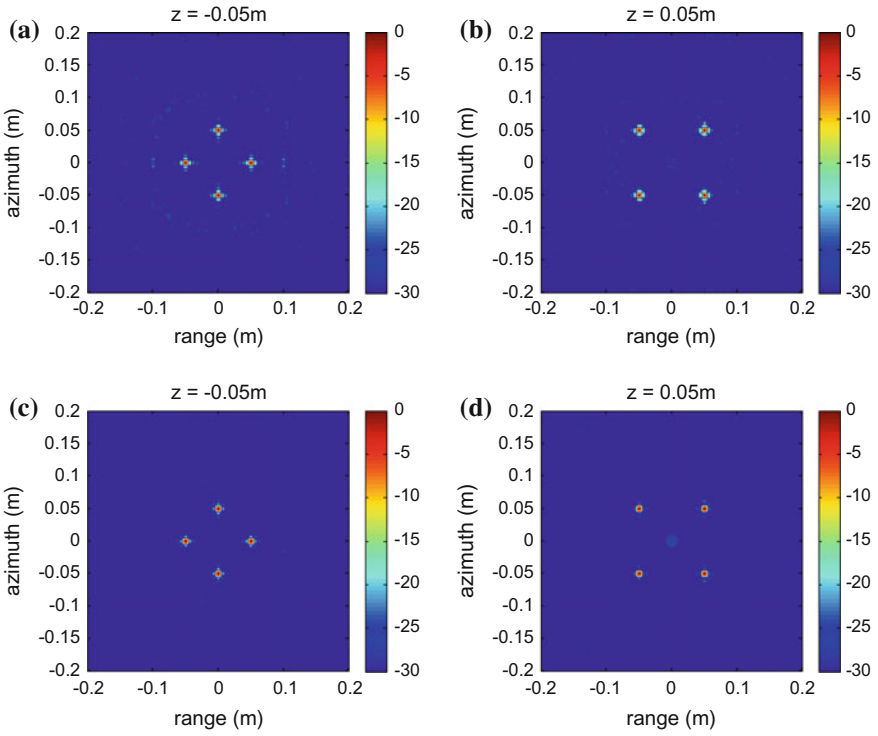



**Fig. 3** The imaging result obtained by (up) single antenna receiving, (down) three antennas receiving, **a**, **c** height z = −0.05 m, **b**, **d** height z = 0.05 m

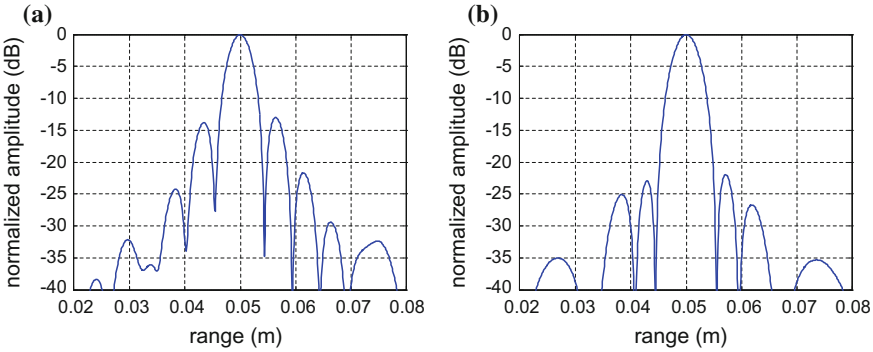



**Fig. 4** The range profile of point target [0.05 0.05 0.05]. **a** amplitude level via single antenna receiving, **b** amplitude level via MAR

can be seen that the images obtained from MAR technique have a higher quality, where both the noise level and the sidelobe level are lower. To further demonstrate this, we plot the range profile of the point [0.05 0.05 0.05], which is shown in Fig. 4. From Fig. 4, we can observe that the first sidelobe level is dramatically decreased and the main lobe does not become wider for image obtained via MAR technique compared with single antenna receiving technique.

## 4 Conclusion

In this paper, a multiple antenna receiving technique for CSAR imaging is proposed. Comparing with the conventional single antenna receiving scheme, the proposed technique could achieve higher quality image with data fusion method, where the sidelobe level and noise level are both lower. Numerical simulations have been performed to verify the effectiveness of the proposed technique.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China under Grants 61371048, 61301265, U1433113.

## References

1. E. Ertin et al., GOTCHA experience report: three-dimensional SAR imaging with complete circular apertures, in *Defense and Security Symposium* (International Society for Optics and Photonics, 2007), pp. 656802–656802
2. B. Zhang et al., A CGRT-CLEAN method for circular SAR three dimensional imaging, in *Communications, Signal Processing, and Systems* (Springer, New York, 2012), pp. 41–52
3. Y. Lin et al., Airborne circular SAR imaging: results at P-band, in *2012 IEEE International Geoscience and Remote Sensing Symposium* (2012), pp. 5594–5597

4. R. Sharma, Analysis of circular aperture SAR image formation, in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers* (2009), pp. 482–487
5. O. Ponce et al., Multibaseline 3-D circular SAR imaging at L-band, in *2009 Conference Record of the Forty-Third Asilomar Conference on Signals, Systems and Computers* (2012), pp. 482–487
6. L.J. Moore, L.C. Potter, Three-dimensional resolution for circular synthetic aperture radar, in *Defense and Security Symposium* (International Society for Optics and Photonics, 2007), pp. 656804–656804
7. H. Balacey et al., Advanced processing sequence for 3-D THz imaging. IEEE Trans. Terahertz Sci. Technol. **6**(2), 191–198 (2016)
8. B. Zhang et al., Terahertz imaging radar with inverse aperture synthesis techniques: system structure, signal processing, and experiment results. IEEE Sens. J. **15**(1), 290–299 (2015)
9. A.F. Yegulalp, Fast backprojection algorithm for synthetic aperture radar, in *Radar Conference. The Record of the 1999 IEEE* (1999), pp. 60–65

# High-Precision Ranging of Ultra-Close Liquid Level

**Mingming Guo, Jinhua Xie and Shuwen Chen**

**Abstract** In the process of high-precision liquid level measurement, radar has to solve the problem of ultra-close liquid level's high-precision survey sometimes, such as the distance of just a few tens of centimeters. For single-channel radar, the echo signal is a series of real data; and its spectrum will become no longer simple when the liquid level is very close. Since for real data, the positive and negative frequency components will superimpose on each other in its spectrum. In this case, the existing high-precision liquid level measurement algorithms are no longer effective. In order to solve this problem, the authors investigate a new high-precision liquid level measurement method. In this method, the echo signal is windowed first; and then a corresponding interpolation algorithm is designed aiming at the window type; finally, iterate the interpolation algorithm, and the high-precision ranging result is obtained. Experiments show that this method can realize the high-precision ranging of ultra-close liquid level. This method can be used for high-precision ranging radar, such as liquid level meter radar, which is of great practical significance.

**Keywords** High-precision ranging · Close liquid level · Single-channel · Real echo signal

## 1 Introduction

High-precision liquid level measurement plays an important role in the various kinds of liquid level automatic control system. In the existing liquid level measurement systems, the characteristics of pressure type liquid level measurement system are the simple principle and low cost, but there are certain limitations in the accuracy of measurement and applications. Ultrasonic level meter has higher ranging accuracy, but its equipment is complicated, which needs to use the catheter to lead ultrasound

M. Guo (✉) · J. Xie · S. Chen
AVIC Leihua Electronic Technology Research Institute, Wuxi 214063, China
e-mail: guo_greeting@126.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_20

into the liquid being measured, so it is more troublesome to install and maintain. Laser ranging system is very strict with the work environment [1].

The microwave level meter system is noncontact, and it is not restricted by liquid state, which can work under the high pressure, high temperature, toxic, corrosive, and viscous liquid environment [1]. In linear frequency modulated continuous wave (LFMCW) radar, the advantages of the low transmitter power, high receiver sensitivity, high range resolution, simple structure and easy to integration makes it suitable for liquid level measurement system [2–5].

This paper is organized as follows. The principle of liquid level measurement of LFMCW radar is introduced first; then the proposed high-precision ranging method of close liquid level is detailed; and then the new method is compared with the traditional high-precision level measurement algorithms; finally, conclusions are given.

## 2 Problem Formulation

LFMCW radar is a kind of radar system that gets the information of target range and velocity through modulating the frequency of continuous wave. In this radar system, the frequency of transmitting signal changes linearly during a modulation period. And the modulation period is much longer than the maximum delay caused by the relative range of radar and target. When radar and target are relatively static, the difference frequency between radar and target is caused only by $\Delta f_r$, which is the relative distance of radar and target. The principle of LFMCW radar is shown in Fig. 1.

Use $\Delta f_r$, we can calculate the target range:

$$R = \Delta f_r \times C/2/K, \tag{1}$$

where, $C$ is the speed of light, $K = B/T_m$ is the slope of linear frequency modulation, $B$ is the bandwidth of frequency modulation, $T_m$ is the modulation period.

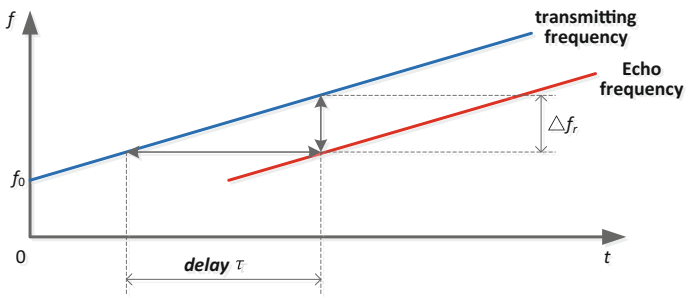



**Fig. 1** LFMCW radar ranging principle

According to the above analysis, the ranging accuracy depends on the accuracy of frequency measurement, which is known as:

$$\Delta f = C/(2B) \tag{2}$$

So the ranging accuracy depends on the bandwidth of radar, and the higher the bandwidth, the more accurate the range measurement.

We can use FFT to estimate the difference frequency roughly, and use a variety of frequency interpolation algorithms for high-precision frequency estimation, such as zero padding, M-Rife [6], Iteration [7], centroid method and so on. However, in the process of high-precision liquid level measurement, radar has to solve the problem of ultra-close liquid level's high-precision survey sometimes, such as the distance of just a few tens of centimeters. For single-channel radar, the echo signal is a series of real data; and its spectrum will become no longer simple when the liquid level is very close. Since for real data, the positive and negative frequency components will superimpose on each other in its spectrum when the liquid level is very close. In this case, the existing high-precision liquid level measurement algorithms are no longer effective. So, we have to study a new solution.

## 3 High-Precision Ranging of Close Liquid Level

The difference frequency signal of single-channel LFMCW radar can be expressed as

$$S = A\ \cos(2\pi f_\tau t) + n(t) \tag{3}$$

where, $A$ denotes the amplitude which is assumed to be a constant for simplicity, $f_\tau = \tau \cdot B/T$ is the difference frequency, $\tau$ is the delay caused by the relative distance between radar and target, $w(t)$ is noise. The additive noise $n(t) \sim N(0, \sigma_w^2)$ refers to the Gaussian white noise.

First of all, the frequency signal is windowed. We take the Hanning window for instance. The coefficients of a Hanning window are computed from the following equation:

$$w(t) = 0.5[1 + \cos(\frac{2\pi t}{T})],\ t \in (-\frac{T}{2}, \frac{T}{2}] \tag{4}$$

where, $T$ is the modulation period. The frequency response of Hanning window can be obtained by FFT, which is expressed as

$$W(n) = \frac{\sin(\pi Tn)}{2\pi n} - \frac{1}{4}[\frac{\sin(\pi Tn)}{\pi(n + 1/T)} + \frac{\sin(\pi Tn)}{\pi(n - 1/T)}] \tag{5}$$

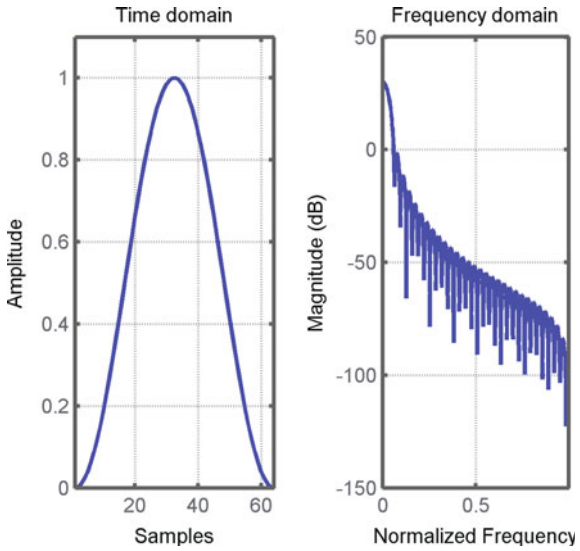



**Fig. 2** 64-point Hanning window

And the time domain and frequency domain of Hanning window are shown as follows (Fig. 2).

Then a corresponding interpolation algorithm is designed. In the discrete-time radar system, we can assume that $1/T = 1$, and the distance from component of the maximum amplitude to 0 is $k$, then we can obtain the maximum amplitude, it is

$$X_0 = \sin(\pi T k)\left(\frac{1}{2\pi k} - \frac{1}{4\pi(k+1)} - \frac{1}{4\pi(k-1)}\right) \tag{6}$$

If the second maximum component is on the left side of $k$, then the second maximum amplitude can be written as

$$X_1 = \sin(\pi T k)\left(\frac{-1}{2\pi(k-1)} + \frac{1}{4\pi k} + \frac{1}{4\pi(k-2)}\right) \tag{7}$$

By (6) and (7), we can obtain that

$$\frac{X_1}{X_0} = \frac{k+1}{2-k} \tag{8}$$

and

$$k = \frac{2X_1 - X_0}{X_0 + X_1} \tag{9}$$

If the second maximum component is on the right side of $k$, then $k$ can be expressed as

$$k = \frac{X_0 - 2X_1}{X_0 + X_1} \tag{10}$$

If the corresponding frequency component of $X_0$ is $k_0$, then the optimal estimated can be expressed as

$$\hat{k}_\tau = k_0 - k \tag{11}$$

In order to reduce the sensitivity of this algorithm for noise, and further improve the measurement precision, we have to iterate this algorithm.

Choose two frequency components near $\hat{k}_\tau$, they are defined as

$$k_1 = \hat{k}_\tau - 0.5,\ k_2 = \hat{k}_\tau + 0.5 \tag{12}$$

Use (5) to calculate the spectrum amplitudes of $k_1$ and $k_2$, respectively defined as $X_1^{(1)}$ and $X_2^{(1)}$, then the new estimated $\hat{k}_\tau^{(1)}$ can be obtained by using (9)–(11). Iterate N times of this method, we will get the optimal estimated $\hat{k}_\tau^{(N)}$.

Finally, the optimal estimated $\hat{R}$ can be obtained as follows:

$$\hat{R} = \frac{C \cdot fs}{2K \cdot N_{FFT}} \cdot \hat{k}_\tau^{(N)}, \tag{13}$$

where, $fs$ is the sampling rate of radar, $N_{FFT}$ is the FFT points, $N$ is iteration times.

## 4 Performance Analysis

In this section, we will first estimate the range of ultra-close liquid level with the proposed method. Then we present the error curves of zero padding, chirp-Z transform, and the new method to compare their detection performance.

We use the method given in (13) to calculate the range of a series of ultra-close liquid levels, and get the mean and standard deviation (Std) of estimated ranges through Monte Carlo experiment. The true ranges of liquid levels, the mean and Std of estimated ranges of different liquid levels are shown in Table 1. The Hanning window is added on the difference frequency signal, the sampling rate is 200 kHz, FFT point is 1024, and slope of linear frequency modulation is 390.62 GHz/s, and Monte Carlo experiment number is 100.

**Table 1** The true value of R and the estimate result under SNR = 20 dB

| True R/cm | Mean of $\hat{R}$/cm | Std of $\hat{R}$ |
|---|---|---|
| 20 | 19.995 | 3.3563e-4 |
| 25 | 24.993 | 3.2775e-4 |
| 30 | 29.993 | 2.9195e-4 |
| 35 | 35.001 | 2.9051e-4 |
| 40 | 39.997 | 2.8116e-4 |
| 45 | 44.993 | 2.6446e-4 |
| 50 | 50.001 | 3.4559e-4 |

We conduct simulations for detecting liquid levels by using the following four methods:

1. Zero padding method
2. M-Rife method
3. Iteration method
4. New method proposed in this paper

In these simulations, the ranging results of the new method we proposed are compared with those of other three methods. The zero padding number is 16 times of signal length. The window used in M-Rife method is Blackman-Harris window. The iteration times of Iteration method and the new method is both 3. The simulation results are shown in Fig. 3.

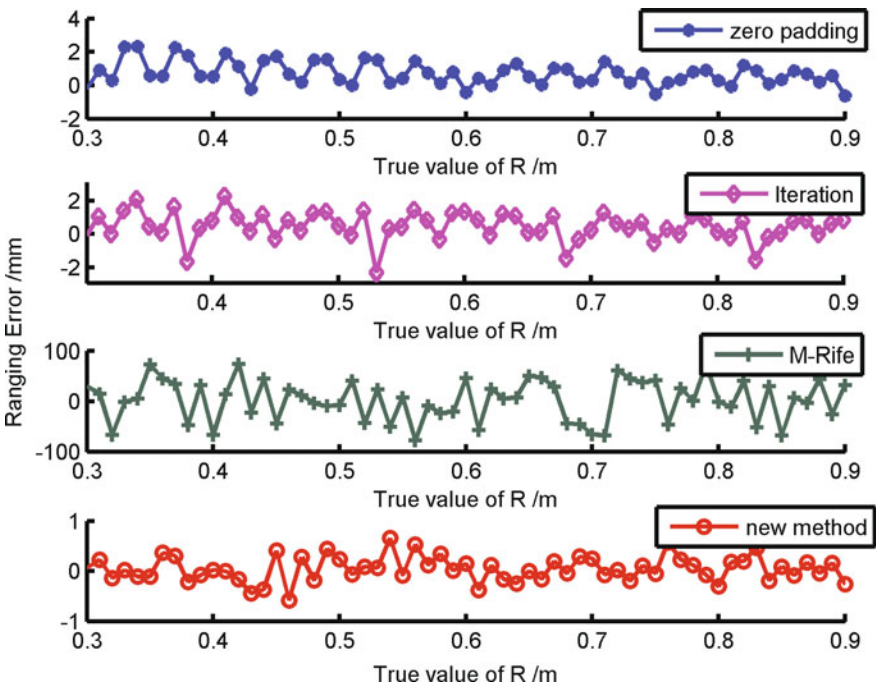



**Fig. 3** Simulation results of 4 methods

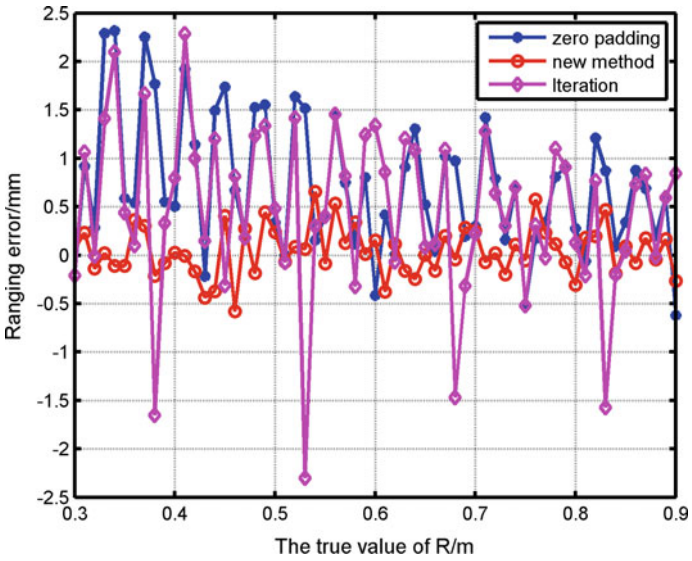



**Fig. 4** Simulation results of the other 3 methods

Figure 3 shows that the ultra-close liquid level ranging precision of M-Rife method is far below the other three methods. In order to further compare the other three methods; we put ranging error curves in one picture, which is shown in Fig. 4.

It is easy to know from Fig. 4 that the ranging precision of the new method is better than zero padding and Iteration method.

## 5 Conclusion

This paper investigates a new high-precision ranging method. Using a window, designing a corresponding interpolation and iteration algorithm, the authors solve the problem of the existing methods' failure in the measurement of ultra-close liquid level. The detecting performance of the new method is demonstrated by the comparison of four methods, which proves that the proposed method has better ranging precision in the measurement of ultra-close liquid level. This method has a good application foreground in the field of high-precision liquid level measurement.

## References

1. G. Qi, High precision FMCW level radar system design and principle. Dalian, DLMU Daliann Maritime University (2001)
2. A.G. Stove, Linear FMCW radar techniques. IEE Proc.-F, Radar Signal Process. **139**(5), 343–350 (1992)

3. P.D.L. Beasley et al., Solving the problems of a single antenna frequency modulated CW radar. Proc. IEEE Int. Conf. Radar **90**, 391–395
4. D.G.C. Luck, *Frequency Modulated Radar* (McGraw-Hill, 1949)
5. A.J. Hymans, J. Lait, Analysis of a frequency-modulated continuous-wave ranging system. IEE Proc. B **107B**, 365–372 (1960)
6. J. Xu, Y. Liu, Quick algorithm for frequency estimation of sinusoid signal of any length. J. Nanjing Univ. Aeronaut. Astronaut. **40**(6), 794–798 (2008)
7. J. Tsui, Digital techniques for wideband receivers. American (2002)

# SAR Image De-noising Based on Nuclear Norm Minimization Fusion Algorithm

**Shuaiqi Liu, Liu Ming, Mingzhu Shi, Xin Qi and Hu Qi**

**Abstract** Synthetic aperture radar (SAR) images play a quite important role in military and environmental monitoring. But the SAR image was greatly affected by coherent noise, which affects its application in the subsequent image analysis. In most of the SAR image de-noising algorithms in hand, the same operation is applied to the whole SAR image, which leads to artificial texture or edge blur. In order to overcome this shortcoming, this paper proposed a new SAR image de-noising method based on nuclear norm minimization (NNM) fusion algorithm. The noisy SAR image is de-noised by two different algorithms, and two de-noising images are fused to final de-noising image based on nuclear norm minimization fusion algorithm. Experimental results show that the proposed algorithm not only effectively improves the visual effect and objective indicators of de-noising image but preserves the local structure of the image better.

S. Liu (✉) · X. Qi · H. Qi
College of Electronic and Information Engineering, Hebei University, Baoding, China
e-mail: shdkj-1918@163.com

S. Liu · X. Qi · H. Qi
Key Laboratory of Digital Medical Engineering of Hebei Province, Baoding, China

L. Ming (✉)
Department of Personnel, Hebei University, Baoding, China
e-mail: annie_liuming@126.com

M. Shi
College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_21

## 1 Introduction

SAR has the advantages of all-weather, all-day work, high precision and abundant information, intuitive real and so on. In recent years, it is widely used in the homeland, hydrology, military and other fields. Due to the coherent imaging mechanism, SAR image contains coherent noise [1–3]. SAR image de-noising algorithms are mainly divided into the following two categories: spatial domain de-noising algorithms and transform domain de-noising algorithms [4]. Spatial de-noising algorithms (including Lee filter [5], Kuan filtering [6], Forst filtering [6], etc.) are based on local statistical characteristics of image, which has a good de-noising effect in the smooth area, but for the area of rich image edge and texture information, it is easy to be over smoothness and lead to image detail information loss; Transform domain algorithms include image de-noising algorithms based on wavelet transform [4, 7] and image de-noising algorithms based on multi-scale geometric transform (Contourlet, Curvelet, Shearlet, etc. [8–11]). Although many amazing results are obtained by these de-noising algorithms, in most of the SAR image de-noising algorithms, the same operation is used to the whole SAR image, which lead to artificial texture or edge blur. In order to overcome this shortcoming, we proposed a new SAR image de-noising method based on nuclear norm minimization fusion algorithm. Our method can both fuse two de-noised images by two different algorithms and de-noise the fused image. So, the new method not only effectively improves the visual effect and objective indicators of de-noising image but also preserves the local structure of the image better.

## 2 Fusion Algorithm Based on Nuclear Norm Minimization

### *2.1 The Theory of Nuclear Norm Minimization*

Generally, the noisy image model is simplified as follows:

$$y = x + n \tag{1}$$

where $y$ is the noisy image, $x$ is the clear image, $n$ is the Gaussian noise with the variance $\sigma_n$.

According to soft thresholds for the singular values, NNM can keep the detail information and low rank structures of image effectively. Firstly, partition the noisy image $y$ and assume the $j$-th block to be $y_j$. Then use the block matching algorithm [12] to find its similar blocks, and then, stack the similar blocks into a group. Deal all of the reference blocks with the mentioned steps and we will obtain a block group matrix named $Y_j$. Consequently, the noise suppression model in formula (1) can be converted into

$$Y_j = X_j + N_j \tag{2}$$

where $X_j$ and $N_j$ are the block matrixes of clear image and noise image, respectively. Because of the low rank of $X_j$, the noise image can be made inpainting by nuclear norm minimization algorithm. The solution to the minimization problem can be expressed as folows:

$$\hat{X} = \arg\min_X \|Y - X\|_F^2 + \lambda\|X\|_* \tag{3}$$

where regular parameter $\lambda$ is positive constant number, $Y$ is the known matrix and $X$ is the wanted low rank matrix approximation to $Y$. $\|\bullet\|_F$ is Frobenius norm, $\|\bullet\|_F = \sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}^2}$, $\|X\|_*$ is the nuclear norm of $X$. The global optimal solution of (3) is

$$\hat{X} = US_\lambda(\sum)V^T \tag{4}$$

where $Y = U\sum V^T$ is the singular value decomposition of the matrix $Y$. Matrixes $U$ and $V$ meet the condition $UU^T = I$ and $VV^T = I$, $\sum$ is the diagonal matrix composed of the singular value of matrix $Y$, and $\sum_{ii} = \sqrt{\lambda_i}$, the values of $\sum_{ii}$ are in descending order. $S_\lambda(\sum)$ is the soft threshold operation of matrix $\sum$, thus each diagonal element $\sum_{ii}$ in $\sum$ should satisfy the following equation.

$$S_\lambda(\sum)_{ii} = \max(\sum_{ii} - \lambda, 0) \tag{5}$$

As shown in problem (5), the nuclear norm minimum function is a convex function, which can be solved rapidly with soft threshold function. So, the algorithm is used widely in low rank approach solution problem.

## 2.2 Image Fusion Algorithm

The image fusion algorithm based on nuclear norm minimization in [13] is used in our paper. We give some reviews of this work. Without loss of generality, set the need fusion image as $A$ and $B$, set the fused image as $F$. Apparently, it is easy to generalize to the multiple image fusion.

Firstly, partition the image $A$ and $B$, and find the nonlocal similar blocks through block matching. Stack the similar blocks into a group, and deal all of the reference blocks with the mentioned steps and we will obtain block group matrixes named $X_A$ and $X_B$. Then compute the sharing similar blocks $L_W^S(r)$ of $X_A$ and $X_B$ [13, 14]. Assume $X_{A(r)}$ to be the block group matrix constructed by the similar blocks belong to $X_A$ in $L_W^S(r)$, $X_{B(r)}$ to be the block group matrix constructed by the similar blocks belong to $X_B$ in $L_W^S(r)$. $X_{A(r)} = U_{X_{A(r)}} \sum^{X_{A(r)}} V_{X_{A(r)}}{}^T$ is singular value decomposition

of $X_{A(r)}$ and $X_{B(r)} = U_{X_{B(r)}} \sum^{X_{B(r)}} V_{X_{B(r)}}{}^{T}$ is singular value decomposition of $X_{B(r)}$. Set $X_{F(r)}$ to be the block group matrix constructed by the similar blocks belong to $X_F$ in $L_W^S(r)$. And then, we can get

$$X_{F(r)}(:,i) = \begin{cases} X_{A(r)}(:,i) & if \sum_{ii}^{X_{A(r)}} > \sum_{ii}^{X_{B(r)}} \\ \left(X_{A(r)}(:,i) + X_{B(r)}(:,i)\right)/2 & if \sum_{ii}^{X_{A(r)}} = \sum_{ii}^{X_{B(r)}} \\ X_{B(r)}(:,i) & if \sum_{ii}^{X_{A(r)}} \leq \sum_{ii}^{X_{B(r)}} \end{cases} \tag{6}$$

where $i = 1, 2, ..., r$. Finally we will get the fused image by using the nuclear norm minimization algorithm. In this paper, the value of $r$ is set to 7.

## 3 The Proposed Method

Generally, SAR de-noising algorithms apply the same de-noised operation to the whole image. Due to the influence of different scales texture, SAR image is flat in some areas, and steep in other areas. Obviously, adopting different de-noising algorithms in different texture regions of SAR image will obtain better result. As is known to all, the Bayesian shearlet shrinkage for SAR image de-noising via sparse representation (BSSR) proposed in [10] has good effect in sparse smooth area; however, it will blur the edge of SAR image in steep texture area. For larger texture steep zones, SAR image de-noising based on generalized nonlocal means in nonsubsample shearlet domain (STGNL) has a better effect, but for the flat area, the algorithm will often bring man-made texture. So we incorporate the two algorithms to get better de-noising effect.

The steps of the proposed method based on nuclear norm minimized image fusion algorithm are as follows:

***Step1***: Apply BSSR and NSTGNL to SAR image and get the de-noising image $A$ and $B$;

***Step2***: Overlap the blocks of the images $A$ and $B$, and obtain the block group matrixes $X_A$ and $X_B$. Fuse the matrixes by NNM and get the fused block group matrix $X_F$;

***Step3***: Reset $X_F$ according to stack order and we will get final de-noised image.

## 4 Experimental Results and Discussion

In order to show the effectiveness of our algorithm, we compare the proposed method (NNM) with Lee filter, NSTGNL proposed in [4] and BSSR proposed in [10]. The results give the advantages of the proposed algorithm. Firstly, the remote sensing

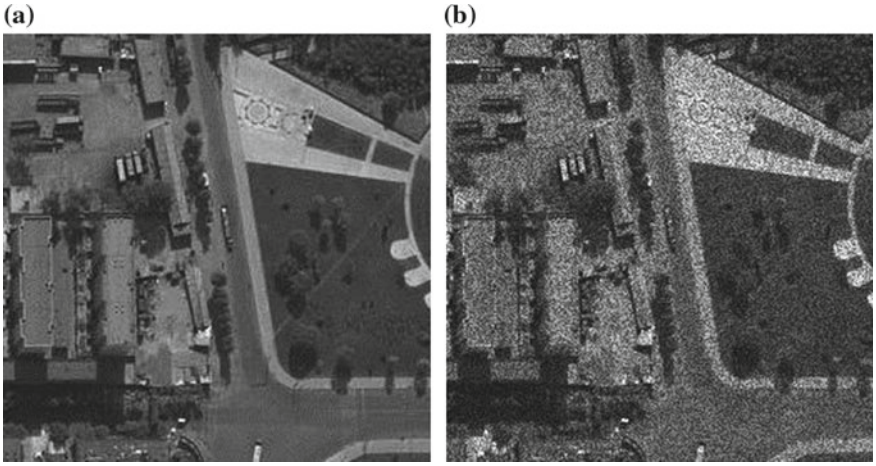



**Fig. 1** The remote sensing SAR image: **a** Original SAR image. **b** noise SAR image

image is tested. Figure 1a is the original image and Fig. 1b is the noisy image with the multiplicative noise in variance of 0.05. Figure 2 shows the experimental results.

From the experimental results, Fig. 2a shows that the image is blurred and it lost a lot of details after Lee filter. And from Fig. 2b, we can see NSTGNL can suppress speckle effectively, but there is still some man-made texture in de-noised image. In Fig. 2c, BSSR blurs the edge of SAR image. Figure 2d shows that the proposed algorithm in this paper has better visual effect than NSTGNL and NSSR.

In order to better display the superiority of the proposed algorithm, four kinds of objective evaluation criteria are given in this paper to reveal the advantage of the algorithm. They are peak signal to noise ratio (PSNR) [10], equivalent numbers of looks (ENL) [10], and edge preservation index (EPI) [10]. The higher PSNR shows better de-noising ability of algorithm, higher ENL shows better de-noising visual effect and higher EPI shows stronger ability to keep the edge. Table 1 shows the objective evaluation values of each algorithm after de-noising the noisy image of different noise variance.

From the objective evaluation shown in Table 1, we can confirm that our algorithm can maintain the structure of the image information better, which makes the EPI of our algorithm higher than other algorithms.

In order to test the de-noised effect of the proposed algorithm in nature SAR image, an electric power line design SAR image of physical network experiment research center of Shanxi province in Fig. 3 is selected to test all the de-noised methods.

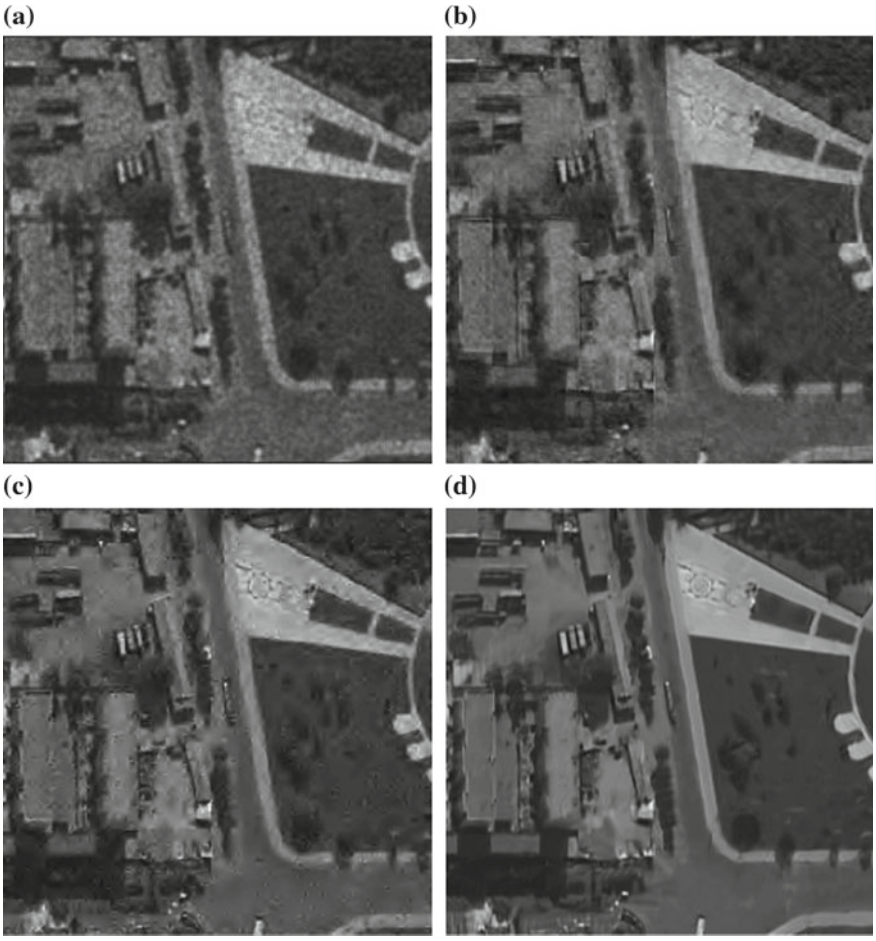



**Fig. 2** Comparison of results on remote sensing SAR image. **a** De-noised result of Lee. **b** De-noised result of NSTGNL. **c** De-noised result of BSSR. **d** De-noised result of ours

Figure 4 is the experimental results. Figure 4a shows that Lee filter has worst effect in dealing with the speckle of SAR image. Figure 4b shows that NSTGNL introduces synthetic texture, and there is some obvious linear texture in the lower left corner of the electric wire. Figure 4c shows that BSSR blurs the edge of SAR image, and the wire in the lower left corner is blurred so violently that it is difficult to identify the lines. Figure 4d shows that the proposed de-noised image has the best visual effect.

In order to illustrate the effect of the de-noised algorithm better, the objective evaluation standard of de-noising algorithms is given in Table 2. It can be seen that all of the indexes of the proposed are the highest, which fully shows the effectiveness of the proposed algorithm.

**Table 1** The performance of different de-nosing methods

| Noise variance | The de-nosing methods | PSNR(db) | ENL | EPI |
|---|---|---|---|---|
| 0.10 | Lee | 18.19 | 7.04 | 0.54 |
| | NSTGNL | 20.37 | 8.52 | 0.85 |
| | BSSR | 20.05 | 8.10 | 0.84 |
| | Ours | 21.01 | 10.16 | 0.92 |
| 0.05 | Lee | 20.32 | 7.40 | 0.58 |
| | NSTGNL | 22.55 | 9.72 | 0.88 |
| | BSSR | 23.78 | 9.56 | 0.87 |
| | Ours | 25.01 | 12.75 | 0.95 |
| 0.02 | Lee | 23.18 | 8.05 | 0.60 |
| | NSTGNL | 25.55 | 12.01 | 0.93 |
| | BSSR | 26.77 | 11.88 | 0.91 |
| | Ours | 29.11 | 14.05 | 0.97 |

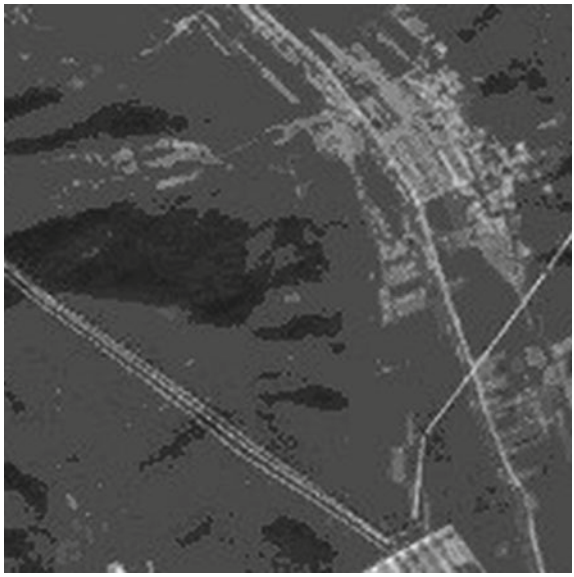


**Fig. 3** The nature SAR image

**Table 2** The performance of de-nosing methods in nature SAR image

| The de-nosing methods | PSNR(db) | ENL | EPI |
|---|---|---|---|
| Lee | 19.33 | 11.04 | 0.85 |
| NSTGNL | 31.65 | 20.57 | 0.95 |
| BSSR | 34.33 | 19.90 | 0.94 |
| Ours | 35.01 | 20.85 | 0.97 |

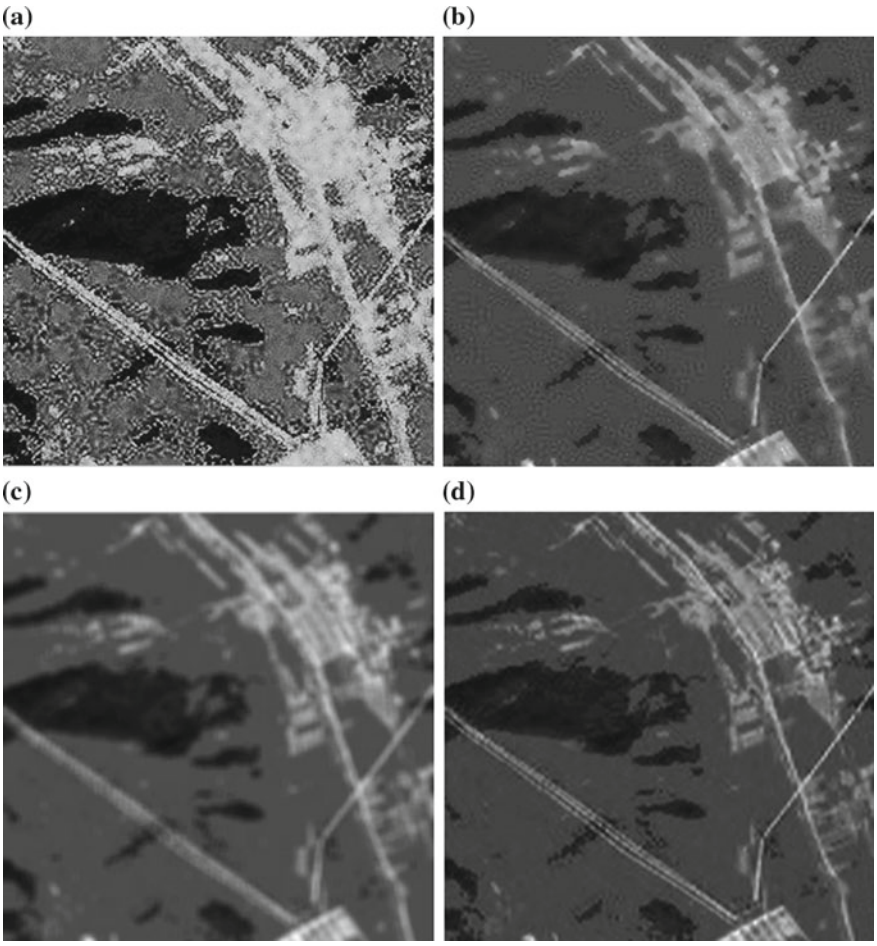



**Fig. 4** Comparison of results on nature SAR image. **a** De-noised result of Lee. **b** De-noised result of NSTGNL. **c** De-noised result of BSSR. **d** De-noised result of ours

## 5 Conclusion

In order to overcome the existing disadvantage of the SAR image de-noising algorithms that it always aims at the whole image with a single algorithm, we proposed a SAR image de-noising algorithm based on nuclear norm minimization fusion algorithm. The algorithm fuses the de-noising image that de-noised by different algorithm. It combines the advantages of different de-noising algorithms and overcomes their disadvantages. We will focus on reducing the algorithm complexity in the future and achieving the final practical application of the algorithm.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (61572063, 61401308), Natural Science Foundation of Hebei University (2014-303), Natural Science Foundation of Hebei Province (F2016201142, F2016201187), Science research project of Hebei Province (QN2016085, ZC2016040), Natural Social Foundation of Hebei Province (HB15TQ015), Science and technology support project of Hebei Province (15210409), National Comprehensive Ability Promotion Project of Western and Central China.

## References

1. J.W. Goodman, Some fundamental properties of speckle. J. Opt. Soc. Am. **6**(11), 1145–1150 (1976)
2. Z.X. Liu, S.H. Hu, Y. Xiao et al., SAR image target extraction based on 2-D leapfrog filtering, in *ICSP2010 2010* (IEEE Press, 2010), pp. 943–1946
3. K.B. Eom, Anisotropic adaptive filtering for speckle reduction in synthetic aperture radar images. Opt. Eng. **50**(5), 97–108 (2011)
4. S. Liu, P. Geng, M. Shi et al., SAR image de-noising based on generalized non-local means in non-subsample shearlet domain, in *CSPS 15* (Springer, Chengdu, China, 2015), pp. 221–229
5. J.S. Lee, Digital image enhancement and noise filtering by use of local statistics. IEEE Trans. Pattern Anal. Mach. Intell. PAMI **2**(2), 165–168 (1980)
6. V. Frost, J. Stiles, K. Shanmugan et al., A model for radar images and its application to adaptive digital filtering of multiplicative noise. IEEE Trans. Pattern Anal. Mach. Intell. PAMI **4**(2), 157–166 (2011)
7. S.G. Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression. IEEE Trans. Image Process. **9**(9), 1532–1546 (2000)
8. G. Chen, X. Liu, Contourlet-based despeckling for SAR image using hidden Markov tree and Gaussian Markov models, in *1st Asian and Pacific Conference on Synthetic Aperture Radar*, Huangshan, China (2007), pp. 784–787
9. J.L. Starck, E.J. Candes, D.L. Donoho, The curvelet transform for image denoising. IEEE Trans. Image Process. **11**(6), 670–684 (2002)
10. S. Liu, S. Hu, Y. Xiao, Bayesian shearlet shrinkage for SAR image de-noising via sparse representation. Multidimension. Syst. Signal Process. **25**(4), 683–701 (2014)
11. S. Liu, M. Shi, S. Hu, Y. Xiao, Synthetic aperture radar image de-noising based on shearlet transform using context-based model. Phys. Commun. **13**(PartC), 221–229 (2014)
12. S. Gu, L. Zhang, W. Zuo et al., Weighted nuclear norm minimization with application to image denoising, in *CVPR 2014* (IEEE Press, Columbus, USA, 2014), pp. 2862–2869
13. S. Liu, T. Zhang, H. Li, J. Zhao et al., Medical image fusion based on nuclear norm minimization. Int. J. Imaging Syst. Technol. **25**(4), 310–316 (2015)
14. D. Guo, J. Yan, X. Qu, High quality multi-focus image fusion using self-similarity and depth information. Opt. Commun. **338**(1), 138–144 (2015)

# Neighbourhood Feature Space Discriminant Analysis for High Range Resolution Radar Target Recognition

**Xuelian Yu, Xuechao Qu, Yuguo Wang, Huaqiong Li and Xuegang Wang**

**Abstract** A new subspace learning method called neighbourhood feature space (NFS) discriminant analysis is proposed for high range resolution radar target recognition. An intra- and extra-class NFS is formed for each sample, and the within- and between-class scatter matrices are redefined according to the point-to-space difference, respectively. An additional weight is introduced to emphasise samples near the class boundaries and de-emphasise samples far from the class boundaries. Both the between-class separability and within-class local geometry preservation are considered to seek a discriminative subspace for classification. Experimental results of the measured high-range resolution profile data demonstrate the effectiveness of the proposed method.

## 1 Introduction

Until now, many state-of-the-art subspace learning methods have been developed and successfully applied to radar target recognition using high range resolution profile (HRRP). Basically, these subspace learning methods can be grouped into three categories. The first category is represented by the principal component analysis (PCA) and linear discriminant analysis (LDA), which are based on the conventional point-to-point (P2P) metric. The second category is represented by the generalised nearest feature line (GNFL) [1], uncorrelated discriminant nearest feature line analysis (UDNFLA) [2] and the nearest feature line embedding [3], which are based on the point-to-line (P2L) metric. The recently proposed nearest feature space analysis (NFSA) [4] and the discriminant NFSA (DNFSA) [4] represent the third category and adopt the point-to-space (P2S) metric.

X. Yu (✉) · X. Qu · X. Wang
University of Electronic Science and Technology of China, No.2006, Xiyuan Avenue, West High-Tech Zone, Chengdu 611731, China
e-mail: xlyu2012@uestc.edu.cn

Y. Wang · H. Li
The Second Research Institute of CAAC, Chengdu 610041, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_22

Generally, the P2L-based methods can perform better than the P2P-based ones by enlarging the representational capacity of limited training samples through feature line construction [1–3], which however is computationally expensive. The P2S-based methods can significantly reduce the computational complexity by constructing a very few feature spaces. In addition, the P2S methods have been shown to outperform the P2L and P2P methods because the P2S connectivity relationship can provide more information to virtually enlarge the training samples [4, 5]. However, both the NFSA and DNFSA ignore the local geometry information embedded in the neighbourhood, which has been proven to be very useful for radar target recognition [6]. Consequently, if some intra-class samples are far away from or some extra-class samples are close to the query point, the P2S relationship constructed in the NFSA and DNFSA might not faithfully reflect the intrinsic structure information of the original samples, which would degrade the recognition performance.

In this work a new subspace learning method called neighbourhood feature space discriminant analysis (NFSDA) is proposed for radar target recognition. For each training sample, an intra- and extra-class NFS is formed, and the intra- and extra-class P2S difference is accordingly defined. The method achieves good discriminative ability by considering both the between-class separability and within-class local geometry preservation. Moreover, a sample weight is introduced to further emphasise samples near the class boundaries and to de-emphasise samples far from the class boundaries.

## 2 NFSDA

Consider a dataset $X = [x_1, \ldots, x_N]$ in $R^D$. Each sample $x_i (i = 1, \ldots, N)$ belongs to one of the $C$ classes $\omega_1, \ldots, \omega_C$. The problem is to find a transformation matrix $\boldsymbol{V}$ that maps each sample $x_i \in R^D$ to a low-dimensional subspace $R^d (d < D)$ by $y_i = V^T x_i$.

### *2.1 NFS and P2S Difference*

For each sample $x_i$, we find its $K_1$ intra-class nearest neighbours, denoted as $x_{w,1}, \ldots, x_{w,K_1}$. The space spanned by these neighbours is called the intra-class NFS of $x_i$, denoted by $F_w(x_i)$. Similarly, $K_2$ extra-class nearest neighbours $x_{b,1}, \ldots, x_{b,K_2}$ are determined to span the extra-class NFS of $x_i$, denoted by $F_b(x_i)$.

Subsequently, we define the intra- and extra-class P2S difference with respect to $x_i$ as

$$\boldsymbol{\Delta}_{w,i} = x_i - x_{i,F_w} \tag{1}$$

$$\boldsymbol{\Delta}_{b,i} = x_i - x_{i,F_b} \tag{2}$$

where $x_{i,F_w}$ and $x_{i,F_b}$ are the projection points of $x_i$ onto its intra- and extra-class NFS, respectively, which can be obtained by [5]:

$$x_{i,F_w} = X_w \left(X_w^T X_w\right)^{-1} X_w^T x_i \tag{3}$$

$$x_{i,F_b} = X_b \left(X_b^T X_b\right)^{-1} X_b^T x_i \tag{4}$$

where $X_w = [x_{w,1}, \ldots, x_{w,K_1}]$ and $X_b = [x_{b,1}, \ldots, x_{b,K_2}]$.

## 2.2 Derivation of NFSDA

First, in our method, the within-class local geometry of each sample $x_i \in R^D$ is characterised by the intra-class P2S difference defined in (1). To preserve this local geometry in low-dimensional subspace $R^d$, we propose to minimise the sum of the intra-class P2S differences by the following:

$$\begin{aligned}\min_V J_1(V) &= \sum_{i=1}^{N} \omega_i \left\| y_i - y_{i,F_w} \right\|^2 = \sum_{i=1}^{N} \omega_i \left\| V^T x_i - V^T x_{i,F_w} \right\|^2 \\ &= \mathrm{tr}\left\{ V^T \left[ \sum_{i=1}^{N} \omega_i (x_i - x_{i,F_w})(x_i - x_{i,F_w})^T \right] V \right\} \\ &= \mathrm{tr}\left\{ V^T \left[ \sum_{i=1}^{N} \omega_i (\boldsymbol{\Delta}_{w,i})(\boldsymbol{\Delta}_{w,i})^T \right] V \right\}.\end{aligned} \tag{5}$$

Second, for classification, we should disperse as far as possible the samples from different classes, especially those that are adjacent to each other in the low-dimensional subspace. Thus, in the NFSDA, we propose to maximise the sum of extra-class P2S differences by the following:

$$\begin{aligned}\max_V J_2(V) &= \sum_{i=1}^{N} \omega_i \left\| y_i - y_{i,F_b} \right\|^2 = \sum_{i=1}^{N} \omega_i \left\| V^T x_i - V^T x_{i,F_b} \right\|^2 \\ &= \mathrm{tr}\left\{ V^T \left[ \sum_{i=1}^{N} \omega_i (x_i - x_{i,F_b})(x_i - x_{i,F_b})^T \right] V \right\} \\ &= \mathrm{tr}\left\{ V^T \left[ \sum_{i=1}^{N} \omega_i (\boldsymbol{\Delta}_{b,i})(\boldsymbol{\Delta}_{b,i})^T \right] V \right\}.\end{aligned} \tag{6}$$

By combining (5) and (6), we can maximise the following objective function:

$$\begin{aligned} J(V) &= \mathrm{tr}\left\{ V^T \left[ \sum_{i=1}^{N} \omega_i (\boldsymbol{\Delta}_{b,i})(\boldsymbol{\Delta}_{b,i})^T \right] V - V^T \left[ \sum_{i=1}^{N} \omega_i (\boldsymbol{\Delta}_{w,i})(\boldsymbol{\Delta}_{w,i})^T \right] V \right\} \\ &= \mathrm{tr}\left\{ V^T \left[ \sum_{i=1}^{N} \omega_i (\boldsymbol{\Delta}_{b,i})(\boldsymbol{\Delta}_{b,i})^T - \sum_{i=1}^{N} \omega_i (\boldsymbol{\Delta}_{w,i})(\boldsymbol{\Delta}_{w,i})^T \right] V \right\} \\ &= \mathrm{tr}\left\{ V^T \left( S_b^{\mathrm{P2S}} - S_w^{\mathrm{P2S}} \right) V \right\} \end{aligned} \tag{7}$$

where

$$S_w^{\mathrm{P2S}} = \sum_{i=1}^{N} \omega_i (\boldsymbol{\Delta}_{w,i})(\boldsymbol{\Delta}_{w,i})^T \tag{8}$$

$$S_b^{\mathrm{P2S}} = \sum_{i=1}^{N} \omega_i (\boldsymbol{\Delta}_{b,i})(\boldsymbol{\Delta}_{b,i})^T \tag{9}$$

are the within- and between-class scatter matrices in the NFSDA, which are redefined according to the intra- and extra-class P2S difference defined in (1) and (2), respectively. $\omega_i$ is a sample weight, which is introduced to emphasise samples near the class boundaries and to de-emphasise samples far from the class boundaries [7], which is defined as

$$\omega_i = \frac{\|\boldsymbol{\Delta}_{w,i}\|}{\|\boldsymbol{\Delta}_{w,i}\| + \|\boldsymbol{\Delta}_{b,i}\|}. \tag{10}$$

Finally, the maximisation problem in (7) can be reduced to an eigenvalue problem.

$$\left( S_b^{\mathrm{P2S}} - S_w^{\mathrm{P2S}} \right) v = \lambda v \tag{11}$$

Transformation matrix $\boldsymbol{V}$ of the NFSDA can be constituted by the $d$ eigenvectors corresponding to the first $d$ largest eigenvalues in (11). Once $\boldsymbol{V}$ is determined, for any data point $x \in R^D$, the transformed feature in low-dimensional subspace $R^d$ is given by $y = V^T x$.

## 3 Experiments

To evaluate the performance of the proposed algorithm, we perform radar target recognition experiments using measured HRRP data. The data are collected from three flying airplanes, namely, An-26, Yark-42, and Cessna Citation S/II. Each

airplane has 260 profiles, and each profile has 256 range bins [6]. Each profile is pre-processed by energy normalisation. For each airplane, *Ntr* profiles are selected for training, and the remaining profiles are used for testing. The value of *Ntr* is set to 20, 10, and 5. At each *Ntr* value, the training and testing processes are executed 20 times, and the obtained recognition rates are averaged.

The performance of the NFSDA is compared with those of the PCA, LDA, GNFL, UDNFLA, NFSA, and DNFSA. Because we only focus on feature extraction, with regards to the classification, the nearest neighbour classifier that uses Euclidean metric is employed for simplicity. To achieve a fair comparison, all results reported here are based on the best-tuned parameters of all the compared methods.

Table 1 lists the recognition rates achieved by the seven methods at each *Ntr* value. The last column lists the average recognition rates of each method under the three cases. Table 1 shows that the proposed NFSDA consistently obtains higher recognition rates than the NFSA and DNFSA and outperforms all the other methods. The results indicate that not only more discriminative power is generated but also the intrinsic locality information among samples is better preserved in the proposed method, which is very helpful in improving the recognition performance.

Parameters $K_1$ and $K_2$ of the NFSDA, are manually set during the experiments. Now, we analyse the effect of the two parameters. Parameters $K_1$ and $K_2$ are assigned in the ranges of 3–10, 3–10, and 3–5 when $Ntr = 20$, 10, and 5, respectively. The experiments are performed for each possible parameter combination. The recognition rates of all combinations are recorded, and the standard deviations of the parameter combinations at each *Ntr* value are listed in Table 2. Because the standard deviations in our experiment are less than 0.54%, we can conclude that the recognition rates of the proposed NFSDA are insensitive to the two parameters.

**Table 1** Comparison of recognition rates (%) of the seven methods

| Ntr | 20 | 10 | 5 | Average |
|---|---|---|---|---|
| PCA | 69.86 | 72.53 | 67.71 | 70.03 |
| LDA | 70.56 | 71.07 | 71.76 | 71.13 |
| GNFL | 71.11 | 68.00 | 64.44 | 67.85 |
| UDNFLA | 73.19 | 74.27 | 70.27 | 72.73 |
| NFSA | 72.36 | 70.27 | 71.63 | 71.42 |
| DNFSA | 74.56 | 73.47 | 69.41 | 72.81 |
| NFSDA | 80.97 | 79.6 | 74.64 | 78.40 |

**Table 2** Effect of parameters $K_1$ and $K_2$ on the NFSDA

| Ntr | 20 | 10 | 5 |
|---|---|---|---|
| Parameter range | 3–10 | 3–10 | 3–5 |
| Standard deviation (%) | 0.54 | 0.32 | 0.47 |

## 4 Conclusion

A new subspace learning method called NFSDA for radar target recognition has been proposed. The within- and between-class scatter matrices are redefined according to the P2S metric, rather than the traditional P2P metric. The experimental results show that the NFSDA achieves better recognition performance than the other methods because NFSDA considers both the between-class dispersity and the within-class local geometry preservation, which provide it with more discriminative power. In addition, the FSDA is insensitive to its two parameters in our experiments.

**Acknowledgements** This work was supported by the Civil Aviation Science and Technology Innovation and Guidance Foundation (Application and Demonstration of Foreign Object Debris Detection on Runways with Multi-Modes Fusion System).

## References

1. Y. Pang, Y. Yuan, X. Li, Generalized nearest feature line for subspace learning. Electron. Lett. **43**(20), 1079–1080 (2007). doi:10.1049/el:20072176
2. J.W. Lu, Y.P. Tan, Uncorrelated discriminant nearest feature line analysis for face recognition. IEEE Signal Process. Lett. **17**(2), 185–188 (2010). doi:10.1109/LSP.2009.2035017
3. Y.L. Chang, J.N. Liu, Y.N. Chen, Hyperspectral image classification using nearest feature line embedding approach. IEEE Trans. Geosci. Remote Sens. **52**(1), 278–287 (2014). doi:10.1109/TGRS.2013.2238635
4. J.W. Lu, Y.P. Tan, Nearest feature space analysis for classification. IEEE Signal Process. Lett. **18**(1), 55–58 (2011). doi:10.1109/LSP.2010.2093600
5. J.T. Chien, C.C. Wu, Discriminant wavelet faces and nearest feature classifiers for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **24**(12), 1644–1649 (2002). doi:10.1109/TPAMI.2002.1114855
6. X.L. Yu, X.G. Wang, Kernel uncorrelated neighbourhood discriminant embedding for radar target recognition. Electron. Lett. **44**(2), 154–155 (2008). doi:10.1049/el:20082251
7. Y.J. Zheng, J.Y. Yang, J. Yang, X.J. Wu, Z. Jin, Nearest neighbour line nonparametric discriminant analysis for feature extraction. Electron. Lett. **42**(12), 679–680 (2006). doi:10.1049/el:20060609

# SAR Imaging Using a PulsON 410 UWB Radar: Simulation Versus Measurement

**Huaiyuan Liu, Chengchen Mao, Xiaofeng Yu and Jing Liang**

**Abstract** In this paper, synthetic aperture radar (SAR) images are investigated theoretically and experimentally for a high two-dimensional (2D) resolution (cm level). In situ measured data of two soda cans at different spatial location are collected using a PulsON 410 ultra-wideband (UWB) radar with a pair of improved directional planar antennas. The images are obtained using two SAR imaging algorithms: Time Domain Correlation (TDC) and range Doppler (RD) respectively, and the results of RD are preferable to images via TDC. Compared to the simulation, the measurement results can provide more accurate targets location information.

**Keywords** SAR image · 2D resolution · UWB radar

## 1 Introduction

By utilizing the movement of a platform to form a larger antenna beam width, synthetic aperture radar (SAR) can extend the coverage of radar detection and dramatically improve image resolution in comparison of real aperture radar [1]. Furthermore, without the limitation of weather condition, such as light and moisture, SAR has played a significant role for targets recognition and imaging [2]. Considering the advantages of UWB technology, such as sufficiently high spatial resolution and strong penetrating ability of materials, SAR can be implemented based on UWB signals to achieve higher range resolution images.

SAR imaging with UWB systems can acquire more detailed scattering features of targets to reveal comprehensive information about objects. Some researchers have investigated SAR imaging using UWB radar [3–5]. Charvat et al. [3] studied the through-lossy-slab radar targets imaging in field measurement environment using a low-power, rail UWB SAR imaging system. Lou et al. [4] obtained the scattering information in the four-dimensional domain of landmine echo in order to extract

H. Liu (✉) · C. Mao · X. Yu · J. Liang
University of Electronic Science and Technology of China, Electronic Engineering, Chengdu 611731, China
e-mail: louis.uestc@outlook.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_23

feature vector for improving radar target detection performance based on an airship-borne UWB SAR system. Le et al. [5] created through-wall SAR images of building interior via the back-projection imaging algorithm using a vehicle-borne UWB SAR system. These previous work focuses on exploiting the performance of UWB SAR imaging system to reveal the feasibility of through-wall targets imaging and recognition. However, they fail to provide the spatial resolution of SAR images and examine practical 2D resolution of images compared to theoretical resolution. To solve this problem, the novelty of this paper is two folds as follows:

1. Both simulated and measured SAR images of two targets at short range (less than 1 m) are first compared to the best of our knowledge. Exploiting the improved directional radiation pattern of broadband planar antennas other than the omni-directional antennas of PulsON, the inherent antenna coupling problem can been alleviated and high spatial resolution (cm level) can be resolved.
2. SAR images of distinct range and cross-range direction are revealed by using two imaging algorithms: TDC and RD. We show that UWB Gaussian signal outperforms UWB LFM in SAR imaging, and the practical resolution is nearly the same as the theoretical resolution.

The rest of this paper is organized as follows. Section 2 formulates the model of SAR and gives an overview of two SAR imaging algorithms. Section 3 introduces measurement setup and data collection by a portable PulsON 410 UWB radar. Section 4 describes the properties of our UWB SAR imaging system, and compares the results of measured data with the theoretical numerical simulations. Finally, conclusions are addressed in Sect. 5.

## 2 SAR Model and Imaging Algorithms

### *2.1 SAR Model*

There are three different operating modes of a SAR system: stripmap, spotlight and scan mode [6]. In this paper, we use broad side-looking stripmap operating mode as a UWB SAR imaging system. Consider a stationary target region composed of a set of point target reflectors with coefficient $\alpha_n$ located at the coordinates $(x_i, y_j)$ $(n = 1, 2, \ldots)$ in the spatial $(x, y)$ domain. Assuming PulsON 410 UWB radar is located at $(0, u)$, a simplified SAR signal $s(t, u)$ is as following [7],

$$s(t,u) = \int_u \int_t \alpha_n p\left[t - \frac{2\sqrt{x_i^2 + (y_j - u)^2}}{c}\right] dtdu, \tag{1}$$

where $p(t)$ is the radar transmitting signal, the variable $t$ is used for fast-time domain, and the variable $u$ identifies the slow-time domain.

**Table 1** PulsON 410 specifications

| Parameter | Value |
|---|---|
| Center frequency | 4.3 GHz |
| Operating band | 3.1–5.3 GHz |
| Pulse type | Gaussian monocycle |
| Pulse duration | 0.5 ns |

Suppose that time duration of transmitting an electromagnetic pulse $p(t)$ is $\tau$, and the range resolution $\delta_r$ is

$$\delta_r = (c\tau)/2. \tag{2}$$

From Eq. (2), the geometric resolution of a stripmap SAR image is limited in the range by pulse duration $\tau$.

The length of synthetic aperture $L_s$ by moving antennas defines the cross-range resolution in a stripmap SAR image, i.e.

$$\delta_{cr} = \beta_{0.5} \cdot R = \lambda/(2L_s). \tag{3}$$

From Table 1, the theoretical range and cross-range resolution of SAR image based on PulsON 410 is 6.82 and 7 cm according to the Eqs. (2) and (3). Where $L_s = 0.5$ m, $\lambda = c/f_c = 7$ cm ($\lambda$ is calculated from the center frequency).

## *2.2 SAR Imaging Algorithms*

### 2.2.1 Time Domain Correlation Imaging Algorithm

The basic principle behind the Time Domain Correlation (TDC) imaging method is simply by correlation to implement the matched filtering processing [8]. Suppose that interested targets are located at a set of spatial sampled points $(x_i, y_j)$. The first step of TDC processing procedure is to correlate the SAR signature at a given grid point $(x_i, y_j)$, and the second step of TDC is to implement signal correlation divided into range and cross-domain. The imaging equation is illustrated as follows:

$$f_{\text{tdc}}(x_i, y_j) = \int_u \int_t s(t, u) p^* \left[t - t_{ij}(u)\right] dtdu, \tag{4}$$

where $p^*(t)$ is the complex conjugate of $p(t)$, and the time delay caused by distance difference between PulsON 410 and target area is

$$t_{ij}(u) = \frac{2\sqrt{x_i^2 + (y_j - u)^2}}{c}. \tag{5}$$

### 2.2.2 Range Doppler Imaging Algorithm

For the reason of algorithm processing time efficiency and imaging precision, range Doppler (RD) imaging algorithm is evolved from the TDC. The rationale of the RD procedure is to perform matched filtering separately in the fast Fourier transforms (FFTs) range and cross-range domains [6]. Suppose that the vertical distance from a given point P at target location to radar moving path is $R_0$, we can obtain the slant range between point P and the PulsON 410 UWB radar is

$$R_p = \sqrt{R_0^2 + (vt_p)^2}, \tag{6}$$

where $v$ is radar platform velocity and $t_p$ is the time at P point.

In RD, the first step is matched filtering for each range SAR signal $s(t, u)$ by range FFTs, i.e.

$$s_{\mathrm{rc}}(t, u; p) = \mathrm{IFFT}\{\mathrm{FFT}[s(t, u; p)] \cdot \mathrm{FFT}[s^*_{\mathrm{ref}}(t, u; p)]\}, \tag{7}$$

where $s(t, u; p) = s(t - 2R_p/c, u)$, $s_{\mathrm{ref}}(t, u; p) = p(t - 2R_p/c)$. $s_{\mathrm{ref}}(t, u; p)$ is range matched filter reference function the same as the transmitting pulse.

In order to obtain the cross-range compression, the range compressed signal $s_{\mathrm{rc}}(t, u; p)$ needs to be processed via cross-range FFT and range cell migration correction (RCMC). The range cell migration for stripmap broad side-looking SAR is illustrated in the following equation:

$$\Delta R = R(t) - R_0, \tag{8}$$

where $R(t)$ indicates the instantaneous slant range of point target P and radar in time $t$. RCMC is usually achieved by interpolation.

After cross-range matched filtering of each cross-range SAR signal and cross-range inverse fast Fourier transforms (IFFTs), the final SAR image of point target P is

$$f_{\mathrm{rd}}(t, u; p) = \mathrm{IFFT}\{\mathrm{FFT}[s_{\mathrm{rcmc}}(t, u; p)] \cdot \mathrm{FFT}[s^*_{\mathrm{refa}}(t, u; p)]\}, \tag{9}$$

where $s_{\mathrm{rcmc}}(t, u; p)$ is the aligned range compressed signal, and $s_{\mathrm{refa}}(t, u; p)$ is corss-range matched filter reference function, respectively. For a Linear frequency modulation (LFM) waveform signal,

$$s_{\mathrm{refa}}(t, u; p) = \exp(j\pi K_a^2 t_p^2), \tag{10}$$

where, $K_a = (-2v^2)/(\lambda R_0)$.

# 3 Measurement Setup

## *3.1 Measurement Environment*

In situ SAR measurements were performed at a laboratory environment area of approximate 4 $m^2$, which is located in University of Electronic Science and Technology of China, Chengdu, China. The SAR data acquisition took place on July 2016. In our experiment, we collected SAR signals of targets along range and cross-range direction (shown in Fig. 1).

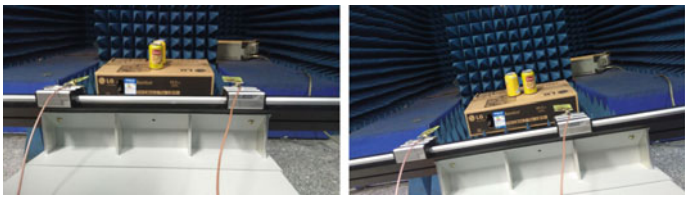
## *3.2 Instruments*

### 3.2.1 UWB Radar

In the measurement, we used a UWB short pulse monostatic radar module PulsON 410, manufactured by Time Domain Corporation, which is chosen as the signal transceiver. Each sample is in step of 61 ps (actually 32 steps of 1.907 ps equals 61.024 ps) and 480 data points are collected for single scan echo in the range profile. The radar emits UWB short pulses at a pulse repetition frequency (PRF) of 10.1 MHz (As a side note, 1–20 MHz are supported). The waveform of UWB signal is centered at 4.3 GHz, occupies more than 2 GHz of bandwidth and achieves an effective RF bandwidth of 1.4 GHz [9]. The main parameters of the PulsON 410 are provided in Table 1.

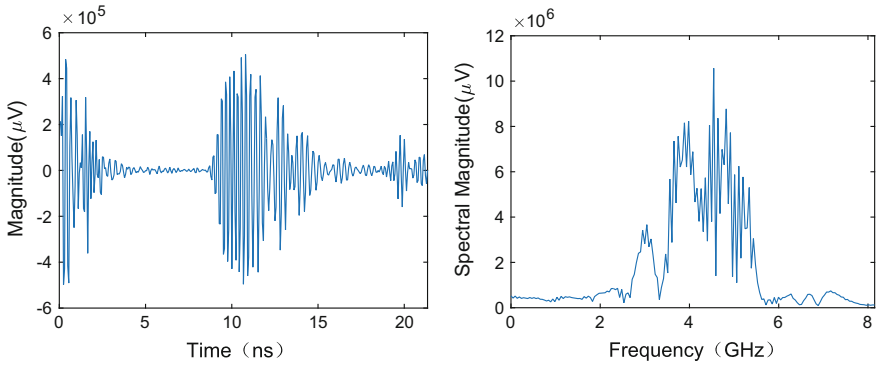
### 3.2.2 UWB Antennas

In order to ameliorate the quality of SAR image, we used two improved third party broadband planar printed quasi-Yagi antennas to replace the original toroidal dipole antennas as transmitting and receiving antennas of PulsON 410. In addition, the initial noise of echo, which is coupling from antenna directly to PulsON 410, is also



(a) range direction (b) cross-range direction

**Fig. 1** The measurement environment for 2 soda cans




(a) time domain (b) frequency domain

**Fig. 2** The received UWB signal

reduced within 2 ns (shown in Fig. 2a) compared to the 5 ns coupling of original antennas. The operating RF band of this directional antenna is from 2.9 to 10.1 GHz with a ratio of about 3.48: 1, E-plane 3 dB beamwidth is 80° and front-to-back ratio is greater than 10 dB [10].

## 4 Measurement and Simulation Results

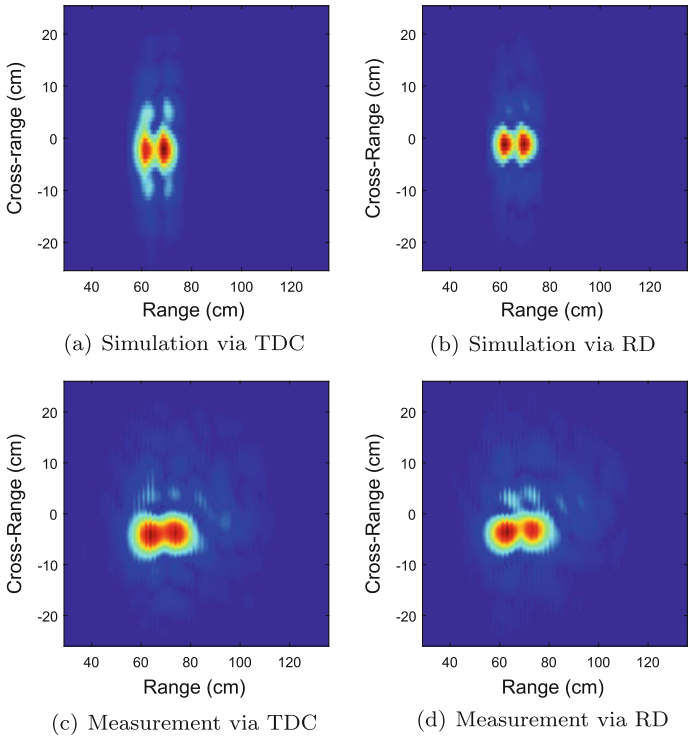
The simulated and measured magnitude of UWB data versus time are post-processed into images via TDC and RD. Additionally, the measured data are calibrated by acquiring background range profile to remove environmental clutter.

### *4.1 Signal of PulsON 410 UWB Radar*

The result of a typical measured UWB signal is shown in Fig. 2, where PulsON 410 is at a distance of 65 cm from a soda can. Figure 2a shows a raw scan echo in range time bin, first ∼2 ns of the echo is caused by energy coupling and SMA coaxial cable. In the procedure of SAR imaging, we used FFTs to get the complex frequency response of the UWB signal of PulsON 410, and Fig. 2b illustrates that the spectrum ranges of the measured UWB signal is 2.2 GHz centered at approximately 4.3 GHz.

### *4.2 Simulation Results of UWB SAR*

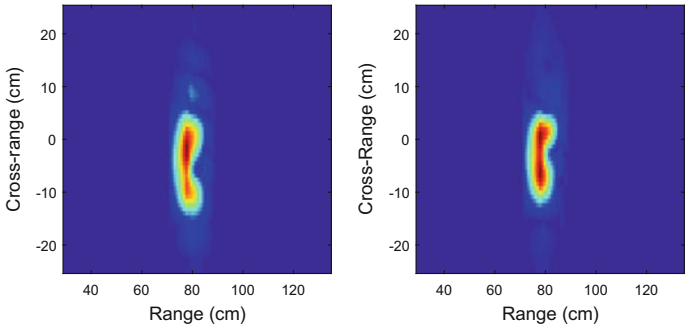
In view of the fact that the limitation between range of radar detection and range resolution, LFM signal was adopted in simulation work, which is different from the pulse type of PulsON 410 (Gaussian short pulse) in measurement. The specific simulation parameters of UWB SAR images are the same as measurement parameters of PulsON 410. In simulation, we used two point targets to examine the 2D resolution of UWB SAR images. The range domain simulated result of two point targets is shown in Fig. 3a, b, and the location of them are (−2, 64 cm) and (−2, 71 cm). The range distance between these two targets is 7 cm. The cross-range domain simulated results are shown in Fig. 4a, b. The location of two point targets are (−7, 79 cm) and (0, 79 cm), that is, the cross-range distance between these two targets is 7 cm.



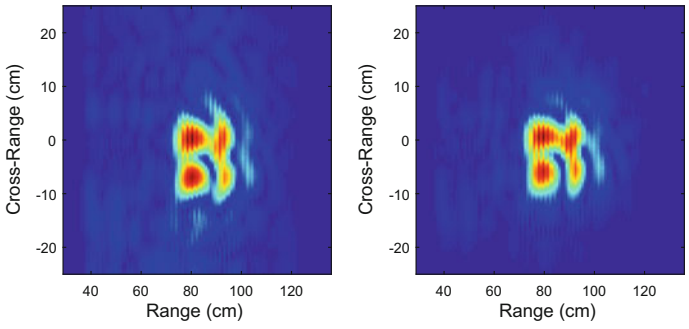

(a) Simulation via TDC (b) Simulation via RD

(c) Measurement via TDC (d) Measurement via RD

**Fig. 3** UWB SAR images for two targets along range direction




(a) Simulation via TDC (b) Simulation via RD




(c) Measurement via TDC (d) Measurement via RD

**Fig. 4** UWB SAR images for two targets along cross-range direction

### *4.3 Measurement Results of UWB SAR*

In the measurement, we used two soda cans with radius r = 6 cm, height h = 11.5 cm as the real targets. We also set the length of synthetic aperture by moving sliding rail to 0.5 m. The two soda cans are located 64 and 71 cm away from the PulsON 410, respectively, which is the same in simulation. The processed images are shown in Fig. 3c, d. Measured range resolution is about 7 cm compared to the theoretical resolution 6.82 cm. Similar to the corss-range location in simulation, two measured soda cans are located at 0 and −7 cm in cross-range direction with the 79 cm range distance to radar (shown in Fig. 4c, d). In both figures, we can observe that cross-range resolution (about 7 cm) of practical UWB SAR images are obtained.

### *4.4 Discussion*

Figures 3 and 4 present the simulation and measurement results of UWB SAR images via two imaging algorithms mentioned in Sect. 2: TDC and RD. First, note that two

soda cans in range domain are successfully reconstructed in Fig. 3c, d. Similar to the simulation of Fig. 3a, b, these two targets can be clearly resolved into two separate images in range direction despite the side-lobe phenomenon. In the aspect of image quality, the results of RD are relatively better than that of TDC owing to the process of RCMC in both simulation and measure situations.

In Fig. 4a, b, the two targets located at a distance 7 cm of each other in cross-range direction cannot be separated clearly. Whereas the measured UWB SAR images (shown in Fig. 4c, d) are able to distinguish these two targets in the same position since the Gaussian type waveform of PulsON 410 is designed based on maximizing resolution [9], which is different from the LFM signal in the simulation. The noticeable defocusing effect of Fig. 4c, d may be caused by insufficient phase compensation in cross-range direction. Additionally, it is not difficult to find that RD is superior to TDC with relatively weak side-lobe in imaging results.

## 5 Conclusion

In this paper, the stripmap SAR model was applied on the PulsON 410 UWB radar measurements of two soda cans at different range and cross-range locations. Then, we presented simulated and measured UWB SAR images based on TDC and RD imaging algorithms, respectively. The practical 2D resolution of images was examined, nearly the same as the theoretical resolution, range: 6.82 cm and cross-range: 7 cm. Compared to the simulation, the measurement results can provide more accurate targets position information because UWB Gaussian signal of PulsON 410 outperforms UWB LFM signal in SAR image. The results of RD are preferable to images via TDC as a result of RCMC processing.

In future work, we may investigate the defocusing of image in cross-range direction and other more accurate SAR imaging algorithms for improving image quality.

**Acknowledgements** Authors would like to thank Prof. Deqiang Yang for providing the testing environment. This work was supported by the National Natural Science Foundation of China (61671138), the Fundamental Research Funds for the Central Universities Project No. ZYGX2015J021, and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

## References

1. J.C. Curlander, R.N. McDonough, *Synthetic Aperture Radar* (Wiley, NY, USA, 1991)
2. Z. Cao, Y. Ge, J. Feng, J. Adv. Signal Process. **2014**(1), 1 (2014)
3. G.L. Charvat, L.C. Kempel, E.J. Rothwell, C.M. Coleman, E.L. Mokole, IEEE Trans. Antennas Propag. **58**(8), 2594 (2010). doi:10.1109/TAP.2010.2050424
4. J. Lou, T. Jin, Z. Zhou, Prog. Electromagn. Res. **138**(2), 157 (2013)
5. C. Le, T. Dogaru, L. Nguyen, M.A. Ressler, IEEE Trans. Geosci. Remote Sens. **47**(5), 1409 (2009). doi:10.1109/TGRS.2009.2016653

6. G. Franceschetti, R. Lanari, *Synthetic Aperture Radar Processing* (CRC press, 1999)
7. M. Soumekh, *Synthetic Aperture Radar Signal Processing* (Wiley, New York, 1999)
8. B. Barber, Int. J. Remote Sens. **6**(7), 1009 (1985)
9. A. Petroff, (2012), pp. 0880–0884
10. J. Qu, D. Yang, S. Liu, H. Zeng, in *2015 10th International Conference on Communications and Networking in China (ChinaCom)* (IEEE, 2015), pp. 17–21

# Soil Moisture Retrieval via Non-singleton Fuzzy Logic with UWB Echoes

**Xiaoxu Liu, Xiaofeng Yu and Jing Liang**

**Abstract** Two model-free algorithms for soil moisture retrieval based on singleton type-1 fuzzy logic system (T1FLS) and non-singleton T1FLS are investigated in this paper. We reformulate the traditional soil volumetric water content (VWC) retrieval problem into a signal recognition issue of ultra wide band (UWB) echoes. We forecast UWB echoes of different VWCs employing both singleton T1FLS and non-singleton T1FLS. Their convergences are measured via root mean square error (RMSE). After that, the forecasting signals in T1FLS are retained as the templates. The testing UWB echoes with an unknown VWC are compared with 5 templates of different VWCs, and classified into one of the VWCs with the minimal RMSE. Monte Carlo simulations not only show that the correct VWC recognition rates are robust at different SNRs, but also indicate that the non-singleton T1FLS is superior to singleton T1FLS in performances of both convergence and VWC recognition rates.

**Keywords** Soil moisture retrieval · Fuzzy logic · UWB

## 1 Introduction

Nowadays, soil moisture retrieval using radars is rapidly developing based on the relations between soil's dielectric properties and radar echoes. However, radar echoes from bare soil are relevant to mounts of interrelated factors such as frequency, polarization, soil moisture, surface roughness and the incidence angle of the incoming microwave [1–3]. Recent studies focus on various of models that reveal the

X. Liu (✉) · X. Yu · J. Liang (✉)
Qingshuihe Campus, University of Electronic Science
and Technology of China No. 2006, Xiyuan Ave, West Hi-Tech Zone, Chengdu,
Sichuan, People's Republic of China
e-mail: llxxljshao@foxmail.com

J. Liang
e-mail: liangjing@uestc.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_24

numerical relationships between soil parameters (roughness, moisture, etc.) and microwave dielectric behaviors. Parts of achievements on soil moisture retrieval are listed. Hallikainen et al. fitted the polynomial expressions [4] and Dobson et al. developed the semi-empirical four-component mixing model [5]. Meanwhile, Ulaby and Batlivala [6] found the optimal incidence angle, 7°–15°, with an ideal radar configuration at C-band. However, new techniques are desired to conquer the restrictions of uncertainty existed in different models.

With the superior abilities on information collection, UWB radars are applied to improve the accuracy of soil moisture retrieval. However, the features of UWB make obstacles on the retrieval. For example, recent studies on UWB ground penetration radar (GPR), show satisfactory performance only in the low frequency. It restricts UWB's bandwidth for information collection, let alone the ignored cost-efficiency. Due to the manifold interactions between electromagnetic fields and various objects, the UWB echoes applied in soil moisture retrieval is an open question by far.

Considering research mentioned above, this paper aims to establish a model-free technique to retrieve soil moisture with UWB. We apply UWB short pulse monostatic radar module PulsON 410 with a wide bandwidth, 2 GHz. Two time-series forecasting methods based on both singleton and non-singleton T1FLS are proposed. During the forecasting, we continuously train the T1FLS's factors aiming to get desired output from the system. We apply root mean square error to measure the deviation of forecasting and compare the forecasting performances of two types of T1FLSs. Then the testing signals are compared to the templates, the forecasting signals, with different VWCs. Testing signals are classified in the templates with the minimal RMSE. The novelty of this paper is threefold:

1. Different from traditional methods considering the interrelations of UWB echoes and soil moisture, a model-free time-series forecasting technique is applied which emphasizes parameters training and testing with the help of templates.
2. The application of T1FLS, to the best of our knowledge, is barely used in processing the soil moisture retrieval. Moreover, T1FLS with a remarkable performance on time-series forecasting is the lightspot because of its few parameters applied in training.
3. We apply two methods, singleton and non-singleton T1FLS, and make comparisons of the performance on forecasting and recognitions in different soil moistures.

The rest of the paper is organized as follows: Sect. 2 illustrates the appliances and method for data collection. Section 3 gives an overview of two T1FLSs for forecasting. Section 4 compares the performance of two methods with the standard of RMSE. Monte Carlo is employed to discuss the recognition rates in different SNRs and soil moistures. Finally a conclusion is made in Sect. 5.

## 2 Measurements Setup

The soil retrieval measurement experiments were continued between March 2015 and April 2015. We choose bare soil, a typical soil texture in agriculture, to verify the feasibility of our retrieval method. The measurement environment for UWB echoes collection is a $50 \times 50$ m bare-soil cropland. Meanwhile, the measured areas with plat surface and open surroundings are helpful to reduce the influence of unrelated factors for soil moisture retrieval.

The UWB radar manufactured by Time Domain Corporation, PulsON 410 is applied in experiments to transmit and receive signals. The main parameters are listed in Table 1. To decease the influence of coupling noise and the strong interfere from the adjacent transmitting signals, the PulsON 410 should keep 0.7 m away from the target [7]. Hence, the radar is hung up with a height of 0.8 m. Time domain reflector FieldScout TDR 300 is applied to calibrate the accuracy of soil moisture retrieval applying UWB echoes of PulsON 410. The probe of TDR 300 is inserted in the soil vertically with a depth of 3 cm. Figure 1 introduces the measurement view using UWB radar with calibration of TDR 300. For bare soil, 9 measurements are taken in a 3-by-3 array with an interval of 15 cm, then an average of 9 measurements is taken as the VWC value. In our experiment, we continually irrigate the measured fields to get the different VWCs as follows: the bare soil VWC 13.7, 21.7, 28.3, 35.0 and 40.9%.

**Table 1** Main parameters of PulsON 410

| Parameters | Values |
|---|---|
| Center frequency | 4.3 GHz |
| Bandwidth UWB | 3.1, 5.3 GHz |
| Power requirements | 5.75–30 V @ 40 W max |
| Spectrum | FCC 15B (−14.5 dBm) |
| Time index | 61 ps |

**Fig. 1** The measurement environment

# 3 Signal Processing and a Review of Singleton and Non-singleton T1FLS

## 3.1 The processing of UWB echoes

The transmitted signal of PulsON 410 is a short pulse with a duration of several nanoseconds shown in Fig. 2a. As the serious interfere of coupling noise, segments of UWB echoes which submerged in noise must be cut. Meanwhile, considering the investigation depth for UWB, we finally set the time index 91 and 490 as the beginning and the end point of the processed UWB signal shown in Fig. 2b.

However, it is still harsh to forecast the UWB echoes precisely due to the lack of the training points (only 400 points for a processed signal). Therefore, the UWB echoes which measure the same VWC are concatenated to construct an aperiodic signal group with sufficient points for training.

Before the forecasting, a severe problem must be considered: Liang [8] validated the feature of UWB with no self-similarity. Therefore, we propose the T1FLS for forecasting with excellent performance of Mackey–Glass chaotic time-series. Before the introduction of forecasting, the fuzzy logic should be listed as a basis.

## 3.2 Singleton and Non-singleton T1FLS

Generally, TIFLS is a nonlinear system with multiple inputs and a crisp output. Both singleton TIFLS and non-singleton TIFLS follow the same structure composed by four sections: fuzzifier, rules, fuzzy inference and defuzzifer.

The fuzzifier is contained by membership functions (MFs), where the Gaussian MF is applied in our system.

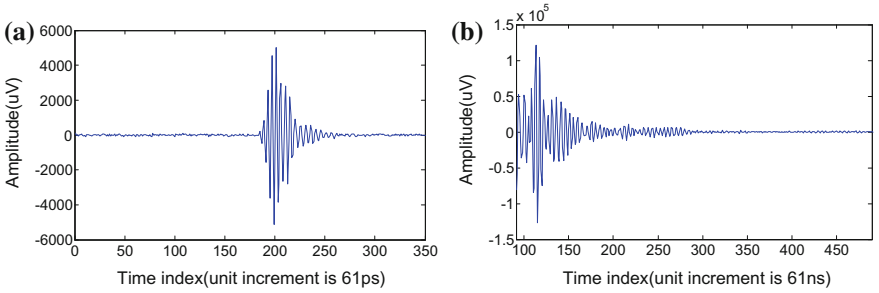



**Fig. 2** Example of UWB echoes. **a** Transmitted signal. **b** The UWB echoes without coupling

$$\mu_{F_k^l}(x_i) = \exp\left(-\frac{\left(x_k^{(i)} - m_{F_k^l}\right)^2}{2\sigma_{F_k^l}^2}\right) \tag{1}$$

(1) shows the antecedent MF function where $l$ denotes the $l$th rule and $x_i$ denotes the ith input.

The fuzzy rule is described as (2) which defines the lth rule with totally M rules.

$$R^l : \text{If } x_1 \text{ is } F_1^l \text{ and } x_2 \text{ is } F_2^l \text{ and } \ldots\ x_p \text{ is } F_p^l \text{ Then y is } G^l \text{ l=1}, \ldots, \text{M} \tag{2}$$

The rules and fuzzifier are connected by the fuzzy inference show in (3) adopted Mamdani implications,

$$\begin{aligned}\mu_{B^l}(y) = &\mu_{G^l} \bigstar \left\{ \left[ \sup_{x_1 \in X_1} \mu_{X_1}(x_1) \bigstar \mu_{F_1^l}(x_1) \right] \bigstar \cdots \right. \\ &\left. \bigstar \left[ \sup_{x_p \in X_p} \mu_{X_p}(x_p) \bigstar \mu_{F_p^l}(x_p) \right] \right\}, \quad y \in Y\end{aligned} \tag{3}$$

where ★ denotes fuzzy operation, t-norm (minimum or product). $G^l$ is the consequent MF described with Gaussian MF as well. Finally, the defuzzification which applies the height defuzzier transforms the complex inference into a crisp output, y.

$$y_h(x) = \frac{\sum_{l=1}^{M} \bar{y}^l \mu_{B^l}(\bar{y}^l)}{\sum_{l=1}^{M} \mu_{B^l}(\bar{y}^l)} \tag{4}$$

With the acknowledgment of similar construction, we introduce the characters of the singleton and non-singleton T1FLS.

### 3.2.1 Singleton T1FLS

Singleton T1FLS is a normal FLS with no uncertainty in antecedent MFs. Describing with the operations in fuzzy logic, it means that the value of antecedent MF is 1 at $x_1$ as well as 0 in other points when the input is $x_1$. Therefore, in (3), $\sup_{x_1 \in X_1} \mu_{X_1}(x_1) \bigstar \mu_{F_1^l}(x_1)$ can be simplified into $\mu_{X_1}(x_1)$ based on fuzzy logic operations. It illuminates that there's no description on uncertainty of input in singleton T1FLS.

Generally, the singleton T1FLS is hard to match the strong fluctuations influenced from the input which means the system is sensitive to noises.

### 3.2.2 Non-singleton Fuzzy Logic

Considering the UWB echoes with noises, the input signals become vague comparing with desired data. In order to improve the accuracy of forecasting, the uncertainty of inputs should be described with fuzzy logic operations. In non-singleton T1FLS, it treats the input as a type-2 fuzzy set which means there's a Gaussian curve, $\mu_{X_k}(x_k)$, to represent the fluctuation around the desired value considering a decided input, $x_k$. Therefore, $\mu_{Q_k^l}$ in (5) should be considered.

$$\mu_{Q_k^l} = \mu_{X_k}(x_k) \bigstar \mu_{F_k^l}(x_k) \tag{5}$$

As the t-norm of p interval type-1 sets are the maximum value, $f^l(x)$ comes from the maximum value of $\mu_{Q_k^l}$. In Gaussian MF, we deduce (6)

$$x_{k,max}^l = (\sigma_{x_k}^2 m_{F_k}^l + \sigma_{F_k}^2 m_{x_k}^l)/(\sigma_{x_k}^2 + \sigma_{F_k}^2) \tag{6}$$

where $m_{x_k}^l$ and $\sigma_{x_k}^2$ are the factors of Gaussian curve aiming to describe the uncertainty of MF. Then we reformulate non-singleton T1FLS as singleton T1FLS with the input, $x_{k,max}^l$.

## 3.3 Time-Series Forecast Using T1FLS

Imitating the technique of forecasting based on T1FLS in Mackey-Glass chaotic time-series, the processed 8000 signal data are divided into training and forecasting data points half-and-half [9]. Each 4 data points makes a group and forecast the next data point in the forecasting.

$$x^{(i)} = [4 \times 1\ input, output]^T = [x_1^{(i)}, x_2^{(i)}, x_3^{(i)}, x_4^{(i)}, x_5^{(i)}]^T\ i = 1, \ldots, 3996 \tag{7}$$

A backpropagation method is proposed to tune the parameters in T1FLS with tuning equations deduced as follows.

$$m_{F_k^l}(i+1) = m_{F_k^l}(i) - \alpha_m [f_s(x^{(i)}) - y^{(i)}][\bar{y}^l(i) - f_s(x^{(i)})] \frac{[x_k^{(i)} - m_{F_k^l}(i)]}{\sigma_{F_k^l}^2(i)} \phi_l(x^{(i)}) \tag{8}$$

$$\sigma_{F_k^l}(i+1) = \sigma_{F_k^l}(i) - \alpha_\sigma [f_s(x^{(i)}) - y^{(i)}][\bar{y}^l(i) - f_s(x^{(i)})] \frac{[x_k^{(i)} - m_{F_k^l}(i)]}{\sigma_{F_k^l}^3(i)} \phi_l(x^{(i)}) \tag{9}$$

$$\bar{y}^l(i+1) = \bar{y}^l(i) - \alpha_{\bar{y}} [f_s(x^{(i)}) - y^{(i)}] \phi_l(x^{(i)}) \tag{10}$$

In non-singleton T1FLS, an additional parameter, $\sigma_{x_k}$ needs to be tuned: (11).

$$\begin{aligned}\sigma_x(i+1) =&\sigma_x(i) - \alpha_{\sigma_x}[f_s(x^{(i)}) - y^{(i)}][\bar{y}^l(i) - f_s(x^{(i)})] \\ &\times \sigma_x(i)\left[\frac{x_k(i) - m_{F_k^l}(i)}{\sigma^2_{x_k}(i) + \sigma^2_{F_k}(i)}\right]\phi_l(x^{(i)})\end{aligned} \tag{11}$$

(8)–(11) include all the parameters in T1FLS which represent the extracted features in UWB echoes. $m_{F_k^l}$, $\sigma_{F_k^l}$, $\bar{y}^l$ and $\sigma_x$ respectively compose the $16 \times 4$ matrix, $M$ and $\sigma$, the $16 \times 1$ vector $\bar{y}$ and the $4 \times 1$ vector $\sigma_X$.

RMSE represent the deviation between the forecasting signals and UWB echoes is regarded as the accuracy of parameters training.

$$RMSE = \sqrt{\frac{1}{N+1}\sum_{k=i}^{i+N}[s(k+1) - f_s(s^{(k)})]^2} \tag{12}$$

where $s^{(k)} = [s(k-3), s(k-2), s(k-1), s(k)]$ is the T1FLS's input and $f_s(s^{(k)})$ is the output. $i$ denotes the beginning of the test and $N+1$ is the total quantity of testing points.

## 4 Simulation of T1FLS and Recognition Rates

### *4.1 Performance of the Forecasting*

As the restriction of reliable VWCs in these experiments, only 5 types of VWCs' echoes (13.7, 21.7, 28.3, 35.0, 40.9%) are applied in forecasting. We average 40 signal groups, each group is concatenated with 20 UWB echoes in the same VWC. A normalization is extremely needed for tuning $\alpha$ that denotes all the $\alpha$ in (8)–(11) with the same value. In T1FLS, each input corresponds to two MFs with the mean, $M_1$ and $M_2$, and the same variance, $\sigma$. They are respectively initialized in $m_x - 2\sigma_x$, $m_x + 2\sigma_x$, $2\sigma_x$ where $m_x$ denotes mean of the data in the signal group and $\sigma_x$ denotes standard deviation of the data. Meanwhile, $\bar{y}$ is initialized randomly between 0 and 1.

With adaptive choices of $\alpha$ and $\sigma_x$, we get amazing forecasting consequents in Fig. 3 which presents partial UWB echo with 40.9% VWC and its forecasting signal (accurately, the signal duration is 24.4ns). the performance is so remarkable that the forecasting signal is basically matched with the UWB echo.

The comparison of RMSE with different VWCs (13.7, 28.3, 40.9% VWC) between singleton and non-singleton T1FLS are shown in Fig. 4. The non-singleton T1FLS with rapid convergence rates and tiny RMSE has a better performance than that of singleton T1FLS.

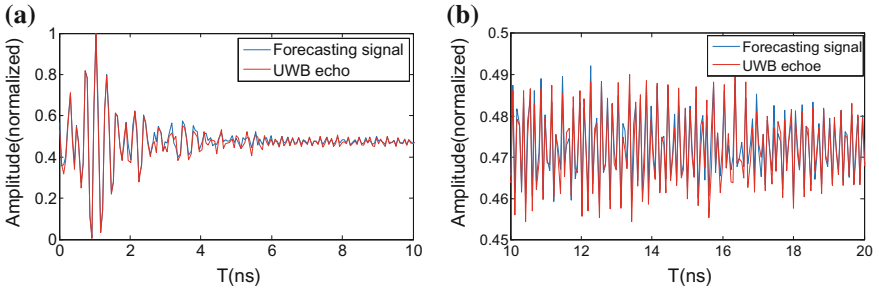



**Fig. 3** Forecasting of UWB echoes with 40.9% VWC. **a** Consequent of the signal forecasting in 0–10 ns. **b** Consequent of the signal forecasting in 10–20 ns

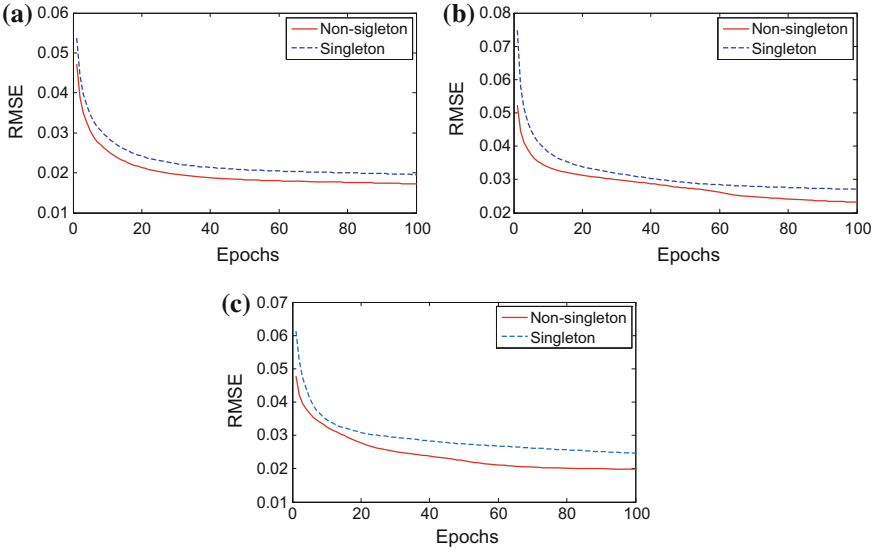



**Fig. 4** Comparison of RMSE in singleton T1FLS and non-singleton T1FLS with different VWCs. **a** RMSE with 13.7% VWC. **b** RMSE with 28.3% VWC. **c** RMSE with 40.9% VWC

## *4.2 Recognition Method and Performance*

The forecasting signals with 5 different VWCs are regarded as templates. The testing signal with unknown VWC is compared with 5 templates respectively and classified into the template associated with the minimal RMSE.

In order to observe the robust of the two methods, the testing signals are added Gaussian white noises and then compare to the 5 templates with the calibration of RMSE. Using a Monte Carlo method, we plot relationship between SNRs and recognition rates with VWC 13.7, 28.3 and 40.9% for simplicity and conciseness. Figure 5

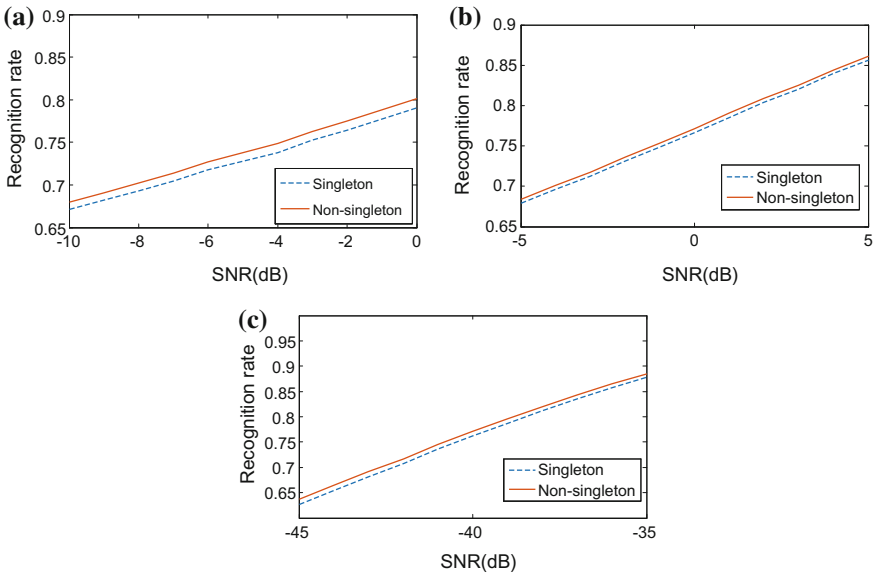



**Fig. 5** Comparison of recognition rate in singleton T1FLS and non-singleton T1FLS with different VWCs. **a** Recognition rate with 21.7% VWC. **b** Recognition rate with 28.3% VWC. **c** Recognition rate with 40.9% VWC

shows that the correct VWC recognition rates via non-singleton T1FLS are higher than those of singleton. Among all these 5 VWCs, each non-singleton T1FLS has a higher recognition rate than that of singleton T1FLS.

## 5 Conclusion

In this paper, we introduce the retrieval of soil volumetric water content (VWC) with UWB short pulse monostatic radar module PulsON 410. Two model-free time-series forecasting methods based on singleton and non-singleton T1FLS are proposed. We forecast the UWB echoes with 5 VWCs (13.7, 21.7, 28.3, 35.0 and 40.9%) applying both two methods and retain the forecasting signals as templates. In recognitions, testing signals with unknown VWCs are compared with the 5 templates and classified into templates with the minimal RMSE. Simulation results show that:

1. The non-singleton T1FLS has a better performance in forecasting than that of singleton T1FLS;
2. In different SNRs, the non-singleton T1FLS obtains higher recognition rates than those of singleton T1FLS.

In future work, more complete soil VWCs will be exploited to increase the accuracy of VWC retrieval. Then a feature extraction method may be proposed in soil moisture retrieval.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (61671138), the Fundamental Research Funds for the Central Universities Project No.ZYGX 2015J021, and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

## References

1. F.T. Ulaby, P.C. Dubois, J. van Zyl, J. Hydrol. **184**(1), 57 (1996)
2. M.S. Moran, C.D. Peters-Lidard, J.M. Watts, S. McElroy, Canad. J. Remote Sens. **30**(5), 805 (2004)
3. A.K. Fung, K.S. Chen, K. Chen, *Microwave Scattering and Emission Models for Users* (Artech House, 2010)
4. M.T. Hallikainen, F.T. Ulaby, M.C. Dobson, M.A. El-Rayes, L.K. Wu, IEEE Trans. Geosci. Remote Sens. (1), 25 (1985)
5. M.C. Dobson, F.T. Ulaby, M.T. Hallikainen, M.A. El-Rayes, I.E.E.E. Trans, Geosci. Remote Sens. **23**(1), 35 (1985)
6. F.T. Ulaby, P.P. Batlivala, IEEE Trans. Geosci. Electron. **14**(2), 81 (1976)
7. A. Petroff, in *2012 IEEE Radar Conference (RadarCon)*, Atlanta (2012)
8. Q. Liang, IEEE Sensors J. **11**(6), 1448 (2011)
9. L.X. Wang, J.M. Mendel, in *IEEE International Conference on Fuzzy Systems* (1992), pp. 1409–1418

# KNN Classification Algorithm for Multiple Statuses Detection of Through-Wall Human Being

**Wei Wang and Dan Wang**

**Abstract** UWB radar with high-range resolution and strong penetration ability can be used to separate multiple human targets in a complex environment. The through-wall human being detection with UWB radar has been relatively mature in the current study. This paper extracts the characteristic parameters which are related to the human targets from the received signals as the sample data. And used machine learning based on the KNN (K nearest neighbor) classification algorithm to identify and classify the through-wall human being status. Experimental results showed that the KNN classification algorithm effectively distinguished three statuses of through-wall human being and reached the prospective goal.

**Keywords** UWB · K nearest neighbor · Through-wall human being · Multiple statues

## 1 Introduction

Through-wall human being detection on ultrawide band (UWB) radar is of interest research which has emerged as one of the preferred choices because of its desirable range resolution and strong penetration power. Meanwhile, the UWB pulses are inherently short, and the short pulses spread their energy over a broad frequency range and exist with extremely positive wall penetration characteristics. At present, UWB technology has accumulated wealthy research achievements in various fields, including target detection, medical monitoring, ranging positioning, through-wall human being detection and tracking [1]. This paper using PulsON 410 UWB radar

W. Wang (✉) · D. Wang
College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin, China
e-mail: weiwangvip@163.com

W. Wang · D. Wang
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_25

in monostatic module is developed by Time Domain Corporation. The experience of through-wall human being detection is mainly using P410 MRM in the uniform environment. And the experimental wall is brick wall of the indoor environment. In this paper, we use classification algorithm of data mining tool to deal with data. WEKA (Waikato Environment for Knowledge Analysis) is a data mining tool. There are various classification algorithms including the decision trees algorithm, naive Bayes algorithm, logistic algorithm, AdaBoost algorithm, artificial neural network algorithm, genetic algorithm, KNN algorithm and so on [2].

Many researchers have focus on through-wall for human being detection based on all kinds of methods. In [3], this paper mainly introduced that the normalized difference square matrix method and reference moving average method with discrete Fourier transform (DFT) as the detection techniques. The experimental results behind gypsum wall and concrete wall have been separately proved for human target detection. In [4], the experimental wall is gypsum wall, and wooden door using UWB radar. And three analytic methods based on normalized difference square matrix method, reference moving average method with discrete Fourier transform (DFT), and empirical mode decomposition (EMD) from Hilbert Huang transform, are introduced for the detection. In [5], the methods based on fast Fourier transform and S transform to explore and identify the human being characteristic of through-wall human detection are mainly introduced. In particular, they extracted the center frequencies of human life signals and located the position of human targets from experimental data. Some research showed that UWB impulse radar system was efficient in capturing human breath and heart-beat movement [6, 7]. Based on these, E. Zaikov explored an experiment using UWB radar for trapped-people detection [8]. In [9], different parts of the human body have different movements when a person is performing different physical activities with the Hilbert–Huang transform (HHT) of through-wall noise are introduced.

Many methods have been proposed for human being detection. However, there are few researches on multiple statuses of human being detection. In this paper, we propose an algorithm which is based on KNN for through-wall human being detection. We also choose another algorithm which is compared with KNN. In Sect. 1, we introduce the research status of through-wall human being detection with UWB radar and classical algorithm based on Weka. In Sect. 2, we introduce the theory of characteristic parameters and classical algorithm of KNN. In Sect. 3, we introduce the experiment and results. In Sect. 4, we introduced the conclusion and outlook.

## 2 Theory

### *2.1 Characteristic Parameter*

In this paper, we select the characteristic parameters which are extracted from the received signal as the condition attributes of the sample data and the three statuses

of are the decision attributes. We could remove the condition attributes with the demand of any different condition. In order to show the characteristics of the signals as much as possible, we selected seven factors as the characteristic parameters which were extracted from the received signals, including the energy of the received signals, the kurtosis, the skewness, the maximum signals amplitude, the variance of the signals, the covariance of the signals and the mean value of the signals. Energy is the total energy of the received signals, the variance can represent the amplitude of the received signals. And the shape of the waveform mainly depended on the kurtosis and skewness. Among them, the kurtosis is the four-order moment of the signals, which reflects the level of the peak of the curve. The skewness is the three moment of the signals, and the symmetry of the value distribution is described. The seven characteristic parameters are defined as follows:

$$\text{Energy of the received signals:}\quad G_r = \int_{-\infty}^{+\infty} |r(t)|^2 dt \tag{1}$$

$$\text{Kurtosis:}\quad K = \frac{1}{\sigma_{|\mathrm{r}|}^4 T} \int_T \left[ \left| r(t) - \mu_{|r|} \right| \right]^4 dt \tag{2}$$

$$\text{Skewness:}\quad S = \frac{\frac{1}{T}\int_T \left[ \left| r(t) - \mu_{|r|} \right| \right]^3 dt}{\left[ \sqrt{\frac{1}{T}\int_T \left[ \left| r(t) - \mu_{|r|} \right| \right]^2 dt} \right]^3} \tag{3}$$

$$\text{Maximum amplitude of the received signals:}\quad r_{\max} = \max_t |r(t)| \tag{4}$$

$$\text{Variance of received signals:}\quad \sigma_{|\mathrm{r}|}^2 = \frac{1}{T} \int_T \left[ |r(t)| - \mu_{|r|} \right]^2 dt \tag{5}$$

$$\text{while}\ \mu_{|r|} = \frac{1}{T} \int_T [|r(t)|] dt \tag{6}$$

$$\text{Covariance:}\quad \mathrm{cov} = \frac{1}{T} \int_T \left[ (|r_1(t)| - \mu_{1|r|})(|r_2(t)| - \mu_{2|r|}) \right] dt \tag{7}$$

$$\text{Mean:}\quad \mu_{|r|} = \frac{1}{T} \int_T [|r(t)|] dt \tag{8}$$

## 2.2 Classification Algorithms and Evaluation

In this experiment, the proposed classification algorithm is evaluated on four ways of using training set evaluation, test set evaluation, cross-validation evaluation and percentage split. First, this paper sets cross-validation in the original training or the training. The evaluation itself has no meaning, but the classifier evaluation results are absolutely optimistic on the training data of any performance. Therefore, cross-validation has become the main evaluation method which is based on the training. In this paper, the default cross-validation of folds is 10. The classification model is evaluated on testing data. Evaluation is given to the test data using KNN classification algorithm. There are two KNN algorithms in Weka: IB1—that is through a neighbor to determine the category of the test samples and IBK—that is through its K neighbors to determine the type of test samples. K is generally judged by experience. The selection of K nearest neighbor samples is determined by a certain distance formula. We selected the distance formula Manhanttan $D(i,j)$ which is used to determine the similarity of the sample data:

$$D(i,j) = \left|X_i - X_j\right| + \left|Y_i - Y_j\right| \tag{9}$$

where $X = (x_1, x_2, \ldots, x_n), Y = (y_1, y_2, \ldots, y_n)$ X, Y represents two sample data respectively, and $n$ is the number of attribute of the sample. The attributes of this paper is 7, $n = 7$. In this paper, we used the ROC (receiver operating characteristic) curve to estimate the performance of the classification algorithm. It describes the relative change between the FPR (False Positive Rate) − TPR (True Positive Rate). In the ROC curve, FPR is used as the X axis, and TPR is used as the Y axis. Compared with the traditional accuracy, ROC analysis is more comprehensive to describe the classification performance of the classification algorithm. And it has the advantages of high credibility, accurate description of the objective, especially not affected by the data environment [10]. The evaluation index of ROC is shown in the following Table 1.

TP is the positive sample which is predicted to be true by the model, FN is the positive sample which is predicted to be false by the model, FP is the false sample which is predicted to be true by the model, and TN is the false sample which is

**Table 1** The evaluation index of ROC

| | | Predicted | | SUM |
|---|---|---|---|---|
| | | 1 | 0 | |
| Actual | 1 | True Positive (TP) | False Negative (FN) | Actual Positive (TP + FN) |
| | 0 | False Positive (FP) | True Negative (TN) | Actual Negative (FP + TN) |
| SUM | | Predicted Positive (TP + FP) | Predicted Negative (FN + TN) | TP + FP + FN + TN |

predicted to be false by the model. The calculation formulas of FPR and TPR are as follows:

$$FPR = \frac{FP}{TN + FP}, \quad TPR = \frac{TP}{TP + FN} \tag{10}$$

## 3 Experimental Process and Results

The UWB radar is P410 of time domain company. The P410 equipment development board mainly composed of FPGA, DSP, network port, UWB transceiver, fan and so on. Net mouth using the UDP protocol is connected with the computer. We could set the signal pulse length, the number of times, the use of the antenna and others by the computer. The center frequency of P410 UWB radar is 4.3 GHz. The material of the wall is brick with the thickness of 23.5 cm. The P410 UWB radar is placed at a distance of 20 cm from the brick wall. The distance between the P410 and the ground is the half of the brick wall. There are three experiments for through-wall human being detection in this experiment environment. The first is no person behind the brick wall. The second is one person standing away from the brick wall. The last is one person walking 2 m away from the brick wall. We collected 500 groups of pulses in each status and a pulse sampling points set for 1000. In this paper, we selected seven factors as the characteristic parameters which were extracted from the received signals, including the energy of the received signals, the kurtosis, the skewness, the maximum signals amplitude, the variance of the signals, the covariance of the signals and the mean value of the signals. These characteristic parameters are normalized to construct the training samples and testing samples. Then, we pre-processed the training data in accordance with the attributes, selected the appropriate K and generated the training model. Last, the testing data was taken into the classification model which has been built to determine the accuracy of classification. The classification accuracy of the cross-validation method is given in Table 2. According to Table 2, we selected the value of K as 1.

In this paper, we randomly selected the decision tree J48 algorithm to compare with the KNN algorithm. The experiments were performed on the same testing data and training data using the two algorithms and the experimental results are showed in Table 3. The accuracy of the KNN algorithm is significantly higher than that of the decision tree J48 algorithm. Meanwhile, the smaller the mean absolute error is, the better the classification performance of the classification algorithm is. It can be clearly seen from Table 3 that the KNN algorithm for the classification of the through-wall human being is better than the J48 algorithm.

**Table 2** Different K corresponded to different classification accuracy

| K = 1 | K = 2 | K = 3 | K = 4 | K = 5 | K = 6 | K = 7 | K = 8 | K = 9 | K = 10 |
|---|---|---|---|---|---|---|---|---|---|
| 44.67% | 41.33% | 44.67% | 40.67% | 42% | 43.33% | 40.67% | 39.33% | 41.33% | 44% |

**Table 3** Experimental results compered with KNN, J48

| Classification algorithm parameters | KNN algorithm | | J48 algorithm | |
|---|---|---|---|---|
| Correctly classified instances | 66 | 88% | 60 | 80% |
| Incorrectly classified instances | 9 | 12% | 15 | 20% |
| Kappa statistic | 0.82 | | 0.7 | |
| Mean absolute error | 0.0871 | | 0.1596 | |
| Root mean squared error | 0.2802 | | 0.2857 | |
| Relative absolute error | 19.6078% | | 35.9049% | |
| Root relative squared error | 59.4412% | | 60.6036% | |
| Total number of instances | 75 | | 75 | |
| AUC (No person status) | 0.93 | | Nothing | |
| AUC (Normal breathing status) | 0.95 | | Nothing | |
| AUC (Walking 2 m away status) | 0.85 | | Nothing | |

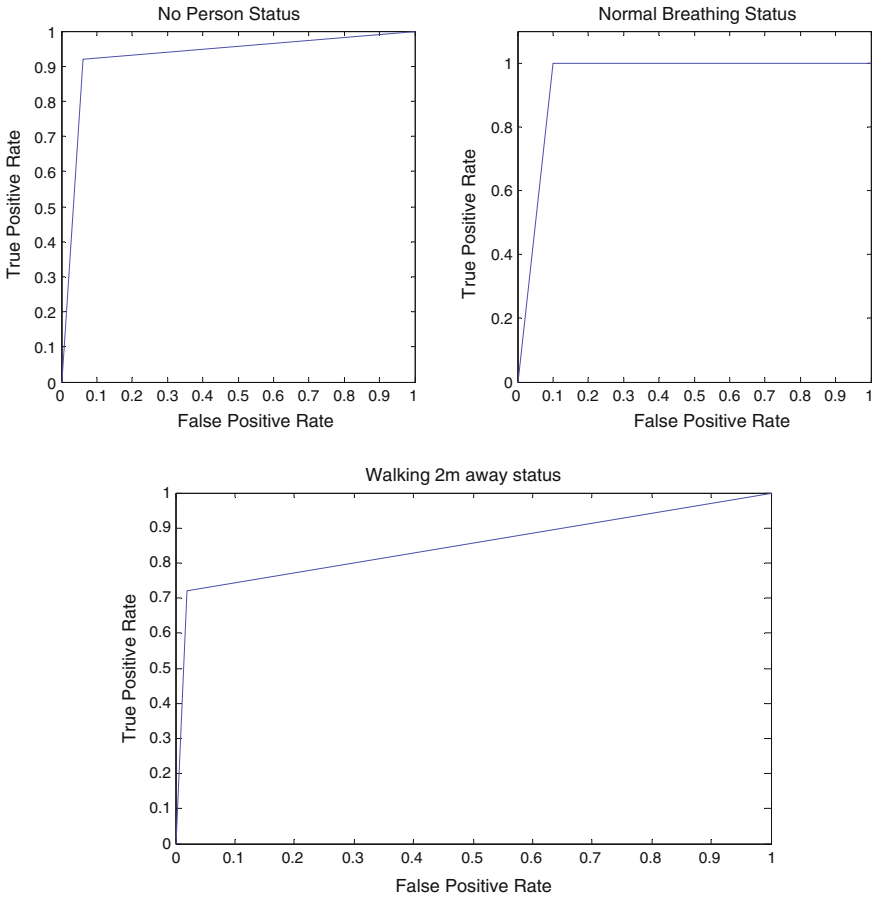



**Fig. 1** ROC of three statuses of the through-wall human being detection

This paper uses the ROC curve to analyze the experimental results. The ROC curve is closer to the upper left corner, it represents that the algorithm has a higher accuracy. As shown in Fig. 1, the three statuses of the through-wall human being using the KNN algorithm obviously absorbed the difference with the three curves. The classified effect achieved the desired goal. The ROC curve of three statuses of the through-wall human being detection using the KNN algorithm is shown as follows.

## 4 Conclusions

We have presented a framework for through-wall human being detection under three statuses with P410 UWB radar on KNN Classification algorithm. The experiments were tested under three statuses for brick wall. The results addressed the fact that the proposed classification algorithm obviously could distinguish between three different statuses, but classification method should be studied for further statuses of human being behind the brick wall or other wall.

**Acknowledgements** This paper is supported by Natural Science Foundation of China (61271411), Natural Youth Science Foundation of China (61501326). It also supported by Tianjin Research Program of Application Foundation and Advanced Technology (15JCZDJC31500) and Tianjin Science Foundation (16JCYBJC16500).

## References

1. B. Zhang, W. Wang, Through-wall detection of human being with compressed UWB radar data. EURASIP J. Wirel. Commun. Netw. **2013**(1), 1–7 (2013)
2. X. Zhao, H. Zhang, Summary: expression classification algorithm and emotional space model. Int. J. Digit. Content Technol. Appl. **6**(3), 37–44 (2012)
3. A. Kumar, Q. Liang, Z. Li, B. Zhang, Experimental study of through-wall human being detection using ultra-wideband (UWB) radar, in *Globecom Workshops*, 2012, pp. 1455–1459
4. A. Kumar, Z. Li, Q. Liang, B. Zhang, X. Wu, Experimental study of through-wall human detection using ultra wideband radar sensors. Measurement **47**(1), 869–879 (2014)
5. J. Li, Z. Zeng, J. Sun, F. Liu, Through-wall detection of human being's movement by UWB radar. IEEE Geosci. Remote Sens. Lett. **9**(6), 1079–1083 (2012)
6. A.S. Bugaev, V.V. Chapursky, S.I. Ivashov, V.V. Razevig, A.P. Sheyko, I.A. Vasi-lyev, Through wall sensing of human breathing and heart beating by monochromatic radar, in *Proceedings of Tenth International Conference on Ground Penetrating Radar*, 2004, vol. 1, pp. 291–294
7. A.G. Yarovoy, L.P. Ligthart, J. Matuzas, B. Levitas, UWB radar for human being detection. Aerosp. Electron. Syst. Mag. **21**, 22–26 (2006)
8. E. Zaikov, J. Sachs, M. Aftanas, J. Rovnakova, Detection of trapped people by UWB radar, in *Microwave Conference*, 2008, pp.1–4
9. P. Vivekanandan, M. Rajalakshmi, R. Nedunchezhian, An intelligent genetic algorithm for mining classification rules in large datasets. Comput. Inform. **2**(1), 49–58 (2011)
10. M. O'Connell, J.L. Szalma, ROC-estimator software and ROC analysis. Proc. Human Factors Ergon. Soc. **57**(1), 1432–1434 (2013)

# Flight Recognition via HRRP Using Fusion Schemes

**Yang Zhang, Xiaofeng Yu, Zhenzhen Duan, Jian Zhang and Jing Liang**

**Abstract** The purpose of this research is to increase the target recognition rate based on high-resolution range profile (HRRP) by the method of information fusion. We fuse the HRRPs of the same target from different radars with varying waves via methods of both the weighted mean and the arithmetic mean. Then, we carry out target recognition with Bayesian classifier. The result shows that the target recognition ratio using the fused HRRP is higher than that of single.

**Keywords** HRRP · Target recognition · Information fusion

## 1 Introduction

Radar automatic target recognition (RATR) is to identify the unknown target from its echoes. As an effective identification method, it has been under widespread study for decades. A high-resolution range profile (HRRP) is the coherent summations of the complex time return from target scatterers in each range cell [1]. Since it contains abundant target structure information, radar HRRP target recognition has received intensive attention from the RATR community [2, 3].

HRRP can be approximated to being made up of a number of individual scatterers, each of which corresponds to a mechanical feature of the targets [4, 5]. An aircraft target may have typically 30 main scatterers. In Fig. 1 it can be seen that the engines, the nose, the main, the tail wing roots, and pylons on the wings have been identified as main scatterers. Each scatterer reflects back the signal toward the radar as a function of the aspect angle to the radar. As HRRP is easy to obtain and can provide high target recognition rate, it plays an important role in modern military.

Y. Zhang (✉) · X. Yu · Z. Duan · J. Zhang · J. Liang
Qingshuihe Campus, University of Electronic Science and Technology of China No. 2006, Xiyuan Ave West Hi-Tech Zone, Chengdu 611731, Sichuan, People's Republic of China
e-mail: xixrui@yeah.net



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_26

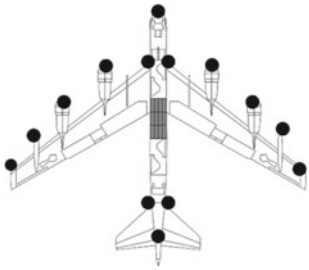


**Fig. 1** Aircraft target showing some of the main scatterers

So far, many recognition algorithms of HRRP have been presented in literature [8, 9]. Du et al. [8] proposed a method for calculating the Euclidean distance in higher order spectra feature space and it avoids calculating the higher order spectra, effectively reducing the computation complexity and storage requirement. The recognition results for measured data show that the power spectrum feature has a better recognition performance than the other higher order spectra features. Guo and Li [9] proposed a differential power spectrum and the product spectrum as the translation invariant. The simulation results showed that the differential power spectrum, and the product spectrum-based features can yield better recognition rates.

Although these above algorithms work well in many cases, a radar always works in a complex electromagnetic environment which contains many noise sources. In reality, the recognition rate is seriously affected by the environment. Zyweck and Bogner [10] refers to radar HRRP target recognition based on information fusion. However, the author did not do the research related to that. With the rapid development of information fusion technology and its application in radar sensors, we believe the target recognition ratio based on HRRP can be improved via information fusion.

In this paper, we fuse the HRRPs obtained by different radars with varying waves after pulse compressing. Then we calculate the length of target and carry out target recognition. The innovation of this paper lies in information fusion. We fuse the HRRPs and the lengths of targets via weighted mean and arithmetic mean. Of course, the recognition ratio after information fusion is higher than that of single radar.

The remainder of this paper is organized as follows. In Sect. 2, we introduce the methods of getting HRR, data fusion and target recognition. Section 3 presents simulation of target recognition based on HRRP by methods of information fusion, and finally Sect. 4 summarizes our investigation.

## 2 Process of Recognition

The whole recognition process is divided into pulse compression, information fusion and target recognition.

## 2.1 Pulse Compression

In order to improve the target recognition ratio, we must ensure that the range resolution is good enough.

The range resolution of conventional pulse radar is

$$\delta_r = \frac{c}{2B} \tag{1}$$

$c$ is the speed of light, and $B$ is the bandwidth of the transmitted waveform. For a simple pulse system, the range resolution is

$$\delta_r = \frac{c\tau}{2} \tag{2}$$

In the pulse compression system, the transmit waveform is modulated in the phase or frequency, and the echo is compressed to meet the condition that the equivalent bandwidth $B_e \gg 1/\tau$. Let $\tau_e = 1/B_e$, Eq. (1) can be changed to

$$\delta_r = \frac{c\tau_e}{2} \tag{3}$$

In the Eq. (3), $\tau_e$ represents the equivalent width of the compressed pulse. Therefore, the pulse compression radar can be used to obtain the range resolution of which the effective transmit pulse width is $\tau_e$ when the transmit pulse width is $\tau$. The ratio of $\tau$ to $\tau_e$ of the system becomes the pulse compression ratio

$$D = \frac{\tau}{\tau_e} \tag{4}$$

An ideal pulse compression system should have a matched filter system. It requires that the transmitted signal has nonlinear phase spectrum, and its envelope is close to the rectangle.

## 2.2 Information Fusion

After obtaining the HRRPs from different radars, we should fuse them with certain methods. In data level, we can choose Bayesian estimation, Kalman filtering, and weighted mean.

Bayesian estimation provides a means for data fusion, which is a common method of information fusion with multi sensor. It let the sensor information combine according to the principle of probability, and the measurement is expressed by the conditional probability. When the observation value of the sensor group is not consistent, we can carry out data fusion on the sensor directly.

However, Bayesian estimation is mainly to predict the magnitude of the next moment by calculating the correlation of time. It is mainly used in radar tracking. And it is often used in conjunction with Kalman filtering. So, we choose weighted mean to fuse the HRRPs.

Weighted average method is the most simple and intuitive method of data fusion. The method is weighted average of the redundant information provided by a set of sensors, and the result is used as the fusion value. Weighted average method can be expressed as

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} k_i x_i \tag{5}$$

and $k_1 + k_2 + \cdots + k_n = n$, $k_i$ is the weight, $n$ is the number of data.

Arithmetic mean is a special case of weighted mean. And the weighted mean can be simplified to the arithmetic mean when $k_i = 1$.

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} x_i \tag{6}$$

## *2.3 Bayesian Classifier*

After the completion of the information fusion, the feature extraction and target recognition are carried out. Here, we introduce a simple and feasible approach: Bayesian classifiers.

Bayesian classifiers are based upon knowing the statistics of each of the target features. If the statistics of each feature comprising the target is known, the probability of determining the identify can be found from integrating the statistical contributions from each of the feature.

For the measurement made of length, L, the respective probabilities of each aircraft having this dimension are $P(L|A)$ and $P(L|B)$, as shown in the Fig. 2.

These probabilities mean that if the actual target identity is known, the length measured has this distribution. The probabilities of the targets being types A and B are defined as $P(A|L)$ and $P(B|L)$, respectively. Using Bayes Formula, and assuming that there is equal likelihood of targets A and B being detected, then the actual probabilities of the target being type A or type B are

$$P(A|L) = \frac{P(L|A)}{P(L|A) + P(L|B)} \tag{7}$$

$$P(B|L) = \frac{P(L|B)}{P(L|A) + P(L|B)} \tag{8}$$

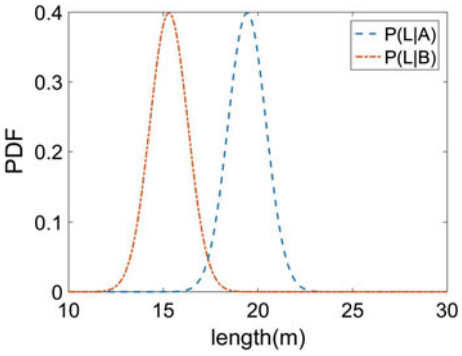



**Fig. 2** Probabilities of two targets

If the probability dense function of a target feature are known for various targets, by measuring the values of the features, the probability that the target is a certain type can be determined.

## 3 Simulation Results

The main work of the research is fusing the HRRPs obtained by two different radars to the same target. Then calculating the length of target and taking classification. The radar parameters in our simulation is listed in Table 1.

### *3.1 Acquisition of HRRP*

The echo data in the simulation obtained by the radar simulation software BSS. We get the echo data first and carry out the pulse compression: taking the echo signal to do quadrature demodulation, and moving the carrier frequency to near the zero frequency. Then taking the demodulation signal to pass through the matched filter for pulse compression.

Assuming that the noise amplitude obeys Rayleigh distribution, the average power of the noise and false alarm probability are used to set the constant false alarm threshold

**Table 1** Parameters in simulation

| Target | Waveform | Pulse width | Bandwidth |
|---|---|---|---|
| AH-64, F-15 | Gaussian, rectangle | 30 μs | 200 M, 500 M |
| Polarization | Wavelength | $P_{fa}$ | SNR |
| Horizontal | 5 cm | 0.00001 | 15 dB |

$$A = \sqrt{-2P_n log(P_{fa})} \tag{9}$$

In the Eq. (9), $P_n$ represents noise average power, $P_{fa}$ is the false alarm probability, and $A$ is the constant false alarm threshold.

After treated with false alarm threshold, we get a high-resolution range profile from the data.

After obtaining the HRRPs, the weighted mean method is used to fuse the HRRPs obtained by different radars. After the fusion, the amplitudes of each sampling point are:

$$x(i) = kx_1(i) + (1 - k)x_2(i) \tag{10}$$

where, $x_1(i), x_2(i)$ are the amplitude of HRRP obtained by different radars at the $i$-th sampling point, $i$ is the sampling point, and $k$ is the weight.

## *3.2 Data Fusion of Different Waves*

This subsection depicts the simulation of fusions for different waves. We choose the 200 M bandwidth. Here, $x_1(i)$ is the echo of radar whose wave is rectangle, and the other is Gaussian. After fusing, we need to calculate the length of target by HRRP. The result is the average of the length calculated from the 300 sets of data.

Table 2 shows that, when k takes 0.43, the length calculated of F-15 after fusion is the most close to the real length (19.43 m).

As can be seen in Table 3, the fusion effect is the best when k takes 0.11 for AH-64. The real length of AH-64 is 15.3 m.

From the above results, we know that the best value of $k$ is different when the target species changes. So, the fusion problem becomes an optimization problem. HRRP does not have linear characteristics, so the optimization of $k$ cannot use linear programming or least square method. Here, we choose a simple and effective method. We let the $k$ change from 0 to 1 with the step size of 0.01, calculating the target recog-

**Table 2** Length of F-15

| k (weight) | 0.10 | 0.20 | 0.30 | 0.39 | 0.40 |
|---|---|---|---|---|---|
| Length (m) | 18.0526 | 18.2009 | 18.7661 | 19.2524 | 19.3171 |
| k (weight) | 0.41 | 0.42 | 0.43 | 0.44 | 0.45 |
| Length (m) | 19.3548 | 19.3989 | 19.4431 | 19.4969 | 19.5377 |
| k (weight) | 0.46 | 0.47 | 0.48 | 0.49 | 0.50 |
| Length (m) | 19.5802 | 19.6353 | 19.6742 | 19.7310 | 19.7690 |
| k (weight) | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
| Length (m) | 20.2133 | 20.5337 | 20.7661 | 20.9324 | 21.1396 |

**Table 3** Length of AH-64

| k (weight) | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 |
|---|---|---|---|---|---|
| Length (m) | 15.5010 | 15.4603 | 15.4491 | 15.4399 | 15.4334 |
| k (weight) | 0.06 | 0.07 | 0.08 | 0.09 | 0.10 |
| Length (m) | 15.4058 | 15.3864 | 15.3778 | 15.3612 | 15.3437 |
| k (weight) | 0.11 | 0.20 | 0.30 | 0.40 | 0.50 |
| Length (m) | 15.3084 | 15.1774 | 15.0403 | 14.9668 | 14.8257 |
| k (weight) | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
| Length (m) | 14.7428 | 14.6946 | 14.6866 | 14.6933 | 14.7253 |

**Table 4** The recognition rate

| k (weight) | 0.10 | 0.20 | 0.30 | 0.39 | 0.40 |
|---|---|---|---|---|---|
| Recognition rate (%) | 67.2 | 72.3 | 66.1 | 69.6 | 72.8 |
| k (weight) | 0.41 | 0.42 | 0.43 | 0.44 | 0.45 |
| Recognition rate (%) | 81.2 | 72.5 | 69.4 | 71.6 | 62.3 |
| k (weight) | 0.46 | 0.47 | 0.48 | 0.49 | 0.50 |
| Recognition rate (%) | 70.1 | 64.7 | 71.3 | 72.1 | 77.5 |
| k (weight) | 0.60 | 0.70 | 0.80 | 0.90 | 1.00 |
| Recognition rate (%) | 64.3 | 62.7 | 67.4 | 63.8 | 59.38 |

nition rate of each value and selecting the value of $k$ which makes the highest target recognition rate. Using Bayesian classifier to carry out target recognition. The probability density functions of AH-64 and F-15 satisfy $x_a \sim N(15.3, 1), x_b \sim N(19.43, 1)$.

It should be noted that the recognition rate only takes the AH-64 and F-15 into account. If the length of target calculated is greater than 23 m or less than 12 m, it will be deemed to be unable to identify. And the recognition rate is the average of the 600 sets of data.

Table 4 shows that the correct recognition rate is the highest when $k$ takes 0.41. Increasing the number of aircraft still can use this method to optimize the value of $k$.

As can be seen from Fig. 3, the recognition rate after fusion is better than that of single radar when the SNR changes from 10 to 20 dB. So, the simple weighted mean can improve the recognition rate of the target when the signal-to-noise ratio is not ideal.

We draw the curve of SNR and recognition rate with 3 different methods of fusion. (weight mean, arithmetic mean, the optimal of two radars). From Fig. 4, we can see that weighted average is superior to the arithmetic average, and arithmetic average is better than anther one. The correct recognition rate of two fusion methods is higher than that of single radar.

Then, we fuse the lengths calculated by different radars and calculate the value of $k$ with the same method as data fusion. The result has shown in Fig. 5. As can be seen, the feature-level weighting is slightly better than the data level weighting.

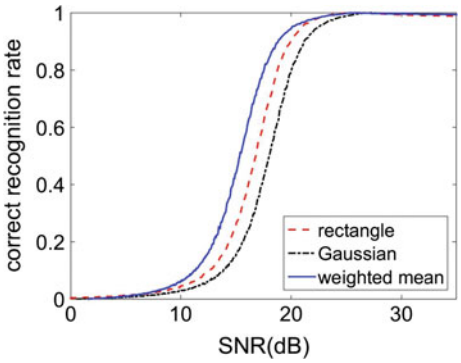



**Fig. 3** The recognition rate before and after fusion

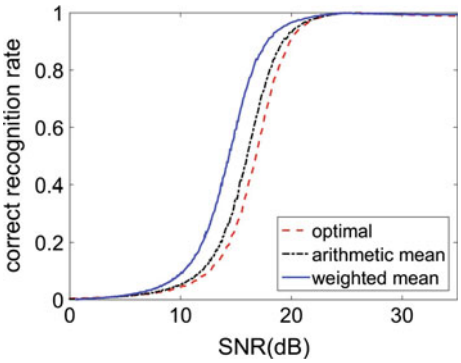



**Fig. 4** The recognition rate for different fusion methods

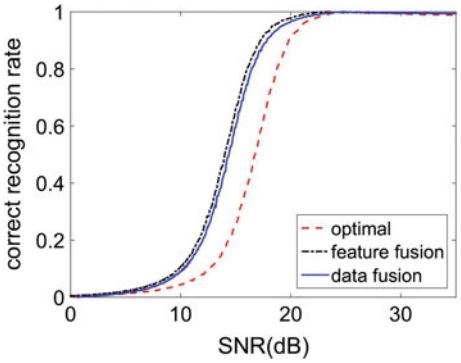



**Fig. 5** The recognition rate for different fusion level

One possible factor is that when the data is fusing, the radar detection threshold also made the corresponding weighted, which leads to the error in calculation of length and target identification. The feature-level fusion directly avoids the fusion of detection threshold, so we obtained a higher recognition rate.

## 4 Conclusion

In this paper, we fused the HRRPs obtained by different radars with varying waves or bandwidths after pulse compressing. Then we calculated the length of target and did target recognition. The recognition rate of the target is higher than that of the single radar when the waves of radars is different. The weighted mean is better than arithmetic mean and fusion in feature level is better than that in data level. Improving the quality of the algorithm and adopting more effective fusion methods can improve the recognition rate further.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (61671138), the Fundamental Research Funds for the Central Universities Project No. ZYGX2015J021, and the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

## References

1. C.R. Smith, P.M. Goggans, Radar target identification. IEEE Antennas Propag. Mag. **35**(2), 27–38 (1993)
2. X.D. Zhang, Y. Shi, Z. Bao, A new feature vector using selected bispectra for signal classification with application in radar target recognition. IEEE Trans. Signal Process. **49**(9), 1875–1885 (2001)
3. D.H. Kim, D.K. Seo, H.T. Kim, Efficient radar target recognition using the music algorithm and invariant features. IEEE Trans. Antennas Propag. **50**(3), 325–337 (2002)
4. D.R. Wehner, *High Resolution Radar*, vol. 44, no. 1, 2nd edn. (Artech House, 1995)
5. H.J. Li, S.H. Yang, Using range profiles as feature vectors to identify aerospace objects. IEEE Trans. Antennas Propag. **3**(3), 261–268 (1993)
6. B. Bhanu, Automatic target recognition: state of the art survey. IEEE Trans. Aerosp. Electron. Syst. **22**(4), 364–379 (1986)
7. R.D. Strattan, Target identification from radar signatures, in *IEEE International Conference on ICASSP Acoustics, Speech, and Signal Processing* (1978), pp. 223–227
8. L. Du, H. Liu, Z. Bao, M. Xing, Radar HRRP target recognition based on higher order spectra. IEEE Trans. Signal Process. **53**(7), 2359–2368 (2005)
9. Z. Guo, S. Li, One-dimensional frequency-domain features for aircraft recognition from radar range profiles. IEEE Trans. Aerosp. Electron. Syst. **46**(4), 1880–1892 (2010)
10. A. Zyweck, R.E. Bogner, Radar target classification of commercial aircraft. IEEE Trans. Aerosp. Electron. Syst. **32**(2), 598–606 (1996)

# The Optimization of Radar Echo Pulse Compression Algorithm Based on DSP

**Yan Wang, Chao Wang and Jie Li**

**Abstract** Radar signal processing is usually involved in a large amount of complicate tasks processing and repeated computations. Hence, DSP chips are always used in this field because of its real-time signal processing capacity. In this paper, an optimization scheme of radar pulse compression algorithm is addressed based on TMS320C6678 DSP platform. First, radar echo pulse compression algorithm is realized on DSP platform according to the MATLAB code. Then we use compile options optimization, lookup table optimization, library function optimization, algorithm optimization, and cache optimization to further increase the code efficiency. Exactly processing times are given before and after every optimization method. It can be seen from the experiment results that the processing time is distinctly decreased after optimization. The optimization methods proposed in this paper are proved to be effective.

**Keywords** DSP · Radar echo · Pulse compression algorithm · Optimization

## 1 Introduction

Radar signal processing becomes more and more important in modern military and civil application fields these years. It is widely used in early warning, target tracking and locking, meteorological observation system, etc. However, a large amount of data usually should be dealt with during the processing procedure. It is necessary that a lot of sampling data should be used and processed by certain kind of algorithms which is always complicated and time consuming [1, 2]. Although general processor has outstanding performance, its volume, power, and application conditions are not adapted to embedded system which is widely used for radar signal

Y. Wang · C. Wang
Tianjin Electronic Information College, Tianjin, China

J. Li (✉)
Zhonghuan Information College Tianjin University of Technology, Tianjin, China
e-mail: themilkyway79@126.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_27

processing. Pulse compression algorithm is the generally used method to settle the conflict between radar resolution and detection range. It is a common way to perform distance gate estimation in radar signal processing. But the large computation of this algorithm is still the key problem in the embedded system.

Lots of methods have been addressed to ensure the low power, small volume, and real-time signal processing at the same time. DSP chip is one of the main solutions to make big computation possible and solve the processing speed problem. TMS320C6678 multicore DSP is a kind of high-end embedded processor with high-performance, multi-tasking, and high-density applications. A number of high-frequency cores are integrated on this chip. Multicore high-speed bus shared by the memory and peripheral devices can improve the efficiency of the entire chip. Compared with the single-core DSP, multi-core DSP has more obvious advantages on parallel processing and computation speed. At present, many international and domestic famous universities, research institutes and manufacturers are bending themselves in exploring and researching the applications on TMS320C6678 in all fields. Several foreign researchers and manufacturers have made grateful success on DSP research work. It still belongs to the beginning status in domestic although many results have been achieved since recent years.

In this paper, we propose an optimization scheme of radar pulse compression algorithm based on TMS320C6678 DSP platform. It can be seen from the experiments that the processing results of optimized pulse compression algorithm on DSP platform are the same as the originals of MATLAB platform. Meanwhile, the processing time is obviously decreased after optimization. It not only verifies the feasibility and reliability of the proposed method but also proves its effectiveness.

## 2 Theory of Pulse Compression

With the development of science and technology, high range resolution radars which can distinguish multi-targets simultaneously and possess better range resolution are taken advantage of in many areas. Doppler frequency is very small compared with the signal bandwidth, so radar echo range can be detected by the matched filter. It can reduce the blindness point of high pulse repetition frequency as well [3].

The pulse compression system can be generally described by time bandwidth production in many cases. When using a simple rectangular pulse signal, pulse width B and bandwidth $\tau$ are generally satisfied $B\tau \approx 1$. The band of narrow pulse is wide, so we must pay the cost when we use broad pulse. That is to reduce the range resolution. If the broad pulse is modulated by frequency or phase, it will have the same bandwidth of narrow pulse. If the bandwidth become to B after modulation, the bandwidth after processing by matched filter of the receiver will be 1/B. This is called pulse compression. Pulse signal with different frequency components has different time delay after processing by the matched filter. Since the pulse

compression may not need the high peak power, high broad pulse power and narrow pulse resolution can be realized as the same time.

High resolution of narrow pulse can be achieved by augmenting the range of radar transmission pulse and peak power. In another word, we can make it by enhancing the carrier bandwidth. In radar system, frequency or phase modulation based on the simple waveform can reach the pulse compression of chirps (LFM) signal and barker signal and so on.

For a given radar system, the range resolution that can be achieved as

$$\delta_\tau = \frac{c}{2B}, \tag{1}$$

where C is velocity of light, B is transmitted signal bandwidth.

To the simple noncoding pulse radar system, the transmitted pulse bandwidth can be expressed as

$$\delta = \frac{cT}{2} \tag{2}$$

In a pulse compression system, the signal is always modulated by phase or frequency. When $B \gg \frac{1}{T}$, $\tau = \frac{1}{B}$. According to Eq. (1), we can get

$$\delta_\tau = \frac{c\tau}{2} \tag{3}$$

Pulse compression ratio (CR) is defined as broad pulse width T which divides pulse width after compression $\tau$.

$$CR = \frac{T}{\tau} \tag{4}$$

After the pulse compression, we can use the radar transmission signal of pulse width T to get the same effect of transmission signal of pulse width $\tau$. Thus, we can simply promote the range resolution of radar pulse system [4].

Assume that the peak power and the resolution range are both the same, the pulse compression CR also can be described as the power of input signal which divides the power after compression system. As $\tau = \frac{1}{B}$, we can get

$$CR = TB \tag{5}$$

From the point of view of time domain, we can use digital signal processing technique to realize the pulse compression.

Assuming the radar echo signal sequence is x(n), the impulse response function of matched filter is h(n). So the output y(n) of matched filter can be modeled as:

$$y(n) = \sum_{k=0}^{N-1} x(k) * h(n-k) = \sum_{k=0}^{N-1} h(k) * x(n-k), \tag{6}$$

where N is the order of filter and the sampling point number of radar transmission pulse signal. In practice, the matched filter is always processed in complex domain, x(n) and h(n) should be all complex sequence.

According to the theory of signal processing, we can use the following formula which can increase the processing speed instead of Eq. (6) to express the pulse compression algorithm in frequency domain.

$$y(n) = IFFT[X(k)H(k)] = IFFT\{FFT[x(n)]FFT[h(n)]\} \tag{7}$$

Since the signal in frequency domain is discrete, we can know from the digital signal processing basic theory that the result of frequency domain discretization leads to the periodization of time-domain signal. Thus, the linear convolution of sequence in time domain becomes the cyclic convolution. According to the sampling theory in frequency domain, we can substitute cyclic convolution for linear convolution when certain condition is satisfied. We suppose the input sequence x (nT) of length M and matched filter coefficient sequence h(nT) of length N. Then the sequence length of time-domain linear convolution is M + N − 1. If we need to perform linear convolution by FFT which can enormously enhance the computation speed, zero padding of each sequence will be done in advance. The length of x(nT) and h(nT) should be M + N − 1 at least.

## 3 Characteristics of TMS320C6678 DSP

The KeyStone architecture DSP TMS320C6678 is an octa-core processor released by TI company in November 2010. The extreme working frequency of each single core is 1.25 GHz. Each single core can perform 40GMAC fixed-point calculations or 20GFLOP floating-point calculations [5]. It can obtain 160GFLOP calculations under 10 W power consumption. Its overall performance has been able to reach five times of other multicore DSPs which has been formally developed. The chip integrates multicore navigator, hardware accelerators, EDMA, GigabitEthernet, USB controller, SRIO, and other useful equipments [6]. It is widely used for communication, radar, navigation, sonar, and electronic countermeasure fields.

C6678 is a multicore DSP which can do parallel processing. Each single core of C6678 has 32 KB program Cache, 32 KB data Cache, and 512 KB level 2 Cache. The memory of each processor can exchange and swap data through High-speed interconnection bus. The equipment on chip which is related to multicore includes hardware message editor, queue manager, packet switching DMA, EDMA module, and so on [7]. The use of the above module content promotes the information data exchange ability between each core substantially.

**Fig. 1** TMDSEVM6678LE platform

The HyperLink of TMS320C6678 chip is a four-Lane SerDes interface. The running peak of each Lane is 12.5 Gbps. The eight DSP cores is divided into four parts [8]. The HyperLink of two DSP crosses and connects with each other in each part. It helps to deal with data exchange between nodes and increase the speed of its own. TMS320C6678 chip contains 64-bit wide DDR3-1333 external memory bus interfaces which can connect to any 8-bit or 16-bit DDR3 chips perfectly. The peak data throughput efficiency can reach 10 GB/s [9].

The C6678 evaluation platform TMDSEVM6678L we used in this paper is shown as Fig. 1. The EVM board consists of free multicore software, CCS integrated development environment and demo codes. It makes the programmer get easier during their development process.

## 4 Algorithm Optimization on DSP

We need to perform the radar pulse compression algorithm on DSP platform before optimization. As a widely used development environment for TI DSP, CCS is used in this paper as well. Experiment results between DSP platform and MATLAB are compared to ensure the reliability of the codes processed by C6678. In order to further improving the processing speed and take full advantage of the on-chip resources of C6678, algorithm optimization is taken into account. An optimization goal should be determined at first. After deciding the optimization goal, current runtime characteristics should be determined to decide whether the implementation can meet the goal. After getting the benchmark data, the capability of DSP needs to be analyzed. Then the optimization direction can be determined. For each optimization field, relative optimization methods can be applied to do optimization.

This is a multi-loop procedure between profile and optimization. If the goal is met, current procedure can terminate. If not, other optimization methods should be used or the goal should be adjusted accordingly.

For the sake of evaluating the optimization result precisely, a relative accurate timing method should be taken to attain the exact time saving. In this paper, on-chip timer is to use to calculate the function implement time by computing the time difference before and after function. There are two counting registers TSCL/TSCH with the same frequency of CPU on C66x DSP core. Either of them represents a 64-bit value. When CPU runs a clock cycle, the register value will add one automatically. Thus, the accurate running time can be achieved with the clock cycle value. These two registers must be initialized before timing. Initialized with zero is common adapted in engineering applications.

## *4.1 Compiler Options Optimization*

There are some main compile options that play an important role in optimizing procedure.

–o3: This option is common used because it is the best and strongest optimization in the file level and its effect is especially obvious. The code cannot be dealt with pipeline until –o3 option has been selected. Some problems will appear with –o3 and it is not suitable for all programs. In that it appears, we can choose the –g option together with –o3 to eliminate most of the problems.

–pm: Optimize at the program level. This option can combine all the files together to optimize. Its main purpose is to eliminate the unused return values and the functions and constants which are not invoked.

–g: It allows symbols to debug. While it is great for debugging, it should not be used in production code. This option can cause a 30–50% performance degradation for control code and should not be used without debug requirement.

According to the effect of different options, –o3, –pm is chosen and –g is excluded. The original processing time is 23991.79 μs while the time reduces to 930.9 μs after compiler options optimization. There is a tremendous heap of processing time. It is mainly because –o3 option enables the pipeline calculation which can take fully advantage of C6678 processing units by parallel.

## *4.2 DSP Library Function Optimization*

For the sake of simplify the development difficulty, TI company affords a lot of commonly used function library which is maximum optimized by assembles. The user can invoke the functions conveniently only by adding the corresponding library file to the project. Complex multiply and convolution is the most time consuming operation during pulse compression. We can use TI optimized

function to enhance the operating efficiency. The float optimized function DSPF_sp_dotp_cplx and DSPF_sp_fir_cplx is chosen for computing the complex multiply and convolution respectively. During the process of calculation, real part is saved in the even position and imaginary part is saved in the odd position. The optimized C code of the functions is afforded and we can change the codes in terms of our specific application situation. After DSP library optimization, the processing time further reduces from 930.9 to 746.076 μs.

## *4.3 Algorithm Optimization*

The optimization method with TI library function has been discussed in advance but it is only a model for ordinary cases. The algorithm can be further simplified for special applications. In the pulse compression algorithm, the former parts of the input array are zero, so there is no need to calculate the convolutions between echo signals and matched filter coefficients. The operation time can be degraded by decreasing the loop count in which it convolutes the signals with zeros. As the afforded C code functions have been optimized by TI inline, we cannot directly change the codes according to the formula. Every inline function should be read and analyzed to ensure the validity of the optimized code. The part of convolution computed in function DSPF_sp_fir_cplx can be expressed as Fig. 2. First, the convolution of the library function multiplies the real and imaginary part of filter h0h1 by the real and imaginary part of array x0x1, x2x3, x4x5, x6x7 one by one. And then multiply h2h3 by x-1x-2, x0x1, x2x3, x4x5. We move coefficient array h to the right end and finish multiplying and adding. After that, we change the x array with a group of eight and repeat the convolution step until we finish the whole signal array. The former part of x array that the signal is zero is not computed, therefore a lot of multiplication and addition operation will be cut down. The processing speed will go up further.

It spends 371.804 μs on this algorithm after optimization. Almost half of the processing time is saved by this method.

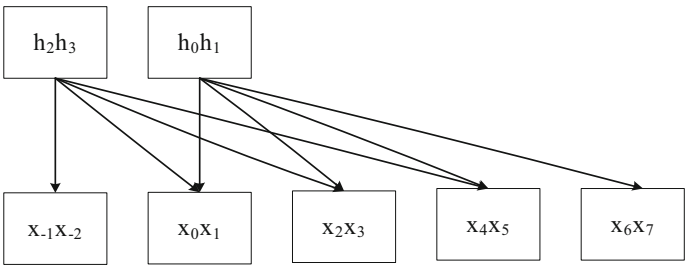



**Fig. 2** Part of the convolution diagram

### 4.4 Lookup Table Optimization

Function call and control code will break the pipeline during the DSP computation procedure. The parallel pipeline computation is mainly aimed at loop codes. There are sin or cos functions calling in the loop code of pulse compression algorithm. Hence, the loop which contains sin or cos function cannot implement pipeline. In this paper, we calculate the cos values with the step of 0.01 rad in advance and save it in an array. In the loop code, the array similar to lookup table is used instead of function call.

In that the parallel pipeline can be implemented in loop code, the processing time after optimization sharply goes down to 186.98 μs. It proves the validity of the method we have proposed.

### 4.5 Cache Optimization

A kind of L1P Cache optimization method is afforded by TI company. It can be used to enhance the cache accessing efficiency [10]. The whole process can be easily reached by the following several steps.

(1) Make use of TI compiler to compile the project with adding the option—gen_profile_info to generate the information which is used for profile.
(2) Run the .out executable file to generate .pdat file which contains analysis information. It is ensured that the program can run to the exit of main function because the .pdat file is formed when the exit function is executed.
(3) We use pdd6x command to generate .prf file with the command line "pdd6x pprofout.pdat –e xx.out –o = pprofout.prf". Then rebuild the project with the .prf file as the profile feedback file and .csv file will be generated automatically.
(4) Use command line "clt6x main.csv lots.csv rare.csv –o pfo.cmd" to generate new .cmd file and add it into the project to optimized memory allocation.

After that, cache optimization is finished. The more is the user code, the more obvious effect is cache optimization. Due to the codes of pulse compression algorithm is not too many, there is little enhancement of time which finally reaches to 184.37 μs after cache optimization.

We can distinctly find the optimization result of every method from Table 1. It can be seen that the processing time decreases from 23991.793 to 184.37 μs and basically satisfies the real-time processing requirement.

**Table 1** Optimization time comparison table

| Optimization method | Time (μs) |
|---|---|
| Original time | 23991.793 |
| After compiler options optimization | 930.9 |
| After library function optimization | 746.076 |
| After algorithm optimization | 371.804 |
| After lookup table optimization | 186.98 |
| After cache optimization | 184.37 |

## 5 Conclusions

Radar pulse compression algorithm is usually involved in a large amount of complicate tasks processing and repeated computations. Based on DSP platform, this paper addressed a serious of optimization methods to reduce the operation time of pulse compression algorithm. Exactly processing times are given before and after every optimization method such as compile options optimization method, lookup table optimization method, library function optimization method, algorithm optimization method, and cache optimization method. It can be seen from the experiments that the processing time is obviously decreased after optimization. It proves the effectiveness of the optimization methods proposed in this paper.

## References

1. Z. Bao, M. Xing, *Radar Imaging Technology* (Publishing House of Electronics Industry, 2005)
2. B. Deng, *Research on the Signal Designing and Processing of Multi-Carrier Phase Coded Radar* (National University of Defense Technology, 2011)
3. X. Ma, *Modern Radar Signal Processing* (National Defense Industry Press, 2013)
4. H. Tang, *Design and Implementation of PC in Radar Signal Processing* (Xidian University, 2014)
5. SPRS691C, TMS320C6678 Multicore Fixed and Floating-Point Digital Signal Processor Data Manual. Texas Instruments, 2012.2
6. SPRUGW1A, KeyStone Architecture Serial Rapid IO (SRIO) User Guide. Texas Instruments, 2011.10
7. TMS320C6000 DSP Enhanced Direct Memory Access (EDMA) Controller Reference Guide, Nov 2006
8. SPRUGW8. "KeyStone Architecture Hyperlink User Guide". Texas Instruments.2011.11
9. L.J. Karam, I. AlKamal, A. Gatherer et al., Trends in multicore DSP platforms. IEEE Sig. Process. Mag. **26**(6), 38–49 (2009)
10. A. Yin, C6000 Code Optimization Using CLT Tool. Texas Instruments, 2013.11

# On the Ergodic Throughput Capacity of Massive MIMO Supported Hybrid Wireless Networks

**Ganlin Zhao and Qilian Liang**

**Abstract** In this paper, we investigate theoretical transmission capacity limit in the uplink hybrid wireless network under infrastructure mode. Massive MIMO is considered to improve network capacity. Multi-user MIMO is preferred over Point-to-Point MIMO for Massive implementation to achieve improved scalability. For ad hoc mode, without infrastructure support, Massive MIMO is not practical to implement in each user device due to the limitation of complexity. Another perspective of this paper is to include the fading effect on capacity. Under favorable propagation condition, Massive MIMO greatly mitigates small-scale fading effect between each user and base station antenna. Outage capacity over large-scale fading channel is derived in both low SNR and high SNR scenarios.

**Keywords** Hybrid wireless networks · Massive MIMO · Infrastructure · Fading · Uplink · Outage capacity

## 1 Introduction

Hybrid wireless network [1] is constructed by placing several wired interconnected base stations in an ad hoc network. The network topology is shown in Fig. 1.

Throughput capacity is the key element to analysis the performance of hybrid networks. A lot of research works has been done. Gupta and Kumar [2] initiatively studied scaling law of a random ad hoc wireless networks. When nodes are randomly placed in the network and they randomly choose a destination, the per-node capacity is shown to be $\Theta(\frac{W}{\sqrt{n \log n}})$ as the number of nodes $n$ tends to infinity, where $W$ is the transmitting rate each node is capable of using a fixed range. In [3], with $b$ base stations and $n$ nodes settings, the author proves that in order to achieve infrastructure

G. Zhao (✉) · Q. Liang
Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76019-0016, USA
e-mail: ganlin.zhao@mavs.uta.edu

Q. Liang
e-mail: liang@uta.edu



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_28

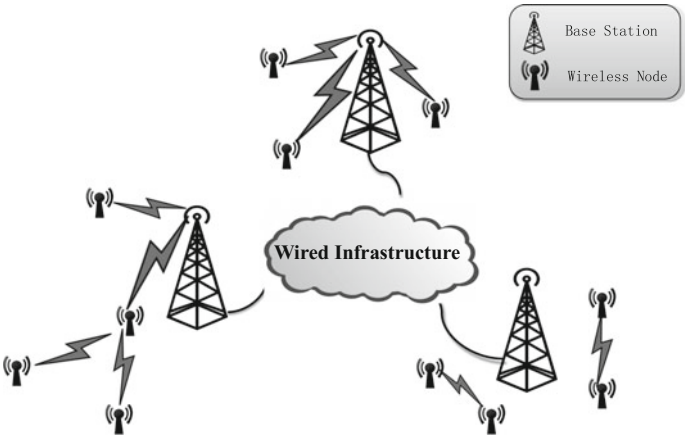



**Fig. 1** Hybrid wireless network

gain, $b$ should grow at least faster than $\sqrt{\frac{n}{\log n}}$ and the maximum throughput scales as $\Theta(bW)$ which increases linearly with the number of base stations. In [4], squared cell model is assumed and fading environment is considered. The per-node outage throughput capacity over Rayleigh fading channel scales as $O(\log(\epsilon^{\frac{b}{n}}\frac{n}{b})W_1)$ under ad hoc mode and $\Theta(\frac{b}{n}\log(\epsilon\frac{n}{b})W_2)$ under infrastructure mode, where $W_1$ is the bandwidth shared by ad hoc transmission and uplink infrastructure transmission and $W_2$ is bandwidth assigned for downlink infrastructure transmission.

In terms of Massive MIMO implementation under hybrid wireless networks, channel capacity could be different from previous scenario. Massive MIMO capacity is studied in [5, 6].

In this paper, considering using Massive MIMO for hybrid wireless network infrastructure to further improve system performance, we focus on studying how the uplink outage capacity scales with number of base stations $b$, number of users $n$, as well as large number of antennas $M$ on the base station. Our analysis relies on deriving the outage capacity in closed form expression including the variables aforementioned.

## 2 Network Modeling

### *2.1 Hybrid Wireless Network Model with Massive MIMO Support*

The following hybrid network model is considered throughout this paper:

1. The network consists of $n$ nodes and $b$ base stations totally.

2. The whole area has size $n$ and is evenly divided into $b$ hexagonal cells with unit density. Thus cell has distance $c = \sqrt{\frac{2\sqrt{3}}{3}\frac{n}{b}}$ between parallel sides.
3. Each cell contains only one base station which is placed in the center. There exists $K = \Theta(\frac{n}{b})$ users circularly uniform placed surround the base station within each cell. Each node has distance $d_k$ to the base station in the same cell.
4. The number of base stations $b = o(\frac{n}{\log n})$ is assumed that $b$ increases at a slower rate than $n$.
5. Multi-user MIMO is preferred in this work. Each user terminal is assumed to have only one antenna.
6. The base station located in each cell has an array of $M$ antennas. The antennas are sufficient to simultaneously serve all $K$ users in the cell. The number of antennas greatly exceeds the number of terminals $M \gg K$.
7. $\mathbf{G}$ is $M \times K$ propagation channel matrix:

$$\mathbf{G} = \mathbf{H}\mathbf{D}_{\beta}^{1/2} \tag{1}$$

Matrix $\mathbf{H}$ has dimension $M \times K$. Each entry is the small-scale fading coefficient between the $K$th user terminal and the $M$th antenna on base station. Matrix $\mathbf{D}_{\beta}^{1/2}$ is a $K \times K$ diagonal matrix. The diagonal entries of $\mathbf{D}_{\beta}^{1/2}$ are normalized large-scale fading coefficients related to the antenna array and $K$th user terminal.

### *2.2 Transmission Modes*

Two transmission mode are used in hybrid networks: infrastructure mode and ad hoc mode. In this paper, we only consider the uplink phase transmission of the infrastructure mode.

## 3 Uplink Outage Throughput Capacity Under Infrastructure Transmission Mode

One of the key characters with base station antennas greatly exceeds the number of user terminals is called favorable propagation [7]. The propagation matrix $\mathbf{G}$ is asymptotically orthogonal as $M \gg K$. When $M \gg K$, column vectors in matrix $\mathbf{H}$ become long and are asymptotically pairwise orthogonal according to random matrix theory [8]. Since small-scale fading coefficients have magnitude of 1, we have $\mathbf{H}^H\mathbf{H} \approx \mathbf{I}$. The propagation matrix is thus can be written as:

$$\left(\frac{\mathbf{G}^H\mathbf{G}}{M}\right)_{M\gg K} = \mathbf{D}_{\beta}^{1/2}\left(\frac{\mathbf{H}^H\mathbf{H}}{M}\right)_{M\gg K}\mathbf{D}_{\beta}^{1/2} \approx \mathbf{D}_{\beta} \tag{2}$$

From Eq. (2) we can observe that long propagation matrix makes that small-scale fading effect greatly mitigated under favorable propagation while large-scale fading coefficient dominates.

The sum rate capacity of uplink MU-MIMO can be obtained from [9] assuming receiver knows the channel:

$$C_{sum\_ul} = \log_2 \det \left( \mathbf{I}_K + \rho_u \mathbf{G}^H \mathbf{G} \right) \tag{3}$$

by applying favorable propagation condition

$$C_{sum\_ulM \gg K} \approx \log_2 \det \left( \mathbf{I}_K + M \rho_u \mathbf{D}_\beta \right) \tag{4}$$

$$= \sum_{k=1}^{K} \log_2 \left( 1 + M \rho_u \beta_k \right) \tag{5}$$

Equation (5) indicates that the sum rate could be represented as individual rate correlated to each terminal. This yields an simplification of linear decoding process by using matched-filter at the BS.

To further derive the outage capacity of the uplink channel, we assume each large-scale fading coefficient $\beta_k$ has the following form [10]:

$$\beta_k = \phi d_k^{-\alpha} \zeta_k \tag{6}$$

where $\phi$ is a constant related to the antenna gain and carrier frequency, $d_k$ is the distance between the base station and $k$th terminal. $\alpha$ is the path loss exponent. $\zeta_k$ is the log-normal shadowing with $10 \log_{10} \zeta_k \sim \mathcal{N}(0, \sigma_k^2)$.

We assume that all user terminal devices transmit at a sum rate $Rbit/s/Hz$, the related outage probability is

$$P_{out} = Pr \left\{ \sum_{k=1}^{K} \log_2 \left( 1 + M \rho_u \phi d_k^{-\alpha} \zeta_k \right) < R \right\} \tag{7}$$

Then to solve the equation above, we explore the outage capacity both in high SNR and low SNR scenario.

### 3.1 Low SNR

At low SNR, we employ $\ln(1 + x) \approx x$ and obtain the following,

$$P_{out} = Pr \left\{ \log_2^e M \rho_u \phi \sum_{k=1}^{K} \left( d_k^{-\alpha} \zeta_k \right) < R \right\} \tag{8}$$

For nodes are circularly uniform placed, we assume the path loss $d_k^{-\alpha}$ is given during solving the equation. Since $\zeta_k$ is log-normal distributed with $10\log_{10}\zeta_k \sim \mathcal{N}(0,\sigma_k^2)$, we could derive $d_k^{-\alpha}\zeta_k$ is also log-normal distributed with $10\log_{10}\left(d_k^{-\alpha}\zeta_k\right) \sim \mathcal{N}(\mu_k,\sigma_k^2)$ where $\mu_k = 10\log d_k^{-\alpha}$.

Applying Fenton–Wilkinson method [11], we approximate the summation of log-normal distributed random variables $d_k^{-\alpha}\zeta_k$ with a new log-normal random variable, say $\chi$ that $10\log_{10}\chi \sim \mathcal{N}(\mu_\chi,\sigma_\chi^2)$. The new parameter of $\chi$ is chosen by,

$$\mu_\chi = \xi^{-1}\left(\frac{\xi^2}{2}\left(\sigma_k^2 - \sigma_\chi^2\right) + \ln\left(\sum_{k=1}^{K} e^{\xi\mu_k}\right)\right) \tag{9}$$

$$\sigma_\chi^2 = \xi^{-2}\ln\left(\left(e^{\xi^2\sigma_k^2} - 1\right)\frac{\sum_{k=1}^{K} e^{2\xi\mu_k}}{\left(\sum_{k=1}^{K} e^{\xi\mu_k}\right)^2} + 1\right) \tag{10}$$

Then the outage probability is,

$$P_{out} = Pr\left\{\chi < \frac{R}{\log_2^e M\rho_u\phi}\right\} \tag{11}$$

The cumulative distribution function (CDF) of the log-normal distributed random variable $\chi$ is given by

$$F(x) = \frac{1}{2}erfc\left(-\frac{\ln x - \mu_\chi}{\sigma_\chi\sqrt{2}}\right) \tag{12}$$

where $erfc(x)$ is the complementary error function.

With the CDF of $\chi$, we have the following

$$P_{out} = \frac{1}{2}erfc\left(-\frac{\ln T - \mu_\chi}{\sigma_\chi\sqrt{2}}\right), T = \frac{R}{\log_2^e M\rho_u\phi} \tag{13}$$

According to [12], tight exponential upper bounds for the complementary error function are derived

$$erfc(x) \le \frac{1}{2}e^{-2x^2} + \frac{1}{2}e^{-x^2} \le e^{-x^2}, x > 0 \tag{14}$$

Applying the approximation to Eq. (13), we got

$$P_{out} \le \frac{1}{2}\left(\frac{1}{2}\left(T'\right)^2 + \frac{1}{2}T'\right) = \epsilon \tag{15}$$

where

$$T' = e^{-\left(-\frac{\ln T - \mu_\chi}{\sigma_\chi \sqrt{2}}\right)^2} \tag{16}$$

By solving Eq. (15) we get

$$T' = \frac{\sqrt{1+16\epsilon}-1}{2} \tag{17}$$

Equation (17) is satisfied when $\epsilon < \frac{1}{2}$. Combining (16) and (17) we got

$$C_\epsilon = \log_2^e M \rho_u \phi exp\left\{\mu_\chi - \sigma_\chi \sqrt{2\ln\left(\frac{2}{\sqrt{1+16\epsilon}-1}\right)}\right\} \tag{18}$$

From Eq. (18) we can see that the capacity is related to $\mu_\chi$ and $\sigma_\chi$ which are defined in (9) and (10)

Since the distance $c = \sqrt{\frac{2\sqrt{3}}{3}\frac{n}{b}}$ between the parallel sides of a cell is bounded by $\Theta(\sqrt{\frac{n}{b}})$ as the system increases. With uniform distribution, the distance $d_k$ between the $k$th user in one cell and the base station is proportional to $c$ and it is bounded by $\Theta(\sqrt{\frac{n}{b}})$ as well. Since $\mu_k = 10\log d_k^{-\alpha}$. Then $\mu_k = \Theta(\log\sqrt{\frac{n}{b}})$. Obviously

$$\sigma_\chi^2 \approx \ln(\frac{1}{K}+1) = \Theta(\log(\frac{b}{n}+1)). \tag{19}$$

We can also derive that

$$\mu_\chi \approx \Theta(\log(\frac{n}{b})). \tag{20}$$

Then, we make the conclusion that the transmission rate under low SNR for Massive MIMO uplink is:

$$R_{ul}^{low} = O\left(M\left((\frac{n}{b})e^{-\sqrt{\log\left(\frac{b}{n}+1\right)\delta_\epsilon}}\right)\right) \quad bit/s/Hz \tag{21}$$

where $\delta_\epsilon = 2\ln\left(\frac{2}{\sqrt{1+16\epsilon}-1}\right)$.

From Eq. (21) we can see that in low SNR scenario, uplink outage capacity scales linearly with number of base stations $M$, which means that increasing number of antenna will always increase throughput capacity. Figure 2 shows uplink outage capacity with different $M$ settings when outage probability $\epsilon = 0.001$.

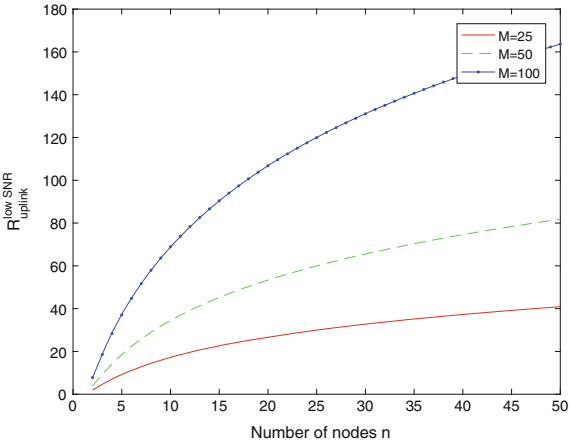



**Fig. 2** Uplink outage throughput capacity under low SNR scenario ($\epsilon = 0.001$)

### 3.1.1 High SNR

In high SNR scenario, we approximate $\log(1 + x) = log(x)$ to Eq. (7),

$$P_{out} = Pr\left\{\sum_{k=1}^{K} \log_2\left(M\rho_u \phi d_k^{-\alpha}\zeta_k\right) < R\right\} \tag{22}$$

Previously we have seen that $\zeta_k$ is log-normal distributed with $10\log_{10}\left(\zeta_k\right) \sim \mathcal{N}(0, \sigma_k^2)$. We define another random variable $L = \log_2 M\rho_u \phi d_k^{-\alpha}\zeta_k$ that $L \sim \mathcal{N}(\mu_L, \sigma_L^2)$ and

$$\mu_L = \log_2(M\rho_u \phi d_k^{-\alpha}) \tag{23}$$

$$\sigma_L^2 = (\frac{\sigma_k}{10\log_{10} 2})^2 \tag{24}$$

The summation of normal distributed random variables is also normal distributed. That is $\hat{L} = \sum_{k=1}^{K} L$, $\hat{L} \sim \mathcal{N}(\mu_{\hat{L}}, \sigma_{\hat{L}^2})$ with $\mu_{\hat{L}} = K\mu_L$, $\sigma_{\hat{L}}^2 = K\sigma_L^2$.

Then we can rewrite Eq. (22)

$$\begin{aligned} P_{out} &= Pr\left\{\sum_{k=1}^{K} L < R\right\} \\ &= Pr\left\{\hat{L} < R\right\} \\ &= \frac{1}{2} erfc\left(-\frac{R - \mu_{\hat{L}}}{\sigma_{\hat{L}}\sqrt{2}}\right) = \epsilon \end{aligned} \tag{25}$$

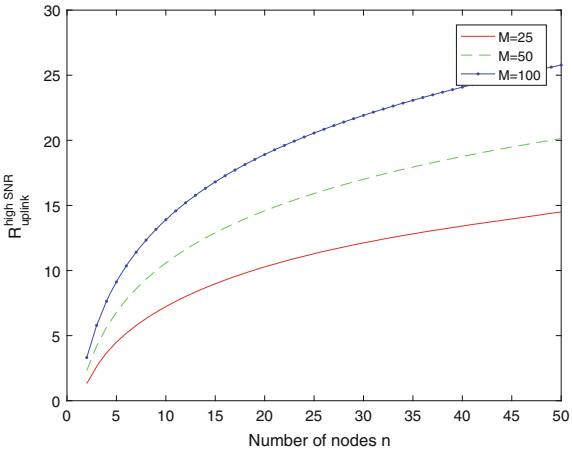



**Fig. 3** Uplink outage throughput capacity under high SNR scenario ($\epsilon = 0.001$)

Apply approximation (14)

$$C_\epsilon = \mu_{\hat{L}} - \sigma_{\hat{L}} \sqrt{2 \ln \left( \frac{2}{\sqrt{1+16\epsilon}-1} \right)}, \epsilon < \frac{1}{2} \tag{26}$$

By analysis of (23) and (24), we can get $\mu_{\hat{L}} = \Theta(\frac{n}{b} \log(M\sqrt{\frac{n}{b}}))$ and $\sigma_{\hat{L}} = \Theta(\sqrt{\frac{n}{b}})$ So the transmission rate under high SNR for Massive MIMO uplink is:

$$R_{ul}^{high} = O\left( \frac{n}{b} \log(M\sqrt{\frac{n}{b}}) - \frac{n}{b}\sqrt{\delta_\epsilon} \right) \quad bit/s/Hz, \tag{27}$$

where $\delta_\epsilon = 2 \ln \left( \frac{2}{\sqrt{1+16\epsilon}-1} \right)$.

Under high SNR scenario, the uplink outage throughput capacity yields logarithmic increase with number of base station antennas $M$. From Fig. 3, we can see that doubling $M$ from 25 to 50 almost has the same capacity gain when doubling $M$ from 50 to 100.

## 4 Conclusions

In this paper, with Massive MIMO implementation in the hybrid wireless network, we derived the theoretical uplink outage throughput capacity for infrastructure mode. Under favorable propagation condition, small-scale fading is averaged out due to asymptotic orthogonality of the propagation matrix with increased base station antenna $M$. Large-scale fading on throughput capacity is examined for both high and

low SNR. Close-form outage throughput capacity is derived. We proved that at low SNR, the infrastructure mode uplink outage capacity is $O\left(M\left(\left(\frac{n}{b}\right)e^{-\sqrt{\log\left(\frac{b}{n}+1\right)\delta_\epsilon}}\right)\right)$ bit/s/Hz. At high SNR, the infrastructure uplink outage capacity is $O\left(\frac{n}{b}\log(M\sqrt{\frac{n}{b}})-\frac{n}{b}\sqrt{\delta_\epsilon}\right)$ bit/s/Hz. Increasing number of antennas at low SNR scenario yields linear capacity gain compared to logarithmic capacity gain at high SNR.

**Acknowledgements** This work was supported in part by National Science Foundation under Grants CNS-0964713 and Office of Naval Research under Grant N00014-13-1-0043.

## References

1. P.T. Oliver Dousse, M. Hasler, Connectivity in ad-hoc and hybrid networks, in *Proceedings of IEEE Infocom 2002*
2. P. Gupta, P.R. Kumar, The capacity of wireless networks. IEEE Trans. Inf. Theory **46**(2), 388–404 (2000)
3. B. Liu, Z. Liu, D. Towsley, On the capacity of hybrid wireless networks, in *Infocom 2003*, vol. 2, pp. 1543–1552, Apr 2003
4. X. Wang, Q. Liang, On the throughput capacity and performance analysis of hybrid wireless networks over fading channels. IEEE Trans. Wirel. Commun. **12**(6), 2930–2940 (2013)
5. H.Q. Ngo, E.G. Larsson, T.L. Marzetta, Energy and spectral efficiency of very large multiuser MIMO systems. IEEE Trans. Commun. **61**(4), 1436–1449 (2013)
6. F. Rusek et al., Scaling up MIMO: opportunities and challenges with very large arrays. IEEE Signal Process. Mag. **30**(1), 40–60 (2013)
7. H.Q. Ngo, E.G. Larsson, T.L. Marzetta, Aspects of favorable propagation in Massive MIMO, in *22nd European Signal Processing Conference (EUSIPCO)*. Lisbon 2014 (2014), pp. 76–80
8. J. Hoydis, S. ten Brink, M. Debbah, Massive MIMO in the UL/DL of cellular networks: how many antennas do we need? IEEE J. Sel. Areas Commun. **31**(2), 160–171 (2013)
9. D. Tse, P. Viswanath, *Fundamentals of Wireless Communication* (Cambridge University Press, 2005)
10. K. Zheng, S. Ou, X. Yin, Massive MIMO channel models: a survey. Int. J. Antennas Propag. **2014**, Article ID 848071, 10 pp. (2014)
11. G.L. Stuber, *Principles of Mobile Communication* (Kluwer Academic Publishers, 2001)
12. M. Chiani, D. Dardari, M.K. Simon, New exponential bounds and approximations for the computation of error probability in fading channels. IEEE Trans. Wirel. Commun. **2**(4), 840–845 (2003)

# Increasing Capacity of Multi-cell Cooperative Cellular Networks with Coprime Deployment

**Hao Liang and Qilian Liang**

**Abstract** A novel deployment for multi-cell cooperative cellular network based on the two-dimensional (2D) coprime array, and analysis on its sum-rate capacity are proposed. Taking advantage of that the 2D coprime array system can provide O (N2) degree of freedom by using only N physical sensors when the second-order statistics of the received data is used, we show that the derivation procedure of average sum-rate capacity for the cooperative cellular network is still valid for the coprimed distributed base stations (BSs) in the non-fading and Rayleigh fading channels. Simulations further validate these theoretical results.

**Keywords** Coprime · Sum-rate capacity · Cellular network

## 1 Introduction

With the forthcoming 500 million connections and explosion of mobile broadband data, heterogeneous networks with a combined approach including improving, densifying and finding better deployment of the cells are becoming the future way to enhance network capacity and performance [1].

Several studies have addressed the cell planning and interference issues in the past. In [2], different symmetric cell deployment strategies have been studied based on inter-site distance and it is concluded that the network capacity can be enhanced through denser deployment of cells. In [3] the authors propose to deploy a massive amount of small cells in order to increase the total capacity and reduce the energy consumption.

H. Liang · Q. Liang (✉)
Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76019, USA
e-mail: liang@uta.edu

H. Liang
e-mail: hao.liang@mavs.uta.edu



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_29

However, when the cells are deployed in a super dense way, the inter-cell interference problem would become dominant. In a high interference scenario with many line-of-sight (LOS) interferers around, keep adding cells regardless of the surrounding environment will not always help to improve the network performance [4].

Hence, finding better deployment of BSs and increasing the level of collaboration would be a better way to achieve higher capacity. In this paper, a novel coprime distributed network is proposed which can increase the degree of freedom of the co-arrays to reduce the number of physical sensor while maintaining the same performance. The work [5] introduced this theorem into multi-dimensions. In [6], another sparse deployment method named nested-array was introduced.

The rest of the paper is organized as follows. In Sect. 2, we give a brief introduction of coprime arrays system and the system model. In Sect. 3, we study the sum-rate capacity of the designed system. In Sect. 4, Monte Carlo simulation result is provided. The conclusion is drawn in Sect. 5.

## 2 Preliminary and Model Description

### *2.1 2D Coprime Co-array*

We will first introduce several concepts about multidimensional lattice, which is the basic knowledge to coprime co-array mentioned in [5].

Given a D × D singular matrix V, consider the set of all D × 1vectors of the form

$$t = Vn \tag{1}$$

where n are integer vectors (vectors with integer entries $n_i$). This set of all is called the lattice generated by V and is denoted as LAT(V). The set of all D × 1 vectors of the form Vx, where x ∈ $[0, 1)^D$ (i.e., where the elements of x satisfy $0 \leq x_i \leq 1$), is said to be the fundamental parallelepiped of V, and is denoted as FPD(V). Similarly the symmetric parallelepiped SPD(V) is the set of vectors Vx with $x \in (-1, 1)^D$. To visualize these definitions, consider the 2D example where V is 2 × 2 and can be written as

$$V = [v_1 \qquad v_2]. \tag{2}$$

Some important points to recall are as follows:

(1) the volume of FPD(V) (area in 2-D) is equal to $| detV |$;
(2) thus, the number of lattice points generated by V per unit volume (lattice density) is equal to $1/| detV |$; larger the determinant, smaller is the density of lattice points;
(3) LAT(V) is the same as LAT(VE) for any integer matrix E with det E = ±1

**Theorem 1** *Coprime Co-arrays in 2D: Assume the 2 × nonsingular integer matrixes M and N are commuting and coprime. Then*

*(1) given any integer vector k, there exist integer vectors $n_1$ and $n_2$ such that $k = Mn_1 - Nn_2$ holds; the co-array contains all integer vectors, if $n_1$ and $n_2$ are allowed to vary over all integer vectors;*

*(2) let the integer vectors $n_1$ and $n_1^{'}$ be restricted to FPD(N) and $n_2$ and $n_2^{'}$ restricted to FPD(M); then*

$$Mn_1 - Nn_2 \neq Mn_1^{'} - N^{'}n_2 \tag{3}$$

*as long as $(n_1, n_2) \neq (n_1^{'} - Nn_2^{'})$;*

*(3) the integer vector $Mn_1$ and $Nn_2$ are distinct when the integers $n_1$ and $n_2$ are such that $n_1 \in FPD(N)$ and $n_2 \in FPD(M)$, unless $n_1 = n_2 = 0$.*

**Theorem 2** *Generating All Integer Vectors in FPD(MN): Assume M and N are commuting, coprime, and nonsigular* D × D *integer matrixes, and consider the difference $k = Mn_1 - Nn_2$. With $n_1 \in FPD(2N)$ and $n_2 \in FPD(M)$, all integers $k \in FPD(MN)$ can be generated. Similarly, with $n_1 \in FPD(N)$ and $n_2 \in SPD(M)$, all integers $k \in FPD(MN)$ can be generated.*

## 2.2 Sum-Rate Capacity for Multi-cell Processing

The ergodic per cell sum-rate capacity is given by [7]

$$C(P) = \frac{1}{L} E[log_2(I_L + PH_L H_L^*)] \tag{4}$$

where P is the transmit power of a single user, and the expectation is taken with respect to fading coefficients $H_L$.

The matrix $H_L H_L^*$ is an $L\times$ L matrix given by

$$[H_L H_L^*]_{m,n} = \begin{cases} a_m a_m^* + b_m b_m^* & m = n \\ b_m a_n^* & n = (m-1) \quad modL \\ a_m b_n^* & n = (m+1) \quad modL \\ 0 & otherwise \end{cases} \tag{5}$$

## 2.3 System Model

Figures 1 and 2 demonstrate the distribution of BSs in a 2D coprime array method described in Sect. 2.1. Where

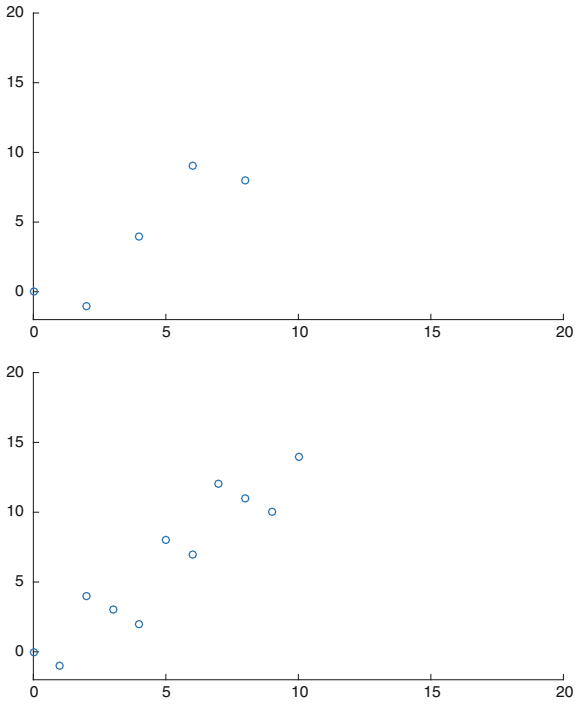



**Fig. 1** The array generated by $Mn_1, n_1 \in FPD(N)$, and $Nn_2, n_2 \in FPD(M)$

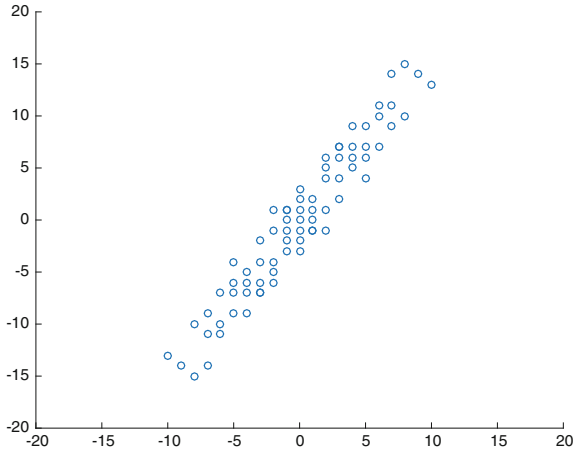



**Fig. 2** The union of elements in $Mn_1 - Nn_2$ and $Nn_2 - Mn_1$

$$M = \begin{bmatrix} 1 & 1 \\ -1 & 4 \end{bmatrix} \quad and \quad N = \begin{bmatrix} 2 & 1 \\ -1 & 5 \end{bmatrix} \tag{6}$$

As, **M - N = I**, M and N are coprime. Every dot in Fig. 1 stands for a physically deployed macrocell and Fig. 2 is the virtual BSs generated by the physical BSs. Every cell has K users. The vector baseband representation of signals received at the BSs is given as

$$y = Hx + n, \tag{7}$$

where H is the channel transfer matrix

$$H = \begin{bmatrix} a_0 & 0 & 0 & \cdots & 0 & b_0 \\ b_1 & a_1 & 0 & \cdots & 0 & 0 \\ 0 & b_2 & a_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & a_{L-2} & 0 \\ 0 & 0 & 0 & \cdots & b_{L-1} & a_{L-1} \end{bmatrix} \tag{8}$$

where $a_m, b_m$ are channel coefficients experienced by the K users from the corresponding M and N BSs.

## 3 Sum-Rate Capacity of Coprime Distributed Cooperative Networks

### *3.1 Invariance of the Difference Co-array*

The difference co-array generated from the coprime array can be expressed as:

$$k = Mn_1 - Nn_2 \tag{9}$$

The correlation between the any two array elements is

$$R(k) = E[x(Mn_1)x^*(Nn_2)] = E[x(n)x^*(n-k)] \tag{10}$$

The received signal vector

$$y_{m,n} = \sum_{k=1}^{k} a(\theta_k)s_k + n_k \tag{11}$$

where $s_k$ denotes the signal waveform vector, $n_k$ denoted a i.i.d. Gaussian noise. $a(\theta_k) = [1, e^{-j\frac{2\pi}{\lambda}u_2 sin(\theta_k)}, \ldots, e^{-j\frac{2\pi}{\lambda}u_{2M+N-1} sin(\theta_k)}]^T$, The covariance matrix of

received signal vector can be expressed as

$$R = Ey(l)y^*(l) = \sum_{k=1}^{k} \sigma_k^2 a(\theta_k)a^H(\theta_k) + \sigma_n^2 \mathbf{I} \tag{12}$$

The covariance matrix will be

$$\hat{R} = \frac{1}{L}\sum_{l=1}^{L} y(l)y^*(l) \tag{13}$$

Taking the average of R,

$$\hat{R}^2 = \frac{1}{MN}\sum_{m=0}^{M}\sum_{n=0}^{n} R \tag{14}$$

The covariance matrix of signal has the same form as $\hat{R}^2$ Therefore, the coprime distributed BSs have a degree of freedom of O(MN), with O(M+N) physical BSs./ subsectionAWGN Channel Assume all fading coefficients in non-fading channels to be 1, all transmission schemes with equal intra-cell power achieve the same throughput.

From (5) we can derive that the uplink average per cell sum-rate without fading is

$$C(P) = \frac{1}{L}\sum_{l=0}^{L} log_{10}(1 + 2KP(1 + cos(2\pi\frac{l}{L}))) \tag{15}$$

The average per cell sum-rate capacity with Rayleigh fading is

$$C(P) = \frac{1}{L}\sum_{l=0}^{L-1} long_{10}(1 + 2KP(m_2 + |m_1|^2 cos(2\pi\frac{l}{L}))) \tag{16}$$

## 4 Simulation Result

Figures 3 and 4 show the channel capacity under the different intra-cell transmission power. Lower bounds and upper bounds are also analyzed with K = 100 per cell. Figure 3 shows lower bound for $\varepsilon = 0.1$ and Fig. 4 shows it for $\varepsilon = 0.5$.

In Fig. 5, we compared the sum-rate capacity between coprime deployment and uniform deployment under Rayleigh fading and non-fading cases with the same number of physical BSs. We observe that the coprime deployment has a 7 times performance improvement comparing to the uniform deployment for both the Rayleigh fading and non-fading cases.

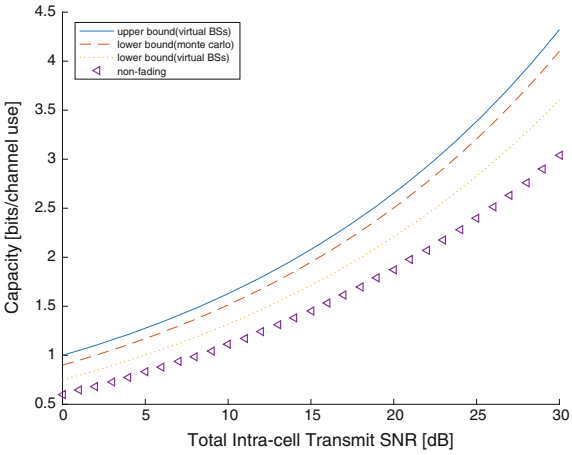



**Fig. 3** Per cell sum-rate capacity for K = 100, $\varepsilon = 0.1$

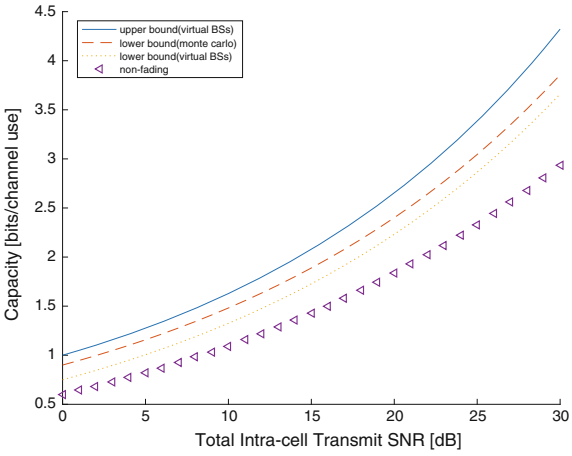



**Fig. 4** Per cell sum-rate capacity for K = 100, $\varepsilon = 0.5$

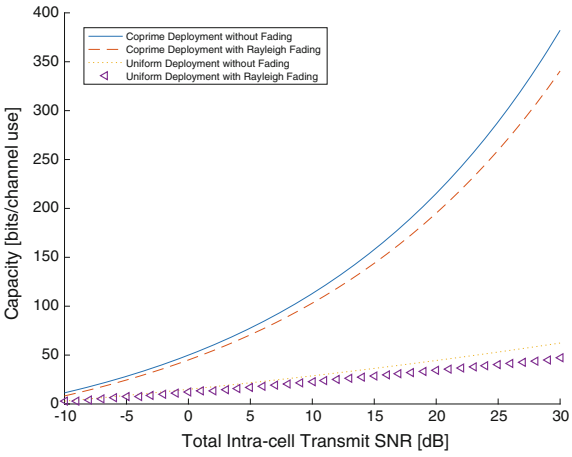



**Fig. 5** Sum-rate capacity comparison between coprime and uniform deployment

## 5 Conclusion and Future Works

In this paper, we apply the coprime array algorithm into BSs depolyment case which is a combination of O(M+N) BSs distributed over nonseparable lattices, whose difference co-array can give rise to a much larger 2D array with O(MN) BSs on the dense lattice. Which means that with M+N physical BSs we can achieve the performance of MN BSs. From the simulation result, we find that the coprime array deployed BSs performs much better than the uniform deployed BSs. Which means we can use less number of BSs to achieve a better performance. Meanwhile, there are still some open question beyond this work, including resource allocation, beamforming, interference, the optimal choice of the generate matrix.

**Acknowledgements** This work was supported in part by National Science Foundation under Grants CNS-0964713, CNS-1247848, and Office of Naval Research under Grant N00014-13-1-0043.

## References

1. Ericsson White Paper, Heterogeneous Networks, Feb 2012
2. M. Gudmundson, Cell planning in manhattan environments, in IEEE Vehicular Technology Conference (Spring, 1992)
3. W. Liu et al., Massive MIMO or small cell network: who is more energy efficient?, in IEEE Wireless Communications and Networking Conference Workshops, Apr 2013
4. Y. Wu, P. Butovitsch, M. Zhang, Capacity upper bound for adding cells in the super dense cellular deployment scenario, in Vehicular Technology Conference, 2014
5. P. Pal, P.P. Vaidyanathan, Theory of sparse coprime sensing in multiple dimensions. IEEE Trans. Signal Process. **59**(9), 3592–3608 (2011)
6. Q. Wu, Q. Liang, Increasing capacity of muti-cell cooperative cellular networks with nested deployment, in IEEE International Conference on Communications, June 2015
7. A.D. Wyner, Shannon-theoretic approach to a Gaussian cellular multipleaccess channel. IEEE Trans. Inf. Theory **40**(6), 1713–1727 (1994)
8. O. Somekh, B.M. Zaidel, S. Shamai, Sum rate characterization of joint multiple cell-site processing. IEEE Trans. Inf. Theory **53**(12), 4473–4497 (2007)

# Channel-Based Collaborative Authentication Scheme for Wireless Sensor Network

**Guangming Han and Ting Jiang**

**Abstract** As the broadcast nature of wireless communications, information security becomes a more challenging issue in contrast to traditional wired channels especially for capabilities limited wireless sensor networks. On the other hand, we can exploit the randomness of wireless channels to authenticate devices by comparing their channel properties. In this paper, we proposed a performance enhanced physical layer authentication scheme for wireless sensor networks by collaboration of multiple sink nodes. A binary hypothesis testing based authentication method is proposed to identify nodes by directly comparing the current and previous channel frequency responses to decide whether the transmitter is malicious node or the legitimate node. Besides, we collaboratively integrate the decisions of different sink nodes to improve the authentication performance. The simulation results show that our proposed collaborative authentication scheme improves detection performance greatly.

**Keywords** Physical layer authentication · Wireless sensor network · Spoofing attack · Binary hypothesis testing

## 1 Introduction

Due to the open nature of wireless channels, there are various attacks in wireless communications, since the intruders can access wireless networks without a physical connection. For wireless sensor networks (WSNs), security is a critical issue for many sensor network applications, such as military target tracking and

G. Han · T. Jiang (✉)
School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China
e-mail: tjiang@bupt.edu.cn

G. Han · T. Jiang
Key Laboratory of Universal Wireless Communications (BUPT), Ministry of Education, Beijing, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_30

security monitoring. To provide security and privacy to small sensor nodes are challenging, due to the limited capabilities of sensor nodes in terms of computation, communication, memory storage and energy supply [1].

Currently, wireless network security mechanisms are largely derived from the idea of traditional wired network, where the security schemes are accomplished above the physical layer of protocol stacks. While recently, it is demonstrated that some physical layer properties derived from wireless channels and devices can be utilized as unique device fingerprints to accurately identify wireless devices [2].

In paper [3], the author utilize received signal strength indicator (RSSI) to detect Sybil attack and this scheme does not burden the WSNs with shared keys to messages. But received signal strength indicator is coarse-grained channel information and fine-grained physical layer authentication property is proposed in [4]. In [5], Xiao et al. proposed a channel based detection scheme to identify Sybil attacks. In common practice, wireless networks are deployed with multiple APs, so that the author proposed a multiple-AP based Sybil detection scheme.

However, the channel estimation results are always corrupted by environmental noise and interference, so that the authentication scheme is not reliable enough [6]. In [7], I/Q imbalance is used as a unique fingerprint to identify transmitters. Besides, the paper proposed an efficient solution in which multiple collaborative receivers are involved in the authentication process to improve authentication performance.

In this paper, we utilize the cooperation of multiple sink nodes to enhance the authentication performance. In wireless sensor networks, there are multiple sink nodes deployed in the network [8], which facilitates collaborative authentication strategy. For instance, many sink nodes may receive an authentication request from the same source due to the broadcast nature of the wireless medium. Thus, the sink nodes will rely on cooperative observations to jointly authenticate the transmitting node to achieve improved authentication reliability. Since the orthogonal frequency division multiplexing (OFDM) systems offer some overwhelming features, such as robustness against the multipath channel, flexible resource allocations and high spectral efficiency, the OFDM scheme is being gradually applied in WSNs [9].

The remainder of this paper is organized as follows. The next section describes our system model. In Sect. 3, the proposed collaborative authentication scheme is described. Section 4 presents the numerical results of our authentication technique. Finally, Sect. 5 concludes the paper.

## 2 System Model

As shown in Fig. 1, the wireless sensor network consists of many sensor nodes and some sink nodes, where multiple sink nodes are deployed in order to improve the lifetime and robustness of WSN. The sink nodes are located in the center of the neighboring sensor nodes so that those sensor nodes can send messages to the sink

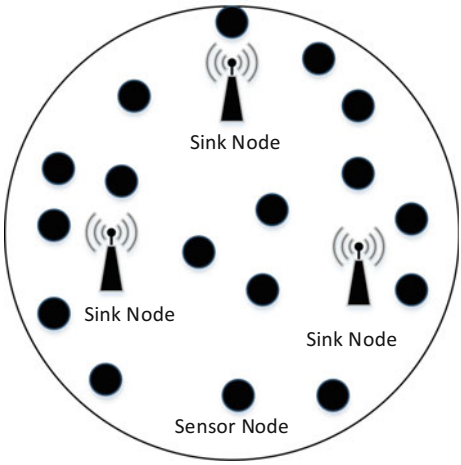



**Fig. 1** System model

nodes through just one hop. The sink nodes collect messages from sensor nodes and act as gateway connecting the WSN and external network.

It should be noted that physical layer authentication is a complementary method to discriminate transmitters, through which we can decrease the work load of up layer authentication schemes and save energy [4]. We assume that the identity of the message is authenticated to be from a legitimate sensor node at the beginning of the transmission by virtue of traditional higher layer authentication schemes.

## *2.1 OFDM Model*

In this paper, we consider an OFDM-based WSN with $N$ subcarriers. The low-pass equivalent OFDM symbol is expressed as

$$x(n) = \frac{1}{N}\sum_{k=0}^{N-1} X(k)e^{j2\pi kn/N}, n = 0, 1, \ldots, N-1. \tag{1}$$

The power of data symbol is $E\left[|X(k)|^2\right] = Var[X(k)] = E_S$ .

The OFDM signal is transmitted over a time-varying frequency selective fading channel, under the assumption that adjacent messages are transmitted within the channel coherence time $T$. So that, the received signal sample in time domain can be written as

$$y(n) = x(n)*h(n) + w(n), \tag{2}$$

where $h(n)$ is the channel impulse response and $w(n)$ is zero-mean white complex addictive Gaussian noise with variance $\sigma_w^2$, i.e. $w(n) \sim N_C(0, \sigma_w^2)$.

## 2.2 Channel Estimation

In the OFDM system, the signal in the receiver will be transferred to frequency domain after precise sampling. In order to estimate the channel, pilot symbols should be inserted into some subcarriers. The received signal in frequency domain can be expressed as

$$Y[k] = H_k X[k] + W_k, k = 0, 1, \ldots, N-1. \tag{3}$$

The subscript means the $k$th subcarrier. The channel frequency response can be estimated by least squares (LS) estimation. In the presence of channel noise, the estimation is corrupted. The MSE of channel estimation is given by [10]

$$MSE^{LS} = \frac{1}{N}\sum_{k=0}^{N-1} E[\left|H_k - \hat{H}_k\right|^2] = \frac{\sigma_w^2}{E_S} = \frac{1}{SNR}, \tag{4}$$

where $SNR = E_S/\sigma_w^2$ expresses the signal noise ratio of the OFDM transmission system.

During the authentication period, the channel estimations are got within the coherent time of the channel. The channel frequency response estimation at subcarrier $k$ is expressed as

$$\hat{H}_i(k) = H_i(k)e^{j\phi_i} + N_i, i = A, T, \tag{5}$$

where subscript $i = A$ denotes the channel frequency response estimation comes from the legitimate sensor node A, while $i = T$ means that the frequency response estimation originates from unknown node that needs to be authenticated.$N_i$ is an additive complex Gaussian error resulting from the receiver thermal noise, which complies for $N_C(0, \sigma^2)$ then $\sigma^2 = MSE^{LS}$. $\phi_i$ is the unknown phase measurement rotation, which complies for uniform distribution $\phi_i \in (0, 2\pi)$.

In this paper, both the amplitude and phase of channel fingerprint will be utilized to clarify the two distinct links [11]. So that, as for the $n$th sink node, the estimation of channel frequency response is

$$\hat{\mathbf{H}}_\mathbf{n} = [\hat{H}_n(k_1), \hat{H}_n(k_2), \ldots, \hat{H}_n(k_M)]. \tag{6}$$

where $k_0$ ,$k_1 \cdots k_M$ are the index of the sample subcarriers and $M$ is the length of channel frequency samples.

## 3 Collaborative Authentication Scheme

In this section, we utilize the channel frequency response got from above and propose a binary hypothesis testing to authenticate the current transmitter. Based on Neyman–Pearson criterion, we make the decision whether the current messages come from the legitimate user or spoofing attacker, and then we collaboratively combine the decisions from different sink nodes to make the final decision to maximize the probability of detecting the attacker.

### *3.1 Hypothesis Testing*

At time $t$, it is assumed that the message originates from sensor node A and the corresponding channel frequency response is $\hat{\mathbf{H}}_{\mathbf{A}}$. At time $t+1$, sink nodes receive another message from a sensor node, we get channel estimation $\hat{\mathbf{H}}_{\mathbf{T}}$. For the $n$th sink node the offset vector $\boldsymbol{\Delta}\hat{\mathbf{H}}_n$ can be defined as

$$\begin{aligned}\boldsymbol{\Delta}\mathbf{H}_n &= \hat{\mathbf{H}}_{\mathbf{T}_n} - \hat{\mathbf{H}}_{\mathbf{A}_n} = \mathbf{H}_{\mathbf{T}_n} - \mathbf{H}_{\mathbf{A}_n} + \mathbf{N}_{\mathbf{T}_n} - \mathbf{N}_{\mathbf{A}_n} \\ &= [\,\Delta\hat{H}_n(k_1) \quad \Delta\hat{H}_n(k_2) \quad \cdots \quad \Delta\hat{H}_n(k_M)\,],\end{aligned} \tag{7}$$

So that, we define $\boldsymbol{\mu}_n = \mathbf{H}_{\mathbf{T}_n}(\mathbf{f}) - \mathbf{H}_{\mathbf{A}_n}(\mathbf{f})$, which is the discrepancy of the original part of the channel frequency response. So that, as for the $n$th sink node, we get the binary hypothesis testing as

$$\begin{aligned}&\mathrm{H}_0: \boldsymbol{\Delta}\mathbf{H}_n = \mathbf{0} \\ &\mathrm{H}_1: \boldsymbol{\Delta}\mathbf{H}_n \neq \mathbf{0}.\end{aligned} \tag{8}$$

Under $\mathrm{H}_0$, which means the latter message comes from legitimate node, while under $\mathrm{H}_1$, which means the latter message is from the attacker.

We get the hypothesis test statistic

$$S_n = \frac{1}{\delta_n^2} \sum_{i=1}^{M} \Delta\hat{H}_n(k_i)^2. \tag{9}$$

From the previous derivation, we know that $\boldsymbol{\mu}_n = \mathbf{0}$ under $\mathrm{H}_0$, so that $S_n$ follows central chi-squared distribution with $2M$ degrees of freedom, which can be represented as $S_n \sim \chi^2_{2M}$. While under $\mathrm{H}_1$, $\boldsymbol{\mu}_n \neq \mathbf{0}$ and $S_n$ follows non-central chi-squared distribution and $S_n$ can be expressed as $S_n \sim \chi^2_{2M}(\lambda)$ with $2M$ degrees of freedom and non-centrality parameter is [12]

$$\lambda = \sum_{i=0}^{M-1} \frac{\boldsymbol{\mu}_n^2}{\delta_n^2}. \tag{10}$$

### 3.2 Collaborative Detection Method

In this section, we use collaborative authentication method to decide whether $H_0$ or $H_1$ is correct. For $n$th sink node, the authentication method is performed by comparing the test statistic $S_n$ and the threshold $T_n$ given by specific false alarm rate. Under $H_0$, we can get the probability density function of $S_n$ [12]

$$P_{H_0}(S_n) = \frac{S_n^{M-1} e^{-\frac{S_n}{2}}}{2^M \Gamma(M)}, \tag{11}$$

where $\Gamma(x) = \int_0^\infty u^{x-1} e^{-u} du$ is Gamma function. Besides, the probability density function of hypothesis $H_1$ is [12]

$$P_{H_1}(S_n) = \frac{1}{2}\left(\frac{S_n}{\lambda}\right)^{\frac{M-1}{2}} e^{-\frac{1}{2}(S_n+\lambda)} I_{M-1}(\sqrt{S_n\lambda}), \tag{12}$$

where $I_{M-1}(\cdot)$ is the first kind Bessel function of order $M$-1.

In order to maximize the detection ability, we use Neyman–Pearson criterion to make the decision. So that, we assume the threshold as $T_n$, then the false alarm rate is

$$P_{FA_n} = P(S_n > T_n | H_0) = \int_{T_n}^{\infty} P_{H_0}(S_n) dS_n = Q_{\chi^2_{2M}}(T_n), \tag{13}$$

where $Q_{\chi^2_{2M}}(\cdot)$ represents the right tail probability for a $\chi^2_{2M}$ random variable. The detailed expression is

$$Q_{\chi^2_{2M}}(T_n) = e^{-\frac{T_n}{2}} \sum_{i=0}^{M-1} \frac{T_n^i}{2^i i!}. \tag{14}$$

According to the Neyman–Pearson criterion, we can get the threshold for given false alarm rate $T_n = Q^{-1}_{\chi^2_{2M}}(P_{FA_n})$. After that, using the threshold $T_n$, we can make decision whether to accept $H_0$ or claim $H_1$. As a result, the detection rate of $n$th sink node is

$$P_{D_n} = P\{S_n > T_n | H_1\} = \int_{T_n}^{\infty} P_{H_1}(S_n) dS_n. \tag{15}$$

As mentioned above, the channel frequency response estimation is corrupted by zero-mean Gaussian error, as a result, the receiver may misidentify the attacker as legitimate user. Then, the receiver will save the false channel frequency response to authenticate the next message, which will probably claim the legitimate user as

attacker. This situation will inevitably degrade the authentication performance and introduce security problems to the network. So that, we choose a strict decision rule to make the final decision.

As described in Sect. 2, we assume that there are $N$ sink nodes deployed in the area and $N_R$ valid sink nodes received the message from the transmitter. So that, the $N_R$ sink nodes considers the current transmitter is legitimate if and only if no sink node claims $H_1$. The final detection rate is

$$P_D = 1 - \prod_{n=1}^{N_R} (1 - P_{D_n}). \tag{16}$$

The overall false alarm rate is

$$P_{FA} = 1 - \prod_{n=1}^{N_R} (1 - P_{FA_n}). \tag{17}$$

In order to evaluate our proposed authentication scheme, as the case in [5], when APs are synchronized, we propose a reference scheme. We assume that once the sink nodes received messages from an unknown sensor node, they will estimate the channel and transmit the channel estimation result to one sink node. We can ignore which sink node the channel vector comes from, and combine them into an extended channel row vector,

$$\hat{\mathbf{H}} = [\hat{\mathbf{H}}_\mathbf{1}, \hat{\mathbf{H}}_\mathbf{2}, \ldots, \hat{\mathbf{H}}_{\mathbf{N}_R}]. \tag{18}$$

It is clear that it is the same case with a single sink node, only with the dimension of channel vector changing from $M$ to $N_R M$. The decision method is similar to single sink node.

## 4 Simulation Analysis

In this section, the performance of proposed enhanced authentication scheme is evaluated by numerical simulations. We consider an OFDM-based wireless sensor network and during the simulation, the OFDM system is employed similar to [13]. The OFDM symbol is modulated with 16-QAM modulation. The number of total subcarriers are 256 and the length of CP are 32. Comb-type pilots with number of 64 are inserted into each OFDM symbol to estimate the channel frequency response. A random Rayleigh channel is utilized with sparse sample-spaced significant taps, uniform power delay profile, where paths on different tap positions are statistically independent with normalized average power. The significant paths connecting the legitimate sensor nodes to sink nodes are assumed to be the same

and the fixed number is 4, while the significant path number between malicious nodes and different sink nodes are randomly chosen from 1 to 6. We use Monte Carlo simulation method and the simulation results are based on 5000 independent realizations of the system.

First, we evaluate the channel estimation performance. From the previous derivation we can see that the channel estimation error is related to the SNR of the communication system. So that, in the simulation, we set SNR of channel as 15 dB. From Fig. 2 we can see that the channel estimation results fit to the true channel frequency response. Unfortunately, we can see from the channel estimation result curve that the channel frequency responses are corrupted.

In order to evaluate our collaborative authentication scheme we set the SNR = 2 and M = 4 to get the receiver operating characteristic curve (ROC). From Fig. 3, we can see that the authentication ability can be improved by utilizing cooperation of multiple sink nodes. The detection rate increases with the number of valid sink nodes. Besides, our proposed scheme has a better performance of detection comparing to the reference method, which views the channel frequency responses of different receivers as one vector to make decision.

As the SNR influence the channel estimation results, we can also find that different SNRs influence the detection rate, so that we evaluate how the detection rate changes with SNRs. Similar to above, we also set M = 4 for each sink node and we set $P_{FA} = 0.001$. Basically, as shown in Fig. 4, the detection rate improves with SNR. What's more, under the same SNR, utilizing multiple sink nodes, the detection rate improves a lot comparing to single node. We can also find that our proposed collaborative authentication scheme has a better performance comparing to the reference scheme, especially when the SNR is low.

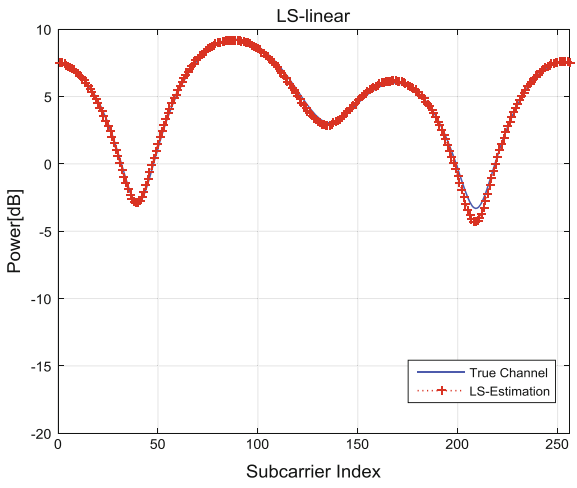



**Fig. 2** Channel estimation results and the true channel frequency response SNR = 15 dB, 4 significant paths

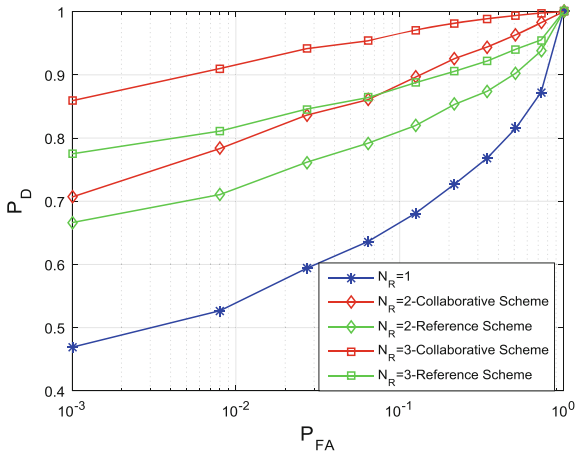



**Fig. 3** $P_D$ versus $P_{FA}$ under M = 4 and SNR = 2 dB

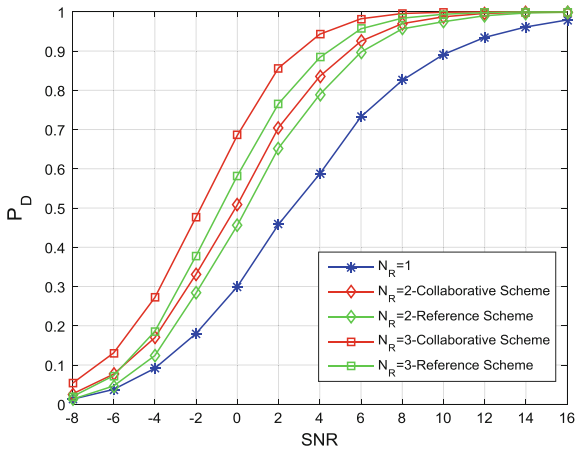



**Fig. 4** $P_D$ versus SNR under M = 4 and $P_{FA}$ = 0.001

## 5 Conclusion

In this paper, we proposed a collaborative authentication scheme for WSN based on collaboration of multiple sink nodes. The binary hypothesis is applied to process the channel fingerprint of frequency response and we use Neyman–Pearson criterion to maximize the detection rate. Through collaboratively combining the detection results from different sink nodes, the detection capability is dramatically enhanced. By means of the simulation, the performance assessment results validated our proposed scheme and showed high authentication accuracy.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (61671075), National Natural Science Foundation of China (61631003), and National Natural Science Foundation of China (61171176).

## References


1. P. Guo, J. Wang, J. Zhu Y. Cheng, Authentication mechanism on wireless sensor networks: a survey (2013)
2. E. Jorswieck, S. Tomasin, A. Sezgin, Broadcasting into the uncertainty: authentication and confidentiality by physical-layer processing. Proc. IEEE **103**, 1702–1724 (2015)
3. M. Demirbas, S. Youngwhan, An RSSI-based scheme for sybil attack detection in wireless sensor networks, in International Symposium on a World of Wireless, Mobile and Multimedia Networks(WoWMoM'06), pp. 5–570 (2006)
4. L. Xiao, L. Greenstein, N. Mandayam, W. Trappe, Fingerprints in the ether: using the physical layer for wireless authentication. IEEE Int. Conf. Commun. 4646–4651 (2007)
5. L. Xiao, L. Greenstein, N. Mandayam, W. Trappe, Channel-based detection of sybil attacks in wireless networks. IEEE Trans. Inf. Forensic Secur **4**, 492–503 (2009)
6. X. Wang, P. Hao, L. Hanzo, Physical-layer authentication for wireless security enhancement: current challenges and future developments. IEEE Commun. Mag. **54**, 152–158 (2016)
7. P. Hao, X. Wang, A. Behnad, Performance enhancement of I/Q imbalance based wireless device authentication through collaboration of multiple receivers. IEEE Int. Conf. Commun. (ICC) 939–944 (2014)
8. P. Pardesi, J. Grover, Improved multiple sink placement strategy in wireless sensor networks. Int. Conf. Futur. Trends Comput. Anal. Knowl. Manag. (ABLAZE) 418–424 (2015)
9. D, Wu, G. Zhu, D. Zhao, L. Liu, Energy balancing in an OFDM-Based WSN, in IEEE 73rd Vehicular Technology Conference (VTC Spring), pp. 1–5 (2011)
10. S. Rosati, G. Corazza, A. Vanelli-Coralli, OFDM channel estimation with optimal threshold-based selection of CIR samples. IEEE Glob. Telecommun. Conf. (2009)
11. F. He, H. Man, D. Kivanc, B. McNair, EPSON: enhanced physical security in OFDM networks. IEEE Int. Conf. Commun. 1–5 (2009)
12. S. Kay, *Fundamentals of Statistical Signal Processing* (PTR Prentice-Hall, Englewood Cliffs, NJ, 1998)
13. F. Liu, X. Wang, S. Primak, A two dimensional quantization algorithm for CIR-based physical layer authentication. IEEE Int. Conf. Commun. (ICC) (2013)

# Key Generation Rate in the Full Duplex Relay Wireless Communication Network

**Lei Chen and Ting Jiang**

**Abstract** Secret key agreement from reciprocal wireless channels has been the hot spot of research which focus on the security at physical layer. In future 5G systems, full-duplex (FD) communication will be one of the key technologies, whose one ability should be security. However, most key agreement models are the half duplex (HD). Due to the FD wireless transmission merit, transmitting simultaneously receives the signal from the other. Therefore, in this paper, we investigate the key generation problem in the FD relay channel, in which there is no direct channel between the key generation terminals. We propose an effective key generation scheme that achieves a substantially larger key rate than that of HD situation. It turns out that, for the application of key agreement, the FD approach enables advantages over the conventional HD setups with less self-interference (SI), but enable advantages under a certain SNR regime while lost the advantages in the high SNR regime and has an upper bound due to the non-negligible SI. Meanwhile, key rate in FD relay network lost its advantage when SI is higher.

**Keywords** Physical layer security · Key generation · FD relay channel

## 1 Introduction

The broadcast nature of wireless communications allows all the users within range to hear the transmission, thus making it vulnerable to various attacks. Exploiting the channel characteristics to ensure wireless network security and extract security key thus has attracted many research interest [1–10, 11]. In [1, 2], the authors summarize the existing works that are based on interpolation, filtering, decorrelation

L. Chen (✉) · T. Jiang (✉)
Key Laboratory of Universal Wireless Communication, Ministry of Education Beijing University of Posts and Telecommunications, Beijing, China
e-mail: buptcl131@bupt.edu.cn

T. Jiang
e-mail: tjiang@bupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_31

and quantization, where they were able to extract secret keys more efficiently. Relay can enhance the security of wireless communication system have shown in [3–5], the existing researches on the physical layer security (PLS) are mostly concentrated on cooperative relay. The key generation from two-way relay channel problem with active attackers was considered in [3, 5], which proposed several interesting schemes to the situation of no direct channel to provide the necessary common randomness. The author in [4] considers the untrusted relays which may connect to third party.

However, most of these works are based on the fact that the nodes who are HD. There are many researches which focus on the PLS for FD nodes [6–8]. The author in [6] consider multi-hop relaying systems, [7] considers a friendly FD relay to increase the secrecy rate. Vogt et al. [9] proposed representations of key generation models in FD modes and analyzed their performances by utilizing the key communication function of secret key agreement, and demonstrate a practical testbed with FD capability in [10]. Inspired by their work, we propose a scheme for the key generation in the FD relay channel. Our work improves the performance of [5]. To the best of our knowledge, rare work has focused on the secret key generation problem under power constraint public communication with FD relay.

The remainder of the paper is organized as follows. In Sect. 2, we introduce the model studied in this paper and discuss the proposed scheme in detail. Simulation results and analysis are presented in Sect. 3. Conclusions are given in Sect. 4.

## 2 System Model and Scheme

### 2.1 System Model

We use the model as shown in Fig. 1. There exists a wireless channel between every pair of terminals in the system except between Alice and Bob. We assume that Alice, Bob are HD nodes with single antenna, while the relay is a FD node with two antennas, one for receiving and the other one for transmitting sinal. More specifically, if Alice transmits signal $X_A$ in a given channel use, at the same time the relay transmits signal $X_{R1}$, thus Relay, Bob and Eve will receive

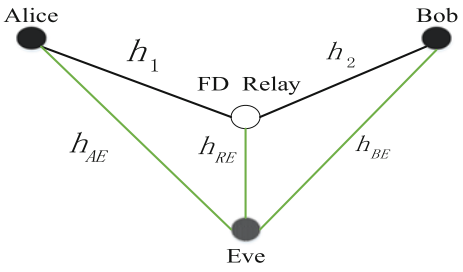



**Fig. 1** Model of FD relay system

$$y_{R1} = h_{AR}X_A + \alpha\sqrt{SNR_{AR}}n_{IA} + n_{R1} \tag{1}$$

$$y_B = h_{RB}X_{R1} + n_B \tag{2}$$

$$y_{E1} = h_{AE}X_A + h_{RE}X_{R1} + n_{E1} \tag{3}$$

respectively, where $h_{AR}$, $h_{RB}$ is the fading coefficient of the channel from Alice to the relay and relay to Bob, $n_{R1}$ and $n_B$ are zero mean Gaussian noise with variance $\sigma^2$ at the relay and Bob, $n_{IA} \sim CN(0, \sigma_{na}^2)$ denote residual SI(RSI) induced by simultaneous transmission and reception at relay, which used in [9, 11]. The parameter $\alpha$ with $0 \leq \alpha \leq 1$ denotes the relation of power levels comparing the RSI and the desired received signal. $h_{AE}$ and $h_{RE}$ are the channel gains between Alice and Eve, relay and Eve, respectively. $n_{E1}$ is the noise at Eve. Moreover, $h_{AR}$, $n_{R1}$, $n_{IA}$, $h_{RB}$, $n_B$ are random variables and independent of each other. No part in the system knows the value of channel gain a priori, but all parts know their distribution. Noise in all channels is independently and identically distributed.

Similarly, when Bob transmits signal $X_B$ and relay transmits signal $X_{R2}$, the relay and Alice receive

$$y_{R2} = h_{BR}X_B + \beta\sqrt{SNR_{BR}}n_{IB} + n_{R2} \tag{4}$$

$$y_A = h_{RA}X_{R2} + n_A \tag{5}$$

$$y_{E2} = h_{BE}X_B + h_{RE}X_{R2} + n_{E2} \tag{6}$$

where $n_{IB} \sim CN(0, \sigma_{nb}^2)$ denotes RSI at relay. The parameter $\beta$ with $0 \leq \beta \leq 1$ is similar to $\alpha$. Besides, the other parameter are similar to the above.

## 2.2 Key Generation Scheme

In this section, we study the key generation problem in HD relay and in FD relay. We assume that all the channels are reciprocal, i.e. $h_{AR} = h_{RA}$ (we denote as $h_1$), $h_{BR} = h_{RB}$ (we denote as $h_2$). But the scheme developed in this paper still works (with a different key) even if this assumption does not exist, as long as the correlation coefficient between the forward and backward channel is not zero [11]. Furthermore, we consider a block fading wireless channel model that means the channel gain remains constant during a period of T, and changes randomly to another independent value in the next one period of T. We assume $h_1 \sim CN(0, \sigma_1^2)$ and $h_2 \sim CN(0, \sigma_2^2)$, which was also used in [5, 11]. We hold that all noises at each terminal are zero mean Gaussian noise with variance $\sigma^2$. Certainly, it will still work even if the channel changes.

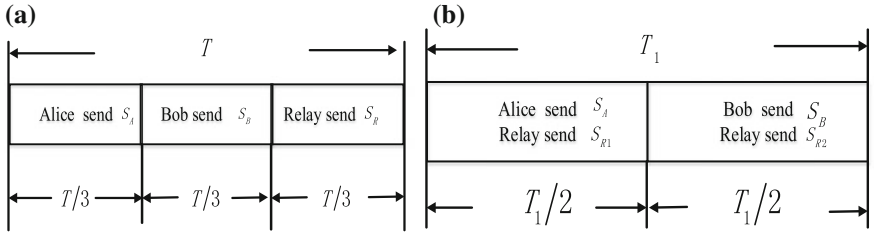



**Fig. 2** **a** Time Frame for HD Relay network. **b** Time Frame for FD Relay network

A. HD Relay Scheme

Figure 2a shows the general time frame for the HD relay network in [5]. Each fading block is divided into three slots which is $T_0$, and we set $T_0 = T/3$. Suppose training sequence Alice sends with power $P_{A,HD}$, Bob with power $P_{B,HD}$ and the relay with power $P_{R,HD}$. Therefore, each training sequence's energy is $\|S_A\|^2 = T_0 P_{A,HD} = TP_{A,HD}/3$, $\|S_B\|^2 = TP_{B,HD}/3$ and $\|S_R\|^2 = TP_{R,HD}/3$, and the total power constraint is

$$P_{A,HD} + P_{B,HD} + P_{R,HD} \le P_T \tag{7}$$

B. FD Relay Scheme

Figure 2b shows the general time frame for the FD relay network. We divide each fading block into two slots each with duration $T_f$, and we set $T_f = T_1/2 = T_0$. Suppose Alice sends training sequence with power $P_{A,FD}$, Bob with power $P_{B,FD}$, and the relay with power $P_{R_1,FD}$ and $P_{R_2,FD}$. Therefore, the energy of each training sequence is $\|S_{R2_{FD}}\|^2 = T_1 P_{R_2,FD}/2$, $\|S_{R1_{FD}}\|^2 = T_1 P_{R_1,FD}/2$,

$\|S_{A_{FD}}\|^2 = T_f P_{A,FD} = T_1 P_{A,FD}/2$, $\|S_{B_{FD}}\|^2 = T_1 P_{B,FD}/2$, and the total power constraint in (7) turns into

$$P_{A,FD} + P_{B,FD} + P_{R_1,FD} + P_{R_2,FD} \le P_T \tag{8}$$

In the first slot, Alice sends a known sequence $S_{A_{FD}}$ to the relay. Meanwhile, the relay sends a known sequence $S_{R1_{FD}}$ to the Bob. The relay and Bob receive

$$Y_R^{(1)} = h_1 S_{A_{FD}} + \alpha\sqrt{SNR_{AR}} N_{IA} + N_R^{(1)}, \tag{9}$$

$$Y_B = h_2 S_{R1_{FD}} + N_B, \tag{10}$$

respectively, then the relay and Bob estimate the channel

$$\tilde{h}_{1,R_{FD}} = \frac{S_{A_{FD}}^T}{\|S_{A_{FD}}\|^2} Y_R^{(1)} = h_1 + \frac{S_{A_{FD}}^T}{\|S_{A_{FD}}\|^2} N_{ZA} \tag{11}$$

$$\tilde{h}_{2,B_{FD}} = \frac{S_{R1_{FD}}^T}{\|S_{R1_{FD}}\|^2} Y_B = h_2 + \frac{S_{R1_{FD}}^T}{\|S_{R1_{FD}}\|^2} N_B \tag{12}$$

where $N_{ZA} = (\alpha\sqrt{SNR_{AR}}N_{IA} + N_R^{(1)}) \sim CN(0, \alpha^2 SNR_{AR}\sigma_{IA}^2 + \sigma^2)$.

In the second slot, Bob sends a known sequence $S_{B_{FD}}$ to the relay. Meanwhile, the relay sends a known sequence $S_{R2_{FD}}$ to Bob. The relay and Alice receive

$$Y_R^{(2)} = h_2 S_{B_{FD}} + \beta\sqrt{SNR_{BR}}N_{IB} + N_R^{(2)}, \tag{13}$$

$$Y_A = h_1 S_{R2_{FD}} + N_A, \tag{14}$$

respectively, then the relay and Alice estimate the channel

$$\tilde{h}_{2,R_{FD}} = \frac{S_{B_{FD}}^T}{\|S_{B_{FD}}\|^2} Y_R^{(2)} = h_2 + \frac{S_{B_{FD}}^T}{\|S_{B_{FD}}\|^2} N_{ZB} \tag{15}$$

$$\tilde{h}_{1,A_{FD}} = \frac{S_{R2_{FD}}^T}{\|S_{R2_{FD}}\|^2} Y_A = h_1 + \frac{S_{R2_{FD}}^T}{\|S_{R2_{FD}}\|^2} N_A \tag{16}$$

where $N_{ZB} = (\beta\sqrt{SNR_{BR}}N_{IB} + N_R^{(1)}) \sim CN(0, \beta^2 SNR_{BR}\sigma_{IB}^2 + \sigma^2)$.

C. General Key Agreement With Relay

Step 1: Alice and the relay agree on a pairwise key $K_1$ using the correlated estimation pair $(\tilde{h}_{1,A}, \tilde{h}_{1,R})$. Step 2: Bob and the relay agree on a pairwise key $K_2$ using the correlated estimation pair $(\tilde{h}_{2,A}, \tilde{h}_{2,R})$. Step 3: The Relay broadcast $K_1 \oplus K_2$. Then Alice and Bob can obtain $K_1$ and $K_2$. They choose the one with smaller size as the common secret key.

D. Analysis of FD relay scheme

In this section, we use the key agreement shown in Sect. 2(C).Using the result from [5], the relay and Alice can establish a pairwise key $K_1$ with a rate from (11) (16):

$$\frac{1}{T_1} I(\tilde{h}_{1,A_{FD}}; \tilde{h}_{1,R_{FD}}) \tag{17}$$

Note that scalar $\tilde{h}_{1,A_{FD}}$ and $\tilde{h}_{1,R_{FD}}$ are two correlated Gaussian variables with zero mean, thus we have [5]

$$I(\tilde{h}_{1,A_{FD}};\tilde{h}_{1,R_{FD}}) = -\frac{1}{2}\log(1-\rho_1^2) \tag{18}$$

where $\rho_1$ is the correlation coefficient of $\tilde{h}_{1,A_{FD}}$ and $\tilde{h}_{1,R_{FD}}$, we can easily get

$$\rho_1^2 = \frac{\mathrm{cov}^2(\tilde{h}_{1,A_{FD}},\tilde{h}_{1,R_{FD}})}{Var(\tilde{h}_{1,A_{FD}})\bullet Var(\tilde{h}_{1,R_{FD}})} = \frac{1}{(1+\frac{\alpha^2 SNR_{AR}\sigma_{IA}^2+\sigma^2}{\sigma_1^2 T_f P_{A,FD}})(1+\frac{\sigma^2}{\sigma_1^2 T_f P_{R2,FD}})}.$$

where $\mathrm{cov}(x,y)$ means covariance between $x$ and $y$, $Var(x)$ means the variance of $x$. Similarly, the relay and Bob establish a pairwise key with a rate $K_2$ from (12) (15)

$$\frac{1}{T_1}I(\tilde{h}_{2,B_{FD}};\tilde{h}_{2,R_{FD}}) = -\frac{1}{2T_1}\log(1-\rho_2^2) \tag{19}$$

$$\rho_2^2 = \frac{\mathrm{cov}^2(\tilde{h}_{2,B_{FD}},\tilde{h}_{2,R_{FD}})}{Var(\tilde{h}_{2,B_{FD}})\bullet Var(\tilde{h}_{2,R_{FD}})} = \frac{1}{(1+\frac{\beta^2 SNR_{BR}\sigma_{IB}^2+\sigma^2}{\sigma_2^2 T_f P_{B,FD}})(1+\frac{\sigma^2}{\sigma_2^2 T_f P_{R1,FD}})}.$$

Thus the final key is

$$R_{final_{FD}} = \frac{1}{T}\min\{I(\tilde{h}_{1,A_{FD}};\tilde{h}_{1,R_{FD}}), I(\tilde{h}_{2,B_{FD}};\tilde{h}_{2,R_{FD}})\} \tag{20}$$

The improvement ratio is

$$\phi = (R_{final_{FD}} - R_{final_{HD}})/R_{final_{HD}} \tag{21}$$

The upper bound for FD relay due to the RSI which is not zero when the SNR tends to be infinite $\rho_{3,up}^2 = \frac{1}{(1+\frac{\alpha^2}{\sigma_1^2 T_f})}$, $\rho_{4,up}^2 = \frac{1}{(1+\frac{\beta^2}{\sigma_2^2 T_f})}$, So (20) turns into

$$R_{final_{FD,up}} = -\frac{1}{2T_1}\min\{\log(1-\rho_{3,up}^2),\ \log(1-\rho_{4,up}^2)\}. \tag{22}$$

## 3 Simulations and Analysis

In this section, we compare the performance of our study of FD relay with the research in [5]. We assume that $T = 6$ s, $\alpha = \beta$, $\sigma_{IA}^2 = \sigma_{IB}^2 = \sigma^2$. For Fig. 3a, b, we set $P_{A,HD} = P_{B,HD} = P_{R,HD} = P_T/3$, $P_{A,FD} = P_{B,FD} = P_{R1,FD} = P_{R2,FD} = P_T/4$. Thus there is equal total energy and transmit power for all nodes.

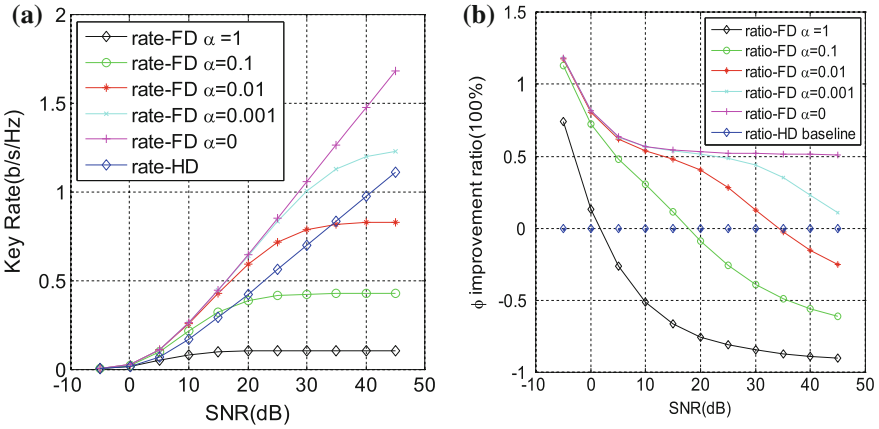



**Fig. 3** **a** Key rates of FD compare with HD versus SNR, 3 **b** Improvement ratio of FD compare with HD versus SNR

From Fig. 3a, we can see that under the power constraint, when the RSI is relatively small, the gain for FD relay has better performance than HD relay, and when SI is perfectly cancelled, the FD have more than 50% improvement than HD relay. While RSI is not small enough, the gain for FD relay have advantage over the HD relay under a certain SNR, but lost its advantage when SNR exceeds a certain threshold. Meanwhile, the key rate increases with SNR, but reach a certain upper bound when SNR is high because of non-negligible RSI.

Figure 3b shows the improvement ratio $\phi$ in (21), we can see that under the power constraint, when the ratio is relatively high in low SNR, and decrease with the SNR due to the RSI. While the SI is perfect cancelled, the ratio is reduced to a certain value with SNR. While RSI is not small enough, the ratio is less than 0, which means lost advantage over HD relay network.

## 4 Conclusions

In this paper, we compare secret-key rate from reciprocal wireless channels for two-way relay with HD and FD capabilities. We proposed an effective key generation scheme for FD relay, which has better performance than the HD relay's under the constraint of RSI when the SNR is not high. In addition, we have analyzed the performance in no SI in particular. The results show an improvement in secret key rate in FD relay over HD two-way relay under a certain SNR when SI is not perfect cancelled, while have more than 50% improvement when SI is perfect cancelled. In the future, we will extend our work in hybrid FD/HD relays network and in MIMO FD network.

**Acknowledgements** This work was supported by National Natural Science Foundation of Chi-na (61671075), National Natural Science Foundation of China (61631003), and National Natural Science Foundation of China (61171176).

## References

1. R. Guillaume, C. Zenger, A. Mueller, C. Paar, A. Czylwik, Fair comparison and evaluation of quantization schemes for phy-based key generation, in *Proceedings of 19th International OFDM Work-shop 2014 (InOWo'14)*, Aug 2014
2. S. Gopinath, R. Guillaume, P. Duplys, A. Czylwik, Reciprocity enhancement and decorrelation schemes for PHY-based key generation, in *IEEE Globecom Workshops (GC Wkshps)*. pp. 1367–1372, Dec 2014
3. T. Shimizu, H. Iwai, H. Sasaoka, Physical-layer secret key agreement in two-way wireless relaying. IEEE Trans. Inf. Forensics Secur. **6**(3), 650–660, Sept 2011
4. R. Guillaume, S. Ludwig, A. Müller, A. Czylwik, Secret key generation from static channels with untrusted relays. IEEE WiMOB, pp. 635–642, Oct.2015
5. H. Zhou, L.M. Huie, L. Lai, Secret Key generation in the two-way relay channel with active attackers. IEEE Trans. Inf. Forensics Secur. **9**(3), 476–488 (2014)
6. J.-H. Lee, Full-duplex relay for enhancing physical layer security in multi-hop relaying systems. IEEE Commun. Lett **19**(4), 525–528, Apr 2015
7. S. Parsaeefard, T. Le-Ngoc, Secrecy rate with friendly full-duplex relay, in *IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1135–1140, Mar 2015
8. S. Parsaeefard, T. Le-Ngoc, Full-duplex relay with jamming protocol for improving physical-layer security, in *IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*, pp. 129–133, Sept 2014
9. H. Vogt, A. Sezgin, Full-duplex vs. half-duplex secret-key generation, in *IEEE International Workshop on Information Forensics and Security (WIFS)*, pp. 1–6, Nov 2015
10. H. Vogt, K. Ramm, A. Sezgin, Practical secret-key generation by full-duplex nodes with residual self-interference, in *20th International ITG Workshop on Smart Antennas (WSA)*. pp. 1–5, Mar 2016
11. A. Sadeghi, M. Zorzi, F. Lahouti, Analysis of key generation rate from Wnel in-band full-duplex communications, *IEEE International Conference on Communications Work-shops (ICC Wkshps)*, pp. 104–109, May 2016

# Performance Analysis for the Improved Topology Updating Mechanism for ZigBee Networks in 5G

**Saichao Li and Jiasong Mu**

**Abstract** The ZigBee network is widely studied and deployed recently because its low cost and simplicity features. However, the power consumption issue needs a further improvement since the application requirements are not fully satisfied. The emerging 5G communication technology is characterized by the smarter devices and the native support for the M2 M communication. On that basis, the 5G terminals are capable of joining the existing ZigBee networks and have the potential to improve the data transmission. In this paper, we investigate the performance of the ZigBee networks in the 5G environment for different scenarios. To make the 5G devices optimize the communication, the improved topology updating mechanism for 2 scenarios are investigated. The performances are evaluated based on the simulation results, it is shown that our scenarios lead to better performances with higher packet delivery ratio, less hop counts from ZigBee devices and fewer packets sent per ZigBee node. And the effect on 5G node mobility is also studied.

**Keywords** ZigBee · 5G · Routing · Performance analysis · Mobility

## 1 Introduction

ZigBee technology is characterized by low cost, low power, low data rate, and simplicity [1]. Although, there are two options for the ZigBee routings, the ZigBee tree routing (ZTR) and ad hoc on demand distance vector (AODV) [2] designed for different requirements. Limited by the insufficient resources in the ZigBee devices, the described routing protocols cannot fully address power consumption issues [3]. Besides optimizing the current network formation and protocols, it is another way that to improve the performance with the help of the deployed environment, which includes the coming 5G communication network.

S. Li · J. Mu (✉)
College of Electronic and Communication Engineering,
Tianjin Normal University, Tianjin, China
e-mail: mujiasong@aliyun.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_32

Due to the extremely higher aggregate data rates and the much lower latencies required, along with other breakthrough technologies, the smarter devices and the native support for the M2 M communication are two of the most important features in the 5G communication network [4]. The smarter devices will be able to have access to different wireless technologies at the same time. Meanwhile, a native inclusion of M2 M communication in 5G involves satisfying three fundamentally different requirements associated with different classes of low data rate services: support of a massive number of low rate devices, sustaining a minimal data rate in virtually all circumstances, and very low latency data transfer.

A sample deployment is shown in Fig. 1. Thanks to the inclusion of M2 M communication function, the 5G terminals which locate within the coverage of the ZigBee networks are able to join the existing ZigBee networks. Compare with the ZigBee nodes, owing to the adequate energy, storage, processing, and bandwidth resources in the mobile devices; they have the potential to improve the data transmission in the ZigBee network. Different from the existing relay or external node, note that the 5G terminals could also access the IP network. That means they are capable of communicating with the sink and management device directly without consuming the ZigBee bandwidth, and the ZigBee packets can be gathered and conveyed in the internet links. Our work is to develop a routing scheme for 5G devices to share the traffic in ZigBee network as much as possible so that the overhead in the ZigBee networks can be lightened effectively.

To the best of our knowledge, this is the first work on the performance analysis for ZigBee networks in the 5G environment. Although the 5G terminal could improve the routing discovery, the related links may not last long because its location keeps changing. Moreover, for the large-scale networks or the nodes with mobility, ZTR has a much better performance [5]. Thus, in this paper, we exploit the ZTR-based method to find the paths via 5G devices.

The remainder of this paper is organized as follows: Sect. 2 introduces some preliminary knowledge in the ZigBee networks. Two scenarios of the heterogeneous networks are described and analyzed in the Sect. 3. Section 4 shows the simulation results and the evaluation of our methods. The conclusion is given in Sect. 5.

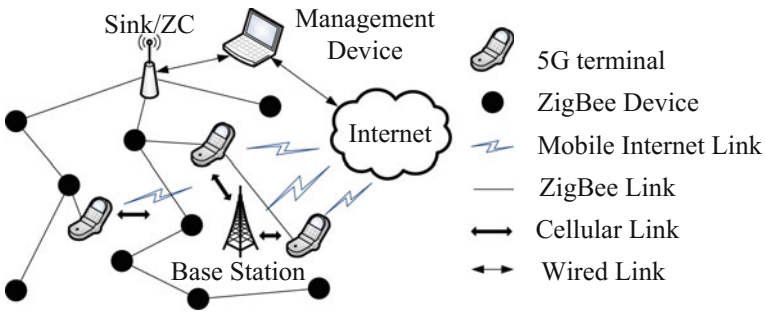



**Fig. 1** The ZigBee network in 5G environment

## 2 ZigBee Networks

### 2.1 Overview of ZigBee

Three device types are defined in the ZigBee specification: ZigBee coordinator (ZC), ZigBee routers (ZR), and ZigBee end devices (ZED). The ZC, which usually works as the sink node and is connected to the management device, is responsible for starting a new network. The ZR, as well as the ZC, is routing capable, while the ZED cannot take part in routing and have to rely on their corresponding ZigBee parent nodes for data transmission. The deployed ZigBee devices self-organize the network, and then the topology changes can be automatically reflected in the network configuration [1].

ZigBee devices support the link quality indication (LQI) measuring whenever they receive a packet. The LQI measurement is a characterization of the strength and/or quality of the frame, the acquiring costs no more extra calculation and communication [1].

### 2.2 Neighbour Table

Based on the ZigBee specification, every ZigBee device maintains a neighbour table, which contains the information of all the nodes in its one-hop neighbourhood. Each neighbour entry should include the network's PAN (Personal Area Network) identifier; node's extended address, network address, device type, relationship, LQI and etc. Additional information such as the depth is also allowed to be added. Entries in the table are created/removed when the node joins/leaves to an existing network. Since the information on the neighbour table is updated every time a device receives any frame from a certain neighbour, the neighbour table can be regarded up-to-date all the time.

## 3 Two Scenarios and the Topology Updating Mechanism

One certain device may have one of the three roles (ZC, ZR or ZED) based on the ZigBee specification. To treat a 5G terminal as an end device is neither efficient nor meet the application requirements. Thus, we have 2 remaining options: to make 5G terminals work as coordinators or routers. We are going to dig up the performances in both scenarios.

- **Scenario 1 (S1)**: **5G Terminals Work as Coordinators**

In this strategy, each 5G terminal, as a ZC, may initialize a unique network. Since the existing ZigBee nodes are not aware of the new network which works in a

different channel, we develop a simple mechanism to make the 5G terminal "visible" for the ZigBee devices.

The 5G terminal acts its ZigBee role starting from the channel detection, during which it tries to find the existing ZigBee PAN and the available channel. If it finds a channel occupied, the channel information is restored. Before performing its ZC duty in a selected channel, the 5G device is required to announce its existence along with its working band in all the occupied channels by broadcasting a notification. When a ZigBee node receives the frame, it accordingly shifts its band after a random time with a preset expectation, and then inquires about its distributed address. The existing nodes decide to join the new networks if a network address with lower depth is provided. Afterwards, the recruit devices broadcast the notification with new PAN and channel information in their original bands. The joining decision-making and announcing procedure repeats if needed so that the new PAN is visible to all the ZigBee devices around its coverage.

To fully explore the potential of the performance improvement, in this scenario, we allow a certain ZigBee node communicates in different PANs by in time channel switching with prior knowledge, which violates the ZigBee specification. The aim of this scheme is to keep the existing links in the initial ZigBee network and prevent the network segmentation caused by the disconnection when some nodes leave the current PAN [33].

By above settings, the scenario may result in a less topology distance to the ZC for some ZigBee nodes and the data transmission benefits from the lower ZTR hops.

- **Scenario 2 (S2)**: **5G Terminals Work as Routers**

On the contrast, 5G nodes join the network as ZRs in this section. When successfully detected the existing ZigBee networks function as routers based on the acquired channel and PAN ID information. Note that this process does not instinctively improve the data transmission because the other ZigBee devices will not utilize the links involving the new participants until the network topology changes.

To make that happen, we also design a simple scheme. The 5G devices are asked to notify its existence and identities to its neighbours periodically. When a ZigBee unit receives the notification, it compares its current depth with the new one linked with 5G device. It updates its network address in the case of a lower depth can be obtained and announcing its address modification. The nodes in the neighbourhood follow the same principle to decide whether they change their addresses [34]. By the explained method, the links via the 5G terminals may be used if it benefits the packet forwarding.

## 4 Simulations and Evaluation

Based on the mentioned designs, the simulations on different scenarios of ZigBee networks with 5G devices are carried out to evaluate their performances. The general network settings for NS-3 are shown in Table 1. To analyze the effect of the

**Table 1** Simulation parameters

| Simulation parameters | Value |
|---|---|
| Network size | 100 m * 100 m |
| Simulation duration | 350 s |
| Association duration | 50 s |
| Number of ZigBee nodes | 200 |
| Deployment | Random (Uniform) |
| Transmission range | 25 m |
| Propagation | Two-ray |
| Data packet size | 1 KB |
| Packet interval | 1packet/s |
| Packet start time | 50–100 s (Uniform) |
| Destination address | Random |
| Number of 5G nodes | 0, 5, 10, 15, …, 50 |
| First appearance | 50–100 s (Uniform) |
| Trajectory | Straight line |
| Average speed | 1.5, 5, 10, 15, 20 m/s |

number of the joining 5G nodes, the number of exotic devices varies from 0 to 50 (step by 5). The standard ZigBee configuration in which the 5G devices only join the ZigBee network as determined device type without any further functions is compared with the S1 and S2 as a reference. To emulate the actual ZigBee application, the destination address is randomly chosen. The packet is transmitted to the ZC in 50% chance, and in another 50%, the destination is selected from other devices in equal probability. Moreover, the packet interval for the ZC is 10 packets per second. The mentioned 3 types of frames stand for the sensing data, the general communication, and the command and control packets respectively. Considering the mobility of 5G nodes, the following mechanism is adopted. Each 5G device begin to take part in the simulation at a random time from 50 to 100 s after the simulation starts. It first appears at a random position on the boundary of the simulation square. The initial moving direction is going into the simulation area and the trajectory is a straight line. In our first simulation, the average speed is 1.5 m/s which is approximately the walking speed of an adult. When reaching the boundary, the 5G node has a 50–50 chance to reflect the trajectory with the same entry and exit angles or escape from the simulation area. In the latter case, a new 5G node appears at a random position on the boundary and act as the mentioned procedure.

We firstly investigate the network performance including the packet delivery ratio, average hop counts and the end-to-end delay, which are shown in Fig. 2. Generally speaking, compared with the standard ZigBee specification, both the two scenarios improve the data transmission as the number of 5G nodes increases and S1 outperforms S2. As shown in Fig. 2a, the packet delivery ratio degenerates as more nodes participate in the network. However, the PDR of S1 and S2 decrease slower and have a similar performance. With regards to the hop counts in Fig. 2b, the 5G nodes may not optimize the transmission link based on the ZigBee

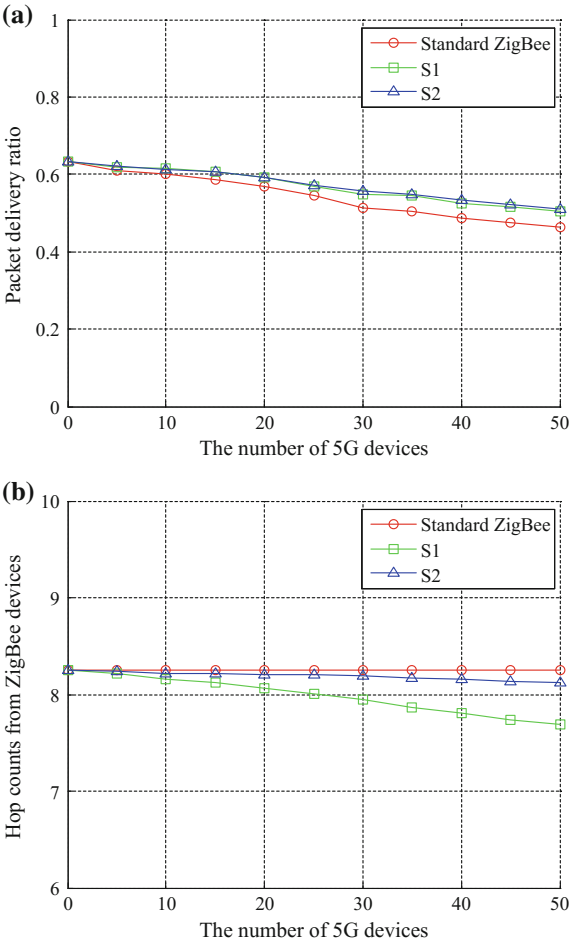



**Fig. 2** Performance evaluation for different scenarios **a** packet delivery ratio **b** hop counts

specification, the hop counts for standard ZigBee remain in a similar level, and the updating mechanisms in S1 and S2 enable the topology updating and result in less hops. Since we focus on the performance of ZigBee devices, only the transmission starts from a ZigBee node is counted (which means that the data sending from a 5G terminal is excluded in the hop counting). S2 slightly reduces the hop counts while S1 has a greater effect.

Although the improvement is limited because the 5G devices only function as common ZigBee devices and their abundant resources are not exploited. The performance analysis shows that our scenarios improve the data transmission in ZigBee networks. S1 and S2 both have better PDRs and less hop counts. The multiple ZCs in S1 lead to greater improvement than S2. Besides its overhead may be too large for ZigBee devices, some other issues also make S1 hard to realize. First, as mentioned, a ZigBee device cannot work in different bands simultaneously. The channel switching seems like a rational solution, but the cycling interval is

impossible to precisely determined, to say nothing of the related energy consumption. Second, the co-existence of different networks may introduce unnecessary inter-channel interference (ICI). Especially in the case of a large amount of 5G terminals, the most legal channels are occupied and the ICI tends to increase. Finally and most importantly, one of the inherent features of 5G device is its mobility. In our previous simulations, the 5G device mobility is low; Fig. 3 shows the performances of different scenarios with growing mobility of 5G devices (The number of 5G devices is fixed 50). It can be seen that the performance of S1 quickly falls and the overheads are also increasing rapidly. And S2 is less sensitive to changes in moving speed. In case of the 20 m/s average speed (the approximate speed of moving vehicles), S2 has a similar PDR as the standard ZigBee while the performance of S1 is about 12.93% worse. The hop counts of 3 scenarios tend to be equal, that is mainly because the high mobility reduces the

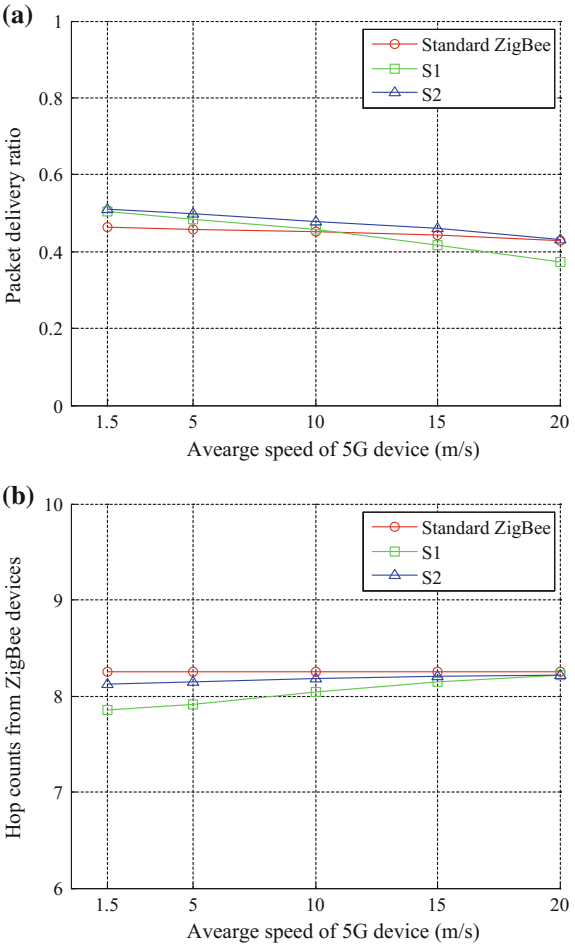



**Fig. 3** Performances on different scenarios with different 5G node mobility **a** packet delivery ratio **b** hop counts

duration of better links. The moving 5G terminal may be within the coverage of the ZigBee networks for only a short while; such an ephemeral network lifetime may extremely declines the benefit. Since the ZC is always regarded fixed in the ZigBee network, the mobile ZC may bring about the frequently changes of the network address for its descendants, the latency, packets sent per ZigBee nodes and the routing overhead are all increasing as the mobility adds, and the results for S1 are much worse than S2. They are unaffordable overheads for the ZigBee nodes with limited resources.

## 5 Conclusions

In this paper, we evaluate the performances on ZigBee networks in 5G environment by adopting the topology updating mechanism we designed. Based on our simulation, it is shown that the 5G devices have the potential to improve the communication in ZigBee networks. By comparing the two mentioned scenarios, the ZR device type is considered a more practical scheme.

For our further work, the routing cost is going to be studied to make a fair comparison. Furthermore, we are going to develop the improved algorithm based on S2 to optimize the path and data transmission by utilizing the resources in 5G devices.

**Acknowledgements** This work is funded by the National Science Foundation of China, (NSFC: 61401310).

## References

1. ZigBee Standard Organization, *ZigBee Specification Document 053474r17* (ZigBee Alliance, San Ramon, 2007)
2. C.E. Perkins, E.M. Royer, Ad-hoc on-demand distance vector routing, in *The Workshop on Mobile Computing Systems & Applications*, pp. 90–100 (1999)
3. H.Y. Tung et al., The generic design of a high-traffic advanced metering infrastructure using ZigBee. IEEE Trans. Ind. Inf. **10.1**, 836–844 (2014)
4. F. Boccardi, R.W. Heath, A. Lozano, T.L. Marzetta, Five disruptive technology directions for 5G. IEEE Commun. Mag. **52**(2), 74–80 (2014)
5. L.J. Chen, T. Sun, N.C. Liang, An evaluation study of mobility support in ZigBee networks. J. Signal Process. Syst. **59.1**, 111–122 (2010)

# Part III
# Wireless Networks

# The Optimization and Research of Primitive HBase Data Storage Based on Wireless Sensor Network

**Xiang Li and Huazhi Sun**

**Abstract** With the spread of wireless sensor network, the increasing number of sensors and the cross-regional spread of wireless sensor network, a large member of sensor data is being produced. Therefore, this paper aims to improve storage and query efficiency based on the data storage structure characteristics of distributed file system HBase. Extenics primitives is used to integrate the heterogeneous data sets which stored in the database of HBase, cross-regional wireless sensor network data and global data storage management directory of double layer distributed storage architecture, thus enhance the efficiency of storage and access. And this paper realize the real-time memory system of data by simulation.

**Keywords** Wireless sensor network · HBase · Primitive

## 1 Introduction

With the development of Internet of Things in China, there are problems for large amount of information in Information technology and as the information storage of the end Internet of Things, wireless sensor network spreads more widely with larger scale and sends greater amount of information [1, 2]. The existing database based on relationship and object model, the storage of complicated data can be solved, but the cost of storage and system consumption are larger [3, 4]. Wireless sensor network has the high real-time requirements for data analysis, but in practical application, many deployment wireless sensor networks lead the current processing

X. Li · H. Sun (✉)
College of Computer and Information Engineering,
Tianjin Normal University, Tianjin 300387, China
e-mail: sunhuazhi@mail.tjnu.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_33

of sensor data [5, 6]. Therefore, regional storage cluster need to be built. Every region has its own storage cluster of data and the global data memory should be built to save information of every region.

## 2 Wireless Sensor Network

The data storage of wireless sensor network is divided into own internal data storage and external application data storage and wireless network is made by sensors and sink nodes [7, 8]. Through the wireless network communication, a large number of sensor nodes deployed within or near monitoring area and a network can be constituted by self-organization [9, 10]. Wireless sensor network collects data collected by sensors through gathering node, and gathering nodes realize the interaction between data and application through gateway [11, 12]. At the same time, wireless sensor network has scale and intensive in deployment, which is the thousands of sensor nodes are deployed within the wider monitoring areas, so as to get large amount of information; and thousands of sensor nodes are deployed within the narrower monitoring areas, so as to get accurate data.

## 3 Primitive HBase Database Storage

Extenics argues that any material things and relationships in the objective world of can be expression by using primitive formalization, computer can be intelligent after dealing with the contradiction between primitive extenics, and solving primitive data storage is the premise to realize intelligent.

Definition: set $U = \{U_1, U_2, \ldots, U_n\}$ as the research object collection (including the collection of materials, things and relationship), set $O = \{O_1, O_2, \ldots, O_m\}$ as the characteristics collection of research object, the value is $Va_{ij}$ of object $U_i$ about characteristic $O_j$, in which the value range of $ij$ is $i = 1, 2, 3, \ldots, n$; $j = 1, 2, 3, \ldots, m$, so $Va_{ij}$ is called $i$ as primitive and $j$ as dimension primitive. When the value of $Va_{ij}$ is empty, the object $U_i$ has no data relationship of corresponding characteristic $O_j$.

HBase is a storage system which is distributed, column-oriented and is suitable for unstructured data storage and can provide algorithm of real-time write and read and random access [13, 14]. The storage system of HBase is column-oriented in the form of tables and aromatically divides the tables into rows and column. Every region contains a subset of all rows of tables, which is identified by Timestamp. The master nodes are responsible for guiding and coordinating multiple regions server, and responding to the clients' requests of reading and writing.

$$Va_{ij} = \begin{bmatrix} & O_1 & O_2 & \dots & O_n \\ U_1 & Va_{11} & Va_{12} & \dots & Va_{1n} \\ U_2 & Va_{21} & Va_{22} & \dots & Va_{2n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ U_m & Va_{m1} & Va_{m2} & \dots & Va_{mn} \end{bmatrix} \quad (1)$$

## 4 The Optimization of Primitive HBase Data Storage

For the technology of Primitive HBase data storage specific optimization, data sharding and data distribution are the two ordinary optimization strategies, this paper chooses the latter, though data sharding unified manage storage all types of data, solving the congestion problem when a single queue is used to accept data, it neglects the different patterns among the multi-source data, impeding the efficiency of inputting data and stream query efficiency. Whereas data distribution optimization is used, consistent hashing algorithm can conquer the monotony of the traditional Hash data distribution algorithms, and solve the problem of data server load is too high [15, 16]. Meanwhile, the adoption of writing parallel data based on the Multi-Threading technology from the actual demand of the development of data communication software can better solve the problem of delay processing of data communication.

HBase store data in the form of tables, each table consists of rows and columns, each column belongs to a particular column family—determined by rows and columns in the table is called an element, the storage unit of each element to save multiple versions of the same data, identified by timestamp—Table 1 illustrates the web www.cnn.com data storage logical view.

In addition to the logical storage model, Table 2 shows the physical storage model of HBase, HBase physical model is stored in the column of sparse matrix row/column, is actually a line integral of the conceptual model and stored in accordance with the column.

**Table 1** The logic view of data storage

| Line keyword | Time stamp | Contents | Anchor | Mime |
|---|---|---|---|---|
| *com.cnn.* *www* | T9 | | "Anchor:cnnsi.com" "CNN" | |
| | T8 | | "Anchor:my.look.ca" "CNN.com" | |
| | T6 | "<html> …" | | "text/html" |
| | T5 | "<html> …" | | |
| | T3 | "<html> …" | | |

**Table 2** The physical storage form of data

| Line keyword | Time stamp | Contents |
|---|---|---|
| *com.cnn.www* | T5 | "<html> …" |
| | T4 | "<html> …" |
| | T3 | "<html> …" |
| Line keyword | Time stamp | Anchor |
| *com.cnn.www* | T8 | "Anchor:cnnsi.com" "CNN" |
| | T7 | "Anchor:my.look.ca" "CNN.com" |
| Line keyword | Time stamp | Anchor |
| *com.cnn.www* | T6 | "text/html" |

## 5 Performance Analysis

Considering that most of the operation is writing and inquiry in the actual application, the optimization of reaction time in this paper is based on the main performance analysis about the HBase data storage in this paper. HBase is column-oriented database, the applying column is the primary key. And column alignment according to the dictionary sequence acquiescently when writing data, the index is built according to $B^+$ tree on the data column, therefore, time complexity of data writing queries is log $N$. When in queried access, clients will firstly accesses the major nodes, then responses to the write and read requests of client through the guidance of major nodes and the coordination of area servers. Therefore the time of inquired access is composed by the major nodes time $T_{master}$ and the guidance of major nodes and coordinating area server time $T_{Area\,Server}$. Set the quantity of inquiry data as $i$ column, the number of area servers as $m$ coordinated by the major nodes, the storage data in every area server as $j$ column, the length of major nodes' content as $\frac{i}{j}$, the length of saved data in area server as $\frac{i}{j*m}$, the total time is $T_{Total}$. Thus, the data access time is obviously shortened after optimizing using primitive HBase data storage.

$$T_{Total} = T_{master} + T_{Area\,Server} = \log(\frac{i}{j}) + \log(\frac{i}{j*m}) < \log(i) \tag{2}$$

In order to verify the efficiency of storage and query of HB and extensibility, using 300 sensor nodes, is single and configuration of five node of the cluster to write and read time, Fig. 1. shows a single cluster and its relation to the sensor nodes is write time. Can be seen from the diagram, the cluster write time significantly lower than the single write time, is efficient to write. Figure 2. shows a single cluster and its relation to the number of sensor nodes access time. Can be seen from the diagram, along with the increase of the number of sensors, the cluster read time significantly lower than the single load time, have the property of efficient reading.

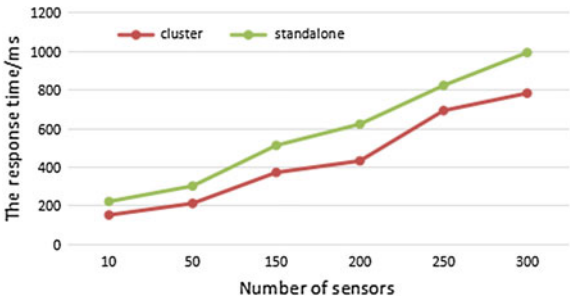



**Fig. 1** The comparison of number of sensors and the writing time

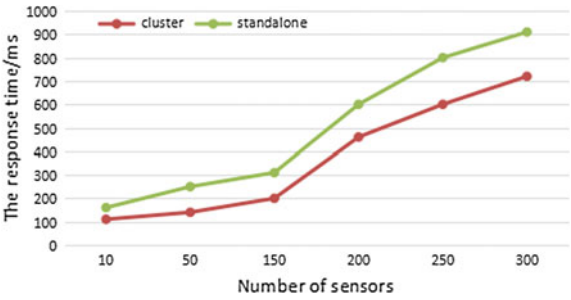



**Fig. 2** The comparison of number of sensors and the read time

## 6 Conclusion

This paper analyses the storage models of big data in the HBase database under the wireless sensor internet environment, offers distributed memory architecture and integrates the pre process flow data. This paper, parallel and equal, stores data into the HBase cluster server by using partition buffer segmentation of different types, optimization design strategy of time point data, which passes the test of the storage efficiency of high efficiency, and the access efficiency.

## References

1. K. Sun, D. Wu, MPC-based delay-aware fountain codes for live video streaming, in *IEEE International Conference on Communications ICC* (2016)
2. R.O. Schmidt, Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag. **34**(3), 276–280 (1986)
3. K. Sun, H. Zhang, D. Wu, H. Zhuang, MPC-based delay-aware fountain codes for real-time video communication. IEEE Internet Things J. (IoT) (2016)
4. N. Tayem, H.M. Kwon, L-shape 2-dimensional arrival angle estimation with propagator method. IEEE Trans. Antennas Propag. **53**(5), 1622–1630 (2005)
5. X. Liu, Z. Li, Information-centric mobile ad hoc networks and content routing: a survey. Article in press (2016)
6. M. Gruteser, D. Grunwald, Anonymous usage of location based services through spatial and temporal cloaking, in *ACM/USENIX MobiSys* (2003)

7. J.F. Gu, W.P. Zhu, M.N.S. Swamy, Joint 2-D DOA estimation via sparse L-shaped array. IEEE Trans. Signal Process. **31**(5), 1171–1182 (2015)
8. K. Sun, D. Wu, Video rate control strategies for cloud gaming. J. Vis. Commun. Image Represent. (2013)
9. Z. Li, Y. Chen, H. Shi, K. Liu, NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. (2016)
10. K. Sun, B. Yan, H. Gharavi, Low complexity content-aware image retar-geting, in *IEEE International Conference on Image Processing (ICIP'2012)* (2012)
11. K. Sun, D. Wu, Video rate control strategies for cloud gaming. J. Vis. Commun. Image Represent. (2013)
12. Z. Li, K. Liu, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11) (2014)
13. T.Q. Xia, Y. Zheng, Q. Wan, X.G. Wang, 2-D angle of arrival estimation with two parallel uniform linear arrays for coherent signals. IEEE Radar Conf. **55**(9), 244–247 (2007)
14. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Article in press (2016)
15. M. Tang, Q. Wu, G. Zhang, L. He, A new scheme of LBS privacy protection. IEEE (2009)
16. K. Sun, B. Yan, Efficient P-frame complexity estimation for frame layer rate control of H.264/AVC", in *IEEE International Conference on Image Processing (ICIP'2011)* (2011)
17. G. Ghinita, P. Kalnis, S. Skiadopoulos, PRIVE: anonymous location based queries in distributed mobile systems, in *Proceedings of International Conference on World Wide Web (WWW '07)* (Banff, Alberta, Canada, 2007), pp. 1–10

# Ant-Colony Based Double Cluster Heads Adaptive Periodic Threshold-Sensitive Energy Efficient Network Protocol in WSN

**Jinyu Ma, Shubin Wang and Yanhong Ge**

**Abstract** In wireless sensor networks, the critical issue is the energy efficient utilization of sensor nodes, which can be enhanced through employing clustering techniques. But cluster head (CH) performs intensive assignments, brings about unbalanced energy dissipation and premature death of nodes. An ant-colony based Double Cluster heads Adaptive Periodic Threshold-sensitive Energy Efficient Network protocol (ADCAPTEEN) is proposed for solving the disadvantages in this paper. In which, a master cluster head (MCH) and a vice cluster head (VCH) are selected in each cluster. MCH is selected randomly as same as APTEEN, and VCH is selected according to the pheromone concentration by MCH. The tasks, data collection, fusion, transition, etc., can be performed by the double cluster heads (DCH). This method can reduce CH's node energy dissipation. Simulation result in OPNET shows that, the proposed algorithm is much better on network lifetime than traditional APTEEN to a large extent.

**Keywords** Wireless sensor networks · Ant colony · Double cluster heads · Network lifetime

## 1 Introduction

Wireless sensor networks (WSN) consisting of a large number of sensor nodes are deployed randomly in monitoring regions, the applications of the WSN have grown rapidly over the recent years. The performance and the quality of WSN are closely related to the type of routing protocol. It is commonly used because cluster routing protocol can promotes effectually energy-saving [1]. The Low Energy Adaptive Clustering Hierarchy (LEACH) [2] protocol is one of the early cluster algorithms. It belongs to proactive network, but it is not suitable for reactive network. The

J. Ma · S. Wang (✉) · Y. Ge
College of Electronic Information Engineering,
Inner Mongolia University, Hohhot, China
e-mail: wangshubin@imu.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_34

Adaptive Periodic Threshold-sensitive Energy Efficient Network (APTEEN) is proposed in [3], it is classified as a hybrid network which preserves the advantages of both the proactive and reactive. APTEEN protocol can cycle to collect data, and respond to emergencies. But APTEEN has shortcomings—direct communication between CH and sink node result in high energy consumption and low efficiency. A Hierarchal Multipath Protocol HMAPTEEN [4] allows multiple routes transmitted data to reach sink node, and decreases the energy consumption and improves the network reliability. But HMAPTEEN selects nodes in each level for routing to the next level, leads to the waste of energy. A Distance Energy Efficient routing Protocol DAPTEEN [5] is proposed. It enhances the network survivability by removing data redundancy. But it considers only distance between nodes within a cluster without considering node residual energy.

The APTEEN protocol is a single cluster head based algorithms, and nodes have same the initial energy. As a result of node energy dissipation is imbalance and node is untimely death because the Cluster Head (CH) needs to perform extra computations. However, the algorithm based on double cluster heads (DCH) can overcome this shortcoming. Among them, the method is commonly used for LEACH. A double heads static cluster DHSC algorithm is proposed in [6]. It takes the thickness of the area into consideration. But it [6] does not take the node residual energy and the dist as the distance from the sink node into consideration. However, a based on double cluster-head topology control algorithm DCCCA [7] considers those factors but complicated.

To sum up, aim to combine the advantages of APTEEN and double cluster heads, this paper proposed an ant-colony based Double Cluster heads Adaptive Periodic Threshold-sensitive Energy Efficient Network protocol or ADCAPTEEN in abbreviation. The proposed protocol can balance energy consumption, and prolong effectively the life cycle of WSN.

The rest of the paper is organized as follows: we provide the protocol related knowledge in Sect. 2. In Sect. 3, the detailed description of the ADCAPTEEN is introduced. The simulation and analysis are introduced Sect. 4. Finally, the conclusions are presented in Sect. 5.

## 2 Related Knowledge

### *2.1 APTEEN Protocol*

APTEEN protocol was developed on the basis of LEACH protocol. Similarly, the protocol assumes CH is selected in randomization. During forming clusters, it will generate a random number in the range 0–1, and compare the number with a threshold, $T(n)$. The node is chosen as a CH for the current round, if generated value $<T(n)$; otherwise, the node maintains continuously as a cluster member (CM) node [2]. The threshold $T(n)$ can be expressed by using Eq. (1).

$$T_n = \begin{cases} \frac{p}{1 - p*\left(r \bmod \frac{1}{p}\right)} & if\ n \in G \\ 0 & otherwise \end{cases}, \quad (1)$$

where $p$ represents the elected probability of the CHs among all the nodes, $r$ represents the number of the current round, $G$ represents the collections of the nodes that have not yet been selected as CH nodes during the first $\frac{1}{p}$ rounds.

Once CH is determined, it declares itself is selected as the CH in this round, and broadcasts the attribute, soft threshold (ST), hard threshold (HT), and count time (CT) parameters. Each node would participation in the cluster which the cluster node has the strongest received signal strength indication (RSSI). After cluster is formed, CM starts sensing continuously, sensed value named sensed value (SV) is saved to an internal variable and transmitted the data to the CH according to the TDMA schedule assigned for it, if the value exceeds HT. And then CM node continues to sense. Sensed value is stored and transmitted again if and only if the sensed value exceeds HT, meanwhile its variation which in the sensed value it switches on its transmitter and transmits exceeds the ST [8]. In addition, the nodes are forced to transmit data to CH, when no data is sensed at intervals as CT because the sensed data does not exceed the threshold value. Where CT is the maximum time interval between two reports sent successively by a node. Since data transmission is periodic, the nodes sensed value is transmitted frequently to CH. It is kind of a necessity for selecting the threshold values and the CT reasonably to reduce the energy consumption.

## 2.2 DCH Protocol

The DCH routing protocol for the WSN, namely, the nodes are grouped into clusters and the routing algorithm allows to selecting a master cluster head (MCH) and a vice cluster head (VCH) in each cluster. Ant-colony based double cluster heads algorithm, which DCHs are selected by ant colony algorithm [9]. The two kinds of cluster heads play different role in the network. And the two cooperate the mutual the coordination, the primary tasks, data collection, fusion, and transition from each cluster to the sink node, are completed. Nodes in the cluster transmit data to MCH, and MCH works to collect and fuse data and sends aggregation data to the same cluster of VCH. VCH aims to transmit data to far sink node. Due to avoid direct communication between with MCH and sink node, energy consumption is reduced to a great degree in the network. The survival time of nodes and the life cycle of the whole network are extended.

## 3 ADCAPTEEN Protocol

In this section, we describe the ADCAPTEEN protocol and list its characteristics. The ADCAPTEEN protocol's executive procedure is periodic. Each round is divided into two phases: cluster formulation and intra-cluster data transmission phase, VCH selection and inter-cluster data transmission phase. The ADCAPTEEN network model can be seen in Fig. 1. Assume in cluster formulation and intra-cluster data transmission phase, the nodes are divided dynamically into several clusters, and MCH is chosen and clusters are formed as same as APTEEN. Assume in VCH selection and inter-cluster data transmission phase, when MCH assigns TDMA schedule for its CM, establish virtual connection between CM and MCH in each cluster (the initial value of pheromone is 0) using simple ant colony algorithm. After CM receive TDMA schedule assigned by MCH, calculate the pheromone concentration in formula (2), the expression is as follows:

$$\tau_{mi}(t) = \frac{Q}{d_{mi}} * (1 - \rho) \tag{2}$$

Pheromone volatilization rate is defined as $\rho$, $\rho = \frac{E_{init} - E_m}{E_{init}}$, it related to send messages of node energy. The distant between CM and MCH is defined as $d_{mi}$.

After all CMs send data to MCHs, when satisfying the conditions for data transmission of the proposed algorithm from CM to MCH, CM sends itself residual energy information to its MCH. And adding a certain amount of pheromone in the corresponding link can be expressed as:

$$\Delta\tau_{mi} = \alpha * \frac{e_i}{e_{init}}, \tag{3}$$

where $e_i$ represents node $i$ current energy, $e_{init}$ represents node $i$ initial energy, $\alpha$ represents regulatory factor.

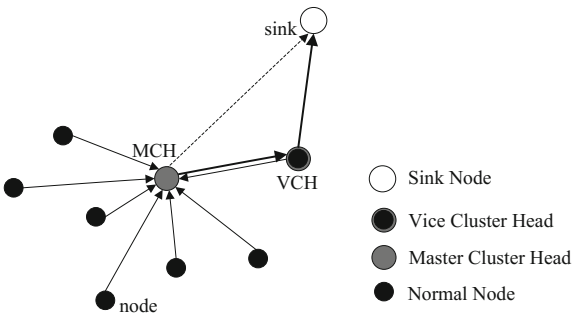



**Fig. 1** ADCAPTEEN network model

Then according to formula (2), (3), we conclude the pheromone concentration expression is as follows:

$$\tau_{mi} = \tau_{mi}(t) + \Delta\tau_{mi} = \frac{Q}{d_{mi}} * \frac{E_m}{E_{init}} + \alpha * \frac{e_i}{e_{init}} \tag{4}$$

The node which has the maximal pheromone concentration is chosen as VCH by the MCH, and VCH is never changed until next round. In this phase, CM sends sensed data to the MCH during its allocated TDMA slot. When the MCH receives the data from all the CMs, it executes data collection and data aggregation, and sends the merging message to the VCH. The VCH receives and transmits the merging message to far sink node.

Figure 2 shows the allocation process. Obviously, owing to nodes communication only between MCH and VCH, and no direct communication between with MCH and sink node. A mountain of communication costs are performed by VCH, energy consumption of MCH is reduced to a great degree. The life cycle of the WSN is prolonged.

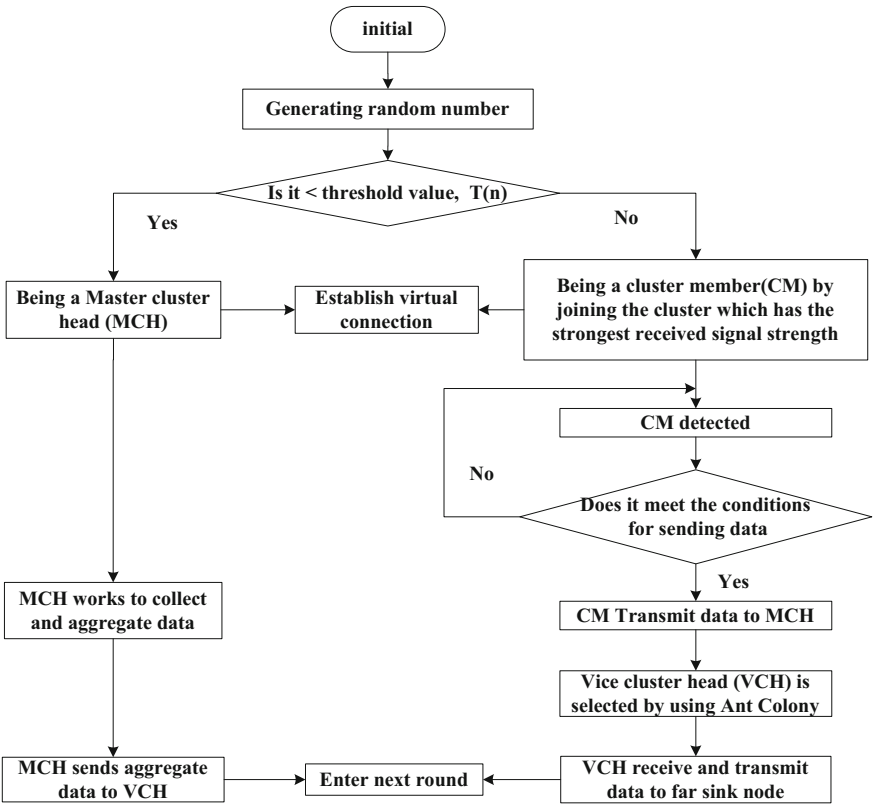



**Fig. 2** The flowchart of the allocation process

The node dies when node energy is exhausted. To avoid MCH is not selected for long time when the number of survival node remains last one. The illusion that lifecycle is prolonged is caused. Actually, no data is transmitted effectively. Under the circumstance, ADCAPTEEN data transmission qualification is described by the node. Once the conditions are met, data is transmitted directly from the node to sink node.

# 4 Simulation and Analysis

## *4.1 Simulation Conditions*

Our simulation model consists of many sensor nodes grouped into clusters with a fixed sink node. We use OPNET to evaluate ADCAPTEEN and compare it to APTEEN. In the experiment, a 10-node network where nodes are randomly distributed between (x = 0, y = 0) and (x = 200, y = 150) with the sink node at (x = 75, y = 175). The simulation parameters are shown in Table 1. CM senses continuously data, we assume detect times (DT) are K in per round, and CMs are forced to transmit data to MCH if CM doesn't transmit data times exceeds CT to its MCH.

## *4.2 Results and Analysis*

The performance of the life circle is compared both ADCAPTEEN algorithm and APTEEN algorithm. It can be seen from Fig. 3. Nodes survival much longer in ADCAPTEEN than APTEEN, because using DCH is more effective than single CH in reducing energy dissipation. So the ADCAPTEEN algorithm is superior to APTEEN protocol in energy dissipation.

Figure 4 shows the energy consumption of the nodes of ADCAPTEEN. Two energy consumption curves of the node 3 and node 6 decreased steadily along with increased simulation time. And the energy consumption curve of the node 1 downward trend, but sudden drop about in the 3500 s. This is because, when the node 1 is the last alive node, it transmits data to sink node directly without cluster

**Table 1** Parameters table

| The parameter name | Value | The parameter name | Value |
|---|---|---|---|
| Initial energy | 0.5 J | HT | 5 |
| Data packet | 4100 bytes | ST | 0.1 |
| Control packet | 100 bytes | CT | 3 |
| P | 0.2 | K | 10 |
| $Q$ | 10 | $\alpha$ | 10 |

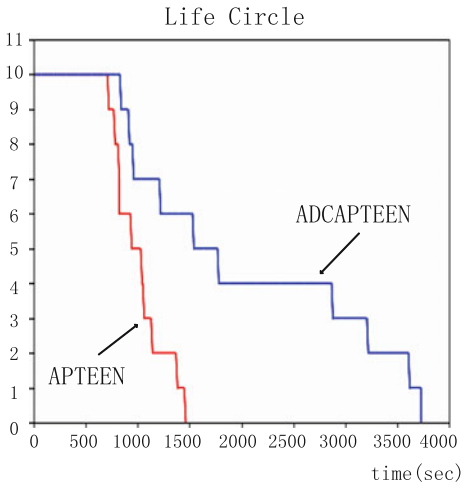



**Fig. 3** Comparison of the life circle

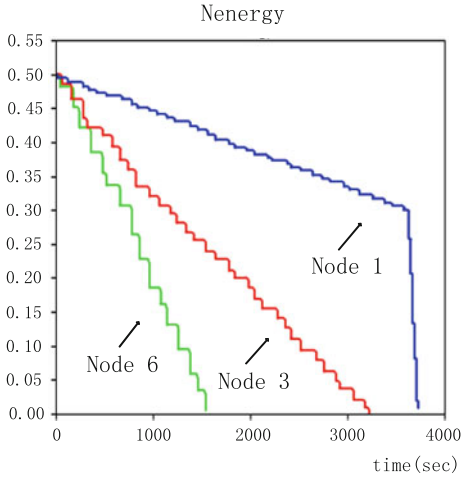



**Fig. 4** The energy consumption of the nodes of the ADCAPTEEN

heads selection. In the transmission data process, the ADCAPTEEN algorithm fully considers the reasonability of the network.

Figure 5 shows that the total numbers of nodes remain alive of the ADCAPTEEN over simulation time. Nodes remain alive in 20 nodes more than in 10 nodes. It illustrates effectiveness of the network, and shows that the algorithm suitable for large-scale network.

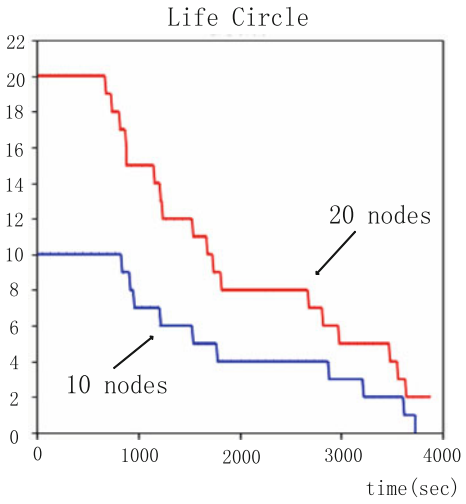



**Fig. 5** The total numbers of nodes remain alive of the ADCAPTEEN

## 5 Conclusion

In this paper, a clustering routing algorithm in wireless sensor networks was proposed based on ant-colony based double cluster heads. Simulation results show that compared with traditional APTEEN algorithm, the ADCAPTEEN reduced greatly in the energy dissipation, and achieved improvement on network life cycle. And simulation results also show that the reasonability and effectiveness of the network.

**Acknowledgements** Shubin Wang (wangshubin@imu.edu.cn) is the correspondent author and this work was supported by the National Natural Science Foundation of China (61261020), and the Natural Science Foundation of Inner Mongolia, China (2016MS0616), and the PetroChina Innovation Foundation (2014D-5006-0603).

## References

1. X. Li, F. Zhou, J. Du, LDTS: a lightweight and dependable trust system for clustered wireless sensor networks. IEEE Trans. Inf. Forensics Secur. **8**(6), 924–935 (2013)
2. P. Nayak, A. Devulapalli, A fuzzy logic-based clustering algorithm for WSN to extend the network lifetime. IEEE Sens. J. **16**(1), 137–144 (2016)
3. S.D. Muruganathan, A.B. Sesay, W.A. Krzymien, Analytical query response time evaluation for a two-level clustering hierarchy based wireless sensor network routing protocol. IEEE Commun. Lett. **14**(5), 486–488 (2011)
4. O. Abouelkhair, A. Elsaadany, Hierarchal energy efficient network protocol for wireless sensing, in *Computer Engineering Conference (ICENCO)* (2014), pp. 128–132
5. Anjali, G. Anshul, Suhali, Distance adaptive threshold sensitive energy efficient sensor network (DAPTEEN) protocol in WSN, in *Signal Processing, Computing and Control (ISPCC)* (2015), pp. 114–119

6. X. Yi, L. Deng, A double heads static cluster algorithm for wireless sensor networks, in *Environmental Science and Information Application Technology (ESIAT),* vol. 2 (2010), pp. 635–638
7. Q. Xuegong, C. Yan, A control algorithm based on double cluster-head for heterogeneous wireless sensor network, in *International Conference on Industrial and Information Systems (IIS)*, vol. 1 (2010), pp. 541–544
8. D. Baghyalakshmi, J. Ebenezer, S.A.V. Satyamurty, Low latency and energy efficient routing protocols for wireless sensor networks, in *International Conference on Wireless Communication and Sensor Computing (ICWCSC)* (2010), pp. 1–6
9. M. Dorigo, V. Maniezzo, A. Colomi, The ant system: optimization by a colony of cooperating agents. IEEE Trans. Syst. Man Cybern. B **26**(2), 29–41 (1996)

# Research on TEEN Routing Protocol in Cognitive Radio Sensor Network

**Yanhong Ge, Shubin Wang and Jinyu Ma**

**Abstract** In order to improve the energy efficiency of cognitive radio sensor network, this paper first introduces the TEEN routing protocol into cognitive radio sensor network. The efficient method is joins the process of idle channel detection in each round of TEEN routing protocol to adapt to the dynamic spectrum environment of cognitive radio sensor network. Simulation result shows that compared with LEACH, TEEN routing protocol increases the energy efficiency and extends life cycle of cognitive radio sensor network.

**Keywords** Cognitive radio sensor network · TEEN routing protocol · Energy efficiency

## 1 Introduction

In recent years, spectrum resources are increasingly scarce with the massive increase in users. The utilization of licensed spectrum is very low resulting in a serious waste of spectrum resources [1]. The proposed of cognitive radio sensor network [2] effectively address the problem of spectrum scarcity. Cognitive wireless sensor network combines the characteristics of sensor network and cognitive technology. Cognitive users are accessing the vacant frequencies opportunistically to improve the spectrum utilization. These free spectrums are licensed to the primary users originally. In cognitive radio sensor network, the selection of routing and spectrum are considered to be the most important research hotspots [3].

Y. Ge · S. Wang (✉) · J. Ma
College of Electronic Information Engineering, Inner Mongolia University, Hohhot, China
e-mail: wangshubin@imu.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_35

There are two routing metrics was proposed to adjust to the cognitive radio sensor network in [4]. The main purpose of this paper is to maximize data rates and minimize delay for a set of user communication sessions. A distributed and efficient cluster-based spectrum and interference aware routing protocol is proposed in [5], the proposed algorithm can well fit into the mobile CR ad hoc network and improve the network performance. Cognitive LEACH (CogLEACH) was introduced in [6], which is a spectrum-aware extension of the LEACH routing protocol. CogLEACH is a spectrum-aware and energy efficient clustering protocol for cognitive radio sensor networks. The clustering routing protocol LEACH is introduced to cognitive radio sensor networks in [7], and adapting the cognitive dynamic spectrum environment.

Low Energy Adaptive Clustering Hierarchy (LEACH) has been introduced into cognitive radio sensor network, however, without taking into account the energy efficiency of the network. One of the drawbacks of the cognitive radio sensor network is that node energy is limited and cannot be added. It is necessary to use an effective routing protocol to improve energy efficiency and extend the survival time of the networks. Reducing the energy consumption in the process of data transmission as far as possible is very important in cognitive radio sensor network.

TEEN routing protocol is the improvement of LEACH, which is added the hard threshold and soft threshold to control the number of data transmission. Introduce TEEN routing protocol into cognitive radio sensor network, compared with LEACH, can save the energy consumption of nodes and extend the life cycle of cognitive radio sensor network. This paper joins the process of available channel detection in each round to adapt the spectral dynamic characteristics of cognitive radio sensor networks is an effectively method. Thus, in each round, nodes perform the available channel detection process, cluster head election process, clustering process and data transmission process ordinal.

The rest of the paper is organized as follows. We introduce the TEEN routing protocol in Sect. 2. In Sect. 3, we describe the method of introduce TEEN routing protocol into cognitive radio sensor networks. In Sect. 4, we use OPNET to modeling and simulation the network and present the simulation results. Section 5 is the conclusion.

## 2 TEEN Routing Protocol

TEEN (Threshold sensitive Energy Efficient sensor Network protocol) is a reactive clustering routing protocol which is improved by LEACH routing protocol. The cluster head (CH) of each cluster collects data from its cluster members. CH is responsible for the integration and processing of data and sends data to the BS or the higher level CH. All of the nodes only need to transmit data to their CH and

only the CHs need to aggregate data is the advantage of the clustering routing protocols. It is energy-saving. All the nodes take turns becoming the CH for the cluster period T in order to evenly distribute the energy consumption. CHs election uses the random selection mechanism. Once the clusters are formed, the CHs allocate time slot for the cluster members in which the cluster members can transmit their data.

Unlike LEACH, in TEEN routing protocol, the CHs broadcast Hard Threshold (HT) and Soft Threshold (ST) to its members to control the quantity of data transmission. HT is set based on the range of interest value of users, only when the current value of the sensed attribute is greater than the value of HT, the nodes will transmit the sensed data to CH in current round. Then, the sensed attribute is stored in the sensed value (SV) which is an internal variable. Nodes sense attribute continuously, only the next value differs from SV by an amount equal or greater than the ST, it will be transmitted. The ST reduces the frequency of data transmission by abandoning the little change of sensed attribute. The value of ST can be changed according to the needs of user. Setting a smaller ST makes the network more accurate, at the cost of increased energy consumption. Thus, users must adjust the size of the ST to control the trade-off between energy efficiency and accuracy.

## 3 TEEN Routing Protocol in Cognitive Radio Sensor Network

Due to the dynamic spectrum characteristics of the cognitive radio sensor network, TEEN routing protocol in cognitive radio sensor network is different from the traditional TEEN routing protocol. In cognitive radio sensor network, available channels of each cognitive node are not fixed. It must ensure that two nodes have at least one common idle channel available to communicate.

The operation of TEEN consists of many periods. Each period consists of many rounds. The process of CHs election, clustering, and data transmission will be restart after the end of each round. In order to adapt to the spectrum variability of cognitive radio sensor network, this paper joins the process of available channel detection at the beginning of each round. Each node has a set of available channel after the process of available channel detection. Such as node i has the available channel sets M(i), cluster node j has the available channel sets M(j). Once the clusters are formed, the cluster heads broadcast a TDMA schedule to cluster members. TDMA schedule gives the order to cluster members when they can transmit data to CHs and avoids conflicts between nodes.

In the current round, it is the time shot of node i, if the node i and cluster head j have a common channel available, node i will send data to cluster head j ordinary as long as the sensed value meets the requirements of HT and ST. If there are no common channel available between the node i and j, node i in this time slot does not transmit data to j and wait for the next round.

Assuming that cognitive radio sensor network consists of N cognitive nodes, P authorized nodes and the set M = {1, 2, …, m} of channels. Operation flow chart of TEEN in cognitive radio sensor networks contains four stages.

Stage 1: Available channel detection

At the stage of available channel detection, each node performs the available channel detection process and list the set of available channel. For example, node i senses the available channel list M(i), which is a subset of M.

Stage 2: Cluster head election

Every node is assigned a random number between 0 and 1, and then computes a threshold formula $T(n)$. If the random number is smaller than $T(n)$, the node decides to become the CH. $T(n)$ is computed as Eq. (1)

$$T(n) = \begin{cases} \frac{P}{1 - P^{*}[r \bmod (1/P)]}; & n \in G \\ 0; & otherwise \end{cases}, \tag{1}$$

where $P$ is the percentage of the nodes becoming CHs in each round, $r$ is the current round number. $G$ is the set of nodes which have not been elected as CHs in the last 1/p rounds. After the end of each round, the cluster head election process will be repeated to ensure each node can be elected as the CHs. It ensures the evenly energy consumption of nodes.

Stage 3: Clustering

After cluster head node is selected, it broadcasts a message to the rest of the nodes to announce itself was elected as the CH. The ordinary nodes decide to join the cluster according to the strongest signal they have received. Then it sends a request message to join the CH.

Stage 4: Data transmission

The nodes sense their environment continuously. Cluster head nodes broadcast HT and ST to the cluster member nodes. In the data transmission stage of TEEN routing protocol, member nodes transmit data to CHs under the requirements of HT and ST within their assigned time slot. Operation Flow chart of data transmission of TEEN is shown in Fig. 1.

After a round time, Stage 1–4 is re-initiated.

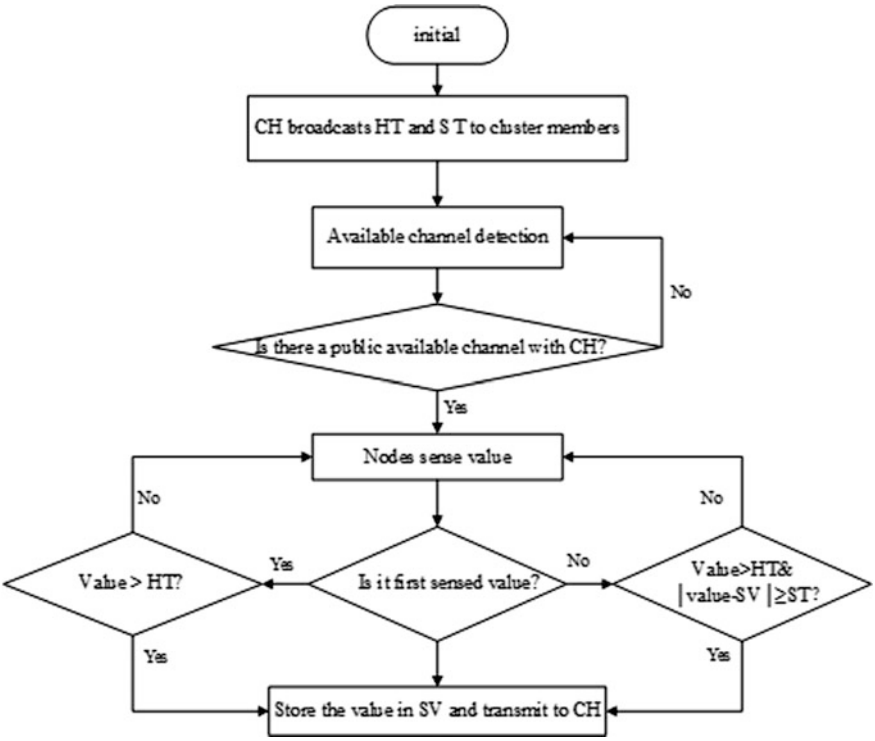



**Fig. 1** Operation flow chart of data transmission of TEEN

## 4 Simulation and Analysis

### *4.1 Modeling and Simulation the Cognitive Radio Sensor Network*

OPNET is simulation software which is used to model and simulate networks. It uses three layers modeling mechanism of network, node and process.

Simulation parameters: There are 100 cognitive nodes and 10 authorized nodes are randomly deployed in a cognitive radio sensor network. The size of the cognitive radio sensor network is 500 m × 500 m. Five available channels are <f1, f2, f3, f4, f5> respectively, the initial energy of each node equals to 50 J, idle probability of channel Pf = 0.3. BS is located at (1000 m, 1000 m) away from the cognitive radio sensor network.

We use the three layers modeling mechanism of OPNET to introduce the cognitive radio sensor network modeling and Simulation. The network model of the cognitive radio sensor network is shown in Fig. 2.

The node model of cognitive nodes and authorized nodes is different. The special spectrum monitoring devices need joined to the node model of cognitive

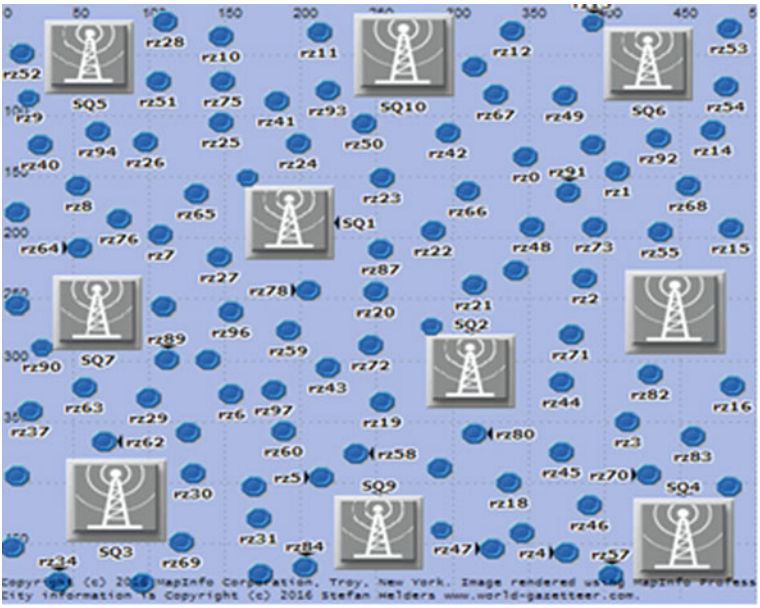



**Fig. 2** The network model of the cognitive radio sensor network

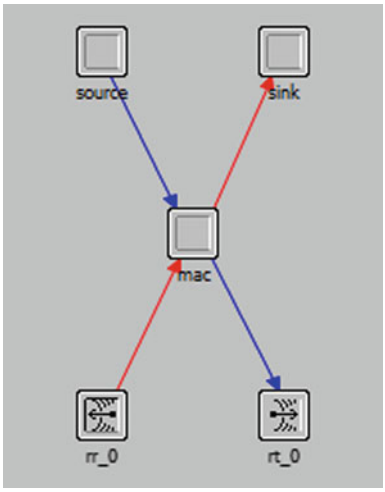



**Fig. 3** The node model of authorized nodes

nodes. Figure 3 shows the node model of authorized nodes and Fig. 4 shows the node model of cognitive nodes.

The process model of TEEN routing protocol in cognitive radio sensor network is shown in Fig. 5.

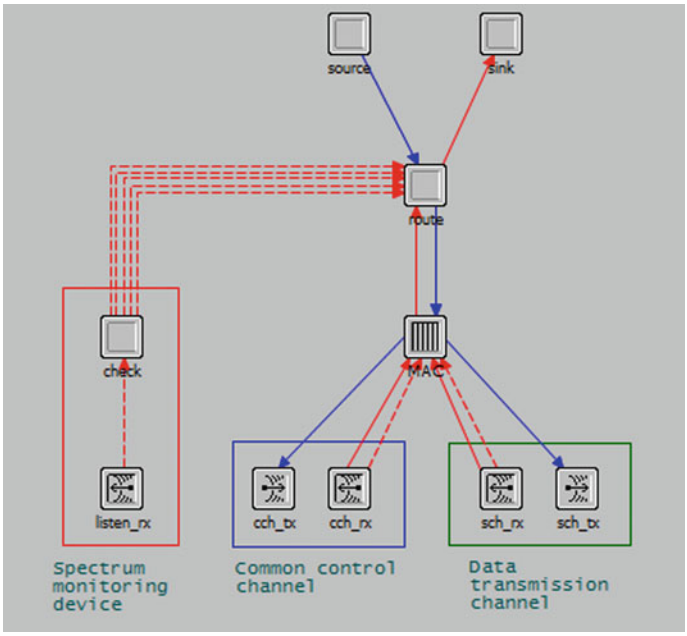



**Fig. 4** The node model of cognitive nodes

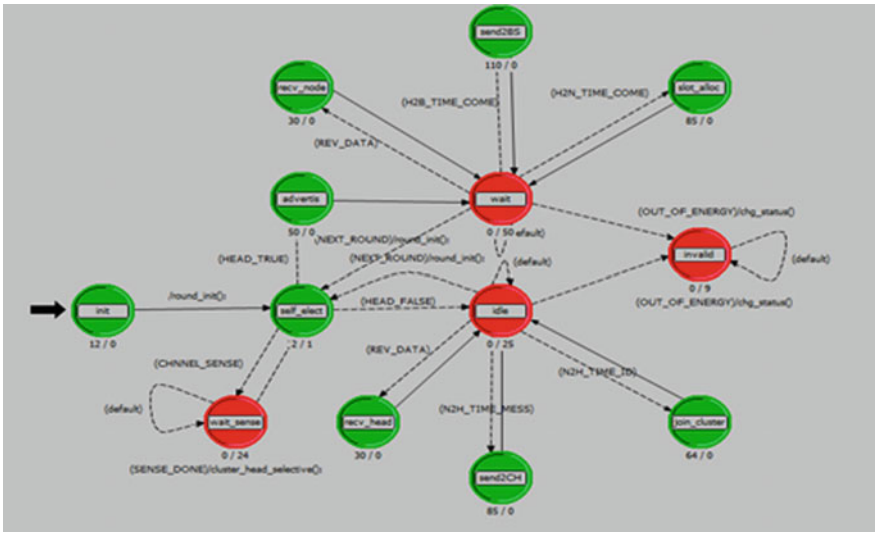



**Fig. 5** The process model of TEEN routing protocol in cognitive radio sensor network

## 4.2 Simulation Results

After modeling and simulation in OPNET, choose one of the cognitive nodes, the energy consumption under LEACH and TEEN routing protocol is shown as Fig. 6. The horizontal axis is the simulation time which unit is second, vertical shaft is the node energy (J).

From the energy consumption simulation diagram of a single node, TEEN routing protocol reduced the energy consumption of single node compared with LEACH. The energy consumption declining with vertical state is the time when node as the cluster head. As is shown in Fig. 6, the energy consumption of the cluster head nodes is much larger than member nodes. Due to the CHs need to fuse and process the data sent by the member nodes.

The network life circle comparison between TEEN and LEACH in the cognitive radio sensor network is shown in Fig. 7. The vertical shaft is the number of live

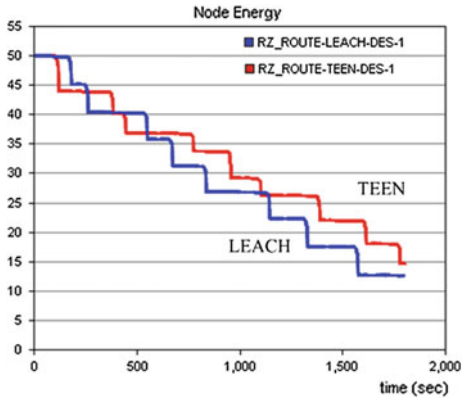



**Fig. 6** Energy consumption of a cognitive node under LEACH and TEEN

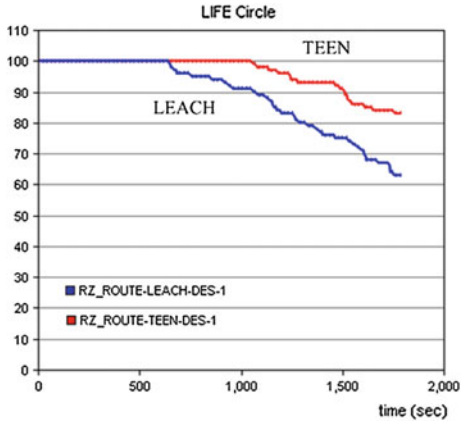



**Fig. 7** Life circle under LEACH and TEEN

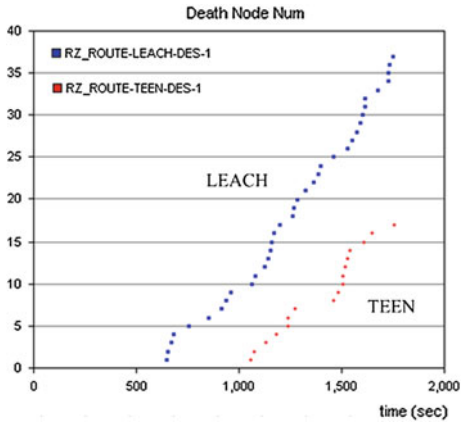



**Fig. 8** Number of death node under LEACH and TEEN

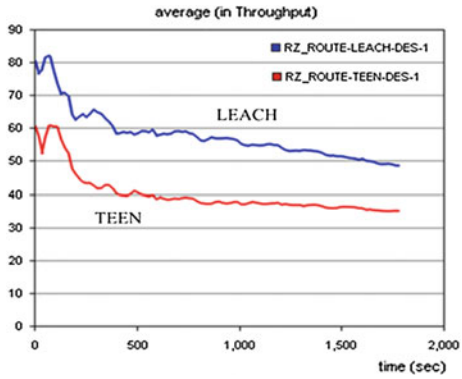



**Fig. 9** Throughput under LEACH and TEEN routing protocol

nodes. Figure 8 is the number of death node comparison between TEEN and LEACH routing protocol. The vertical shaft is the number of death nodes.

From Fig. 7, we can see that the life circle of cognitive radio sensor network under TEEN far longer than under LEACH. As is shown in Fig. 8, about after 600 s, LEACH began node dies. However, after about 1000 s, TEEN began nodes dies. We can know that compared with LEACH, TEEN routing protocol reduced the node energy consumption and delayed the time of node death. From the comparison between the TEEN routing protocol and LEACH in cognitive radio sensor networks, we can come to a conclusion that introduce TEEN routing protocol into cognitive radio sensor network, compared with LEACH, reduces the number of death node and extends the life cycle of network.

Throughput represents the number of data transmission per unit time in the network. Due to the limited of HT and ST, TEEN routing mechanism reduces unnecessary data transmission times. Figure 9 is the throughput comparison between TEEN and LEACH in cognitive radio sensor networks. It shows that throughput of TEEN lower than LEACH.

## 5 Conclusion

This paper first introduces TEEN routing protocol to cognitive radio sensor networks through joins the process of idle channel detection in each round. Due to the limited of HT and ST, TEEN routing mechanism reduces unnecessary data transmission times. Compared with LEACH, it improved the energy efficiency and delayed the time of node death to extend the network life cycle.

**Acknowledgements** Shubin Wang (wangshubin@imu.edu.cn) is the correspondent author and this work was supported by the National Natural Science Foundation of China (61261020), and the Natural Science Foundation of Inner Mongolia, China (2016MS0616), and the PetroChina Innovation Foundation (2014D-5006-0603).

## References

1. D. Cavalcanti, S. Das, J. Wang, K. Challapali, Cognitive radio based wireless sensor networks, in *Proceedings of 17th International Conference on Computer Communications and Networks* (2008)
2. O.B. Akan, O. Karli, O. Ergul, Cognitive radio sensor networks. IEEE Netw. 216–227 (2009)
3. G. Shah, O. Akan, Spectrum-aware cluster-based routing for cognitive radio sensor networks, in *IEEE International Conference on Communications (ICC)* (2013), pp. 2885–2889
4. A. Cagatay Talay, D. Turagy Altilar, ROPCORN: routing protocol for cognitive radio ad hoc networks, in *International Conference on Ultra Modern Telecommunications and Workshops* (2009), pp. 1–6
5. A.C. Talay, D.T. Altilar, United nodes: cluster-based routing protocol for mobile cognitive radio networks. IET Commun. 2097–2105 (2011)
6. M. Eletreby, M. Elsayed, M. Khairy, CogLEACH: a spectrum aware clustering protocol for cognitive radio sensor networks, in *International Conference on Cognitive Radio Oriented Wireless Networks and Communications (CROWNCOM)* (2014), pp. 179–184
7. N. Panahi, H.O. Rohi, A. Payandeh, M.S. Haghighi, Adaptation of LEACH routing protocol to cognitive radio sensor networks, in *International Symposium on Telecommunications* (2012), pp. 541–547

# Research on Trust Sensing Based Secure Routing Mechanism for Wireless Sensor Networks

**Danyang Qin, Shuang Jia, Jingya Ma, Yan Zhang and Qun Ding**

**Abstract** For network routing protocol's lower security issues, Trust Sensing based Secure Routing Mechanism (TSSRM) with the ability to resist various attacks is proposed, at the same time we present an optimized routing algorithm, which will provide reliable guarantee to improve the network routing protocol security performance. The simulation results show that TSSRM could realize the low cost and high security level of wireless sensor networks.

**Keywords** Wireless sensor networks · The optimal route · Security · QoS metrics · Trust degree

## 1 Introduction

Smart city is more and more popular in modern society since the development of Internet of Things has always increasing trend [1]. Wireless sensor networks have played an important part in promoting various services of smart city since the characters of low energy consumption, rapid deployment and self-organizing. Therefore, some researchers have made many studies on smart city based on wireless sensor network technologies. The routing protocol often suffers from various attacks

D. Qin (✉) · S. Jia · J. Ma · Y. Zhang · Q. Ding
Key Laboratory of Electronics Engineering, College of Heilongjiang University, Heilongjiang Province, Harbin 1500840, China
e-mail: qindanyang@hlju.edu.cn

S. Jia
e-mail: shuang_jia@qq.com

J. Ma
e-mail: majingya659@163.com

Y. Zhang
e-mail: zhangyanheida@126.com

Q. Ding
e-mail: qunding@hlju.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_36

since the nodes' mobility. To prevent the wireless sensor networks from the attacks of malicious nodes, different security routing protocols have been studied for many years. However, the traditional routing protocols depend mainly on key management and identification schemes, which do not apply to the wireless sensor networks.

Trust management is a good way to solve the security problems of wireless sensor networks [2], however, the conventional routing protocol based on trust is difficult to ensure the security of information transmission. This paper proposed a trust sensing based secure routing mechanism for wireless sensor networks to address the challenges mentioned above. And the simulation results show that TSSRM cannot only achieve high security for wireless sensor networks, but also reduce the routing overhead significantly.

## 2 Routing Algorithm

### 2.1 System Model

The graphical models will be adopted to analyze the routing problems for wireless sensor networks. We will treat $G_H(V, E_H, r)$ as physical map, where $V$ represents the set of directed wireless physical links, the set of vertices $V$ represents sensor nodes in the network, $E \subseteq V \times V$ represents a subset of node relationship. For each $e(i,j) \in E$, the node $i$ is the assessing unit and node $j$ is the assessed one, $j$ represents the trust degree of node $j$ for node $i$, the trust degree $dt$ of any node contains direct trust $dt$ and indirect trust $it$. Similarly, $G_R(V, E_R, r)$ represents route map of route choice and message transmitting. Therefore, the routing issues could be seen as the choice of the optimal route on a weighting physical map $G_H(V, E_H, r)$ by utilizing the routing metric $r$. The routing metric $r$ consists of trust degree and QoS metric, which is limited by the trusted map $G_T(V, E_T, t)$ and corresponding QoS map $G_Q(V, E_Q, q)$. All the optimal paths $p^*$ form the optimal route map $G_R^*(V, E_R, r)$ at last.

### 2.2 Trust Computation of Nodes

A lightweight computation way is presented to assess the trust degree of sensor nodes for wireless sensor networks with $\alpha + \beta = 1$, $\alpha > 0$, $\beta > 0$:

$$t(i,j)^l = \alpha \times dt(i,j)^l + \beta \times \frac{\sum_{(k \in C_j, k \neq i)}^{n} it(k,j)^l}{n-1}, \tag{1}$$

where $t(i,j)$ denotes the trust degree of node $j$ for node $i$. $dt(i,j)$ denotes the direct trust degree. $it(k,j)$ denotes the suggestions provided by node $k$ which is part of the neighbor nodes' set $C_j$ of node $j$. $n$ stands for the number of neighbor nodes and $l$ denotes the serial number of the assessment records. $\alpha$ and $\beta$ represent weight factor

related with security policies, the trust degree $t$ is subject to $0 \le t \le 1$. The direct trust degree is

$$dt(i,j)^l = \gamma_1 \times dt_{P(j)}(i,j)^{l-1} + \gamma_2 \times dt_{N(j)}(i,j)^{l-1} + ids(i,j)^l, \tag{2}$$

where $dt_{P(j)}(i,j)^{l-1}$ represents the direct trust degree of node $j$ for node $i$ according to the good behavior of node $j$ in old days, $dt_{N(j)}(i,j)^{l-1}$ represents the direct trust degree of node $i$ for node $j$ according to the bad behavior of node $i$ in old days. $\gamma_1$ and $\gamma_2$ respectively correspond to the self-adapting exponential decay time factor of positive and negative evaluation. $ids(i,j)^l$ represents the evaluation for the current behavior of node $j$ using intrusion detection system. $ids(i,j)$ is

$$ids(i,j) = \begin{cases} P(j), & 0 < P(j) < 1 \\ 0, & uncertainty \\ N(j), & -1 < N(j) < 0 \end{cases}, \tag{3}$$

where $P(j)$ and $N(j)$ denote the positive and negative evaluation for the node $j's$ behavior respectively. A self-adapting exponential decay factor $\gamma$ will be utilized for the trusted computation method to handle on-off attacks [3]; its expression is as follows:

$$\gamma = \begin{cases} \gamma_1 = e^{-\rho_1 \times (t_c - t_{c-1})}, & dt_{P(*)} \\ \gamma_2 = e^{-\rho_2 \times (t_c - t_{c-1})}, & dt_{N(*)} \end{cases}, \tag{4}$$

where $t_c$ represents the current moment and $t_{c-1}$ denotes the moment that the last interaction occurred. The process of indirect trust evaluation is shown as follows:

$$\sum_{(k \in C_j, k \ne i)}^{n} it(k,j)^l = \sum_{(k \in C_j, k \ne i)}^{n} dt(i,k)^l \times dt(k,j)^l \tag{5}$$

The trust links are adopted to assess the indirect trust degree of wireless sensor nodes in this model. $dt(i,k)$ represents the direct trust degree of the node $k$ for node $i$. $dt(k,j)$ denotes the direct trust degree of the node $j$ for node $k$. An inconsistent examine mechanism is proposed to deal with bad mouthing attacks [4], which is given by (6)

$$ic(i,j)^l = \frac{\sum_{(k \in C_j, k \ne i)}^{n} dt(i,k)^l \times dt(k,j)^l + dt(i,j)^l}{\sum_{(k \in C_j, k \ne i)}^{n} dt(i,k)^l + 1} \tag{6}$$

Since the received suggestions may contain wrong messages supplied by bad mouthing attacks, our trust model takes use of a threshold $\varepsilon$ to estimate whether the message is questionable for each suggestion. If $\left| dt(k,j)^l - ic(i,j)^l \right| > \varepsilon$, the suggestion will be abandoned.

## 2.3 Trust Computation of Paths

A source node will evaluate the trust degree of the route when it is ready to transmit messages to the terminal node via the multi-hop communication. The trust degree of route $p$ can be calculated by (7) when we select the most credible route decided by the highest product value of all trust degree traverse the route.

$$t(p) = \prod (\{t(i,j)|i,j \in p, i \to j\}), \tag{7}$$

where the node $i$ and node $j$ are neighbor nodes. Node $j$ is the next hop of node $i$.

We can also select the most credible route decided by the highest minimum trust degrees of intermediate nodes in the route. The trust degree of route $p$ can be expressed as follows:

$$t(p) = \min(\{t(i,j)|i,j \in p, i \to j\}) \tag{8}$$

## 2.4 Routing Metrics

Routing metrics may contain the trust degree and QoS metrics that are used for improving service quality in practical applications. Therefore, the routing metric of path $p$ will be expressed as $r(p) \triangleq (t(p), q_1(p), q_2(p), \ldots, q_n(p))$, where $q_1(p), q_2(p) \ldots q_n(p)$ is different QoS metrics of path $p$. Since the different computation methods of routing metric can affect the outcome of route choice. So a mathematical theory called semiring [5] is introduced in this paper to avoid interfering with the design of routing algorithm.

**Definition 1** The semiring of trust degree is an algebraic expression $\left(T, \oplus_T, \otimes_T, \overline{0}_T, \overline{1}_T, \leq\right)$, which $T$ is a set of trust degree. $\oplus_T$ and $\otimes_T$ denote an operational character to connect with trust degree along a route and converge trust degree traverse routes, respectively.

**Definition 2** The semiring of QoS is an algebraic expression $\left(Q, \oplus_Q, \otimes_Q, \overline{0}_Q, \overline{1}_Q, \leq\right)$, which $Q$ is a set of QoS metrics. $\otimes_Q$ and $\oplus_Q$ denote an operational character to connect with QoS metrics along a route and converge QoS metrics traverse routes, respectively.

## 2.5 Routing Choice

The choice of the optimal path will be described using semiring theory. For example, we must choose the most trusted path in networks, the optimal path is given

$$p^{*}(v_1, v_n) = \oplus_T [t(p(v_1, v_k)) \otimes_T t(p(v_k, v_n))], \tag{9}$$

where $k \in p(v_1, v_n)$. There is an assumption that the trust degree of path is calculated by (7), $\otimes_T$ represents "$\times$", $\oplus_T$ represents "max(*)".

Then, the routing algorithm considering various routing indicators is proposed. We can see the details of the routing algorithm in Algorithm 1.

**Algorithm 1** Routing algorithm

(1) Process Initialization
(2) $G_R^*(V, E_R, r) = v_n$, $v_n$ is the terminal node
(3) Add $v_n$ to $V^*$, $V^*$ denotes the set of nodes which have the optimal paths to $v_n$
(4) **while** $V \neq V^*$ **do**
(5) **for all** node $v_i \in V - V^*$ **do**
(6) Sort $r\left(p\left(v_i, \overrightarrow{v_n}\right)\right)$
(7) Obtain $r(p(v_i, v_n)) \square (q_0, q_1, ..., q_m)$
(8) where $q_0 = t\left(p\left(v_i, v_n\right)\right)$
(9) **for all** $v_k \in \Gamma(v_i)$ **do**
(10) **if** $t(v_i, v_k) \otimes_T t\left(p\left(v_k, v_n\right)\right) \geq t\left(p\left(v_i, v_n\right)\right)_{\text{th}}$ **then**
(11) Add $\left(v_i, p\left(v_k, v_n\right)\right)$ to $P_{Q_0}^*\left(v_i, v_n\right)$
(12) **end if**
(13) **end for**
(14) **if** $P_{Q_0}^*\left(v_i, v_n\right) = 0$ **then**
(15) $v_i$ is remove out of the network
(16) continue;
(17) **end if**
(18) **for** $j = 1$; $j < m$; $j{+}{+}$ **do**
(19) $P_{Q_j}^*\left(v_i, v_n\right) = \oplus_{Q_j} p_{Q_{j-1}}^*\left(v_i, v_n\right)$
(20) where $P_{Q_{j-1}}^*\left(v_i, v_n\right) \subseteq P_{Q_{j-1}}^*\left(v_i, v_n\right)$
(21) **end for**
(22) **if** $P_{Q_m}^*\left(v_i, v_n\right) = 0$ **then**
(23) $v_i$ is remove out of the network
(24) continue;
(25) **else**
(26) Add $v_i$ to $V^*$
(27) Add $P_{Q_m}^*\left(v_i, v_n\right)$ to $G_R^*(V, E_R, r)$
(28) **return** $p_R^*\left(v_i, v_n\right)$, $p_R^*\left(v_i, v_n\right) \subseteq P_{Q_m}^*\left(v_i, v_n\right)$
(29) **end if**
(30) **end for**
(31) **end while**
(32) END Process

## 3 The Scheme of TSSRM

A secure routing mechanism is proposed in this section. Figure 1 is a exemplification on the process in which source node ($v_0$) searches the optimal path to reach the terminal node ($v_{11}$). The detail process of TSSRM is shown as follows.

S1. $v_0$ initializes the process of trust derivation and transmits a demand message to its neighbors (eg, $v_2$) when $v_0$ is ready to transmit message to $v_{11}$. $v_2$ that receives, the trust demand message will primarily examine whether the same demand has been received. If yes, the demand will be promptly abandoned. If not, $v_2$ will spread trust demands to its every neighbor node.

S2. Every node except the assessing nodes should examine whether the assessed node is its neighbors once receiving the trust demand packet. If not, it keeps silent. Otherwise, the node ($v_1$ , $v_3$ and $v_6$) may unicast a reply to the assessing node ($v_0$) via the existing inverted paths.

S3. After $v_0$ receives the suggestions given by the neighbor nodes of assessing node, $v_0$ calculates trust degree and estimates whether the assessed node $v_2$ is credible on the basis of the required trust route constraints. Similarly, a node $v_0$ can find a credible forwarding set ($v_2$,$v_3$) and send routing demands to the nodes in it.

S4. The intermediate trust node that receives routing demands will send a reply to $v_0$ if it has the optimal route to $v_{11}$. Then, $v_0$ can obtain the optimal route to $v_{11}$. Go to S6. If not, the intermediate trust nodes in the network will repeat S1–3 to look for the next credible node.

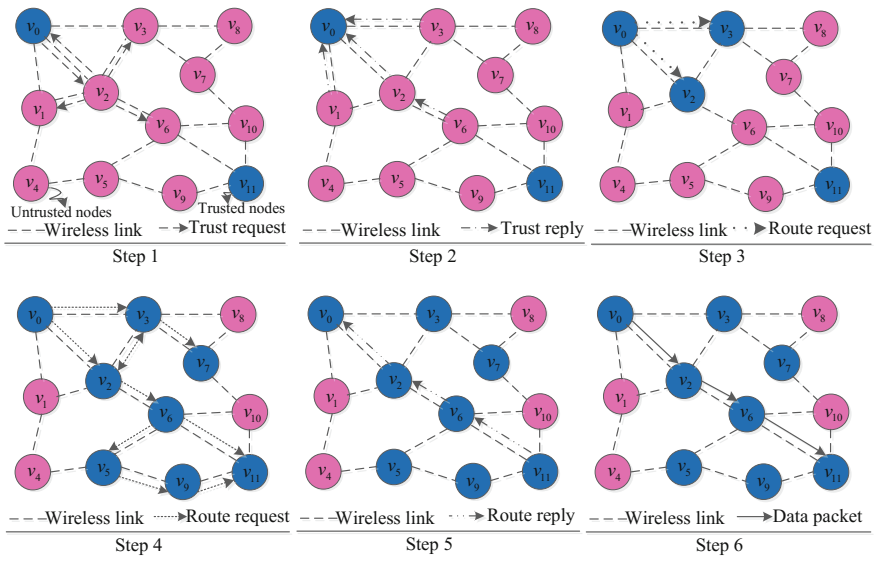



**Fig. 1** Routing process

S5. $v_{11}$ will send a reply to $v_0$ via the selected reverse route when the route demand hit the target.

S6. Finally, $v_0$ sends a packet to $v_{11}$ via the optimal route.

# 4 Performance Evaluation

## *4.1 Experimental Setup*

The simulation results of TSSRM will be analyzed by NS2 simulator. w represents the proportion of malicious nodes in the simulation. All of the default experiment parameters are shown in Table 1.

## *4.2 The Effect of Attacks on TSSRM*

Figure 2 shows that the trust degree usually increases with time if there is no abnormal phenomenon (from 30 to 70 s). But the trust degree will decline as the malicious nodes activate on-off attacks. A self-adapting exponential decay time factor is utilized for TSSRM to handle on-off attacks at this moment. Therefore, the figure shows that TSSRM could provide better resilience for on-off attacks comparing with trust assessment process without considering the time decay factor.

**Table 1** Experiment parameters

| Parameters | Value |
|---|---|
| Simulation time | 500 s |
| Monitoring area | 200 m × 200 m |
| Number of sensor nodes | 100 |
| Communication distance | 40 m |
| Message internal | 5 s |
| Length of message | 100bytes |
| Routing protocol | GPSR and BAR |
| Original trust degree | 0.5 |
| Distrust internal | [0, 0.4] |
| Error probability of detection | 0.1 |
| The proportion of malicious nodes | 20% |
| $P(a)$ | 0.01 |
| $N(a)$ | −0.01 |
| $\gamma_1, \gamma_2$ | 0.90, 0.99 |
| $\alpha, \varepsilon$ | 0.7, 0.15 |
| $t(p), t(p)_{\text{th}}$ | min(*), 0.4 |

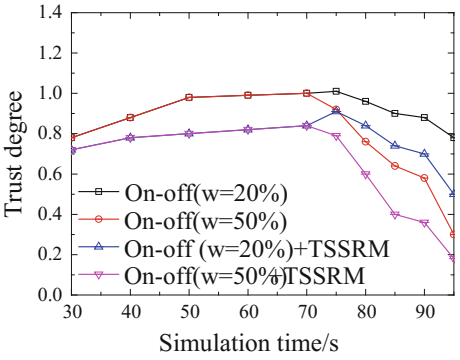



**Fig. 2** The effect of on-off attacks

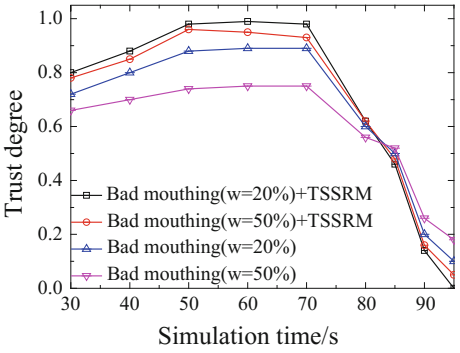



**Fig. 3** The effect of bad mouthing attacks

An inconsistent examine mechanism is proposed to provide better resistance against bad mouthing attacks in Fig. 3. In the simulation, bad mouthing attackers provide positive/negative suggestion about normal/malicious behavior. Thus, the trust degree is much lower when evaluating normal nodes' behavior (from 30 to 70 s) for bad mouthing attacks. Worse bad mouthing attacks (the proportion of bad mouthing attacker is equal to 50%) may result in greater impact on trust assessment process than the attacks which are not so bad (the proportion of bad mouthing attacker is equal to 20%). On this occasion, the inconsistent examine mechanism could wipe out many wrong suggestions since they usually have a obvious distinction (high or low) from those provided by normal nodes. At last, the accuracy of the trust assessment will be improved by applying the inconsistent examine mechanism.

## *4.3 The Effectiveness and Security of TSSRM*

Figure 4 shows that flooding has a large number of broadcast and replay packets, so its routing overhead is much higher than that of the other three schemes. For

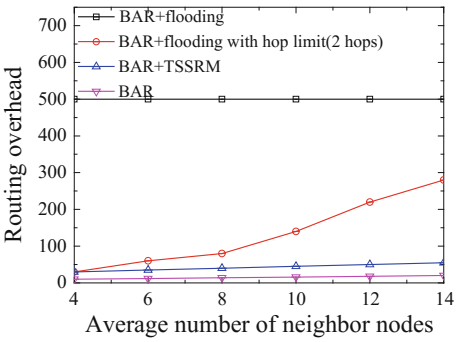



**Fig. 4** Routing overhead

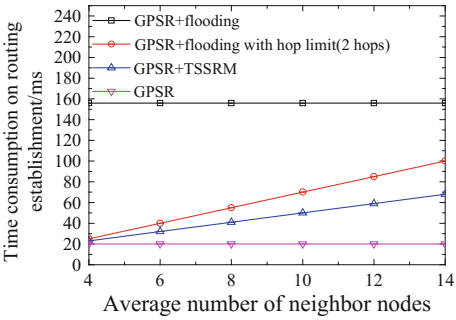



**Fig. 5** Time consumption on routing establishment

example, our mechanism will save 79.4% of routing overhead compared with flooding with 2-hop limit as the number of neighbors equals to 14.

Figure 5 shows the different time consumption to achieve routing setup between TSSRM and other schemes. Clearly, the delay performance shows more obvious advantage over other schemes in routing overhead. For example, our mechanism will save 32.3% of time as the number of neighbors equals to 14 compared with the flooding with 2-hop limit.

Figure 6 shows that more malicious nodes will cause more damage for wireless networks. TSSRM can provide a good way to these attacks. The average packet

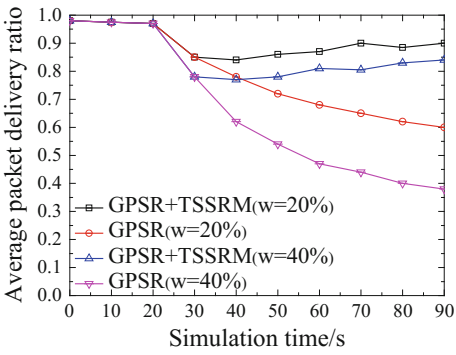



**Fig. 6** The effect of malicious nodes

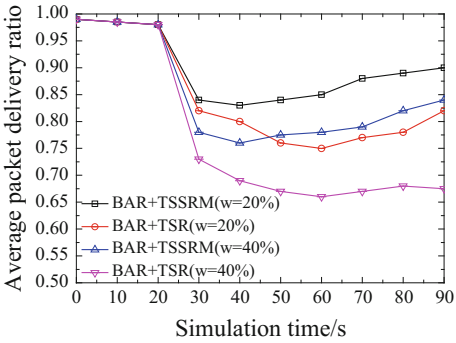



**Fig. 7** The effect of indirect trust

delivery ratio increase gradually by introducing TSSRM into the existing routing protocols; because TSSRM will rapidly activate a path update process to obtain a credible path as it detects malicious intermediate nodes in the former route.

TSR [6] is trust evaluation mechanism that only considers direct trust. Figure 7 shows that the average packet delivery ratio will reduce significantly when a malicious node activates attacks in wireless sensor networks (from 20 s). TSSRM could improve network security compared with TSR when many errors detected incidents could take place in the simulation scenario, since TSSRM considers direct trust and indirect trust, which will provide much resistance against error detected incidents.

## 5 Conclusions

This paper presents a trust sensing based secure routing mechanism to handle common network attacks. An optimized routing algorithm has been proposed, which considers the trust degree and other QoS metrics. Simulation results show that TSSRM can provide much better security compared with the traditional trust mechanism. Future research will study the error detected system in wireless sensor networks, which will improve the security of the wireless sensor networks.

**Acknowledgements** This work was supported in part by the National Natural Science Foundation of China (61302074, 61571181), Natural Science Foundation of Heilongjiang Province (QC2013C061), Modern Sensor Technology Research and Innovation Team Foundation of Heilongjiang Province (2012TD007), Postdoctoral Research Foundation of Heilongjiang Province (LBH-Q15121), Postgraduate Innovation Research Foundation of Heilongjiang University (YJSCX2016-019HLJU).

## References

1. J.Y. Feng, Y.Q. Zhang, G.Y. Lu, W.X. Zheng, Securing cooperative spectrum sensing against ISSDF attack using dynamic trust evaluation in cognitive radio networks. Secur. Commun. Netw. **8**(17), 212–2 (2015)
2. T. Watteyne, K.S.J. Pister, Smarter cities through standards-based wireless sensor networks. IBM J. Res. Dev. **55**(1–2), 7–10 (2011)
3. B.S. Sahil, R. Arnab, N.M. Kanti, Trust evaluation based on node's characteristics and neighbouring nodes' recommendations for WSN. Wirel. Sens. Netw. **6**(8), 157–172 (2014)
4. H. Xia, Z. Jia, X. Li, L. Ju, E.H.M. Sha, Trust prediction and trust-based source routing in mobile ad hoc networks. Ad Hoc Netw. **11**(7), 2096–2114 (2013)
5. G.D. Devanagavi, N. Nalini, R.C. Biradar, Trusted neighbors based secured routing scheme in wireless sensor networks using agents. Wirel. Pers. Commun. **78**(1), 1–28 (2014)
6. A. Falahati, M. Akbari, A fault-tolerant cooperative spectrum sensing algorithm over cognitive radio network based on wireless sensor network. Wirel. Sens. Netw. **3**(3), 83–91 (2011)

# Heuristic Algorithm of Lifetime Maximization for Wireless Sensor Network

**Danyang Qin, Songxiang Yang, Yan Zhang, Jingya Ma and Qun Ding**

**Abstract** Wireless sensor network (WSN) is an important branch of modern communication system, and it plays a significant role in human life and production. Due to the number of sensor nodes, wide distribution area and complex environment, the energy consumption caused by battery charge or replacement is much higher than that by redeployment. To prolong the effective work duration, a network lifetime maximization algorithm is proposed to reduce the calculating time and maximize the lifetime at the same time. In the mathematical model, the combinations of sensor deployment, activity scheduling, data routing, and sink mobility are considered. Simulation results show that the proposed algorithm can effectively extend the lifetime of WSN and improve the network performance as well.

**Keywords** Wireless sensor network · Node deployment · Activity scheduling · Data routes · Sink mobility

D. Qin (✉) · S. Yang · Y. Zhang · J. Ma · Q. Ding
Key Laboratory of Electronics Engineering, College of Heilongjiang University, Heilongjiang Province, Harbin 150080, China
e-mail: qindanyang@hlju.edu.cn

S. Yang
e-mail: songxiang_yang@foxmail.com

Y. Zhang
e-mail: zhangyanheida@126.com

J. Ma
e-mail: majingya659@163.com

Q. Ding
e-mail: qunding@hlju.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_37

# 1 Introduction

WSN is an important branch of modern communication system,and it has been widely used in various fields. However, all aspects of its performance need to be improved since it is subject to its own characteristics and application of special restrictions. WSN consists of small battery powered sensor nodes with limited energy resources. Hence, energy efficiency is a key design issue to improve the life span of the network [1]. A failure of an actor may cause the network to partition into disjoint blocks and would thus violate such a connectivity requirement. As actors deployed in harsh environment it is difficult to replace actor nodes, so we need to reposition actor nodes [2].

At present, the researches on the maximum lifetime of the network node mainly focus on the control of energy consumption. For example, the researchers construct a redundant node decision model for heterogeneous WSN, which do not need the geographic location information, and then useful lifetime maximization for partial coverage conserve (ULMPCC) in wireless sensor networks [3] is proposed. The researchers tend to put forward a targeted solution to an aspect of energy consumption, such as the balance of routing energy consumption in wireless sensor networks [4], and on balancing energy consumption in wireless sensor networks [5]. In this paper, two algorithms are proposed to solve the problems of node deployment, activity scheduling, data routing, and sink mobility.

# 2 Mathematical Model and Algorithm of the Network Lifetime

The aim of the optimum mathematic model is to solve the problems of WSN design. The optimal solution of the sensor deployment, activity scheduling, mobile sink, and data routing problem (SAMDP) is obtained by solving the linear programming formula. Two practical heuristic algorithms are proposed to solve the linear programming formula, namely period iteration heuristic and sequential assignment heuristic.

## *2.1 Mathematical Model*

In order to make the problem more imaginative, two different periods of the WSN are compared in Fig. 1.

Figure 1 shows the position of sensors and mobile sinks in two periods, as well as the data routing information indicated by arrows. The figure shows that some sensors are active in two periods, while others are standby in one period, and active

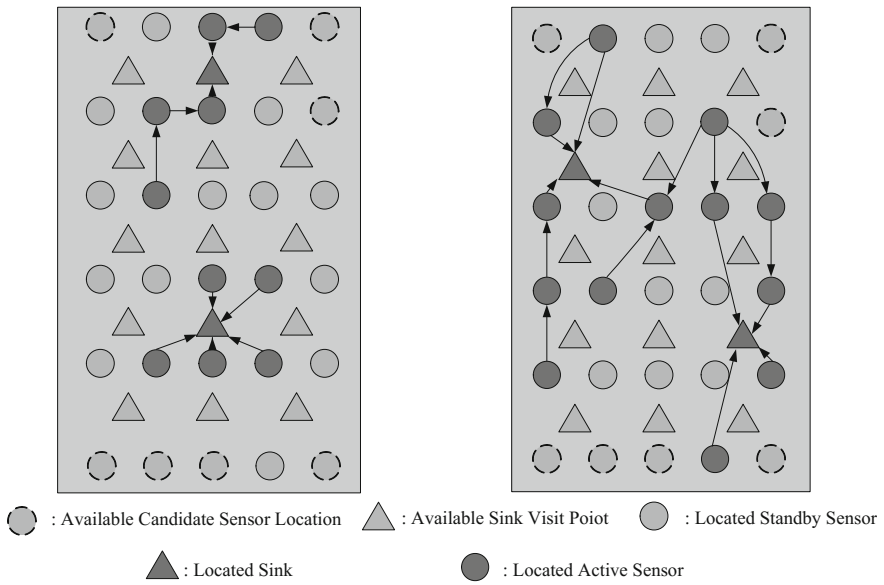



**Fig. 1** Sensor network with sensors, mobile sink, and data routing

in another period. As the result of coverage requirements, each point needs at least two sensors to achieve the description in Fig. 1. It can be seen that some sensors send the collected data to the sink directly. However, some of the collected data is sent to the adjacent sensors which act as relay nodes.

MILP mathematical model is constructed as follows:

Step 1 Assuming that there are $K$ points in the WSN, and $P$ points are selected from these to arrange sinks in each period $t$ of the WSN lifetime $T$. Then, the number of sinks is obtained, as shown in (1):

$$\sum_{\vartheta \in N} z_{\vartheta t} = P \quad t \in T \tag{1}$$

Step 2 The number of active sensors required for the normal operation of the network is shown in (2):

$$\sum_{r \in R} \sum_{i:\, k \in K_{ir}} q_{irt} \geq d_k \quad k \in K, t \in T \tag{2}$$

Step 3 The flow balance equation is constructed, which indicates the sum of the data transmission between the sensors and the sinks (or adjacent sensors) in the whole network, as shown in (3):

$$\sum_{s\in R}\sum_{j:\, i\in S_{js}} x_{jsirt} + h_r a_{irt} = \sum_{\vartheta\in N_{ir}} y_{ir\vartheta t} + \sum_{s\in R}\sum_{j\in S_{ir}} x_{irjst} \quad i\in S, r\in R, t\in T \tag{3}$$

Step 4 According to the sensor energy consumption $f_{ir}$, the total energy consumption budget $B$ is calculated, as shown in (4):

$$\sum_{i\in S}\sum_{r\in R} f_{ir} p_{ir} \le B \tag{4}$$

Step 5 In the whole network lifetime, the upper limit of energy consumed by each sensor is set as the (5):

$$\sum_{\substack{t\in T\\ i\in S, r\in R}} \left( c^s a_{irt} + c^r \sum_{s\in R}\sum_{j:\, i\in S_{js}} x_{jsirt} + \sum_{s\in R}\sum_{j\in S_{ir}} c^t_{ij} x_{irjst} + \sum_{\vartheta\in N_{ir}} c^t_{i\vartheta} y_{ir\vartheta t} \right) \le E_r \tag{5}$$

Step 6 Set a few constraints to reduce the calculation error:

$$\sum_{r\in R}\sum_{i:\, \vartheta\in K_{ir}} y_{ir\vartheta t} \le M z_{\vartheta t} \quad \vartheta\in N, t\in T \tag{6}$$

$$\sum_{s\in R}\sum_{j\in S_{ir}} x_{irjst} \le M q_{irt} \quad i\in S,\ r\in R,\ t\in T \tag{7}$$

$$\sum_{s\in R}\sum_{j:\, i\in S_{js}} x_{jsirt} \le M q_{irt} \quad i\in S, r\in R, t\in T \tag{8}$$

(6), (7), and (8) indicate that if sensors are not active or deployed, there cannot be any data inflow or outflow.

Step 7 In order to save battery energy of sensors, we put the sensors which are not deployed in a dormant state, as shown in (9).

$$q_{irt} \le p_{ir} \quad i\in S, r\in R, t\in T \tag{9}$$

The values of parameters in the above formulas are shown in (10) and (11):

$$w_t, a_{irt}, y_{ir\vartheta t}, x_{irjst} \geq 0 \tag{10}$$

$$z_{\vartheta t}, p_{ir}, q_{irt} \in \{0, 1\} \tag{11}$$

Step 8 The lifetime of wireless sensor network is defined as (12):

$$\max \sum_{t \in T} w_t \tag{12}$$

It should be noted that the results of formulas include sensor deployment, activity scheduling, mobile sink, data routing, and the sum of the data flow. After getting the accurate transmission time and total amount of data transmission, we can make the sensors in the work schedule to be active, and the sensors without transmission task to be dormant.

Here are the meanings of parameters. $B$ means the sensor placement budget. $R$ means the class of sensors. $P$ is the number of sinks. $S$ is the locations of candidate sensor. $N$ is the locations of sinks. Sensor($i$, $r$) means that class $r$ sensors are deployed at point $i$. $K_{ir}$ is the point covered by sensor($i$, $r$). $p_{ir}$ indicates whether sensor($i$, $r$) is deployed or not. $c^r$ is the unit data reception cost. $y_{ir\vartheta t}$ is the amount of flow from sensor($i$, $r$) to sink deployed at point $\vartheta$ in period $t$. $d_k$ is the coverage requirement of point $k$. $h_r$ is the data generation rate of class $r$ sensors. $E_r$ is the battery energy of class $r$ sensor. $f_{ir}$ is the energy cost of sensor($i$, $r$). $K$ is the set of covered points. $a_{irt}$ is an auxiliary parameter that replaces $w_t q_{irt}$. $q_{irt}$ indicates that whether sensor($i$, $r$) is active in period $t$ or not. $z_{\vartheta t}$ indicates that whether a sink is deployed at point $\vartheta$ in period $t$ or not.

## 2.2 Algorithm Analysis

Due to the time constraints, commercial solutions such as Gurobi 4.0 or Cplex 11.0 cannot obtain the accurate result of SAMDP formula. Two practical heuristic algorithms are proposed to find the ideal result in reasonable computation time, namely period iteration heuristic algorithm and sequential assignment heuristic algorithm.

### 2.2.1 Period Iteration Heuristic Algorithm

The difficulty of getting accurate results of SAMDP formulas is that there are a large number of binary variables in the model, which are $z_{\vartheta t}, p_{ir}, q_{irt}$. It can simplify the subset of the model and reduce the number of binary variables by removing the impossible cases.

In order to reduce the number of binary variables, $p_{ir}$ is regarded as the initial value, and the network lifetime in different situations can be obtained by changing the value of $p_{ir}$ in the Gurobi. In the periodic iterative heuristic algorithm, we select the lowest value of $T$ as far as possible, and name it $T^*$. $\phi$ is used to represent the number of iterations, and it is assigned to 1 at the beginning of the operation. The aim of this new parameter is to limit the number of period. The successive variable $w_t$,$x_{irjst}$ and $y_{ir\vartheta t}$ keep nonzero values when period value is no more than $\phi$, i.e., $t \leq \phi$. In other words, the values of $w_t$,$x_{irjst}$ and $y_{ir\vartheta t}$ will be zero, when period value is more than $\phi$.

A mathematical model is constructed after determining the variable value. In this mathematical model, only $p_{ir}$ relates to the locations of sensors, therefore, the analysis of this model is relatively easy. The model can be solved by using Gurobi in a short time, then the optimal results or approximate optimal results are obtained. After getting the optimal solution of the $\phi$ th iteration model, the next iteration process will be carried out, i.e., $\phi = \phi + 1$. The network lifetime would prolong with the increasing of $\phi$ when the network is large enough. However, the network lifetime will no longer occur significant changes when the number of iteration reaches a certain value in the actual analysis process. At this point, the algorithm is considered to obtain the optimal solution. The periodic iterative heuristic algorithm will be formally summarized in Algorithm 1.

**Algorithm 1** First, we set $DIF = O^{(\phi)} - O^{(\phi-1)}$. It represents the difference between two iterations of the network lifetime. A parameter $\varepsilon$ which indicates the accuracy of calculation is introduced, and its specific value is based on the different network settings. At the beginning of the operation, let $\phi = 1$,$DIF = 100$. When $DIF > \varepsilon$ and $\phi > 1$, the upper and lower bounds will be confirmed. Then, Gurobi solver runs the $\phi$ th model, and the process can be sped up through the upper and lower bounds. $O^{(\Phi)}$ represents the optimal target value of solver. Operations will be carried out until $DIF < \varepsilon$, at this time $\phi = T^*$. Finally, the final result and the corresponding target value are obtained.

### 2.2.2 Sequential Assignment Heuristic Algorithm

There is a logical relationship among the design problems of WSN. The locations of sensors are a prerequisite to set the activity scheduling. In addition, mobile nodes need to collect data from the active sensor, so the location of the mobile node and data routing can be determined after setting the work scheduling. Usually, we determine the best location of sensors before setting the activity scheduling of the deployed sensors. Finally, mobile nodes and data routing are identified.

Three sub problems are proposed for hierarchical WSN design. The first sub-problem S1 is to determine the locations of the sensors. Specific description of the S1 is shown in (13), and its constraints are shown as (14)–(18).

$$S_1 \leftarrow \max \sum_{k \in K} u_k \tag{13}$$

$$u_k \leq \leq \sum_{r \in R} \sum_{i:\, k \in K_{ir}} p_{ir} \quad k \in K \tag{14}$$

$$\sum_{r \in R} \sum_{i:\, k \in K_{ir}} p_{ir} \geq d_k + 2 \quad k \in K \tag{15}$$

$$\sum_{i \in S} \sum_{r \in R} f_{ir} p_{ir} \leq B \tag{16}$$

$$u_k \geq 0 \quad k \in K \tag{17}$$

$$p_{ir} \in \{0, 1\} \quad i \in S, r \in R \tag{18}$$

A new parameter $u_k$ is defined to represent the number of sensors which can cover the $k$ point, and $k \in K$. In order to enhance the flexibility of the activity scheduling, we deploy more sensors as much as possible to maximize the objective function and $u_k^{'}s$. Constraint (14) is the definition of variable $u_k$. Constraint (15) can ensure that there are enough sensors deployed near the $K$ point, so at least $d_k + 2$ sensors can be used to monitor the environment of $k$ point. The aim of constraint (16) is to limit the budget. Constraints (17) and (18) provide the nonnegative parameters and the value of binary variables respectively.

By solving S1, the locations of sensors can be determined. Then two other sub-problems S2 and S3 are proposed to solve the problems of activity scheduling, mobile sink location and data routing information.

The mathematical model of S2 is composed of (1), (2), (3), (5), (6), (7), (8), (9), (10), (11), (12). The value of the parameter $p_{ir}$ can be obtained by solving S1, then $p_{ir}$ and $z_{lt}$ are regarded as the initial values. The sensor activity scheduling parameter $q_{irt}$ can be obtained by the Gurobi algorithm. The mathematical model of S3 is composed of (2), (3), (4), (5), (6), (7), (8), (9), (10), (11), (12). The $p_{ir}$ in the S1 and the $q_{irt}$ in S2 are regarded as the initial values to solve the sub problem S3. Then, we can get $z_{lt}$ which is the position parameter of mobile sink. In the subsequent iteration process, it is taken as a new initial value into the S2 operation.

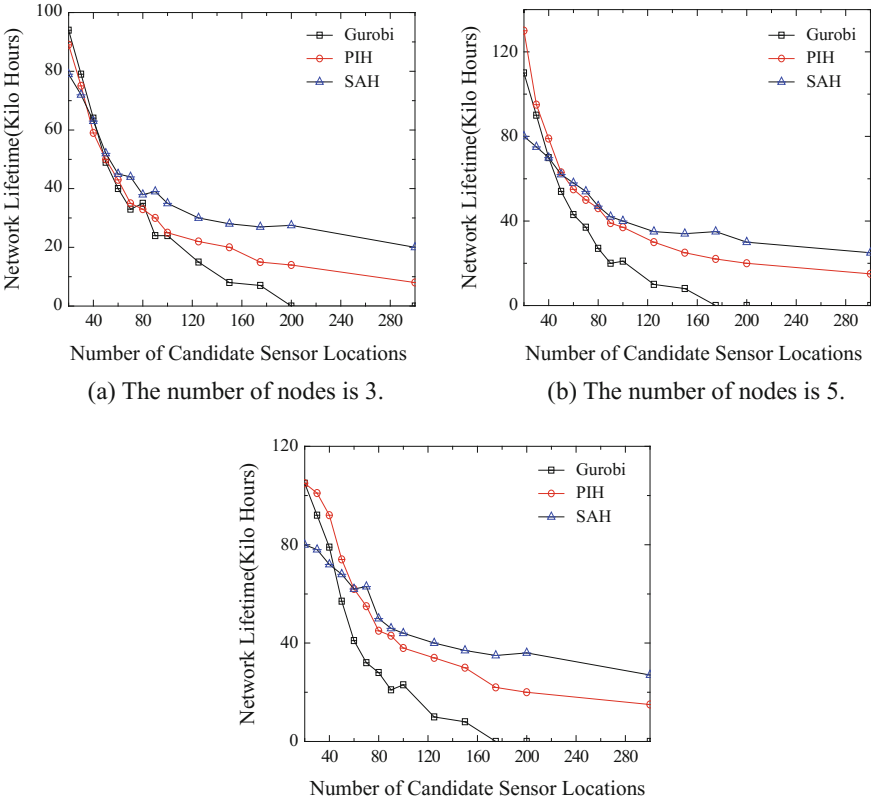
Theory of the cycle iteration in period iterative heuristic algorithm can improve the efficiency. The idea of periodic iteration is to find the minimum number of iterations which can make the maximum of the network lifetime. This is the core of two heuristic algorithms. Finally, it should be pointed out that $Z_{\vartheta\phi} = Z_{\vartheta(\phi-1)}$, $\vartheta \in N$. The specific steps of the sequential assignment heuristic algorithm are shown in Algorithm 2, where $L_2$ and $L_3$ are representative of the network lifetime of S2 and S3, and $L$ represents the network lifetime.

**Algorithm 2** First, the position parameter $p_{ir}$ is determined by solving S1. At the beginning of the algorithm, we set $\phi = 1, DIF_1 = 100, DIF_2 = 100$. Among them, $DIF_1 = L_3 - L, DIF_2 = L_3 - L_2$. Then we set the calculation precision parameters $\varepsilon_1$ and $\varepsilon_2$. When $DIF_1 > \varepsilon_1$ and $DIF_2 > \varepsilon_2$, the values of $q_{irt}$ and $L_2$ are obtained by

solving S2. S3 can be solved by S1's $p_{ir}$ and S2's $q_{irt}$, so as to obtain $L_3$ and $z_{lt}$. $z_{lt}$ is taken as the new initial value for S2's model, and the multiple operation continues until $DIF_2 < \varepsilon_2$. At the beginning of iterative process, we change the value of $q_{irt}$, and recalculate the mathematical model of S3. Then the lifetime value of $L_3$ is given to $L$, i.e., $L = L_3$. The iterative process continues until $DIF_1 \leq \varepsilon_1$, and the maximum value of network lifetime will be obtained at this time, and the final output of the network is $L$.

## 3 Simulation Results and Performance Analysis

This paper uses NS2 to complete the simulation of the mathematical model. This section evaluates the performances of two heuristic algorithms by comparing them with the Gurobi method. First, the PIH and SAH algorithms are simulated by NS2. In each test, the running time of three algorithms is defined within 3 h. Then we count the maximum network lifetime that is obtained by three different methods.




(a) The number of nodes is 3.

(b) The number of nodes is 5.

(c) The number of nodes is7.

**Fig. 2** PIH, Gurobi and SAH corresponding to the maximum network lifetime when the nodes are different

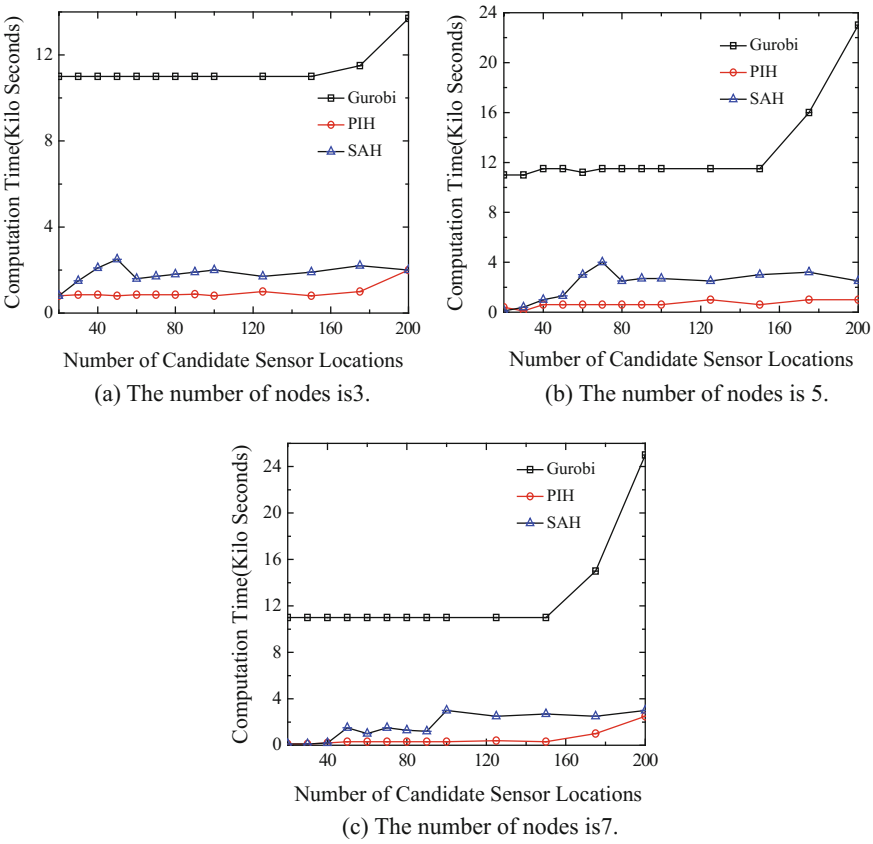
### 3.1 Performance Analysis of Network Lifetime in Different Ways

In order to evaluate the performances of two heuristic algorithms, we simulate the average network lifetime of each algorithm in five cases, as shown in Fig. 2. From the three images, it can be seen that the maximum network lifetime value of Gurobi is higher than that of PIH and SAH in the case of 20, 30, 40 candidate sensors. The maximum value of two heuristic algorithms are higher than that of the Gurobi when the number of candidate sensors is larger than 50. Moreover, the heuristic algorithms have more obvious advantages for large networks. The network lifetime of PIH is longer than that of SAH when the number of candidate sensors is 20 and 30. The network lifetime of PIH and SAH is relatively close when the number of candidate sensors is less than 90. In summary, it can be said that the performances of SAH and PIH are better than that of Gurobi in terms of network lifetime in most cases.

### 3.2 Average Calculation Time Analysis of Different Methods

In each simulation process, the running time of Gurobi, PIH and SAH are limited in 3 h. The results will be immediately recorded if they obtain the optimal solution within 3 h, and the next experiment will begin.

In Fig. 3a, we can see the average computation time of Gurobi, PIH, and SAH when the number of nodes is 3. Two heuristic algorithms need less computation time than Gurobi. After several rounds of iteration, Gurobi cannot get the effective solution of the model, while PIH can rapidly converge and get the result. In most experiments, SAH's computation time is longer than that of PIH. More accurately, computing time of PIH and SAH is almost same when the number of candidate sensors is 20 and 200. In other cases, PIH's computation time is 1.71–4.01 times as that of SAH' computation time. Figure 3b, c show the average computation time of Gurobi, PIH, and SAH, when the number of nodes is 5 and 7. It is observed that the performances of PIH and SAH are similar to the situation with 3 nodes. Gurobi is difficult to get a good solution within limited computing time for large networks. Therefore, the performance advantages of PIH and SAH are more obvious for large networks.




(a) The number of nodes is3. (b) The number of nodes is 5.

(c) The number of nodes is7.

**Fig. 3** The average computation time of Gurobi, PIH and SAH with different number of nodes

## 4 Conclusions

This paper achieves the maximum wireless sensor networks lifetime based on the research of sensor deployment, activity scheduling, and data routing protocol and sink mobility. First, we combine the sensor deployment, activity scheduling, data routing protocol, and mobile sink node routing decision to put forward the idea of mathematical optimization. Then, the results of SAMDP are obtained by two heuristic algorithms. The simulation results show that the performances of the two heuristic algorithms are better than that of Gurobi when the computation time is limited. The calculation ability of PIH and SAH can also cope with large network according to the data obtained in this paper.

**Acknowledgements** This work was supported in part by the National Natural Science Foundation of China (61302074, 61571181), Natural Science Foundation of Heilongjiang Province (QC2013C061), Modern Sensor Technology Research and Innovation Team Foundation of Heilongjiang Province (2012TD007), Postdoctoral Research Foundation of Heilongjiang Province (LBH-Q15121), Postgraduate Innovation Research Foundation of Heilongjiang University (YJSCX2016-019HLJU).

## References

1. S. Sahoo, B.M. Acharya, Energy efficient routing protocol in wireless sensor network. Int. J. Eng. Sci. Technol. **3**(7), 1–4 (2011)
2. M. Younis, K. Akkaya, Strategies and techniques for node placement in wireless sensor networks: a survey. J. Ad. Hoc. Netw. **6**(4), 621–655 (2008)
3. H.Z. Wang, F.Z. Meng, Z.Z. Li, Useful lifetime maximization for partial coverage conserve in wireless sensor networks. J. Xi'an JiaoTong Univ. **43**(43) (2009)
4. X.G. Zhang, Z.D. Wu, The balance of routing energy consumption in wireless sensor networks. J. Parallel Distrib. Comput. **71**(7), 1024–1033 (2011)
5. F. Bouabdallah, N. Bouabdallah, R. Boutaba, On balancing energy consumption in wireless sensor networks. IEEE Trans. Veh. Technol. **58**(6), 2909–2924 (2009)

# A Design on Data Acquisition System of Gas Wells Based on Heterogeneous Network

**Chen Meng, Shubin Wang, Yan Wu and Mingliang Zhang**

**Abstract** Aiming at improving the efficiency in collecting the data of gas wells' working environment, in accordance with surveillance camera system, we hereinafter try to state in this paper that how ZigBee network will be used to collect production data, to upload surveillance video through WiFi network, and how ZigBee nodes will be utilized to get the data of temperature, oil pressure, casing pressure, flow rate, and other gas wells data. Different nodes data will be uploaded to the System Gateway Center with a coordinator node. With the aid of GPRS, all data will be conveyed to the remote data center through the public network. The operator interface of gateway can display the data and the current time of each gas well. The built-in webcam is designed to collect video information in the gas wells transmitted to the terminal devices. So the operators will understand the site situation via Android APP or the Web page connected to WiFi network. And APTEEN algorithm is built to adapt real-time environment for emergency response as well as application environment for continuous data acquisition.

**Keywords** Heterogeneous network · ZigBee · WiFi · GPRS · Remote transmission

## 1 Introduction

The sophistication of data acquisition equipment of oil and gas wells is the key for measuring the level of national energy development. However, generally speaking, due to uneven distribution of gas wells and poor terrain environment, the data acquisition for the gas wells becomes more difficult. Low power consumption was often stressed in the conventional data acquisition with wireless equipment, but the role of the bandwidth was usually neglected [1, 2].

C. Meng · S. Wang (✉) · Y. Wu · M. Zhang
College of Electronic Information Engineering, Inner Mongolia University, Hohhot, China
e-mail: wangshubin@imu.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_38

With the help of both WiFi and ZigBee, we would like to build heterogeneous network which will be clarified in this paper. And GPRS is mainly for sending information to the data center over a long distance [3, 4]. In this way, the real-time data collection and video transmission can be successfully achieved. That is, ZigBee network is used to collect data and lower power consumption, while WiFi network is used to transmit network video and to extend bandwidth. It will become much easier for the operators to receive video signal via Android software and Web pages designed in this system.

Integrated with APTEEN algorithm, the active and reactive strategies are applied in the system, with settings of hard threshold and soft threshold. It will be adapted to the periodic and emergent data acquisition environment [5, 6]. When the sensor data acquisition system reaches the hard threshold, it will give out alarm and be switched to the video interface.

## 2 System Design

### *2.1 ZigBee Nodes Design*

The APTEEN algorithm is applied in ZigBee nodes. When the nodes network successfully, the collected data will be uploaded to the system gateway center with coordinator nodes step by step. The ZigBee nodes use CC2530 microprocessor implanted with Z-Stack. The nodes use SMA interface, 2.4 GHZ frequency antennas and the 5 V DC sockets of the 5.5 * 2.1 mm for power supply. The power supply circuit can be switched between 5 and 3.3 V. The balun circuit used in antenna interface is to convert the different signal to the single ended signal.

DS18B20 sensors in the system are to measure the temperature. The S pressure sensors are to measure the oil pressure and the casing pressure, and the raindrop sensors are to measure the flow. The four types of sensors are connected with the node through IO port. 10 V DC is needed for the pressure sensors to supply power. With the request of accuracy of 0–2 mv, amplifiers are necessary. The system block diagram and the hardware diagram of ZigBee nodes are shown in Figs. 1 and 2.

### *2.2 Gateway Upper Computer Design*

The gateway upper computer hardware will display the data and real time on the LCD screen and encapsulate data to the GPRS module which can transmit it to the remote data center through public network. The gateway center can give out alarm which can be switched to the video interface. The WiFi network uploads surveillance video received by the Android APP or the Web page.

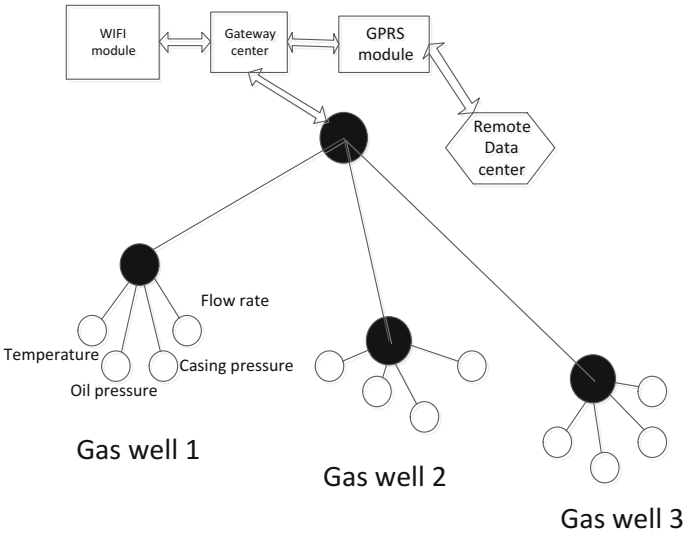



**Fig. 1** System block diagram

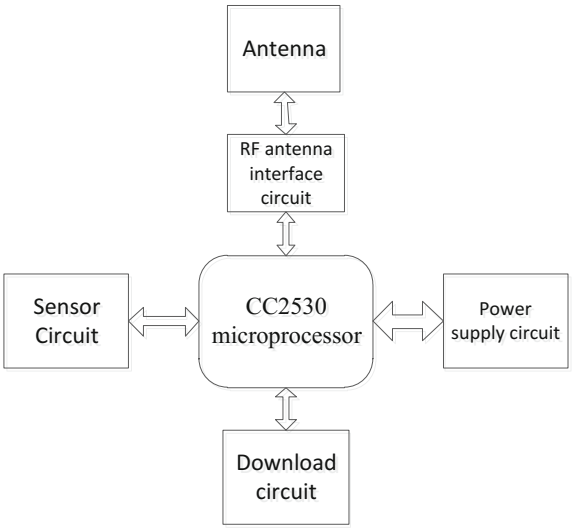



**Fig. 2** Zigbee nodes hardware diagram

In STM32F103VET6 processor, the serial port 1 is used to connect with the CC2530 to receive data, while serial port 2 is used to connect with GPRS module to send data. The WiFi module RT5350F on the gateway is to transmit the real-time video. The pull-up mode is set in the reset circuit. The crystal circuit provides 8 MHz frequency for the processor. The SDIO interface circuit is used to communicate with the SD card storing the font displayed on the LCD. The buzzer

circuit is used for alarming, in which the transistor is used for amplifying the current to drive the buzzer. The LCD interface circuit drives the LCD screen by simulating the 8080 interface which issues the command through controlling the norflash on the LCD. The battery circuit supplies power to the RTC by controlling the button battery. It will keep the RTC running after power off for accurate timing. The camera interface collects the video signals through simulating the IIC bus. The power supply circuit can switch back and forth between 5 and 9 V through flipping the switch. The Ethernet interfaces are used to connect the local area networks and wide area networks using wire node.

SIM800A GPRS module is designed connected with main processor through the serial port.

The hardware diagram of gateway and part of the system PCB figure are shown in Figs. 3 and 4.

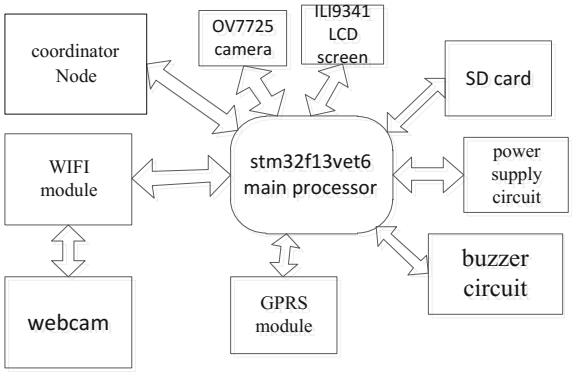



**Fig. 3** Gateway hardware diagram

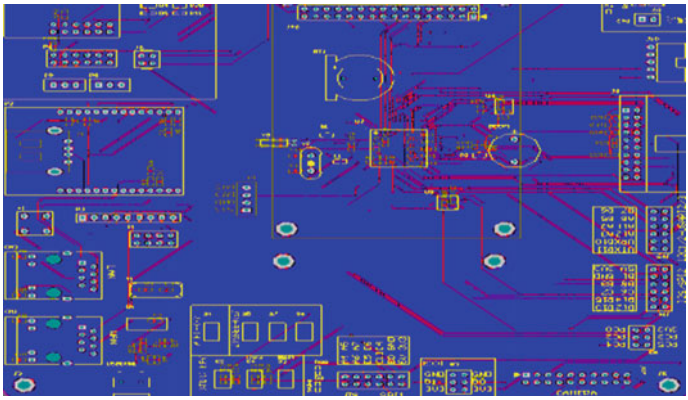


**Fig. 4** Part of the system PCB figure

# 3 Software Design

## 3.1 ZigBee Data Acquisition

The data acquisition software design is divided into terminal nodes and the coordinator node software design.

Z-Stack starts to execute in the main() function located in Zmain folder. In the main() function, the OsalInitTask() function is used to initialize each layer task. When the network device is recognized as the terminal node, the osal_start_timeEx() function is executed, which specifies the time circle of the proactive transmission.

The terminal nodes software design combines the active and reactive strategies of the APTEEN algorithm. The reactive strategy is used for the response when the data exceeds the hard or soft threshold, and the active strategy is used for the response of the periodic event. First, the system sets the hard threshold macro threshold and the soft threshold macro sthreshold in the sending data function (SampleApp_Send_PtoP_Message()) of the terminal nodes. When the data collected by the terminal nodes is greater than the hthreshold, it will be stored in the local variable SV, and be transmitted through the AFDataRequest() function writing by this paper. When the difference between the later received data values and the local variable SV is greater than the soft threshold macro, the terminal node will call the AFDataRequest () function for data transmission. If these two requirements are not met, the system will be delayed by the periodic function (osal_start_timeEx()) before calling the AFDataRequest() function for data transmission. This paper writes the SampleApp_MessMSGCB() function to upload the received data. When it identifies the events sent by the AFDataRequest(), the coordinate node will transmit the received data to the gateway center.

In the coordinator node software design, SampleApp_MessMSGCB() function is built to upload the received data. When it identifies that the events sent by the AFDataRequest(), the coordinate node will transmit the received data to the gateway center. The terminal nodes flow diagram is shown in Fig. 5.

## 3.2 Gateway System

The gateway system is used to display the current time and the data transmitted by coordinator nodes, to send the data to the GPRS module for remote transmission, to give out the alarm, and to switch the interface between main interface and video interface. The gateway system software design includes the main interface and the video interface software design.

The gateway uses simulated 8080 interface to control the LCD screen. This paper call the LCDInitial () function for LCD screen initialization, and the ILI9341_GraScan () function for screen scanning. After that, it will call the ILI9341_Clr() function to clear the screen. Finally, the ILI9341_DisString_EN_CH () function will be called to display the initial screen. The system uses the RTC real-time clock to display the

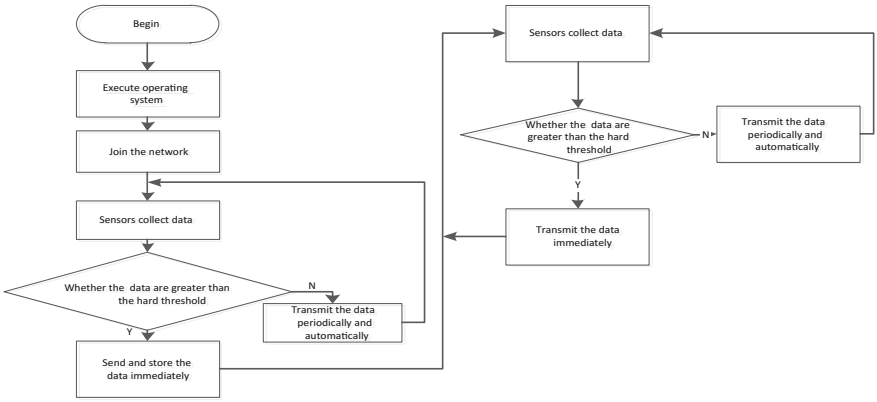



**Fig. 5** Terminal nodes flow diagram

real time. On one hand, it calls RTC_CheckConf () function to match the clock frequency and to enable RTC seconds interruption, but on the other hand, it will call the TimeShow () function to display the RTC clock. The buzzer here serves as an alarm. By writing the BEEPGPIO_Conf () function, the system initializes the IO port pin and the output rate of the buzzer. If the received data reach the hard threshold, the system will display the alarm on the LCD screen and give out an alarm.

The system calls the Ov7725_GPIO_Conf () function to initialize the camera's bus and the memory, while the VSYNC_Initial() function is to initialize the field signal. When the field signal generates twice the high electric level, it means that a frame image has been transmitted already and the system can enter the interruption service function to collect the video in the memory. The system will call the ImagDisp() function to collect the video and display the video on the LCD screen. Finally, the field signal will be cleared by the system. The main processor of gateway flow diagram is shown in Fig. 6.

## *3.3 GPRS and WiFi Network*

GPRS module introduced in this paper is to send AT instructions and wait for the feedback information of the module to complete the setting of the GPRS module.

AT command is to detect the connection of GPRS, the AT+CPIN? command is to detect the status of the SIM card, and the AT+CIPSTART command is to set the IP address of the server. If the server port connection is successful, the module will return a sign of Connection OK. AT+CIPSEND command is aimed to send data to the remote server at last.

The WiFi module uses OPENWRT operating system to configure the network. We set IP address of the WiFi module by modifying the network file in the

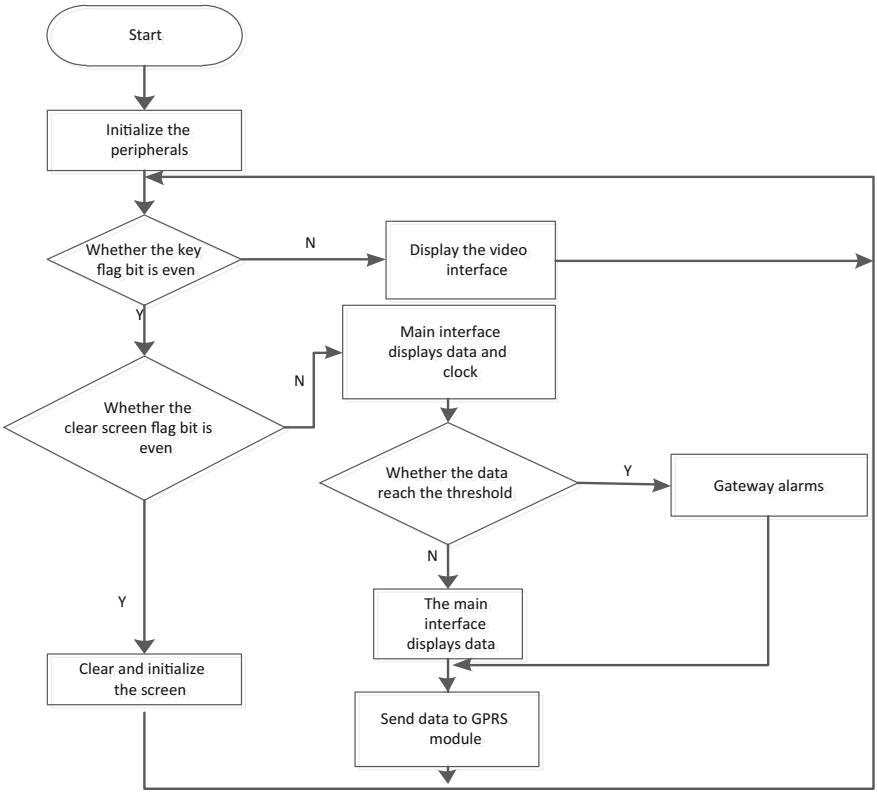



**Fig. 6** Main processor of gateway flow diagram

etc/config/ directory. Owing that the USB interface is used in the network camera, configuring make menuconfig is needed to format and mount the camera. Furthermore, the mipg-streamer software is added to support the video format of the camera. Finally, with the help of Android software and the Web page designed by us, the operators are able to receive and display the video signal by accessing the IP address.

## 4 System Testing

The hardware and software design of the whole system is independently researched and developed by the authors. The gateway center integrates the coordinator Node, video cameras, the WiFi module, the GPRS module, and so on. The physical map of the system and receiving the video signal through the mobile phone client are shown in Figs. 7 and 8.

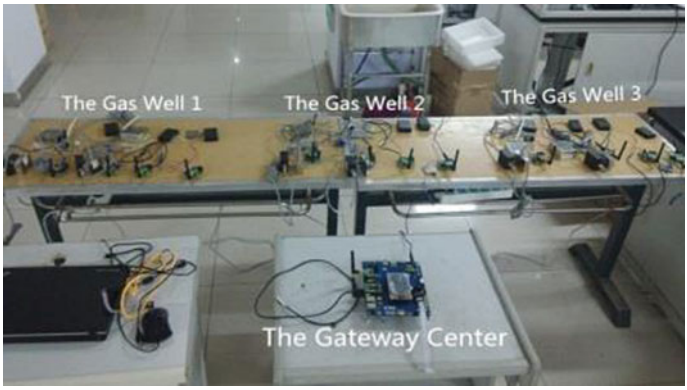



**Fig. 7** Physical map of the system

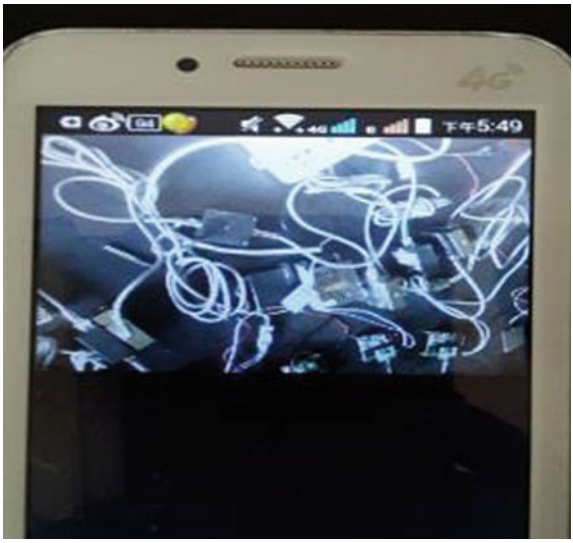



**Fig. 8** Physical map of receiving the video signal

GPRS module built in the system is to send the data to the remote data center via the public network. The receiving interface is shown in Fig. 9. The Opress, Cpress, Temp represent the oil pressure, casing pressure, and the temperature, respectively. Kingview database software is applied to receive and store the data. The display interface of the database is shown in Fig. 10.

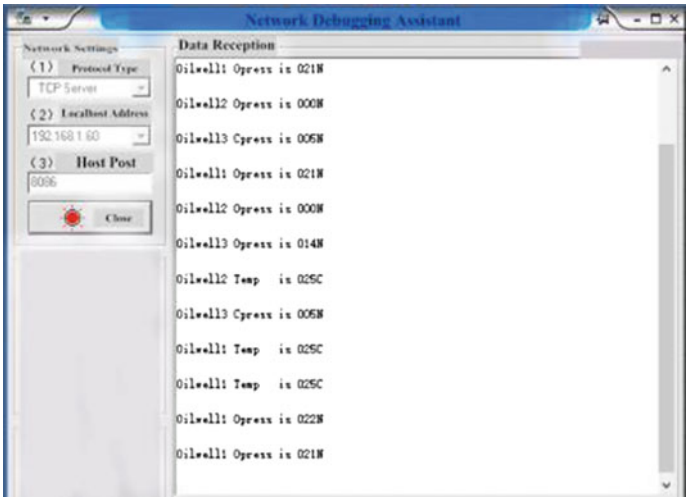



**Fig. 9** The receiving interface

| Casing pressure 1 | Casing pressure 2 | Casing pressure 3 | Temp 1 | Temp 2 | Temp 3 | Oil pressure 1 | Oil pressure 2 | Oil pressure 3 |
|---|---|---|---|---|---|---|---|---|
| 7.0 | 14.0 | 19.0 | 29.0 | 13.0 | 27.0 | 13.0 | 17.0 | 17.0 |
| 7.0 | 14.0 | 19.0 | 29.0 | 13.0 | 27.0 | 13.0 | 17.0 | 17.0 |
| 7.0 | 14.0 | 19.0 | 29.0 | 13.0 | 27.0 | 13.0 | 17.0 | 17.0 |
| 7.0 | 14.0 | 19.0 | 29.0 | 13.0 | 27.0 | 13.0 | 17.0 | 17.0 |
| 7.0 | 14.0 | 19.0 | 29.0 | 13.0 | 27.0 | 13.0 | 17.0 | 17.0 |
| 9.0 | 10.0 | 12.0 | 27.0 | 20.0 | 25.0 | 22.0 | 2.0 | 11.0 |
| 11.0 | 11.0 | 15.0 | 28.0 | 25.0 | 22.0 | 20.0 | 10.0 | 21.0 |
| 11.0 | 11.0 | 15.0 | 28.0 | 25.0 | 22.0 | 20.0 | 10.0 | 21.0 |
| 11.0 | 11.0 | 15.0 | 28.0 | 25.0 | 22.0 | 20.0 | 10.0 | 0.0 |
| 11.0 | 11.0 | 15.0 | 28.0 | 25.0 | 22.0 | 20.0 | 10.0 | 0.0 |
| 11.0 | 11.0 | 15.0 | 28.0 | 25.0 | 22.0 | 20.0 | 10.0 | 0.0 |
| 24.0 | 2.0 | 1.0 | 22.0 | 21.0 | 5.0 | 1.0 | 4.0 | 4.0 |
| 24.0 | 2.0 | 1.0 | 22.0 | 21.0 | 5.0 | 1.0 | 4.0 | 4.0 |
| 24.0 | 2.0 | 1.0 | 22.0 | 21.0 | 5.0 | 1.0 | 4.0 | 4.0 |
| 24.0 | 2.0 | 1.0 | 22.0 | 21.0 | 5.0 | 1.0 | 4.0 | 4.0 |

**Fig. 10** The display interface of the database

## 5 Conclusion

To increase the efficiency in the data acquisition of gas wells' working environment, to fulfill camera surveillance, the system based on heterogeneous network is built in this essay. The ZigBee network is built to collect data. The WiFi network is applied to transmit real-time video and the GPRS is for remote transmission of data

through the public network. The gateway interface can display the data of each gas well and the current time. It can give out alarm and be switched to the video interface. Video signal can be received as well via the Android software and Web pages. The paper states that APTEEN algorithm is applied to upload the data periodically or to deal with any emergencies.

**Acknowledgements** Shubin Wang (wangshubin@imu.edu.cn) is the correspondent author. This work was supported by the National Natural Science Foundation of China (61261020), and the Natural Science Foundation of Inner Mongolia, China (2016MS0616), and the PetroChina Innovation Foundation(2014D-5006-0603).

## References

1. P. Yi, A. Iwayemi, C. Zhou, Frequency agility in a ZigBee network for smart grid application, in *Proceedings of the Innovative Smart Grid Technologies* (2010), pp. 1–6
2. J. Tao, G. Noubir, S. Bo, WiZi-Cloud: application transparent dual ZigBee-WiFi radios for low power Internet access, in *Proceedings of IEEE INFOCOM* (2011), pp. 1593–1601
3. W.Z. Guan, L.B. Wu, X.W. Wu, L. Han, Design of remote seismic monitoring system based on GPRS. J. Geodesy Geodyn. **3**, 91–92 (2013)
4. Y.Y. Gao, Temperature measurement system based on STM32. J. Instrum. Monit. Anal. **1**, 16–18 (2010)
5. J. Huang, G. Xing, G. Zhou, R. Zhou, Beyond Co-existence: Exploiting Wi-Fi white space for ZigBee performance assurance, in *Proceedings of IEEE International Conference on Network Protocols*, vol. 18 (2010), pp. 305–314
6. W. Li, Y. Zhu, T. He, WiBee: building Wi-Fi radio map with ZigBee sensor networks, in *Proceedings of IEEE INFOCOM* (2012), pp. 2926–2930

# Equipment Maintenance Material Warehousing Based on Double-Layer Nested Internet of Things

**Peng Chen, Xiangjun Song, Deliang Liu, Yaozhou Liu and Wanling Li**

**Abstract** Aiming at the practical requirements of equipment maintenance material warehousing (EMM), the paper introduces double-layer nested Internet of Things into its management. First, the double-layer nested Internet of Things architecture is analyzed in the application of EMM warehousing. Then the paper designs the business model and describes the function made up by external layer and local layer. Based on these models, the implementing method of the Internet of Things overall is designed. At last, the method on the analysis of the law of the EMM consumption is studied.

**Keywords** Equipment maintenance material · Warehousing · Internet of things · Double-layer nested

## 1 Introduction

Equipment maintenance material (EMM) is defined as the least replaceable unit, the warehousing of which plays significant role in each parts as supplying, allocating, storing, distributing, and consuming, is the important interchanging node connecting the supplying upstream and the distributing downstream. With the increasing of exercising and training, the category and amount of new equipments rises. As a result, the EMM supply store amounts of tasks increase sharply, and EMM standards continue to change dynamically based on the consumption of a variety of complex operational condition. Traditional centralized storage and manual allocation of storage management mode have been difficult to adapt to the

P. Chen (✉) · X. Song · Y. Liu · W. Li
Institution of Mechanical Engineering, Shijiazhuang, China
e-mail: beimingke@163.com

D. Liu
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: liudeliang82@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_39

new demand, warehousing management of intelligent, information technology is imminent.

In recent years, the technology of Internet of Things is introduced into warehouse management, gradually making it developed from the original manual warehousing management to the modern warehouse management system Integrated with mechanization, automation, integrated, and intelligent [1, 2]. At present, for the maintenance equipment warehouse management is still in a single storage object, decentralized storage mode makes inventory and consumption information very difficult sharing in the nationwide, leading each warehouse storage to excessive with a large number of dull [3]. The factory does not master the EMM in warehouse quality condition, so life expires EMM cannot be rotated, resulting that no material can be used in the face of a large number of obsolete inventory [4, 5].

This paper proposed a centralized management method of the information of warehouses each dispersed in around through Internet of Things, connecting distributed storage to a network system, which covers the storage information and warehousing information related EMM storage, consumption, allocation, etc. This paper designs storage system architecture under the condition of Internet of Things, constructs the warehouse management module and the allocation/destruction management module, and studies the equipment consumption rule of statistical methods, by exploring the EMM management methods in the future.

## 2 Double-Layer Nested Internet of Things Architecture

According to the storage business needs of EMM, double-layer nested Internet of Things [3] constructs three platforms, such as item management, environmental management, safety management, forming the EMM warehouse and Internet of Things. The double-layer nested Internet of Things architectures marked up by the sensing layer operating in the form of library networking as nationwide and the communication layer operating in the form of Internet of Things with four sub-layers and two supporting architectures. The perception layers namely local layer and the communication layer are the external layer. As shown in Fig. 1.

Local layer of the Internet of Things is used to accomplish the information exchanging of the EMM in local warehouse, which integrates a series of new technology. Its item management platform adopts the RFID technology perception layer key technologies; its environmental management platform adopts temperature and humidity sensor, a water immersion sensor, smoke sensor, and sensing layer key technologies; its security management platform adopts video surveillance, infrared alarm, electronic access control, electronic patrolling for the key technology of the sensing layer, and supplemented by a variety of network technology; while it also is assisted by bus technology, database technology, information system integration technology build warehouse of the Internet of Things. These years, with the developing of the EMM warehouse building, most of the local layer.

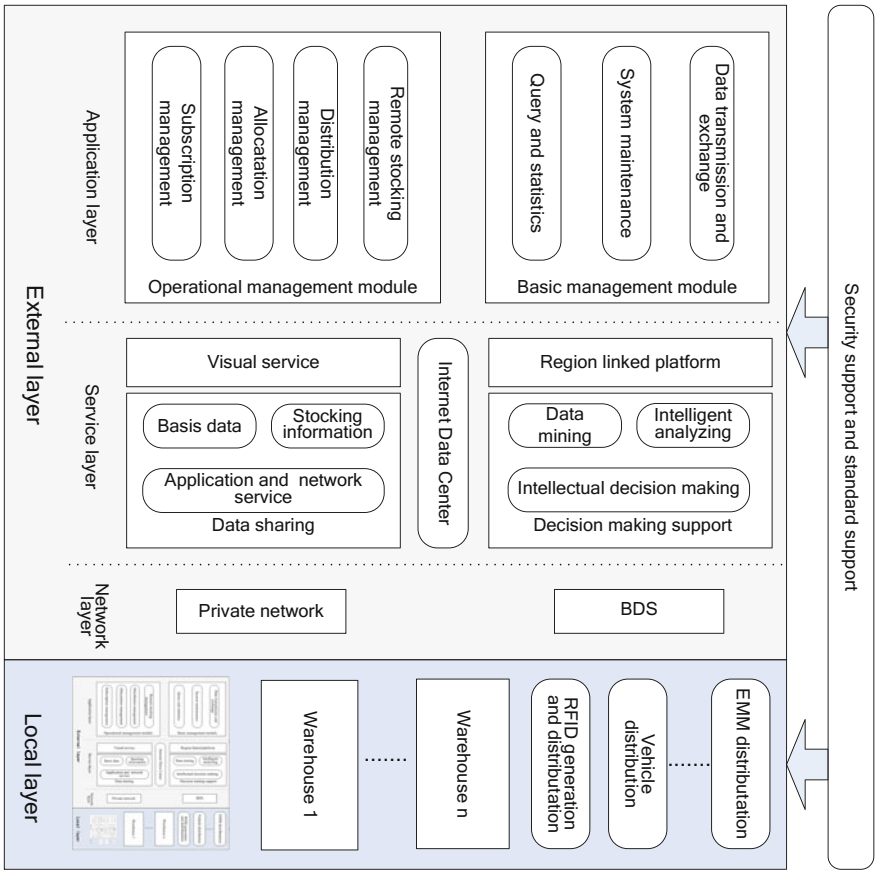



**Fig. 1** The double-layer nested Internet of things architecture

External layer of the Internet of things, that is, the national reserves of EMM, collecting information from the sensing layer, sending the RFID and video information of the warehouse and other information to the national EMM Internet of things through a variety of network technology. National EMM Internet of Things can master the single items of the current state and warehouse environment and security status information in the nationwide, and the information were analysed and processed. It could achieve nationwide storage business management and sharing and provide the help for the realization of the material efficient transportation.

The double-layer nested Internet of Things architecture has the characteristics of reasonable structure, well subsystem independence, strong expansibility, and easy to popularize.

# 3 EMM Warehousing Management Based on the Double-Layer Nested Internet of Things

## *3.1 Business Model*

The business background of the double-layer nested intelligent warehouse management system is about EMM dispatching national wide. The material species diversity and reserving warehouses geographical dispersion decide the complexity of the system. Here, we take a nationwide EMM storage and logistics application for example to interpret the system business model. According to EMM management based on the tertiary structure of the headerquaters level, the relay level, foundation level, at all levels of EMM management departments have the reserve warehouse. The EMM management departments at all levels responsible for the procurement, allocation transportation and traffic summary statistics work. Each warehouse is responsible for the storage of EMM business, including warehousing entry, warehousing output, maintenance, rejection, inventory, and other business.

The national wide EMM Internet of Things adopts a centralized computing model, with the establishment of data centers and the EMM management department at all levels use the system through hierarchical authorization. In this paper, take the procurement of warehousing business, for example, the business model is described in Fig. 2. Procurement and warehousing entry operations is initiated by the EMM management department. First, the EMM management department needs to enter the purchase order. After that, according to the purchase checklist, the allocation checklist is generated, and then it sends the allocation checklist to the warehousing which needs EMM.

After receiving the allocation checklist of the warehouse, the storage can generate warehousing entry checklist based on the allocation checklist, and then the warehouse administrator completes the actual storage operation according to the inventory, as shown in Fig. 3.

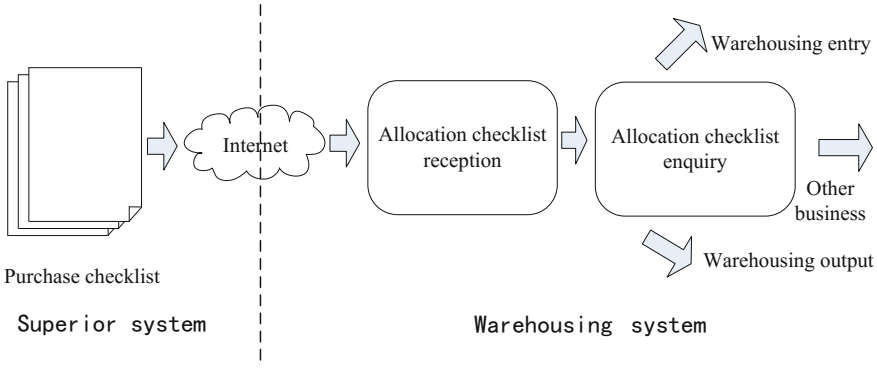



**Fig. 2** The allocation and reception business model

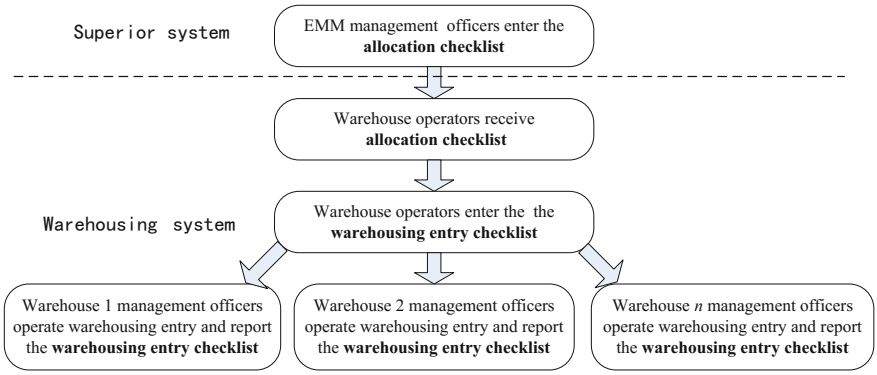



**Fig. 3** The warehousing entry business model

Detailed business process description is as follows:

(a) EMM management officers login the national EMM management information system, input purchase checklist.
(b) EMM management officers enter the allocation checklist, and send it to the warehouse management staff.
(c) the warehouse operators received requisition. Login the EMM warehouse management system and receive requisition.
(d) the warehouse operators enter the warehouse entry checklist and send the warehouse entry checklist to the warehouse administrator according to the allocation checklist.
(e) warehouse management officers complete the operation.

## *3.2 Function Model*

(1) National wide EMM warehouse system

The national EMM warehousing system is mainly composed of the following 10 modules, as shown in Fig. 4.

The following parts of the function of the module are introduced:

(a) allocation management: the main function is to manually receive the higher orders issued by the superior, and is able to be transcribed or transcribed change to send to other departments subsequently.
(b) procurement management: the main function is to develop procurement plans and procurement tasks.
(c) equerry and statistics: the main function is for inventory, warehousing entry, warehousing output, allocation, procurement, and other business statistics and inquiries.

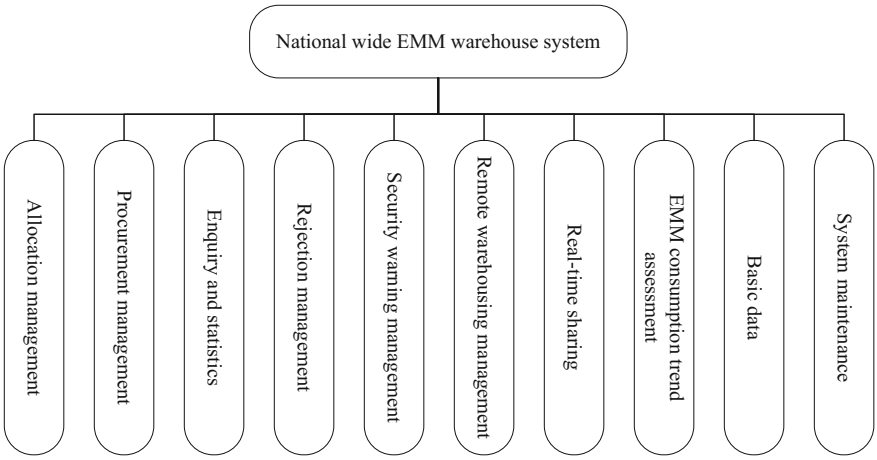



**Fig. 4** The function division of the national wide EMM warehouse system

(d) rejection management: according to the assessment of the quality of the short-period EMM, apply for repertory replacement if it exists overtime items.
(e) security warning management: the main function is to manage the inventory, the period of validity and so on.
(f) early warning management:
(g) remote warehousing management: one function is to monitor junior warehousing management for superior management department; another function is for codetermining when the super-expensive EMM items need warehousing entry or output.
(h) real-time sharing: it can avoid EMM repeatedly disposition and make it feasible to exchange EMM for emergency requirements.
(i) EMM consumption trend assessment: according to EMM consumption trend, the headquarters manager is able to decide the procurement plan with more accuracy.

(2) Local EMM warehouse management system

The function of warehouse management system is divided into 14 functional modules, as shown in Fig. 5.

The following parts of the function of the module are introduced:

(a) allocation checklist reception management: the main function is to receive allocation checklist from the national system.
(b) warehouse entry management: it is responsible for the storage of materials management, which is able to be divided into two major categories of business operations, that is, warehouse entry with or without allocation checklist.

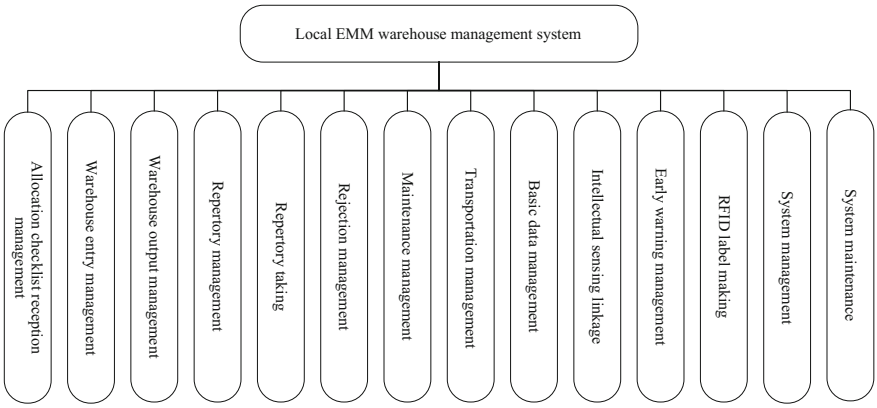



**Fig. 5** The function division of the local EMM warehouse management system

(c) warehouse output management: it is responsible for the delivery management, which is divided into two types of business operations, that is, warehouse output with or without allocation checklist.
(d) early warning management: it is responsible for the monitoring of the EMM items' inventory and valid date, and the warning of the upper and lower limits of the temperature and humidity. library and non-allocation of a single library operations.
(e) inventory management: it is responsible for statistics and equerry the inventory of materials, the warehousing materials entered and output.
(f) label production: it is responsible for the production and corresponding management of RFID labels.
(g) intelligent sensing linkage: according to the actual situation of warehouse, it can plot the location-based warehouse plan view, and then the warehouse managers can acquire the real-time data captured by the various sensing devices. When it encountered an emergency, it can automatic alarm and realize the visualization of the alarm data.

## *3.3 Implementing Method of the Internet of Things Overall Architecture*

The national wide Internet of Things is the outer layer of the double-layer nested Internet of things, and it is a system used by all levels of EMM management department. National system includes web server cluster, switching equipment, database server cluster, etc., and constitutes a typical three B/S architecture to support the business needs of all levels of material management.

The local EMM warehouse is the inner layer of the double-layer nested Internet of Things. Reserve library business management includes management of EMM items, in charge of management of the storehouse, the material security, and warehouse environment state. So, double-nested networking layer could accomplish the management of goods, safety, and environment through the comprehensive use of RFID, one or two-dimensional barcode, video surveillance equipment, infrared leaching of the radiation sensor, electronic fence, water alarm, temperature, and humidity sensor. The overall architecture is shown in Fig. 6.

Item management subsystem, referred to as the EMM management platform, is mainly used for the storage of goods in the warehouse management. It is mainly include by the system and item management module, including warehousing management, library management, inventory management, inventory management, maintenance management, scrap management, etc. Security management subsystem could aggregate the video surveillance system, electronic fence, electronic access control, and electronic patrolling to realize warehouse safety management. With a combination of warehouse material management information system, it can display on the page in a graphical way. Environmental management subsystem applies temperature sensors, humidity sensors, water immersion sensors, smoke sensors, etc., to carry out real-time control of the state of the warehouse environment.

## 3.4 Analysis of the Law of the Consumption of EMM

EMM is often allocated for different equipments in the form of "maintenance unit", in accordance with the EMM standards and each equipment delivered from the factory is equipped with a corresponding ID. According to the different types of

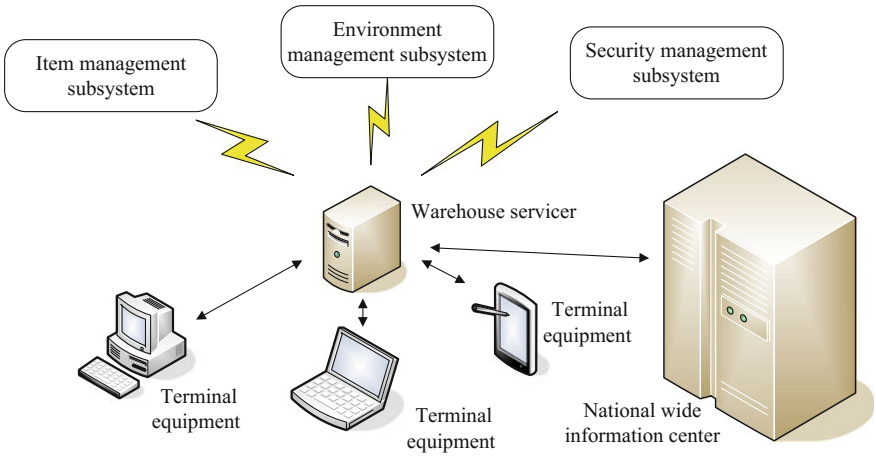



**Fig. 6** The overall architecture of the Internet of Things for EMM warehousing

equipments, the system summarizes the EMM of each warehouse through the data center. Then according to the division of a certain period of time, it operates inventory statistics for corresponding equipments and draws the curve, which is a EMM consumption curve for some specific equipments. A statistical model is to be established to fit the EMM consumption curve of a large number of equipment samples, which can be used to obtain the model of EMM consumption. In the follow-up of the statistical process, this model is constantly revised and can guide the equipment standard setting.

## 4 Conclusion

Internet of Things has significant perspective in EMM warehousing, as it is a strategic emerging technology. The paper introduced double-layer nested Internet of Things into its management, and analyzed the double-layer nested Internet of Things architecture in the application of EMM warehousing. After designing the business model and describe the function, the method on the analysis of the law of the consumption of EMM is studied. The research is definitely able to improve the informatization level on EMM warehousing and upgrade the equipment maintenance ability.

## References

1. J. Zhen, Q. Gan, Design and implementation of intelligent warehouse management system based on RFID. Electron. Des. Eng. **22**(7), 8–10 (2014)
2. Y. Yang, L. He, J. Zhai, Storage location optimization in digital warehouse management system. Ind. Control Comput. **29**(4), 107–109 (2016)
3. F. Zhao, X. Shi, Strategy analysis of enhancing military material containerization construction. Packag. Eng. **34**(9), 127–130 (2013)
4. Z. Li, K. Liu, Y. Zhao, et al., MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Let. **18**(11), 1915–1918 (2011)
5. Z. Li, Y. Chen, H. Shi, et al., NDN-GSM-R: a novel high-speed railway communication system via named data Networking. EURASIP J. Wirel. Commun. Netw. **48**, 1–5 (2016)
6. X. Liu, Z. Li, P. Yang, et al., Information-centric mobile AD Hoc networks and content routing: a survey. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.04.005
7. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference counteraction scheme. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.02.009

# Steiner-Tree-Based 2-Cut-set Network Coding Subgraph Algorithm in Wireless Multicast Network

**Feng Wei and Weixia Zou**

**Abstract** To improve throughput and decrease delay in the wireless multicast networks, this paper focuses on the alteration from routing tree to network coding subgraph. A Steiner-tree-based algorithm (STBNC) is proposed to form a 2-cut-set network coding subgraph. Random linear network coding can be employed in the outcome topology. Simulation results show that in terms of power cost and delay, the algorithm in this paper involves better performance than traditional D algorithm in ultra-dense situation with large amount of nodes and destinations. The algorithm utilizes the flexibility of multi-antenna channels in 60 Ghz.

**Keywords** Network coding subgraph · Steiner tree · Minimum power cost · Cut-set · 60 GHz

## 1 Introduction

With the evolution of wireless communications, people pay more attention to the rate and delay of a service. In academic prospect for 5G wireless communications, data peak rates require to exceed 10 Gb/s with access network latencies of the order of 1 ms [1, 2, 3]. Technical methods, such as massive MIMO, ultra-dense network, and millimeter wave frequency [2, 3], enable the wireless channel becomes more flexible to fulfill the increasing needs. Especially in 60 GHz, the sufficient path loss reduces the interference between different nodes by beamforming [3, 4, 5, 6]. However, simultaneous promotion in rates and delay performance always result in consuming more network resources, such as power cost. How to meet the rate and delay need in wireless networks with minimum cost becomes an urgent problem.

F. Wei (✉) · W. Zou
Key Laboratory of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing 100876, People's Republic of China
e-mail: weifeng_luck@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_40

Minimum cost multicast transmission can be expressed as constrained optimal problem [7]. When the cut-set value is 1, the solution of minimum cost multicast problem is Steiner tree [7, 8]. [9] Presented here is a polynomial-time algorithm that provides tradeoff between cost and run time efficiency.

Network coding subgraph is the solution of minimum cost multicast problem when the cut-set value is greater than 1 [10, 11]. Special network coding structures can be employed to form a coding subgraph. The butterfly structure was utilized with backpressure scheduler in [12, 13] to create subgraph. Another practical coding structure is COPE [14], which matches the broadcast characteristics of wireless channel. The grail structure exits in 2-unicast acyclic networks [15], but not suitable for multicast networks with all the flows sharing a common source node.

The approach to solve the minimum cost multicast problem produces optimal network coding subgraph. However, due to the large amount of computation, the optimal algorithm is difficult to achieve. The other way is to find substitution. [16] proposed a D algorithm method to create coding subgraph. [17] added coding structure on multicast routing tree. The ideas in [16, 17] can be utilized in the alteration from routing tree to coding subgraph to decrease the network cost and delay when the throughput is specified.

In this paper, we propose a Steiner-tree-based algorithm to create network coding subgraph with the cut-set value 2 in 60 GHz wireless multicast network. By adding new paths to the Steiner tree, the algorithm achieves a smooth transition from 1 cut-set routing tree to 2 cut-set subgraph.

The contents of the article are as follows: the second part will introduce the features of 60 GHz wireless channel, which are footstones for our algorithm. Algorithm will be introduced in the third part with some brief explanations. The fourth part is the simulation result. We will compare our algorithm with the other four algorithms, including that in [16, 17], and the last part is conclusion.

## 2 Model

In the wireless multicast scenario, it is assumed that each node operates in 60 GHz frequency with multiple antennas to perform beamforming, which can effectively isolate interferences. The multicast session is composed of a source node $s$ and a destination point set $D$, $s \notin D$. Source $s$ transmits information through a number of wireless nodes to each point in $D$. The model will involve the features of 60 GHz wireless channel and the minimum power analysis for our algorithm.

### *2.1 Features of 60 GHz Wireless Channel*

High loss and interference isolation are two main characteristics of 60 GHz wireless channel [18, 19]. As it is shown in Fig. 1, transmission in 60 GHz will endure a

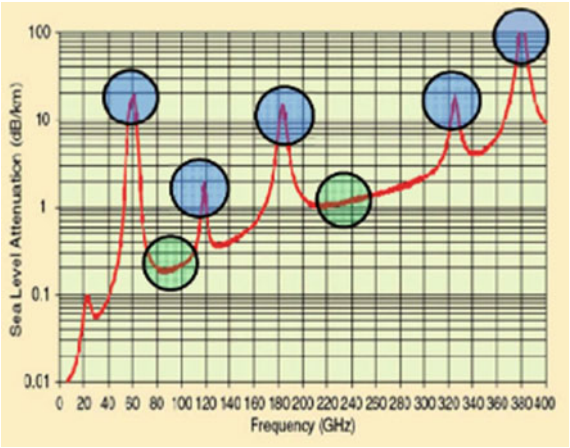



**Fig. 1** Atmospheric absorption across mm-wave frequencies in dB/km [18]

local maximum value of atmospheric absorption. High loss of channel makes 60 GHz not suitable for long distance transmission, but it inspires the beamforming in 60 GHz to effectively isolate the interference in the network.

With the aid of beamforming, the channel gain in 60 GHz frequency of node i can be expressed as follows [6]:

$$G_i(\theta) = \begin{cases} G_i^{(\max)} & \text{if } |\theta| \le \omega_i \\ G_i^{(\min)} & \text{if } |\theta| \le \omega_i \end{cases}, \tag{1}$$

where $\theta \in [-\pi, \pi)$ is the angle off the boresight direction, $\omega_i$ is the beam width of the main lobe, $G_i^{(\max)}$ and $G_i^{(\min)}$ are the array gain of main and side lobes, respectively. For simplification, we ignore the interference between different node, it means $G_i^{(\min)} = 0$. The minimum transmission power from node $i$ to node $j$ with beamforming is as follows [16]:

$$P_{ij}^{\theta} = G_i(\theta)\gamma_j N_0 \frac{\theta}{2\pi} d_{ij}^{-\alpha}, \tag{2}$$

where $\gamma_j$ is the signal to noise ratio received by node $j$; represents the system noise; $d_{ij}$ denotes the distance between node $i$ and node $j$. $\alpha$ is a channel attenuation coefficient. For simplification, the value of $\theta$ is fixed to a same $\theta_{min}$ for every sending node.

## 2.2 Power Cost Analysis

The goal of our algorithm is to set up a 2-cut-set value coding subgraph based on 1-cut-set value routing tree, and reduce the power cost of the network.

Given a destination $d$, there are two ways to increase the cut-set value from 1–2. One is to create a new path, the other is to increase the capacity of each channel in the existing path. Creating a new path will lead to power addition, but double the capacity of an existing channel would be more costly.

Suppose $C_0$ is the capacity of the existing path:

$$C_0 = \log(1 + \frac{P_r}{N_0}), \tag{3}$$

where $P_r$ is the power at the receiver. After doubling the capacity, the power at the receiver becomes $P'_r$.

$$2C_0 = \log(1 + \frac{P'_r}{N_0}) \tag{4}$$

Then,

$$P'_r = 2P_r + \frac{P_r^2}{N_0} = 2P_r + \gamma P_r \approx \gamma P_r \tag{5}$$

Equation (5) means that if doubling the capacity of a channel, the receiving power should be increased to $\gamma$ times. Consider Eq. (2), the transmitting power should also be increased to $\gamma$ times when the other conditions of the channel are unchanged.

## 3 Algorithm and Analysis

Before introducing our Steiner-tree-based network coding algorithm (STBNC), we will clarify some concepts at first.

Path is a sequential combination of channels from source to a destination. Connection is a sequential combination of channels from one node to another. Positive direction is the direction along the path from the source to a destination. Negative direction is the direction along the path from a destination to the source. Full information nodes are the source node and the nodes of 2-cut-set value. All the full information nodes compose full information set.

### *3.1 Algorithm*

The whole algorithm can be divided into two parts: The first part is building up Steiner tree, which can be replaced by that in [9], and the other part is creating network coding subgraph on Steiner tree.

1. At the beginning, all the destinations are unmarked.
2. Calculate the distance from source $s$ to unmarked destinations(s). Choose the destination whose distance from source $s$ is shortest. Mark this destination and attain its path as a branch of Steiner tree.
3. Change the positive direction cost of all the channels in the path selected in step 2 to zero, and the negative direction cost infinite.
4. Repeat step 2 and 3 until all destinations are marked. Then we get Steiner tree, and each destination $d_i$ has its $Path_i$ in this tree. Now, the in-degree of every node except the source in Steiner tree is 1.
5. Add $s$ to full massage set F. Reset all the destination to unmarked. Reset the cost of all the channels.
6. For every unmarked destination $d_i$, increase the cost of the connection $L_i$ from nearest full massage node in F to $d_i$ by $\gamma$ times. For destination $d_j \neq d_i$ set the positive direction cost of every channel $ch$ in $\{ch|ch \in Path_j, ch \notin Path_i\}$ to zero, and the negative direction cost infinite. Recalculate the distance from source $s$ to $d_i$.
7. Choose the destination whose distance from source $s$ is shortest. Mark this destination and attain its new path into coding subgraph. Find nodes of in-degree 2 in the new path, and add them into full massage set F.
8. repeat Steps 6, 7, and 8, until all the destination is marked. Then we get Steiner-tree-based network coding subgraph.

## *3.2 Analysis*

In this part, first, we will illustrate how the algorithm is reducing the power cost, and then give brief explanations for the algorithm.

Reason for algorithm reducing power cost: In the stage of building up Steiner tree, the positive destination cost of channels along the paths from source $s$ to marked destinations is zero. Then, for an unmarked node, the zero cost channels can be reused in the process of finding its own path. Similarly, in the stage of building up network coding subgraph, more zero cost connections can be reused. Thus, the algorithm reduces the power cost.

Explanation 1: After the algorithm, the cut-set value between source $s$ and each destination is 2.

When the Steiner tree is attained, the cut-set value between source $s$ and each destination is 1. The process of creating network coding subgraph brings a new connection from a full massage node for every destination. Then every destination holds a path from the source and a new connection from a full massage node. The cut-set value between source and a full massage node is 2, and then, the cut-set value between source $s$ and each destination is 2. ■

Explanation 2: The node of in-degree value 2 in the subgraph is a full massage node.

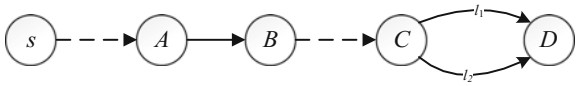



**Fig. 2** A structure in subgraph for apagoge

Suppose there is at least one 2-in-degree node whose cut-set value is 1. Then the structure in Fig. 2 will appear in the subgraph: the in-degree of node D is 2, but the cut-set value from source $s$ to $D$ is 1. Let channel $AB$ be the bottleneck. The solid line represents a connection of 1-capacity, and the dotted line represents a connection whose throughput is no less than 1.

In the following statement, we will see the structure in Fig. 2 could not exist in the coding subgraph generated by STBNC algorithm.

All the 2-in-degree nodes are produced in the process of creating coding subgraph. As STBNC algorithm generates one path at one time, let two paths through $D$ are formed in order of $P_1$ and $P_2$, So $P_2$ must be created in subgraph stage. Suppose path $P_1$ possesses $l_1$, path $P_2$ possesses $l_2$. The following are the only two cases.

Suppose $P_1$ is generated in Steiner tree stage, and $P_2$ is generated in coding subgraph stage. Then $P_1$ and $P_2$ could not be the two paths to the same destination, as the STBNC algorithm form 2-cut-set value structure for every destination. So $P_1$ and $P_2$ are of different destination. When selecting channel for $P_2$, the cost of channels on $P_1$ is set to zero. So there is no need to select channel $l_2$ when the cost of $l_1$, is zero. $l_2$ cannot exist.

Suppose both of $P_1$ and $P_2$ are generated in coding subgraph stage. When selecting channel $l_2$ for $P_2$, the cost of $l_1$ could not be zero. This case can only happen when the corresponding Steiner tree path of $P_2$ possesses $l_1$. Then, $D$ is on $P_2$ and its corresponding Steiner tree path at the same time. So $D$ can get the same information as the destination. It can be inferred that the cut-set value from source $s$ to $D$ is 2, which conflicts with 1-cut-set hypothesis. ■

## 4 Simulation

The simulation scenario consists of 100 nodes of random distributions in a unit area. The STBNC algorithm will be compared with NCBM [16], NCM [17], D algorithm and Steiner tree algorithm in terms of power cost and delay in 2-cut-set multicast wireless networks operating in 60 GHz.

Figure 3 is the simulation result of relative power cost. The dotted line in Fig. 3 represents double cost of corresponding algorithm as the 1-cut-set topology should transmit 2 times to achieve throughput of 2.

In the 2-throughput multicast scenario, STBNC perform better than D algorithm and NCBM, as its main goal is to reduce the power cost. NCBM is based on D algorithm and it forms a 2-cut-set graph. In terms of power cost, Steiner tree has the best performance, but we will see its weakness in terms of delay.

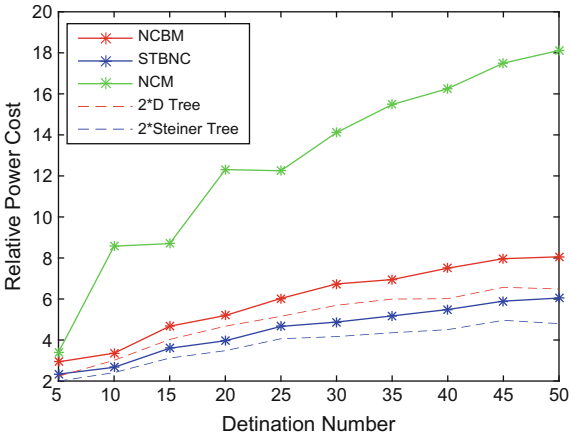



**Fig. 3** Relative Power Cost

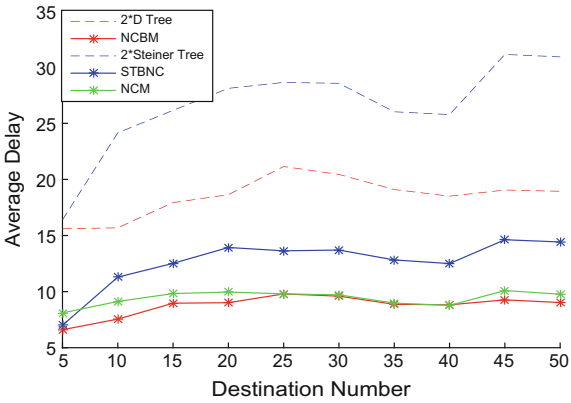



**Fig. 4** Average delay

In Fig. 4, when the throughput is 2, NCBM performs best in terms of delay, while NCM is close to it. STBNC algorithm is better than D algorithm and Steiner tree, as the cut-set value of STBNC is 2. As it shows, STBNC outperforms traditional D algorithm in both power cost and delay in ultra-dense situation.

In summary, STBNC is a smooth transition from Steiner tree to 2-cut-set value coding subgraph. Compared to Steiner tree, STBNC makes a sacrifice on power cost in change of better performance on delay. STBNC is not the best either in power cost or delay performance, but STBNC gains a favorable tradeoff between them and provides an attractive substitute for traditional D algorithm.

## 5 Conclusion

This paper presents an algorithm for establishing the coding subgraph of cut-set 2. The algorithm adds channel on the Steiner tree to gain low power cost network coding subgraph. In terms of power cost and delay, our algorithm has better performance than traditional D algorithm in multicast case of throughput 2. The idea of adding channel to routing tree can be employed to alter existing protocol and the method used in finding the 2-cut-set value coding subgraph could be extended to attain coding subgraph of 3 or more cut-set value.

**Acknowledgements** This work was supported by the 863 Program of China under Grant No. 2015AA01A703, the Fundamental Research Funds for the Central Universities under grant No. 2014ZD03-02,NSFC (No. 61571055), fund of SKL of MMW (No. K201501).

## References

1. J.G. Andrews, S. Buzzi, W. Choi et al., What will 5G be? IEEE J. Sel. Areas Commun. **32**(6), 1065–1082 (2014)
2. P.K. Agyapong, M. Iwamura, D. Staehle et al., Design considerations for a 5G network architecture. IEEE Commun. Mag. **52**(11), 65–75 (2014)
3. R. Baldemair, T. Irnich, K. Balachandran et al., Ultra-dense networks in millimeter-wave frequencies. IEEE Commun. Mag. **53**(1), 202–208 (2015)
4. J. Choi, Iterative methods for physical-layer multicast beamforming. IEEE Trans. Wirel. Commun. **14**(9), 5185–5196 (2015)
5. S. He, Y. Huang, H. Wang et al., Leakage-aware energy-efficient beamforming for heterogeneous multicell multiuser systems. IEEE J. Sel. Areas Commun. **32**(6), 1268–1281 (2014)
6. M. Di Renzo, Stochastic geometry modeling and analysis of multi-tier millimeter wave cellular networks. IEEE Trans. Wirel. Commun. **14**(9), 5038–5057 (2015)
7. L.H. Sahasrabuddhe, B. Mukherjee, Multicast routing algorithms and protocols: a tutorial. IEEE Netw. **14**(1), 90–102 (2000)
8. M. Charikar, C. Chekuri, T. Cheung et al., Approximation algorithms for directed Steiner problems. J. Algorithms **33**(1), 73–91 (1999)
9. S. Ramanathan, Multicast tree generation in networks with asymmetric links. IEEE/ACM Trans. Netw. (TON) **4**(4), 558–568 (1996)
10. D.S. Lun, N. Ratnakar, R. Koetter, et al., Achieving minimum-cost multicast: a decentralized approach based on network codingin, in *Proceedings IEEE 24th Annual Joint Conference of the IEEE Computer and Communications Societies*, vol 3 (IEEE, 2005), pp. 1607–1617
11. K. Bhattad, N. Ratnakar, R. Koetter, et al., Minimal network coding for multicast, in *IEEE Proceedings of International Symposium on Information Theory ISIT,* (2005), pp. 1730–1734
12. A. Eryilmaz, D.S. Lun, B.T. Swapna, Control of multi-hop communication networks for inter-session network coding. IEEE Trans. Inf. Theor. **57**(2), 1092–1110 (2011)
13. L. Chen, S.H. Low, M. Chiang, et al., Cross-layer congestion control, routing and scheduling design in ad hoc wireless networks (2006)
14. S. Katti, H. Rahul, W. Hu et al., XORs in the air: practical wireless network coding. ACM SIGCOMM Comput. Commun. Rev. **36**(4), 243–254 (2006)
15. C.C. Wang, N.B. Shroff, Pairwise intersession network coding on directed networks. IEEE Trans. Inf. Theor. **56**(8), 3879–3900 (2010)

16. J. Wang, Y. Li, X. Wang, Network coding based multicast in internet, in *2007 International Conference on Parallel Processing Workshops (ICPPW 2007)* (IEEE, 2007), pp. 44–44
17. D. Jiang, Z. Xu, W. Li et al., Network coding-based energy-efficient multicast routing algorithm for multi-hop wireless networks. J. Syst. Softw. **104**, 152–165 (2015)
18. T.S. Rappaport, J.N. Murdock, F. Gutierrez, State of the art in 60-GHz integrated circuits and systems for wireless communications. Proc. IEEE **99**(8), 1390–1436 (2011)
19. T.S. Rappaport, S. Sun, R. Mayzus et al., Millimeter wave mobile communications for 5G cellular: it will work! IEEE Access **1**, 335–349 (2013)

# A Novel Space Information Network Architecture Based on SDN

**Gaoling Chen, Xiao Peng, Chenglin Zhao and Fangmin Xu**

**Abstract** The Space Information Network (SIN) is a full-spatial, full-time, full-frequency, multi-users-oriented information network which has the characteristics of complexity, heterogeneous, and openness. In this paper, we propose a software-defined space information network (SDSIN) to solve above-mentioned problems. This architecture based on the core idea of Software-Defined Network (SDN) that separate control and forwarding panel, and takes full use of the global coverage properties of Geostationary Orbit (GEO), powerful computing capacity of ground station, the predictability and regularity of constellation and the forwarding capacity of inner-satellite links (ISLs). Thus the network can allocate and optimize network resources from a global perspective so as to achieve flexible and efficient network configuration and management, and realize direct and effective control of spatial information network.

**Keywords** Space information network · Software-defined network · Virtualization · Routing · Resource scheduling

G. Chen (✉) · C. Zhao (✉) · F. Xu (✉)
Key Lab of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing, China
e-mail: chengaoling@bupt.edu.cn

C. Zhao
e-mail: clzhao@bupt.edu.cn

F. Xu
e-mail: clzhao@bupt.edu.cn

X. Peng (✉)
The State Radio Monitoring_Center Testing Center, Beijing, China
e-mail: pengxiao@srtc.org.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_41

## 1 Introduction

Space information network is a network system based on space platform as the carrier which can acquire, transmit, and processing space information in real time. As the national important infrastructure, the space information network provides service to long ocean voyages, emergency rescue, navigation, air transport, and aerospace measurement and to support earth observation of high dynamic, broadband transmission real-timely, can support explorations of very long range, reliable transmission of large time delay at the same time, and thus to expand human science, culture, production activity in space, ocean, and deep space, is a hot spot in the global scope. The characteristics of complexity, heterogeneity, and dynamic of the space information network have brought enormous challenges to the research of space information network. SDN is a kind of implementation of network virtualization, its core technology OpenFlow separate network into control plane and data plane, so as to achieve flexible and efficient network configuration and management, can realize effective control of space information network directly. This paper presents a space information network architecture based on SDN, make the space network has high availability and reliability, at the same time to ensure that the system has good scalability, reusability and portability, the development and application of space information network architecture is of great significance.

## 2 Design of Space Information Network Architecture

Software-defined space information network architecture is designed by adopting the idea of SDN, the traditional network is separated into control plane and data plane. Space information network can be divided by the GEO satellite layer, LEO satellite layer and ground from a physical point of view. It can be divided into application layer, control plane, and data plane logically as well, GEO satellite and ground take the role of controller, and the low earth orbit (LEO) and medium earth orbit (MEO) satellite are responsible for data forwarding. Using global coverage characteristic of the GEO satellites, GEO satellites mainly are responsible for the entire network topology information acquisition; the ground controller can construct forward strategy of load balance and routing algorithm on the basis of network topology information [1], realize equilibrium distribution of the business in the whole space network and optimal routing of data flow. The core design of software-defined space network includes two aspects, namely design of software-defined space network controller system, design of routing and resource scheduling strategy. The following respectively expounds the design principle of two parts:

## 2.1 Design of Software-Defined Space Network Controller System

The architecture can be physically divided into three layers: space-based backbone network, data forwarding layer, and ground harbor. space missions call all kinds of network resources in the form of software programming through the northbound interface (NBI), while the control plane is responsible for link discovery, topology management, strategic planning, entry issued, etc., through southbound interface (SBI), communications and contacts between multiple controllers accomplished through the east–west interface, and data plane only responsible for simple data transfer.

(1) *Introduction of network operating system*: Network operating system is the core of the whole architecture, it can be divided into physical resources abstraction layer, virtual slice layer, network control, and management layer and resource scheduling and business adaptation layer from down to up. Figure 1b shows that the control plane can be subdivided into resource scheduling and business adaptation layer, network control and management sub-layer, virtual slice sub-layer and physical resources abstraction sub-layer.

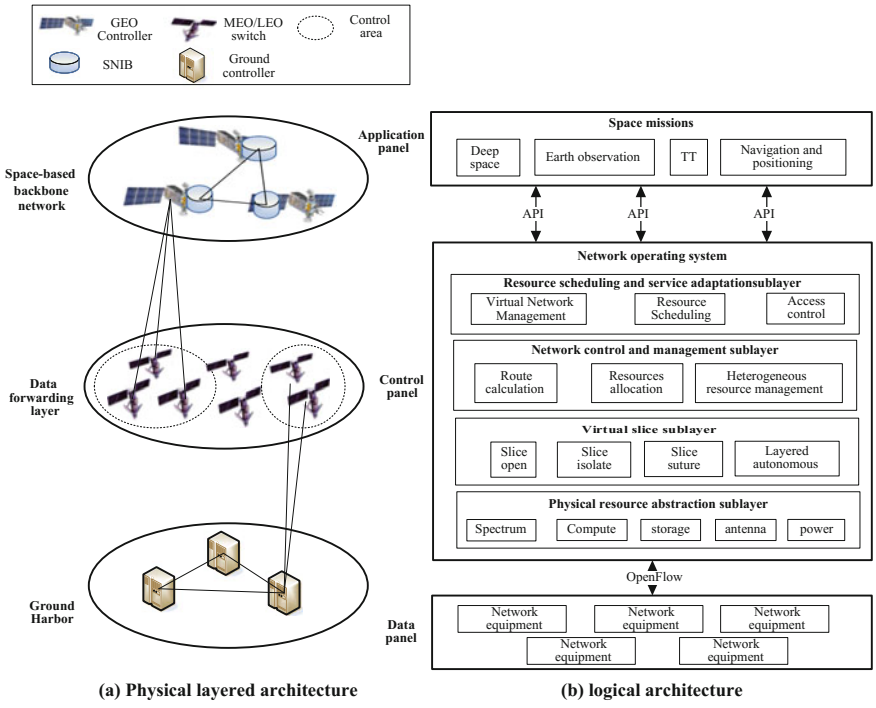



**Fig. 1** Space information network architecture

(a) *Physical Resources Abstraction Sub-layer*: Take advantages of network resources virtualization methods, depending on the mission objectives and needs, abstracts multidimensional resources (antenna, power, frequency, computing, storage, etc.) of the underlying facilities to build multidimensional space resource pool, which can facilitate the achievement of limited space resources maximum utilization.

(b) *Virtual Slice Sub-layer*: Space Information Network Resources Systems at the request of each space virtual network through abstraction and the isolation mechanism allocate underlying physical space in the form of slices to different virtual network to meet their requirements. The underlying space network nodes and resources can be more than one slice share. Each slice run different control logic and high-level policy, and independently of each other by applying Isolation mechanism. The controller can manage a plurality of spatial nodes, a plurality of network space experiments can be run simultaneously on the same underlying physical network in slice-based control mode.

(c) *Network Control and Management Sub-layer*: Network control means superimposed virtual space network will be software-programmable manner unified control. By applying this scheme, the space network resources and computing resources, storage resources and other resources are unified together and on-demand delivery. Various virtualization space network devices to work together under the unified control of space resource management platform, through the virtual space network between space nodes, achieving spatial network resource virtualization.

(d) *Resource Scheduling and Business Adaptation Layer*: After abstracting the underlying network service to a variety of available resources, we need to study the approach to use of resource and resolution mechanisms to resource conflict. We designed Space Information Network Resource Sharing Mechanism based on SDN according to priority and arrival time of tasks by combining dynamic resource scheduling method and static scheduling, which provide as much as possible high-quality support services for the business on the basis of ensuring efficient utilization of network resources.

(2) *Accurate storage of topology information*: In this paper, with reference to the Network Information Base (NIB) [2], we proposed SNIB concept. Take advantage of periodicity and predictability of satellite constellation operation, the constellation cycle is divided into several time slices and in each time slice network topology can be viewed as a virtual topology [3], namely, a fixed topology snapshot of information managed by the SNIB, in order to allow the controller have a more detailed understanding of the network status and achieving more precise control of network, the SNIB entities classes illustrated in Fig. 2.

As shown in Fig. 2, solid lines represent inheritance, while dashed lines correspond to referential relation between entity instances. The numbers on the dashed

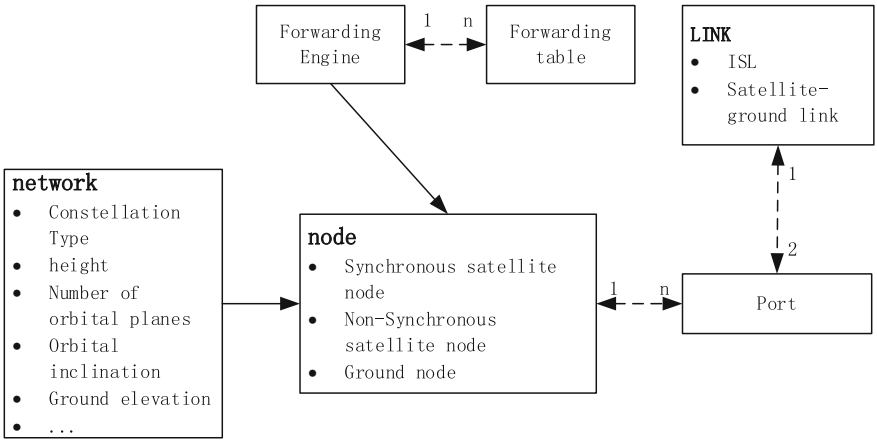



**Fig. 2** The default network entity classes

lines show the quantitative mapping relationship (e.g., one Link maps to two Ports, and two Ports can map to the same Link). Nodes, ports and links constitute the network topology.

All entity classes inherit the same base class providing generic key-value pair access. All typed entities have a common base class limited to generic key-value pair access. The NIB provides multiple methods for the control logic to gain access to network entities. It maintains an index of all of its entities based on the entity identifier, allowing for direct querying of a specific entity. It also supports registration for notifications on state changes or the addition/deletion of an entity.

- Node Entity: used to define different type of node like synchronous satellite nodes, non-synchronous satellite nodes, and ground node;
- Network Entity: used to define the parameters of the satellite constellation, including height, orbital plane, orbital inclination, ground elevation, etc.
- Link Entity: used to define corresponding attributes of satellite link like physical layer protocol, MAC protocol, bandwidth usage, throughput, and load, etc.

## *2.2 Design of Routing and Resource Scheduling Strategy*

In the proposed space information network architecture, ground controllers calculate the best path for each stream based on different optimization goal, to provide differentiated services for different business goals to generate offline flow table; Provide low latency service for emergency important applications, to provide high-bandwidth services for video applications, the most common do my best service for the general application[4]; then through the "soft switch" approach to reduce the data jitter brought by link switching.

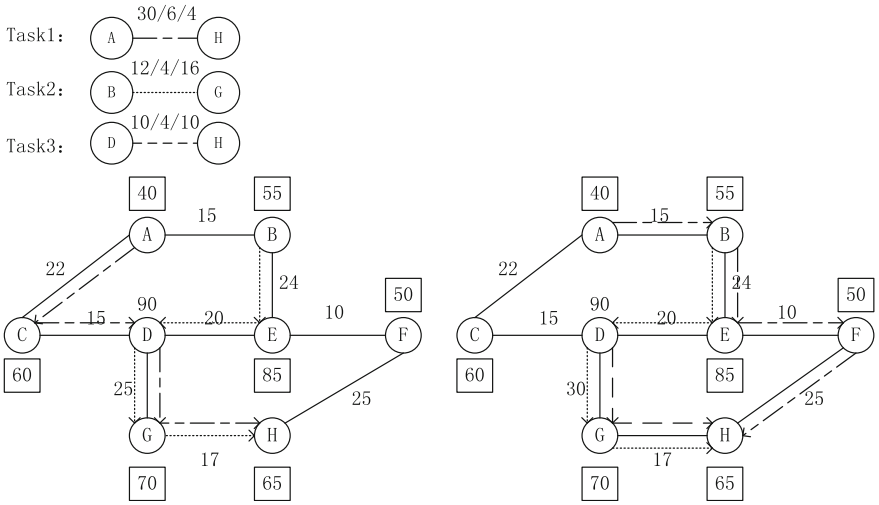



**Fig. 3** Flow-table generation schematic drawing

The flow table can be calculated through multi-objective optimization algorithm [4] by quantizing the demand of flow into latency and bandwidth two indicators [5]. Latency can be divided into processing delay and propagation delay, the former can be simplified as CPU processing capacity of the satellite nodes, and the latter can be simplified to hops.

As shown in Fig. 3, the solid line indicates the current topology, the number inside the box represents the CPU processing capacity of the node, the number on the solid line represents the current total bandwidth of the route; three digital for specific tasks represents CPU demand, the maximum allowed number of hops and the link bandwidth respectively. In Fig. 3a, the transmission path of the first task of is "A-C-D-G-H", the transmission path of task 2 is "B-D-G-H", and task 3 is rejected because the optional path "D-G-H" and "D-E-F-H" cannot meet the bandwidth and latency requirements of task 3, so the controller recalculates the transmission path, the result is shown as Fig. 3b.

Multi-objective optimization algorithm considering the use of satellite application resources and network resources, can optimize the satellite application and network resource and improve the system performance under the condition of increasing in access to network applications.

## 3 Benefits

With respect to terrestrial networks, space-based network has some particularities which affect network performance: Restricted function of satellite and other reasons resulting in poor on-board processing capability; dynamic topology makes routing

instability [6]; Inconsistent communication technologies and protocols weaken the collaboration capabilities of nodes; in terms of various mission requirements, lacking of effective methods to scheduling different types of nodes flexibly to accomplish complex space missions depending on mission requirements currently. Taking the characteristics of space-based network into account, the traditional ground-based networks architecture theory cannot fully migrate to the network space, especially the dynamic routing algorithm, the access load balancing, etc. In this paper, the architecture can be a very good to solve the above problems.

### *3.1 Heterogeneous Network Convergence*

The core idea of separating control and forwarding panel [7] makes unified management of all network equipment possible in the case of heterogeneous network; in addition, the entire satellite system can be programmable from the physical layer to the network layer using SDN and SDR [8] simultaneously. Furthermore, SDN abstracts forwarding and routing table, integrates various levels of network configuration information, makes it possible to handle and response a variety of network protocols simultaneously, so that the problem of heterogeneity of network protocols get a good solution. Therefore, the problem of heterogeneous network interconnection can be handled in the future of space information network.

### *3.2 Flexible Resource Scheduling*

The new architecture abstracting the underlying facilities multidimensional (antenna, power, frequency, computing, storage, etc.) resources, utilizing network resources virtualization methods, to build multidimensional space resource pool depending on the mission objectives and needs, which maximize the utilization of space resources easily; in addition, the control panel has a global view of network resources and achieve a programmable control, thereby control panel can allocate and schedule network resources according to different services demand from upper layer, and ultimately achieve the goal of making the entire network can be controlled, measured, and scheduled, deploying new services rapidly, enhancing utilization of network resources, ensuring quality of service.

### *3.3 Flexible, Scalable Network Control*

Centralized control characteristics of SDN simplify and standardize the processing functions of the satellite nodes, complex network control, and computing functions are deployed in the ground controller, network configuration can be quickly

deployed and updated in this architecture. In addition, GEO controller can be generated global network routing and configuration policies based on a global network view, and then combined with regularity and predictability of satellite network topology, the controller can deploy global routing and configuration policies in advance which can adapt the time-vary network topology flexibly. Thus, relative to traditional network, the new network control architecture is more flexible and scalable.

### *3.4 Energy Efficient*

MEO/LEO satellites are only responsible for simple forwarding and configuration function, all the complex control logic fully deployed on the ground, the costs of designing and realizing satellites can be greatly reduced. In addition, the strategy that uses the GEO satellite takes the responsibility of finding global view reduce the number of ground station required, and then reduces the overall investment of the system.

## 4 Conclusion

This article discusses the architecture of software-defined space information network, expounds the principle of the space network database and the main function of the control plane, put forward routing and resource scheduling strategy in view of the satellite network, and contrasts the advantages of this architecture in theory. The architecture is the well solved problem of heterogeneous network integration and can provide more flexible resource scheduling and network control, and at the same time reduce the system's overall investment.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China (61431008).

## References

1. T. Benson, A. Akella, D. Maltz, Unraveling the complexity of network management, in *Proceedings of the 6th USENIX Symposium. Networked System Design Implementation* (2009), pp. 335–348
2. T. Koponen, M. Casado, N. Gude, et al., Onix: a distributed control platform for large-scale production networks, in *Usenix Conference on Operating Systems Design and Implementation* (USENIX Association, 2010), pp. 351–364
3. J. Feng, L. Jiang, Y. Shen, et al., A scheme for software defined ORS satellite networking, in *2014 IEEE Fourth International Conference on Big Data and Cloud Computing (BdCloud)* (IEEE, 2014), pp. 716–721

4. M. De Sanctis, T. Rossi, L. Rizzo, M. Ruggieri, G. Codispoti, Optimization of ACM algorithms over Q/V-band satellite channels with the Alphasat Aldo Paraboni P/L, in *Proceedings of the IEEE Aerospace Conference 2015*, Big Sky (Montana), March 2015
5. G. Song, M. Chao, B. Yang et al., Research on multi-path QoS routing strategy for the satellite network. Lecture Notes Electr. Eng. **187**, 289–298 (2013)
6. L. Bertaux, S. Medjiah, P. Berthou et al., Software defined networking and virtualization for broadband satellite networks. IEEE Commun. Mag. **53**(3), 54–60 (2015)
7. D. Kreutz, F.M.V. Ramos, P. Esteves Verissimo et al., Software-defined networking: a comprehensive survey. Proc. IEEE **103**(1), 10–13 (2014)
8. A.W. Mast, Reconfigurable software defined payload architecture that reduces cost and risk for various missions, in *Aerospace Conferenc, 2011 IEEE* (IEEE, 2011), pp. 1–5
9. T. Rossi, M.D. Sanctis, E. Cianca, et al., Future space-based communications infrastructures based on high throughput satellites and software defined networking [C], in *IEEE International Symposium on Systems Engineering* (IEEE, 2015)

# The Impact of Distributed Generation Parallel Operation on Smart Grid

**Yufan Lei, Guanglin Han and Yanqun Wang**

**Abstract** This paper introduces the concept of distributed generation resources and smart grid, and analyzes the impact of distributed generation parallel operation on three aspects of network loss, voltage and relay protection.

**Keywords** Smart grid · Distributed generation parallel operation · Network loss · Relay protection

## 1 Introduction

Distributed generation resources (DGRs) has the advantages of energy saving, power generation, flexible local consumption, etc., and is a useful complement to the centralized power supply mode. Parallel operation is a common operation mode of distributed power supply, which has a great impact on smart grid. It needs to strengthen research and reduce the negative impact, so that the positive role has been fully played. Therefore, the analysis of the impact of distributed generation resources grid for smart grid has a significant contribution to the construction of a more perfect, reliable and secure smart grid system.

## 2 Distributed Generation Resources

DGRs usually refers to small modular of the power from several kilowatts to 50 MW, independent power of compatible with the environment, owned by the power sector, users or third parties, to meet the specific requirements of the

Y. Lei
Academy of electronics and communication engineering, Tianjin Normal University, Tianjin 300387, China

G. Han (✉) · Y. Wang
Tianjin Normal University, Tianjin 300387, China
e-mail: hgltj@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_42

power system and users [1]. DGRs emit Electrical energy cannot be directly supplied to the AC load, so it must be connected to a certain interface to parallel operation.

## 3 Smart Grids

Smart grid uses communication technology and information technology to optimize the transmission and distribution of electricity from suppliers to consumers. Generally speaking, smart grid is combining advanced sensing quantity measuring technology, information and communication technology, decision analysis technology, automatic control technology and power electronics technology, and is the new modern power grid which is highly integrated with the power grid infrastructure [2, 3].

## 4 The Impact of DG Parallel Operation on Smart Grid

The wide application of DGRs will have a great impact on the traditional power system, including the power quality, system reliability, relay protection and so on. It will be better to guide the construction of smart grid through the study of the impact of DGRs on the power quality of smart grid. This paper mainly analyzes the impact of the DG parallel operation on the network loss, voltage and relay protection [4].

### *4.1 The Impact of DG Parallel Operation on Network Loss*

It is inevitable for distribution network to produce loss in the transmission and distribution process, even centralized power supply often with the net loss of 10%, but DG parallel operation is able to reduce net loss. Taking a 33-bus test system for example, as shown in Fig. 1.

There is parallel of DG and distribution network in the nodes 25 of the test system. Test results are shown in Table 1.

Table available, Interconnection capacity increase can significantly reduce the active power loss.

In the node 27, selecting leading power factor 0.8, lagging power factor 0.8, and the unit power factor of three cases and people. Test results: the effect of the reactive power of the lagging power factor is better than that of the unit power factor and the leading power factor. And the reactive power can compensate the load, and then significantly reduce the reactive current of the line, to achieve the effect of reducing the network loss.

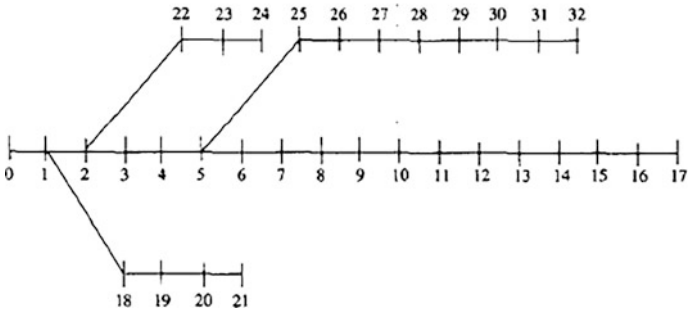



**Fig. 1** 33 bus test system

**Table 1** Network loss of DGRs with different grid connected capacity

| Grid connected capacity/kW | Voltage/kV | Active power loss/kW |
|---|---|---|
| 90 | 0.9440 | 212.8153 |
| 60 | 0.9432 | 215.9470 |
| 30 | 0.9426 | 219.3762 |
| 0 | 0.9415 | 222.1083 |

In addition, if the distributed power supply capacity is the same, and the position of the distribution network parallel is different, the degree of network loss will not be the same. In the test system, the node 0 is the balance node, the closer the distance is, the worse the voltage support effect is, and the bigger the network loss is.

## *4.2 The Impact of DG Parallel Operation on Distribution Network Voltage*

The impact of DG parallel operation on distribution network voltage is mainly including the voltage fluctuation and the voltage distribution.

### 4.2.1 The Impact of the Voltage Fluctuation

The impact of DG parallel operation on voltage fluctuation depends on the specific circumstances. If the DG and the load coordination operation, the DG will restrain the voltage fluctuation; if not, the voltage fluctuation will be more obvious. If the DG access location, capacity and control is not reasonable, it will increase the load flow on distribution lines, increase the difficulty of the adjustment of the distribution network voltage, and increase the frequency of voltage fluctuation.

In order to direct the processing of various types of conventional node, and provide more convenient conditions when transform the algorithm of a variety of different kinds of DG, at present, Newton-Raphson method is widely used in the power flow calculation processing. The method has good convergence ability, and it is also suitable for the power flow calculation of distribution network. Therefore, the Newton-Raphson method is used for power flow calculation for each node type [5].

To the treatment for PQ and PV node type of DG technology can be directly used Newton-Raphson method for power flow calculation. To the treatment for PQ (V) node type of DG, it is discussed for synchronous motor and asynchronous motor.

For the synchronous generator, such as the hidden pole synchronous generator, it issued the reactive power Q and active power P:

$$Q = \frac{E_q V}{x_d}\cos\delta - \frac{V^2}{x_d}, P = \frac{E_q V}{x_d}\sin\delta \tag{1}$$

In the formula, $E_q$ refers to the electric potential of the hidden pole generator in no-load, and it is desirable for a certain value; V is the voltage at the terminal of the hidden pole generator; $x_d$ is the synchronous reactance of the generator set; $\delta$ is the power angle. When the active power is given to a certain value, the relationship between the reactive power and the generator terminal voltage can be derived from the formula (1):

$$Q = \sqrt{(\frac{E_q V}{x_d})^2 - P^2} - \frac{V^2}{x_d} \tag{2}$$

When the active power is a certain value, the output value of the reactive power of the motor can be calculated by the formula (2).

For the asynchronous generator without excitation system, it can establish the excitation system through from the power grid absorb reactive power. Make an equivalent circuit for an asynchronous generator. Where $x_1$ is the stator leakage reactance, $x_2$ is the rotor leakage reactance, $x_m$ is the excitation reactance, $r_2$ is rotor resistance, and s is slip, ignoring the stator resistance. The formula is derived from these parameters:

$$V = \sqrt{\frac{-P[s^2(x_1+x_2)^2 + r_2^2]}{r_2 s}}, Q = -[\frac{V^2}{x_m} + \frac{P(x_1+x_2)}{r_2}] \tag{3}$$

When the output value of active power is a certain value, the relationship among reactive power and active power and terminal voltage can be derived from the formula (3):

$$Q = -\frac{V^2}{x_m} + \frac{-V + \sqrt{[V^2 - 4P^2(x_1+x_2)^2]}}{2(x_1+x_2)} \tag{4}$$

When the reactive power is negative, it is explained that the asynchronous generator absorbs reactive power from the grid.

For the PI node type, because the constant active power output and access to the grid current is known, the reactive power of DG output can be calculated by the relationship among voltage, current, active power and reactive power:

$$Q = \sqrt{I^2V^2 - P^2}r \tag{5}$$

### 4.2.2 The Impact the Voltage Distribution

The voltage distribution model can be established according to simple distribution line of containing distributed generation and superposition theorem of circuit. Assume that the voltage of the beginning of circuit is U, the rated voltage of circuit is U, and d is any point on the circuit, so the corresponding voltage of d point can be calculated by the upper model:

$$U_d = U - \Delta U = U - \Delta U_{DG} - \Delta U_{sd}$$

In voltage distribution calculation, there is a need to consider the impact of DG and system power supply on distribution line, therefore, when the access distribution network the voltage loss includes two parts: the loss caused by DG ($\Delta U_{DG}$) and system power ($\Delta U_{sd}$).

The DG parallel operation will affect the voltage distribution on the feeder. Generally speaking, if incorporated position is fixed, voltage support increases with increase of total output, voltage level is also increased; if the total output is fixed, the closer distance power system bus, the greater the impact on the voltage distribution. If the DG parallel load is larger, the system voltage should be reduced reasonably, and the voltage of local node should be increased appropriately.

## *4.3 The Impact of DG Parallel Operation on Relay Protection*

The distribution network is a multiple power network after DG parallel operation. Once a fault occurs, the fault will form a fault current which for current level to cause certain impact. The capacity of DG and reactance value determines the extent of the impact.

When there is a large current produced and flows through in the distribution network, temperature of equipment rise very easy to burn out the internal components, and quickly spread to the surrounding sound components, ultimately resulting in damage of the equipment. In order to solve the problem usually installed fuse protector, when large current flows through, fuse automatically

burned, and isolate the fault zone, so that avoiding damage to the ambient circuit. After the DG is connected, the protection devices must have the accurate direction. Obviously, the existing protection devices can't meet this requirement, the method of use directional relay to replace the fuse is not appropriate [6–8].

Main feeder select three-stage current protection mode to ensure when the feeder fault occurs, to restore power supply at the shortest time. The DG parallel operation connected to the line is divided into two sections, there may be a side of the double power supply, one side of the single power supply situation, so reducing the sensitivity of the line protection action, and even appear the rejection phenomenon. In addition, the DG parallel operation may also cause the protection malfunction of the circuit, or lead to Instantaneous quick break protection malfunction of nearby line, loss of selectivity.

## 5 Conclusion

Distributed generation parallel operation have significant impact on the voltage, network loss, relay protection, electrical reliability, etc. The rapid development of smart grid provide favorable support for access and stable operation of DG, but it is still necessary for distributed large-scale generation to access network in-depth study and reduce the adverse impact on the grid.

## References

1. H. Xuehao, Problems on distributed-generation resources and its networking. Electr. Eng. **10**, 1–5 (2004). (in Chinese)
2. Y. Yu, P. Luan, Basic philosophy of smart grid. J. Tianjin Univer. **05**, 377–384 (2011)
3. Z. Liu, To build the robust and modern smart grid. State-Own. Enterp. **11**, 26–27 (2011)
4. Z. Gao, X. Wu, Study of influence of distributed generation integration on distribution network loss. Jilin Electr. Power **31**(2), 17–19 (2013)
5. W. Su, Z. Lin, X. Li, W. Hu, L. Wang, Research on impact of distributed generation on static voltage stability of distribution networks. Electr. Meas. Instrum. **14**, 41–46 (2014)
6. W. Shi, P. Song, C. Liu, Power supply reliability of distribution network including distributed generation. East China Electric Power **35**(7), 37–41 (2007)
7. X. Sui, Q. Wang, Y. Yang, et al., Research on impact of distributed generation parallel operation on stability of distribution networks. Sichuan Electric Power Technol. **33**, 31–34 (2010)
8. J. Pang, X. Xia, M. Fang, Impact of distributed generation to relay protection of distribution system. Relay **35**(11), 5–8 (2007)

# Part IV
# Coding, Encryption & Algorithm Design

# Study on the Signal Synthetic Method Based on STD

**Jin Luo, Cheng Wang and Xi Wen**

**Abstract** As the complexity and functionality of electronic equipment increases over time, testing of modern day electronics has brought about the need for more complex stimulus and measurement capabilities. STD (Signal and Test Definition) standard provides the accurate definition for signals and the libraries of predefined basic signals which may be reused and extended. The signal synthetic method based on STD is illustrated in this paper. To support the signal synthetic mechanism, a signal composition component is developed, that enabled the automatic generate and synthesis of basic signals and related to the real measurement facilities. The description of signal composition components based on IDL (Interface Definition Language) is given out.

**Keywords** STD · Signal model · Signal synthetic method

## 1 Introduction

Signal synthetic mechanism was important for the signal-oriented ATS (Automatic Test System) software as it provided a generic reconfigurable and extensible capability of the test software system. STD standard provides the accurate definition for signals and the libraries of predefined basic signals which may be reused and extended. Signal synthetic mechanism based on STD standard discussed in this paper that provided the signal synthetic mechanism by creating a required signal from the existing signal in either the Basic Signal Components (BSCs) or Test

J. Luo (✉) · C. Wang
Department of Missile Engineering, Ordnance Engineering College, Shijiazhuang, China
e-mail: zoe_luojin@163.com

X. Wen
Army 63961, Beijing, China
e-mail: WenXi_4825@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_43

Signal Framework (TSF) defined in STD standard. As the function of the present signal generators and digital waveform generators were completely, signal synthetic mechanism discussed in this paper used the software synthetic method to combine the required signal, and transfer the required signal to the certain instrument to output the synthetic signal.

A key feature of the signal synthetic method is the ability to dynamically modify the signal using the signal attributes so that signals can be combined to form complex signals usable across all test platform. To support the signal synthetic mechanism, a signal composition component was developed and its interface description based on IDL was given out in this paper. The signal composition interface was an internal interface to transfer synthetic signal in ATS software platform across the whole signal lifecycle phase. The composition component permitted signal synthetic operations to be embedded in any object-oriented environments, an example program based on VC development environment was illustrated to show the running mechanism of the signal composition component.

## 2 Signal Synthetic Mechanism Based on STD Standard

STD standard is a tester independent standard for test signals and measurement definition, it provides a collection of objects and associated interfaces that used to describe signal measurement and stimulus components that are instrument independent. These include Basic Signal Components (BSCs) and Test Signal Framework (TSF). BSCs provided the basic signal components for the complex signal composition. TSF defined the domain-specific signal models that made up from BSCs or other TSF signals, it is the extendibility mechanism that allows the creation of additional signal class.

A basic signal component is shown as Fig. 1. Each BSC is described by its class name, class type, attributes, default value and control interfaces. As Fig. 1 shown,

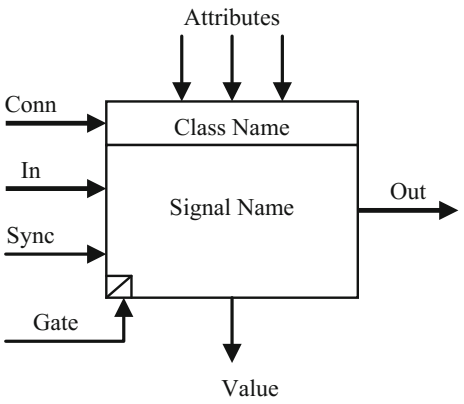



**Fig. 1** BSC's diagram

*In* denoted the input signal of BSC; *out* represented the output signal of BSC, it's also the output port for the synthetic signal of BSC; *Sync* denoted the Sync signal of BSC; *Gate* denoted the gate signal of BSC; *Conn* was used to connect different BSCs through their attributes; *value* is the return value when BSC was used to be a sensor.

BSCs are used to build signal models, which define the required signal. A signal model can contain a single BSC to define a simple signal or combined BSCs to define a more complex signal. As an example, a simple AC waveform is the periodic sinusoid signal defined in STD as basic signal components. If AC signal had more parameters, it would need the combination of several different sinusoid signals defined in BSCs for accurate simulation.

In basic signal component, attributes are associated with the generated signal and are used to define the characteristics or behaviors of that signal. Basic signal components are combined through their input, output and other attributes to build a complex signal component model, namely, signal model. The signal model realized the signal control with its output attribute interface, which named output signal interface usually. The change of the input signal interface state and the different of the output signal interface state could be distinguished by their effect on the behaviors, output signal and other signal models. There are three signal states (*Stopped*, *Ready,* and *Active*) and three methods (*Stop*, *Run*, and *Change*) to realize the signal states transfer. Used the signal control mechanism, the signal components in a signal model can be controlled and combined as a whole, and it's the realization foundation of the BSC synthetic mechanism.

## 3 Signal Synthetic Realization

### *3.1 Realization of the Signal Composition Component*

STD defines combiners that perform arithmetic operations on multiple input signals to combine them into a single output. These include Sum, Product and Difference. All BSC classes used to define signals are derived from one of the base classes, whose interface are *ISignalFunction*, *ISignal*, *IResource*, *IPulseDefns*, and *IPhysical*. *ISignal* component had three methods (*Run*, *Stop* and *Change*) to realize the transfer of signal states. *ISignalFunction* class is the base class of all BSCs, it derived six interface defined in STD standard, that are *ISensor*, *IEventFunction*, *ISource*, *IConditioner*, *IDigital* and *IConnection*. *IConditioner* interface provided the basic mathematics model and physics description of signal synthesizer. To transfer the synthetic signal in the ATS software platform, the *IComposition* interface is designed in this paper.

*IComposition* interface was an internal interface to transfer synthetic signal in ATS software platform across the whole signal lifecycle phase, it permitted signal

synthetic operations to be embedded in any object-oriented environment. The description of *IComposition* interface based on IDL is shown as following.

```
interface IComposition: ISignalFunction
{
enum {COMPOSITION_BASE = (SIGNALFUNCTION_BASE + 256)};
[id(COMPOSITION_BASE + 1), helpstring("method ComposeSignal")]
HRESULT ComposeSignal(BSTR composeMethod, IUnknown*signal1,
IUnknown*signal2);
[propget, id(COMPOSITION_BASE + 2), helpstring("property sampleNumber")]
HRESULT sampleNumber([out, retval] long *pVal);
[propput, id(COMPOSITION_BASE + 2), helpstring("property sampleNumber")]
HRESULT sampleNumber([in] long newVal);
[propget, id(COMPOSITION_BASE + 3), helpstring("property sampleData")]
HRESULT sampleData(long index, [out, retval] double *pVal);
[propget, id(COMPOSITION_BASE + 4), helpstring("property acqfrequency")]
HRESULT acqfrequency([out, retval] double *pVal);
[propput, id(COMPOSITION_BASE + 4), helpstring("property acqfrequency")]
HRESULT acqfrequency([in] double newVal);
[propget, id(COMPOSITION_BASE + 5), helpstring("property signal_fre")]
HRESULT signal_fre([out, retval] double *pVal);
[propput, id(COMPOSITION_BASE + 5), helpstring("property signal_fre")]
HRESULT signal_fre([in] double newVal);
};
```

The *ComposeSignal* method provides a generic synthetic signal interface that could be used to support other STD signals. The attributes of compose method are associated with the *IEngine* interface and are used to define the characteristics of the synthetic signal, such as signal frequency, sampling rate and sampling numbers, etc.

## 3.2 Running Mechanism of the Signal Composition Component

A synthetic signal waveform model showing the combiner addition of two periodic sinusoidal waveforms is shown in Fig. 2. One periodic sinusoidal named signal A with attributes: frequency is 1 kHz, amplitude is 2 V, phase is 0 rad. The other periodic sinusoidal named signal B with attributes: frequency is 2 kHz, amplitude is 2 V, phase is 0 rad. We used *ISum* interface derived from *ICombiner* interface to describe the signal synthetic process.

The following VC program shows the key test flow of the above signal model.

```
spSum- > putref_In(0,spTwoWireOut);
//the signal interface for TwoWire
spSum- > putref_In(1,spSinusoidOut);
```

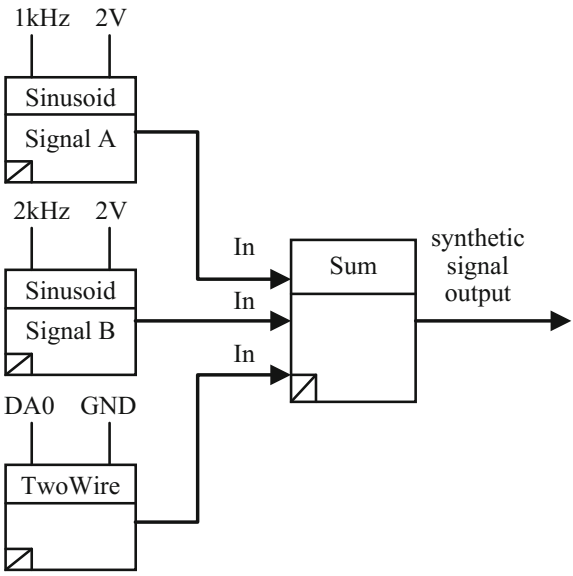



**Fig. 2** Example model diagram for signal composition

```
//the signal interface for Signal A
spSum- > putref_In(2,spSinusoidOut2);
//the signal interface according to Signal B
spSum- > get_Out(0,&spSumOut);
spSumOut- > Run(0);
//signal stimulus running
```

Run the "spSumOut- > Run(0)" process to transfer the Run event to ISum interface with the connection point technology. The connection point technology is a feedback mechanism to allow the object to discover its interface calling function. A connection has two parts, one is the calling object named connection point, the other is the interface realization object named sink object. In the synthetic signal component, the connection point is the IComposition interface, the sink object is the ISum interface. After the ISum interface received the Run event, it responsed to the Run event as Fig. 3 shown.

At first, ISum interface checked whether the input connection of ISum interface is correct. The input connection included the Input port connection, Gate port connection and Sync port connection of the signal interface. If the input connection is correct, ISum interface object would have a signal interface based on IConnection to describe the signal ports, and have plural signal interfaces for signal combination. Transfered the signal interface of input connection of ISum to IComposition interface, and synthesize the signals used the ComposeSignal method. Acquired the

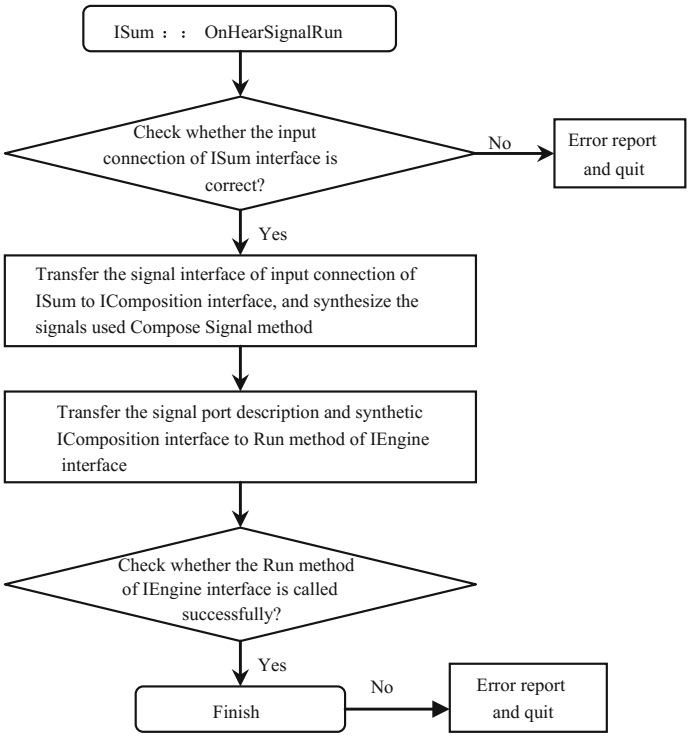



**Fig. 3** Flow diagram for ISum interface's response for run event

signal port description and connection interface of the ICompostion signal interface, and transfer them to IEngine interface of the real-time engine. Then we could realize the physics signal by calling the Run method of IEngine interface.

When combining STD periodic waveforms, it is important that the direct digital synthesizers are synchronized to each other to ensure the correct waveform samples are combined. So we can realize the ComposeSignal method of the IComposition interface as the following steps: first, set up the sampling rate of the synthetic signal according to its attributes, then conformed two sampling data for both input signals according to their sampling data and signal description that named data1 and data2, in the end, generated the sampling data of the synthetic signal according to the signal synthetic method and two sampling data of input signals.

A waveform viewer shown in Fig. 4 provided a dynamic simulation of the combiner addition of two periodic sinusoidal waveforms. The modification of the waveform attributes results in an instantaneous update of the simulated waveform. The objective was to achieve the same result using hardware.

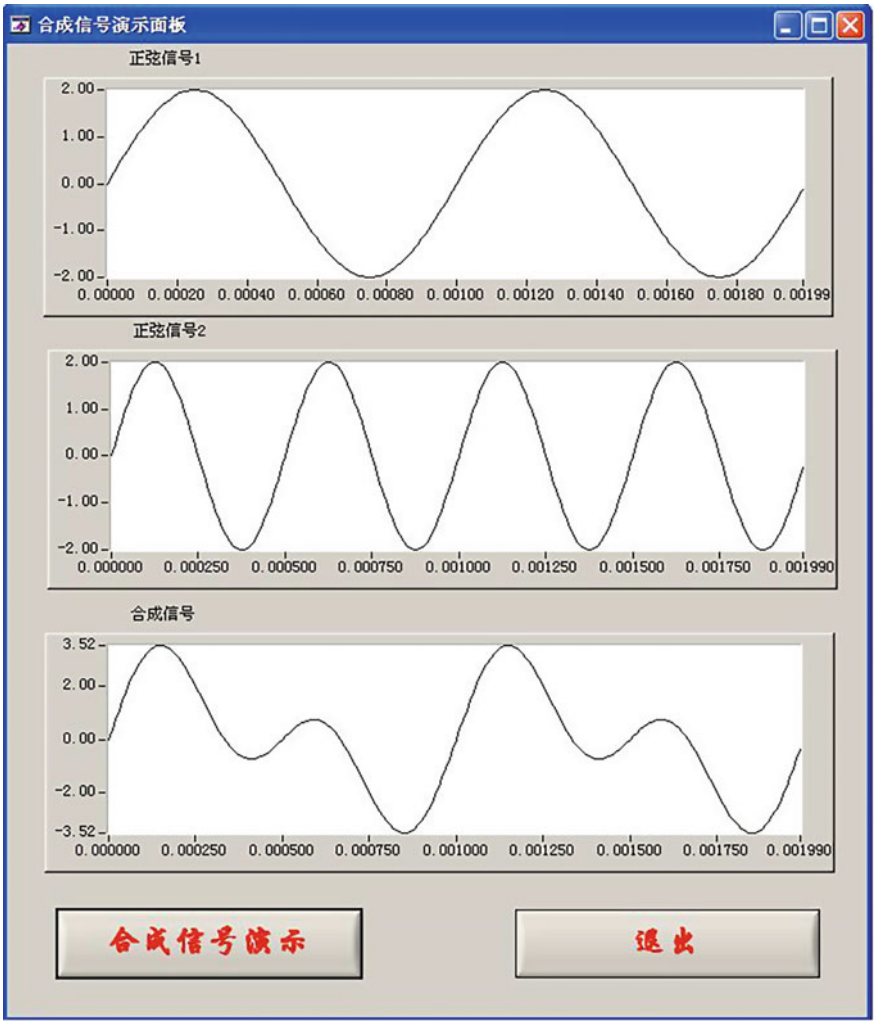



**Fig. 4** Waveform simulated by software

## 4 Conclusion

STD standard is now well established as a standard for defining signals, and has become the required standard for all future test systems. How to synthesize signals into a required complex signal was important for the signal-oriented ATS. This paper researched the signal synthetic technology based on STD standard, and gave out the realization of the signal composition component. The signal composition component permitted signal synthetic operations compliant STD standard to be embedded in any object-oriented environments, and provided the ability to

dynamically modify the signal using the signal attributes so that signals can be combined to form complex signals usable across all test platform. An example based on VC program showing the combiner addition of two periodic sinusoidal waveforms is explicated in this paper to illustrate the running mechanism of the signal composition component.

## References

1. IEEE Std 1641™-2004, IEEE Standard for Signal and Test Definition. Institute of Electrical and Electronics Engineers, Inc
2. C. Baker, A. Hulme, Using an FPGA based system for IEEE 1641 waveform generation. AUTOTESTCON Proc., pp. 180–185 (2010). IEEE
3. C. Gorringe, IEEE std 1641—a revised signal and test definition standard. AUTOTESTCON Proc., pp. 241–245(2009). IEEE
4. A. Hulme, Physical signals, events and digital streams—their relationship and how they effect signal function in IEEE 1641. IEEE AUTOTESTCON Proc., pp. 311–316 (2009). IEEE

# Parameter Tuning for Bees Algorithm on Continuous Optimization Problems

**Xin Zhang and Xunyu Cheng**

**Abstract** Bees algorithm belongs to swarm intelligence category. It is a good algorithm for dealing with continuous optimization, though its main disadvantage is its six parameters being defined by users. Parameter setting is a notorious problem in swarm intelligence. This paper aims to tackle this issue by a parameter tuning strategy. Parameters are extensively studied with different combinations. Moreover, a popularly used numerical function set is taken in experiment with different dimensions. Experimental results are discussed and analyzed. The best ten parameter combinations are identified in terms of average number of function evaluations reaching global optima. They are useful for users to solve their real-world problems.

## 1 Introduction

The so-called optimization refers to satisfy some optimized measure of a system (or function), and drive some performance index of a system to achieve maximum or minimum. It finds a set of parameter values under the condition of satisfying certain constrains. Optimization problem according to the nature objective function, the nature of the constraint function, and the value of optimized variables can be divided into many types. Besides, depending on the different nature, each type of optimization problem has its specific solving method. The main method to solve the optimization problem is to establish mathematical model and solve the optimal strategy.

X. Zhang (✉) · X. Cheng
College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin, China
e-mail: xinzhang9-c@my.cityu.edu.hk; tjnumark@126.com

X. Cheng
e-mail: chengxunyu1314@163.com

X. Zhang · X. Cheng
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_44

With the expanding the application and requirement, the theory of optimum research also get great development. According to the optimization mechanisms and behavior, that is commonly used in current engineering optimization algorithm mainly include classical algorithm, constructive algorithm, improved algorithm [1, 2], algorithm based on dynamic evolution of system [3], hybrid algorithm [4], swarm intelligence algorithm [5, 6], and so on.

Swarm intelligence algorithms can be applied to deal with supply chain optimization [7], radar waveform design [8], and routing problems in communication networks [9]. These algorithms can also be connected with finite element method to enhance the efficiency of electrical devices [10, 11].

Bees algorithm (BA) is a kind of modern heuristic intelligent search algorithm [12]. It has many of the characteristics of modern optimization algorithm. This paper attempts to study the effect of algorithmic parameters of BA. Although BA has shown its usefulness in several applications, it contains six parameters to be predefined by users. A parameter tuning strategy is designed in this paper to find the best parameter combination. Extensive simulation is conducted on a set of numerical optimization functions.

Section 2 in this paper introduces standard bees algorithm. Section 3 explains the tuning strategy for studying parameters of bees algorithm. Experimental results are reported in Sects. 4 and 5 draws the conclusion.

## 2 Bees Algorithm

This section describes standard bees algorithm.

As mentioned above, BA is a metaheuristic algorithm summarized by the foraging behavior of honey bees in natural to find the optimal solution. It consists of seven steps. In step one, honey bees are placed randomly in the search space. Step two is in charge of evaluating the fitness of the sites visited by the scout bees. Step three is to find out the highest fitness sites. In step four, bees are called "selected bees" in the highest fitness sites. We should select the sites visited by them in order to search their neighborhood. According to the fitness associated with the sites they are visiting, in step five we can choose directly the needful bees by conducting searches in the neighborhood of the selected sites and assigning more bees to search near to the best sites. In step six, the next bee population sites will select from the each patch only the bee with the highest fitness. In step seven, the remaining bees in the population are assigned randomly around the search space scouting in order to new potential solutions. Steps two to seven are repeated when we meet a stopping criterion. At the end of each iteration, the colony will have two parts to its new population representatives from each selected patch and other scout bees assigned to conduct random searches.

Algorithm 1 shows the pseudocode of BA in its simplest form, where $f(\cdot)$ is objective function, $D$ is problem dimension, and $\mathbf{x}^{min}$ and $\mathbf{x}^{max}$ are respectively the lower bound and upper bound. $n_b$, $n_{br}$, $n_e$, $n_{er}$, $n_s$, and $s_{lim}$ are algorithmic parameters to be

defined by users. Setting up the algorithm requires lots of parameters, in other words, lots of scout bees, lots of sites selected out of the visited sites, lots of best sites out of selected sites, lots of bees recruited for best sites, lots of bees recruited for the other selected sites, initial size of patches which includes site, and its neighborhood and stopping criterion.

**Algorithm 1** Pseudocode of the BA method

**Require:** $f(\cdot)$, $D$, $\mathbf{x}^{min}$, $\mathbf{x}^{max}$, $n_b$, $n_{br}$, $n_e$, $n_{er}$, $n_s$, $s_{lim}$
**Ensure:** The best solution obtained by the algorithm;
1: Randomly create a solution population;
2: Evaluate fitness values of solutions;
3: **repeat**
4: Select sites for neighborhood search;
5: Recruit bees for selected sites (more bees for best e sites)and evaluate fitness;
6: Select the fittest bee from each patch;
7: Assign remaining bees to search randomly and evaluate their fitness;
8: **until** stopping criterion is met

On the contrary, the steps five and six can also use the fitness values to determine the probability of the bees being selected. Searches in the neighborhood of the best sites witch represent more promising solution are made more detailed by recruiting more bees to follow them than the other selected bees. As we all know, there is no such a restriction in nature, what the aim about the restriction is introduced is that reduce the number of points to be explored.

BA was hybridized with slop angle computation and hill climbing methods for dealing with scheduling problems [13]. Yuce et al. extended BA to multiobjective optimization for supply chain problems [7]. Xu et al. proposed an adaptive BA for fuel economy optimization [14], where adaptive feature means adaptively adjust the range of local search instead of algorithmic parameters. Although BA has been studied and modified to deal with different applications, its algorithmic parameters are still an annoying thing for users and an unresolved issue.

## 3 Parameter Tuning Strategy

The bees algorithm contains six parameters in total. As above, they are the number of best sites $n_b$, the number of foragers per best site $n_{br}$, the number of elite sites $n_e$, the number of foragers per elite site $n_{er}$, the number of scouts $n_s$, and the stagnation limit per scout $s_{lim}$. Parameter $s_{lim}$ designates the stagnation status of a solution. If it could not be improved in a continuous search steps, it was considered as a local optima or the stagnation of population; scout was sent out searching a new position. Usually, this parameter is a non-sensitive one. Thus, it is set to a constant in this article as $s_{lim} = 10$.

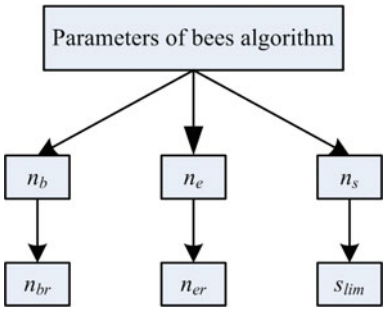



**Fig. 1** Two-level illustration of parameters of bees algorithm

Let us expound the relation of the remaining five parameters. Given a set of feasible solutions, they are first classified to two groups: best sites and non-best sites, where $n_b$ is the size of best site group. The others in solution set are considered as bad ones and would be replaced by randomly created new solutions, where $n_s$ is the size of this group. Hence, the size of solution set is $n_b + n_s$. Elite sites are selected from best sites, that is $n_e \leq n_b$. The colony size of honey bees is $(n_b - n_e)n_{br} + n_e n_{er} + n_s$. Thus, the parameters of BA can be tuned from two levels (see Fig. 1) based on the idea of divide and conquer. $n_b$, $n_e$, and $n_s$ belong to the first level. $n_{br}$, $n_{er}$, and $s_{lim}$ belong to the second level.

As to the three parameters of the first level, $n_b$ should be greater than $2n_e$; otherwise, the algorithm is in high risk of premature convergence and easily trapped in local optima. It is hard to determinate which one should be larger between $n_b$ and $n_s$. Large $n_s$ would impair colony adaptation resulting in slow convergence, whereas low $n_s$ may cause the same effect as large $n_e$. It is assumed that a proportion relation exists among $n_e, n_b$, and $n_s$.

As to the three parameters of the second level, $s_{lim}$ is fixed at 10 since it is a less insensitive parameter than the others. This point is also reported in [8, 15]. Because $n_{br}$ determines the number of scouts foraging around best chosen solutions regardless of elite ones, it is reasonable to assume that proportion relationship exists between $n_b - n_e$ and $n_{br}$. Similarly, $n_e$ and $n_{er}$ also have proportion relationship.

In short, parameter tuning for BA is based on the assumption that proportional relations exist between algorithmic parameters. Besides parameters, we have to discuss the effect of problem difficulty. Different optimization problems may require different parameter settings of BA. It is of high demand that parameter tuning should be verified on extensive problem types to gain a meaningful conclusion. To meet this requirement, a set of continuous optimization functions is taken in numerical experiment. Moreover, the functions are tested with different dimensions so that a diverse problem set is simulated.

## 4 Numerical Experimental Evaluation and Analysis

In this section, the bees algorithm with different parameter settings is applied to deal with an extensive optimization function set with different dimensions.

### *4.1 Parameter Combinations*

The proportions chosen for $n_e$, $(n_b - n_e)$, and $n_s$ are given in Table 1. This table shows 16 combinations of the three parameters. In numerical experiment, each proportion is instantiated by three cases. Take proportion 1:1:1 for example. Three test cases are $(1{:}1{:}1) \times 1$, $(1{:}1{:}1) \times 2$, and $(1{:}1{:}1) \times 3$. Hence, these cases are $n_e = 1, n_b = 2, n_s = 1$, $n_e = 2, n_b = 4, n_s = 2$, and $n_e = 3, n_b = 6, n_s = 3$.

The ratios chosen for $n_e$ versus $n_{er}$ are 1:2, 1:4, 1:6, and 1:8. The ratios for $(n_b - n_e)$ versus $n_{br}$ are 1:1, 1:2, and 1:3. Given $n_e$ or $n_b$ are fixed, $n_{er}$ and $n_{br}$ can then be fixed based on the ratios. As $n_b$ are usually greater than twice of $n_e$, its ratio is larger than $n_{er} : n_e$ for a balancing exploration and exploitation.

### *4.2 Test Function Set*

Eight continuous optimization functions are taken as benchmark which are widely used in benchmarking metaheuristic approaches [2, 16]. The functions are (1) sphere function, (2) axis parallel hyper-ellipsoid function, (3) Rosenbrock's function, (4) Rastrigin's function, (5) Griewank function, (6) Ackley's function, (7) Weierstrass function, and (7) Levy function. All functions are scalable to any dimension. $f_1$ and $f_2$ are unimodal functions, while the others are multimodal functions. Generally, multimodal functions are more difficult to handle than unimodal ones. The functions are tested with different dimensions. That is $D \in \{5, 10, 15, 20\}$ in our experiment.

The instances are totally 3 cases × 16 proportions × 4 ratios × 3 ratios × 8 functions × 4 dimensions = 18432. Each instance is executed 51 independent runs with $MFE = 5000D$ for each test function. This is a commonly used setting to gain an average overview of the performance of an evolutionary algorithm. The source code of BA can be obtained from the author upon request.

**Table 1** Proportions chosen for $n_e$, $n_b - n_e$, and $n_s$

| $n_e : (n_b - n_e) : n_s$ | $n_e : (n_b - n_e) : n_s$ | $n_e : (n_b - n_e) : n_s$ | $n_e : (n_b - n_e) : n_s$ |
|---|---|---|---|
| 1:1:1 | 1:1:2 | 1:1:4 | 1:1:6 |
| 1:2:1 | 1:2:2 | 1:2:4 | 1:2:6 |
| 1:3:1 | 1:3:2 | 1:3:4 | 1:3:6 |
| 1:4:1 | 1:4:2 | 1:4:4 | 1:4:6 |

### 4.3 Experimental Results

After finishing the simulation, it is found that most parameter combinations are able to find the global optimum of test functions. Hence, average number of function evaluations (FEs) is taken to assess which parameter combination is better than another.

First, parameter combination with the minimum number of FEs reaching global optimum over all functions and all dimensions is identified as the best. The parameter setting is $n_e = 2$, $n_b = 6$, $n_{er} = 12$, $n_{br} = 4$, $n_s = 2$, and $s_{lim} = 10$. The average of FEs of this combination is 4.608E4 for eight test functions with four dimensions. This suggests that if a user faces a problem with 5–20 variables, this parameter combination should be used as default setting.

Second, average of FEs needed by each parameter combination is shown in Fig. 2. This figure is plotted in parameter combination order. Observed from this figure, FEs decrease initially along with the increase of parameter values; on the other hand, after combination 250, there are roughly four positions having similar FEs. The best one is the combination 293 and the second is the combination of 500. Moreover, after combination 400, variance of FEs increases heavily indicating that the algorithm becomes unstable. Since parameter combination is ordered in ascend order, thus the results show that small or large $n_e, n_b, n_s$ values are less effective than those in the middle. Furthermore, many combinations perform similarly on the benchmark. Hence, the first ten combinations are listed in Table 2 for providing a reference to users.

Third, parameter combination versus average FEs are plotted in ascending order with respect to average FEs values as shown in Fig. 3. It can bee seen from this figure that the curve increases slowly before combinations 250 and quickly after that. And a sharp increase is observed after combination 500. Thus, parameter tuning

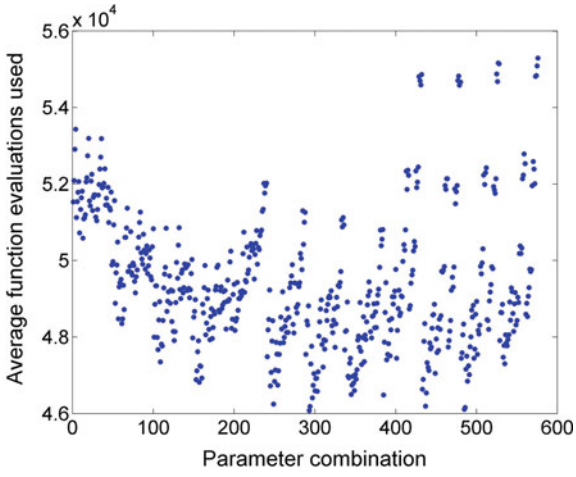



**Fig. 2** Average function evaluations costed by each parameter combination

**Table 2** Best ten parameter combinations on the benchmark

| Number | Parameter combination |
|---|---|
| 1 | $n_e = 2, n_b = 6, n_{er} = 12, n_{br} = 4, n_s = 2, s_{lim} = 10$ |
| 2 | $n_e = 3, n_b = 9, n_{er} = 18, n_{br} = 6, n_s = 3, s_{lim} = 10$ |
| 3 | $n_e = 3, n_b = 9, n_{er} = 18, n_{br} = 6, n_s = 6, s_{lim} = 10$ |
| 4 | $n_e = 2, n_b = 6, n_{er} = 12, n_{br} = 4, n_s = 4, s_{lim} = 10$ |
| 5 | $n_e = 3, n_b = 9, n_{er} = 12, n_{br} = 6, n_s = 3, s_{lim} = 10$ |
| 6 | $n_e = 2, n_b = 8, n_{er} = 8, n_{br} = 6, n_s = 2, s_{lim} = 10$ |
| 7 | $n_e = 2, n_b = 8, n_{er} = 12, n_{br} = 6, n_s = 2, s_{lim} = 10$ |
| 8 | $n_e = 3, n_b = 6, n_{er} = 6, n_{br} = 3, n_s = 3, s_{lim} = 10$ |
| 9 | $n_e = 2, n_b = 6, n_{er} = 16, n_{br} = 4, n_s = 4, s_{lim} = 10$ |
| 10 | $n_e = 3, n_b = 9, n_{er} = 12, n_{br} = 6, n_s = 6, s_{lim} = 10$ |

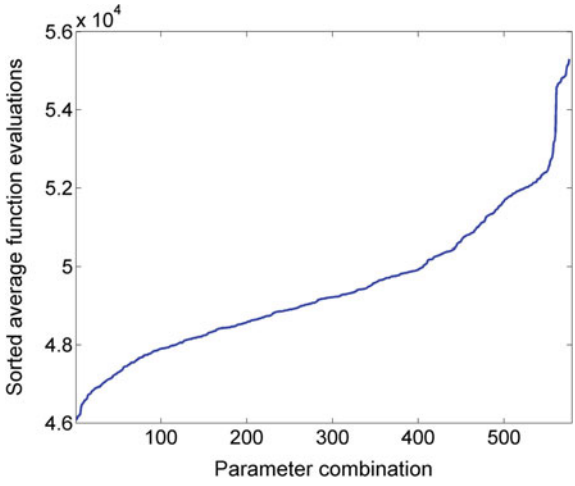



**Fig. 3** Average function evaluation curve in ascending order for all parameter combinations

experiment is meaningful as effective combinations are included in the middle. From this point of view, we can draw creditable conclusion.

## 5 Conclusion

Bees algorithm (BA) is a good paradigm showing good performance in previous researches; whereas it contains six parameters being defined by practioners. This reduces users' interest to utilize this algorithm. This paper aims to handle parameter setting problem of BA. A parameter tuning strategy is designed which consists of $3 \times 16 \times 4 \times 3 = 576$. Eight numerical functions are taken as benchmark with four dimensions. Analyzing from the simulation results, the best parameter combination is identified in terms of minimum average number of function evaluations. Besides, nine other combinations are also reported which also present good performance.

**Acknowledgements** This research was supported in part by the National Science Foundation of China (61603275) and the Applied Basic Research Program of Tianjin (15JCYBJC51500).

## References

1. X. Li, G. Yang, Artificial bee colony algorithm with memory. Appl. Soft Comput. **41**, 362–372 (2016)
2. Y. Lou, S.Y. Yuen, Non-revisiting genetic algorithm with adaptive mutation using constant memory. Memet. Comput. **8**(3), 189–210 (2016)
3. R. Liu, J. Fan, L. Jiao, Integration of improved predictive model and adaptive differential evolution based dynamic multi-objective evolutionary optimization algorithm. Appl. Intell. **43**(1), 192–207 (2015)
4. Y. Lou, J. Li, Y. Shi, L. Jin, Gravitational co-evolution and opposition-based optimization algorithm. Int. J. Comput. Intell. Syst. **6**(5), 849–861 (2013)
5. H. Sharma, J.C. Bansal, K.V. Arya, X.S. Yang, Levy flight artificial bee colony algorithm. Int. J. Syst. Sci. **47**(11), 2652–2670 (2016)
6. Z. Wu, X. Xia, B. Wang, Improving building energy efficiency by multiobjective neighborhood field optimization. Energy Build. **87**, 45–56 (2015)
7. B. Yuce, E. Mastrocinque, A. Lambiase, M.S. Packianather, D.T. Pham, A multi-objective supply chain optimisation using enhanced bees algorithm with adaptive neighbourhood search and site abandonment strategy. Swarm Evolut. Comput. **18**, 71–82 (2014)
8. X. Zhang, X. Zhang, A study of artificial bee colony variants for radar waveform design. EURASIP J. Wirel. Commun. Netw. 2016(1): Article No 13 (2016)
9. X. Zhang, X. Zhang, C. Gu, A micro-artificial bee colony based multicast routing in vehicular ad hoc networks. Ad Hoc Netw. (2016), https://doi.org/10.1016/j.adhoc.2016.06.009
10. X. Zhang, H.L. Li, S.L. Ho, W.N. Fu, A multi-slice finite element model including distributive capacitances for wireless magnetic resonant energy transfer systems with circular coils. IEEE Trans. Magnet. **49**(5), 1857–1860 (2013)
11. X. Zhang, S.L. Ho, W.N. Fu, Fast algorithm to obtain the torque characteristics with respect to load angle of synchronous machines using finite element method. IEEE Trans. Magnet. **50**(11), 1–4 (2014)

12. D.T. Pham, A. Ghanbarzadeh, E. Koc, S. Otri, S. Rahim, M. Zaidi, The bees algorithm—a novel tool for complex optimisation problems, in *Intelligent Production Machines and Systems*, ed. by D. Pham, E. Eldukhri, A. Soroka (Elsevier Science Ltd, Oxford, 2006), pp. 454–459
13. B. Yuce, D.T. Pham, M.S. Packianather, E. Mastrocinque, An enhancement to the bees algorithm with slope angle computation and hill climbing algorithm and its applications on scheduling and continuous-type optimisation problem. Prod. Manuf. Res. Open Access J. **3**(1), 3–19 (2015)
14. S. Xu, F. Yu, Z. Luo, Z. Ji, D.T. Pham, R. Qiu, Adaptive bees algorithmłbioinspiration from honeybee foraging to optimize fuel economy of a semi-track air-cushion vehicle. Comput. J. **54**(9), 1416–1426 (2011)
15. K. Diwold, A. Aderhold, A. Scheidler, M. Middendorf, Performance evaluation of artificial bee colony optimization and new selection schemes. Memet. Comput. **3**(3), 149–162 (2011)
16. Z. Wu, T.W. Chow, Neighborhood field for cooperative optimization. Soft Comput. **17**(5), 819–834 (2013)

# Two-Step Damage Detection Method for Large and Complex Structures

**Yong-jun Li, Li-yuan Ma, Shi-long Li and Tian-hui Wang**

**Abstract** A damage detection method for complex structures is proposed based on model updating and modal flexibility curvature difference in this paper, which includes two steps. At first, the model updating is used for preparatory detection. In the step, detection problem is transformed into an optimization problem by a nonlinear least-square objective function with bound constrains and the trust-region approach is used to solve the minimization problem and identify the damaged unit group. Second, the method based on modal flexibility curvature difference is used to detect the location of structural damage accurately. In order to verify the effect, simulation and experiment on a framework structure are carried out. The proposed method is applied for damage detection. The simulation and experimentation results show that the proposed method is feasible for damage detection on complex structures.

**Keywords** Large and complex structure · Model updating · Trust-region · Modal flexibility curvature difference · Damage detection

## 1 Introduction

Many damage detection methods for structure are put forward in recent years. But, because of synthetical effects such as measurement noise, modeling error, input information inadequacy, and sensor amount limitation, most research of damage detection method is far away from practical application and only suitable for simple structure [1]. Up to now, there are barely methods for large and complex framework structures. In this paper, dynamical test data is used for damage detection. A two-step method is used to detect the damage of large and complex structure based on model updating and modal flexibility curvature difference. At first, the model updating is used for preparatory detection and identify the damaged

Y. Li (✉) · L. Ma · S. Li · T. Wang
Mechanical Engineering College, Shijiazhuang 050003, China
e-mail: 3051481422@qq.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_45

unit group. Second, the method based on modal flexibility curvature difference is used to detect the location of structural damage accurately.

## 2 Damage Detection Based on Model Update

The damage detection based on model updating is commonly divided into two steps. First, the finite element model (FEM) is updated with tested modal data of healthy structure. Second, the tested modal data of damaged structure is used to update the FEM. Then the physical parameters which reflect the damage location and degree are identified by the process [2]. The common methods based on model updating are realized by transforming the damage detection problem into constraint optimization problem [3]. So to seek for an optimized method with quick convergence and high robustness for damage detection is the hotspot.

The first step of the damage detection method put forward by the paper is a damage detection based on model updating technology. A nonlinear least-square object function with bound constrains is put up to transform the damage detection into an optimization problem. The trust-region method with global convergence is used to solve the optimization problem, which possess higher robustness and reduce iterative steps. So it can take less time to identify the stiffness reduction coefficient (SRC). So rapidly detection is accomplished.

### *2.1 Basic Theory*

The dynamic characteristic formula of FEM can be presented as follows:

$$K\varphi^{(j)} - \lambda_j M_0 \varphi^{(j)} = 0, \tag{1}$$

where $K$ is the global stiffness matrix; $M_0$ is the global mass matrix (suppose that no change in mass matrix before and after damage); $\lambda_j$ and $\varphi^{(j)}$ are respectively the characteristic value and characteristic vector of the $j$th order; $j = 1, 2, \ldots, N_m$, $N_m$ is the modal order number;and $\lambda_j = (2\pi f_j)^2$, $f_j$ is the self-vibration frequency ($Hz$).

The damage of the structure can cause the reduction of stiffness. So the global stiffness matrix after damage can be presented as follows:

$$K = K_0 - \sum_{i=1}^{N_e} \theta_i K^{(i)}, \tag{2}$$

where $K_0$ is the global stiffness matrix before damage, $K^{(i)}$ is the expand order matrix of the $i$th unit of the FEM. $\theta_i$ is the SRC vector with $N_e$ dimension,and $N_e$ is the number of FEM unit.

### 2.1.1 Object Function

To compare the approach degree of calculated modal with FEM and tested modal, the following object function is put forward.

$$\min_{\theta \in \Re^{N_e}} \left\{ f(\theta) = \frac{1}{2} \|r(\theta)\|^2 = \frac{1}{2} \sum_{i=1}^{N} r_i^2(\theta) : l < \theta \le u \right\} \tag{3}$$

$$r(\theta)_{N \times 1} = \left[ (r^f(\theta))^T_{N_m \times 1}, (r^{\varphi}(\theta))^T_{N_m \times 1} \right]^T, \tag{4}$$

where $l, u \in \Re^{N_e}$ are the top and bottom boundaries of $\theta$. All the elements of $l$ and $u$ are set with 0 and 1 for detection requirement.

$$r_j^f(\theta) = \left| 1 - \frac{f_j(\theta)}{\hat{f}_j} \right| \tag{5}$$

$$r_j^{\varphi}(\theta) = 1 - \frac{((\hat{\varphi}^{(j)})^T \varphi^{(j)}(\theta))^2}{((\hat{\varphi}^{(j)})^T \hat{\varphi}^{(j)})((\varphi^{(j)}(\theta))^T \varphi^{(j)}(\theta))}, j = 1, 2, \ldots, N_m, \tag{6}$$

where superscript "^" presents tested modal data; the vector $r^f(\theta)$ with $N_m$ dimension is used to weigh the approach degree of different frequencies; the vector $r^{\varphi}(\theta)$ with $N_m$ dimension is used to weigh the approach degree of different vibration shapes, which includes MAC (Modal Assurance Criterion) value.

Expand the $i$th element of object function in first-order Taylor progression that $\theta$ is near $\theta_k$,

$$r_i(\theta) \approx r_i(\theta_k) + \nabla r_i(\theta_k)^T (\theta - \theta_k), i = 1, 2, \ldots, N, \tag{7}$$

so the following can be obtained:

$$\begin{aligned} f(\theta) &= \frac{1}{2} \sum_{i=1}^{N} r_i^2(\theta) \approx \frac{1}{2} \sum_{i=1}^{N} \left[ r_i(\theta_k) + \nabla r_i(\theta_k)^T (\theta - \theta_k) \right]^2 \\ &= f(\theta_k) + g_k^T (\theta - \theta_k) + \frac{1}{2} (\theta - \theta_k)^T H_K (\theta - \theta_k) \end{aligned} \tag{8}$$

$$g_k = \nabla f(\theta_k) = J(\theta_k)^T r(\theta_k) \tag{9}$$

$$H_k = \nabla^2 f(\theta_k) = J(\theta_k)^T J(\theta_k), \tag{10}$$

where $J(\theta_k)$ is the Jacobian matrix of object function vector $r(\theta)$ at $\theta_k$; $g(\theta_k)$ and $H(\theta_k)$ are respectively grads vector and Hessian matrix of object function $f(\theta)$ at $\theta_k$,"$\nabla$" is derivative operator. It can be seen that when the least-square form presented in Eq. (3) is adopted, the grads vector and Hessian matrix are simplified

forms which are presented with Jacobin matrix in first-order bias-derivative form. It overcomes the calculation of Hessian matrix in iterative, which greatly increases the calculation efficiency.

### 2.1.2 Solution of Object Function

The trust-region method is not only rapidly locally convergent, but also globally convergent. It is used to solve the nonlinear least-square problem with boundary constraint in order to identify SRC. The convergence criterion is defined as (1) $|(f(\theta_k) - f(\theta_{k-1}))/f(\theta_{k-1})| < \varepsilon$ or (2) $\left\|D(\theta_k)^2 g(\theta_k)\right\|_\infty < \varepsilon$ and $H(\theta_k)$ is a positive definite matrix. The essential problem of trust-region algorithm is how to efficiently adjust the radius $\Delta$ in order to ensure the object function $f(\theta)$ to quickly drop. $\varepsilon$ is set as $1.0 \times 10^{-6}$ in this paper.

## 2.2 Numerical Simulation and Experiment

### 2.2.1 Modal Experiment on a Framework

The framework is set as 25 knots. Each knot is of three dimensions to get three-dimensional modal. 15 knots are chosen as the measurement point, which includes 28 freedom. The distribution of freedom degree is shown in Table 1.

The framework is hung with rubber rope in "free" state. The stimulus is a hammer, single stimulation. The sample rate is 10 kHz. The sample length in a measurement point is 2048. The serial number of measurement points and experiment scene are shown in Fig. 1.

### 2.2.2 FEM and Model Updating of Framework

In order to get the FEM close to practice structure, three-dimensional substantial FEM is constructed. Because the numerical model is too complex and large, the number of unit, knot, and freedom degree is very large. Hence, the first step of damage detection divides the model into groups. Each group is considered as a integer for damage detection. According to model experiment data, an amendment to the model is made, mainly in welding part.

**Table 1** Distribution of freedom degree for measurement point

| Measurement point | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Freedom degree | x, z | x, z | y | y | x, z | x, z | x, y | x, y | x, y | x, y | y, z | y, z | x, z | x, z | x, z |

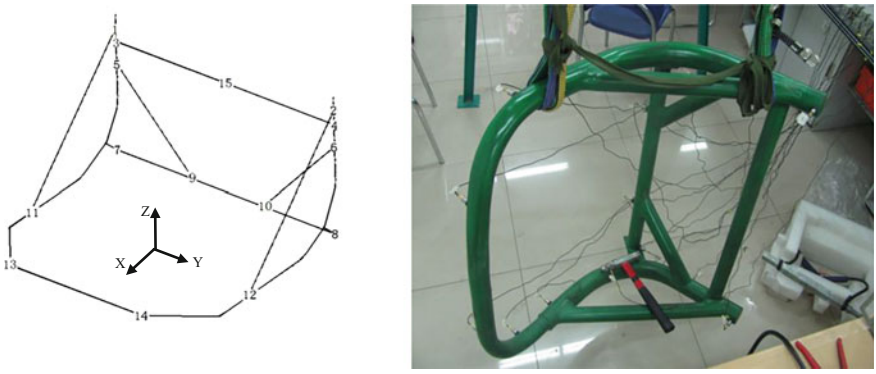
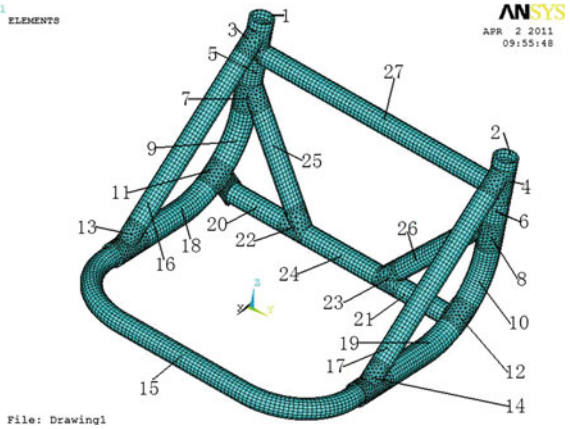



**Fig. 1** Test point number and scene

**Fig. 2** FEM of framework and unit groups




The updated FEM and unit groups are shown in Fig. 2.

The comparison between calculation modal with FEM and tested modal is shown in Table 2. It can be seen that the two modals are very close. It can be concluded that the FEM is exact and can be used for damage detection.

**Table 2** Comparison between calculation modal with FEM and tested modal

| Modal order | Tested frequency(HZ) | Calculation frequency(HZ) | Error(%) | MAC |
|---|---|---|---|---|
| 1 | 112.139 | 112.52 | 0.34 | 0.926 |
| 2 | 156.911 | 149.713 | −4.59 | 0.874 |
| 3 | 267.166 | 254.80 | −4.63 | 0.937 |
| 4 | 292.486 | 298.63 | 2.10 | 0.934 |
| 5 | 438.155 | 444. 175 | 1.37 | 0.903 |
| 6 | 528.56 | 521.25 | −1.38 | 0.891 |

### 2.2.3 Numerical Simulation for Damage Detection

(1) Damage condition configuration

Two damage conditions are configuration. One is single damage (condition 1), and the other is multi-damage (condition 2). The condition 1 supposes that 30% damage (stiffness drop 30%) happens in the 24th unit group. The condition 2 supposes that 30% damage happens in the 24th and 27th unit groups. Table 3 shows the frequency change before and after damage. In single damage condition, the frequency change is not evident except the 4th and 5th order. In multi-damage condition, the 3th, 4th, and 5th order frequency changes greatly. So it is not enough to detect damage only with frequency.

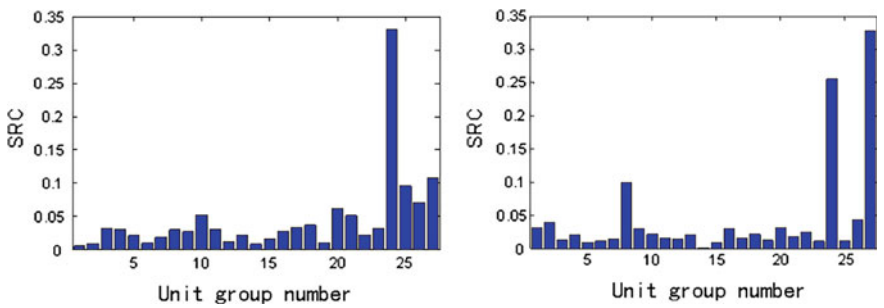
(2) Damage detection

The SRC of each unit group calculated with the aforementioned method is shown in Fig. 3.

Figure 3a shows that the SRC changes greatly at 24th unit group in condition 1. Figure 3b shows that the SRC of 24th and 27th unit group changes greatly. The unit

**Table 3** Modal frequency change before and after damage

| Modal order | Healthy condition (Hz) | Damage condition 1 | | Damage condition 2 | |
|---|---|---|---|---|---|
| | | Frequency (Hz) | Error (%) | Frequency (Hz) | Error (%) |
| 1 | 112.52 | 111.28 | 1.1 | 111.745 | 0.69 |
| 2 | 149.713 | 148.174 | 1.03 | 148.06 | 1.11 |
| 3 | 254.80 | 252.15 | 1.04 | 248.71 | 2.39 |
| 4 | 298.63 | 291.413 | 2.42 | 291.379 | 2.43 |
| 5 | 444.175 | 433.637 | 2.37 | 430.192 | 3.15 |
| 6 | 521.25 | 517.833 | 0.65 | 515.66 | 1.07 |




(a) condition 1 (b) condition 2

**Fig. 3** Damage detection result

groups with outstanding SRC are just the damaged unit group. This shows that the model updating method by this paper is effective.

### 2.2.4 Experimental Validation

(1) Damage condition configuration

Two damage conditions are set in framework:

Condition 1: the 24th unit group is cut in 1/4 part from left to right in a half of radius depth as shown in Fig. 4 (a = 12 mm).

Condition 2: the 24th and 27th unit groups are cut in 1/4 part from left to right in a half of radius depth.

The tested frequency and MAC of the framework before and after damage is shown in Tables 4 and 5.

From Tables 4 and 5, the frequency and MAC change slightly after damage. It is difficult to identify the damage of framework with frequency and MAC.

(2) Damage detection

The SRC of each unit group calculated with the aforementioned method is shown in Fig. 5.

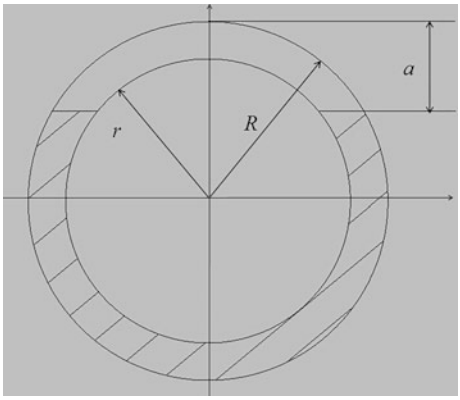


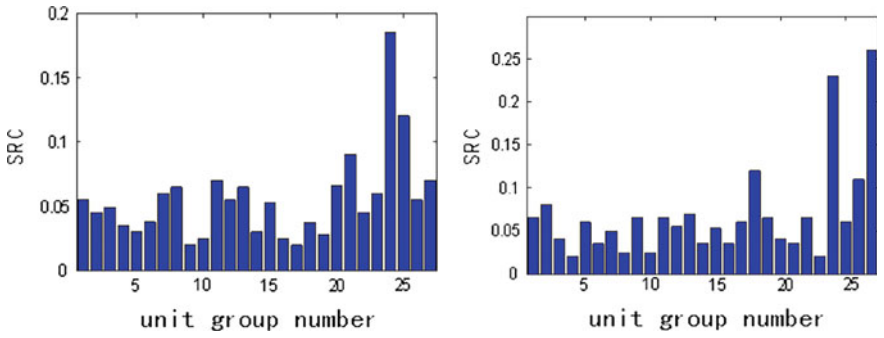

**Fig. 4** Section sketch map after damage

**Table 4** Frequency comparison of framework before and after damage (Hz)

| Modal order | Healthy condition | Damage condition 1 | | Damage condition 2 | |
|---|---|---|---|---|---|
| | | Frequency | Error (%) | Frequency | Error (%) |
| 1 | 112.139 | 110.57 | 1.4 | 102. 45 | 8.64 |
| 2 | 156.911 | 152.86 | 2.58 | 151.772 | 3.28 |
| 3 | 267.166 | 249.23 | 6.71 | 244.835 | 8.36 |
| 4 | 292.486 | 279.525 | 4.43 | 268.39 | 8.24 |
| 5 | 438.155 | 434.16 | 0.91 | 415.327 | 5.21 |
| 6 | 528.56 | 515.78 | 2.42 | 512.227 | 3.09 |

**Table 5** MAC comparison of framework between tested vibration shape and updated model

| Modal order | Healthy condition | Damage condition 1 | | Damage condition 2 | |
|---|---|---|---|---|---|
| | | MAC | Error(%) | MAC | Error(%) |
| 1 | 0.926 | 0.915 | −1.19 | 0.814 | −8.64 |
| 2 | 0.874 | 0.887 | 1.49 | 0.838 | −4.12 |
| 3 | 0.937 | 0.926 | −1.17 | 0.783 | −16.44 |
| 4 | 0.934 | 0.902 | −3.43 | 0.805 | −13.81 |
| 5 | 0.903 | 0.865 | −4.21 | 0.761 | −15.73 |
| 6 | 0.891 | 0.843 | −5.39 | 0.907 | 1.79 |




(a) condition 1 (b) condition 2

**Fig. 5** Damage detection result

The SRC of damaged unit group breaks greatly in Fig. 5, so the damaged unit group is identified.

## 3 Damage Detection Based on Modal Flexibility Curvature Difference

The damage detection for framework in Sects. 2.2.3 and 2.2.4 of this paper only can detect the damaged unit group, where the location is not accurate. The modal flexibility curvature difference is applied to locate the damaged element of the framework.

Research by Pandey and Biswas [4] shows that modal flexibility matrix is more sensitive than modal frequency and vibration shape. Based on the former research, the paper puts forward a local damage detection method—modal flexibility curvature difference method.

### 3.1 Modal Flexibility Curvature Difference

The modal flexibility curvature difference put forward in this paper calculates the modal flexibility curvature matrix $CF_u$ and $CF_d$ before and after damage according to the following:

$$CF_j^u(i) = \frac{F_j^u(i-1) + F_j^u(i+1) - 2F_j^u(i)}{2l_i^2} \tag{11}$$

$$CF_j^d(i) = \frac{F_j^d(i-1) + F_j^d(i+1) - 2F_j^d(i)}{2l_i^2}, \tag{12}$$

where $F_j^u(i)$ and $F_j^d(i)$ are, respectively, the $j$th column and the $i$th row element before and after damage. The difference of modal flexibility curvature matrix is gotten as shown in Eq. (13):

$$\Delta CF = CF_u - CF_d. \tag{13}$$

Sum of the elements of each row in $\Delta CF$ is gotten by addition and are divided by the knot number to get average, which consist a new column matrix as follows:

$$\Delta\{CF\}_{average} = \{\Delta_1, \Delta_2, \cdots, \Delta_n\} \tag{14}$$

$$MFCD = \Delta_i / \Delta_{\max}, \tag{15}$$

where $i = 1, 2, \ldots, n$ is the number of knot or test point, and $\Delta_{\max}$ is the max of $\Delta_i$.

*MFCD* is one-to-one corresponding with knot location, whose value reflects change degree of modal flexibility curvature after damage. When the local damage happens, the stiffness drops and local flexibility increase, which enlarge the modal flexibility curvature of damaged location. If the modal flexibility curvature difference *MFCD* between the two adjacent knots is far bigger than the others, the element between the two knots is the damaged element.

### 3.2 Damage Detection Experiment

The 27th unit group in Fig. 2 is divided into 15 units. 16 sensors are set from left to right to get the local modal in x and z directions before and after damage as shown in Fig. 6. The practice damage of the 27th unit group in condition 2 is set in 1/4 location from left to right, which is between the 4th and 5th knots. The curve in Fig. 7 is the MFCD value of each knot after damage.

From Fig. 7, it can be seen that the *MFCD* between the 4th and 5th knots breaks evidently which indicates the damage location.

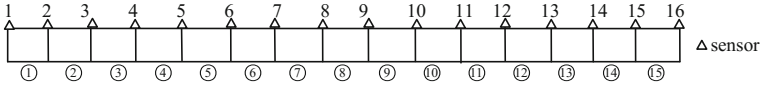



**Fig. 6** Partition of the 27th unit group and sensors allocation

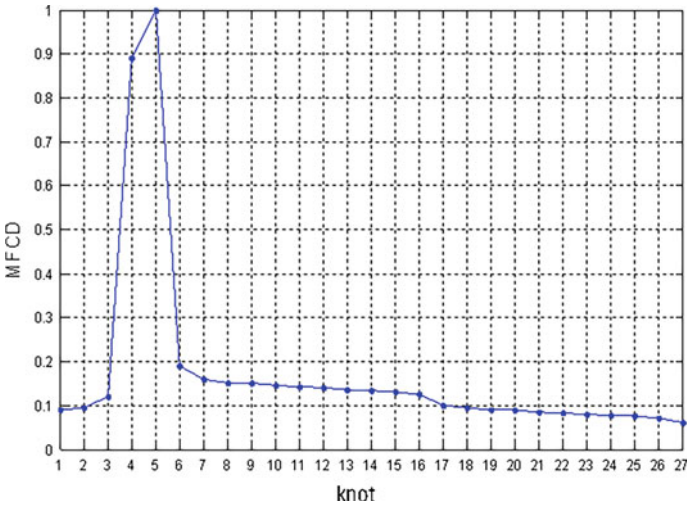



**Fig. 7** Damage detection with *MFCD*

## 4 Conclusion

A two-step damage detection method for large and complex structure is put forward in this paper based on model updating and modal flexibility curvature difference. The numerical simulation and experiment on a framework show that the method is efficient in damage detection, which provides a new idea for damage detection of large and complex structure.

## References

1. Q. Wei-lian, H. Dong-mei, Two-step damage diagnosis methods of large and complex structure. World earthquake Eng. **19**(2), 72–78 (2010)
2. E.P. Carden, F. Fanning, Vibration based condition monitoring a review. Struct. Health Monit. **3**(4), 355–377 (2004)
3. X. Jin-peng, Study on the method for dynamic optimum design of structures based on trust region method. Beijing: Tsinghua University (2006)
4. A.K. Pandey, M. Biswas, Damage detection in structures using changes in flexibility. J. Sound Vib. **169**(1), 3–17 (1994)

5. S. Hassiotis, G.D. Jeong, Identification of stiffness reduction using natural frequencies. J. Eng. Mech. ASCE **121**(10), 1106–1113 (2005)
6. N.R. Conn, N.I.M. Gould, PhL Toint, *Trust-Region Methods* (MPS/SIAMSeries on Optimization, SIAM and MPS, 2000)
7. M.A. Branch, T.F. Coleman, Y.A. Li, Subspace, interior, and conjugate gradient method for large-scale bound-constrained minimization problems. SIAM J. Sci. Comput. **21**(1), 1–23 (1999)
8. Z. Li, K. Liu, Y. Zhao et al., MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)
9. Z. Li, Y. Chen, H. Shi et al., NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. **48**, 1–5 (2016)
10. X. Liu, Z. Li, P. Yang, et al., Information-centric mobile ad hoc networks and content routing: a survey. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.04.005
11. Z. Li, L. Song, H. Shi, et al. Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016), http://dx.doi.org/10.1016/j.adhoc.2016.02.009

# The Dynamic Encryption Method Based on ECG Characteristic Value

**Huiqian Wang, Tong Bai, Yu Pang, Wei Wang, Jinzhao Lin, Guoquan Li, Qianneng Zhou, Zhangyong Li and Xiaoming Jiang**

**Abstract** Body area network (BAN) is a key technology of solving remote medical, where protecting security of vital signs information is a very important technique requirement. This paper presents a data security scheme based on ECG characteristics, investigating methods of the circuit-level data encryption and dynamic key refreshment due to the obtained ECG signal characteristic value. The proposed scheme is constructed from the inherent features of the BAN system, adopting vital information from security protection, which has advantages of high strength, low cost, and easy implementation. Therefore, the scheme helps to design low-power sensor nodes and provides theoretical support and engineering implementation.

**Keywords** Body area network · ECG · Characteristic value · Encryption

## 1 Introduction

Almost since the birth of BAN, it has caused a strong interest in academia and industry all around the world. BAN can obtain the vital signs in natural state, supporting the remote clinical diagnosis, emergency treatment, and health information service. For the user, the vital signs information is extremely private. Different from the common wireless sensor network, security is one of the main factors in the application of BAN, which requests to support the identity authentication, the privacy protection, and the information completeness.

H. Wang · T. Bai (✉) · Y. Pang (✉) · W. Wang · J. Lin
G. Li (✉) · Q. Zhou · Z. Li · X. Jiang
Chongqing University of Posts and Telecommunications, Chongqing, China
e-mail: baitong03@126.com

Y. Pang
e-mail: ppyu78@gmail.com

G. Li
e-mail: ligq@cqupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_46

Since 2005, the research on the security of BAN has carried out a series of methods. It first presents a method whose key is distributed beforehand. When a sensor node sends the information, only the sensor nodes in the same system or the intelligent terminals can decrypt the received information. H. Chan puts forward a multipath reinforcement method, which uses pre-distribution of random number and can strengthen the security between the nodes, effectively resisting the small attack [1]. However, the algorithms of encryption and decryption mentioned before require a large number of operations, leading to the power consumption increasing rapidly. Considering the requirements of the extremely low-power BAN nodes, the BAN system is not suitable to use the key pre-distribution scheme.

The second method uses the characteristics of the BAN channel as the key, which can realize a one-time one key. This method can randomly generate symmetric key between the sides of communication in the physical layer, totally resisting the eavesdropping outside the $\lambda/2$. The paper [2] generates the key in the receiver according to the value calculated of received wireless channel signal strength. Because of the inconsistent distance of the eavesdropper and the receiver, the signal intensity values will be different and the eavesdropper is unable to decrypt the data. The scheme using BAN channel encryption has been widely studied in recent years, but its keys are generated according to the signal intensity in the receivers. This method is too simple which may fail when the transmitters have the same distance with the eavesdropper and the receiver.

The third method generates the keys according to the physiological signs information [3–8]. Different from the general sensor network nodes, the BAN nodes' parameters are the human physiological signs, which show different features according to the different individuals. Even for the same individual, the values are not constant appearing to change slowly over time. Therefore, these parameters can be used as keys, especially the ECG signal containing much information. The paper [3] extracts the R-wave of ECG to calculate the heart rate variability as individual identification information and generate the keys for the whole BAN system in the initial connection stage of wireless nodes. This method has the advantage of no extra overhead when distributing keys and no strict time synchronization [9].

Although the method mentioned above is suitable for BAN system, some characteristic parameters are too approximate to being distinguished with tiny changes over time, which may lead that the encryption intensity is not enough. Usually, it needs a long time to generate the initial key, for instance using the heart rate variability (HRV) generally requires at least 5 min for the ECG signal acquisition.

In this paper, we design a method whose keys are based on the QRS complex (the central and most visually obvious part of the ECG signal under normal conditions), which has high encryption intensity and lower power consumption because this scheme is constructed from the inherent features of the BAN system.

## 2 The Encryption Method

### 2.1 The Generation of Key

After detecting the QRS waves of $t$ seconds ECG signal, the characteristic value can be calculated as

$$\lambda_i = R^i_{amp} \times T^i_{QS} \tag{1}$$

$$\lambda = \frac{\sum_{i=1}^{n} \lambda_i}{n}, \tag{2}$$

where $R^i_{amp}$ is the amplitude of the $i$th R-wave within the $t$ seconds signal and $T^i_{QS}$is the $i$th QS-wave time width within $t$ seconds of the ECG signal and $n$ is the number of R-waves within $t$ seconds shown in Fig. 1. The parameter $\lambda_i$ is calculated by multiplication of $R^i_{amp}$ and $T^i_{QS}$, and the final characteristic value is the $\lambda$ mean of $\lambda_i$.

Then the characteristic value $\lambda$ is quantified as the key to be the initial value of the linear feedback shift register (LFSR) with $N$-order. The generated key is broadcast to every BAN nodes, and then they can use the key to decrypt the data. The method regenerates the key and repeats the above steps every $T$ seconds.

### 2.2 The Encryption Flow

In this paper, the value of $T$ could be 200–400 s, and the value of $t$ could be 4 s. Using the LFSR to produce the stream cipher has advantages of the fast conversion speed and simple hardware implementation. Here we use the 16-order LFSR to

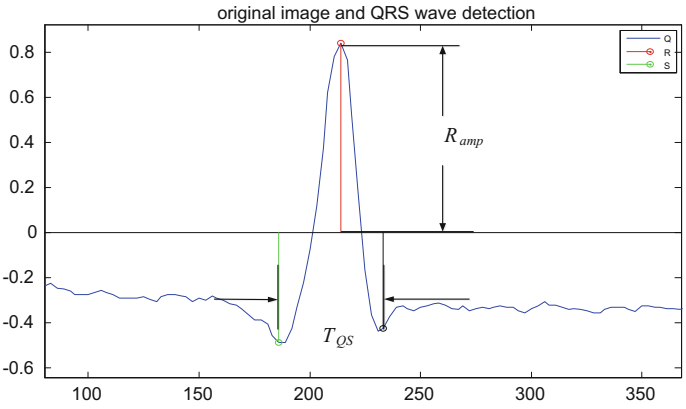



**Fig. 1** ECG signal and QRS localization results

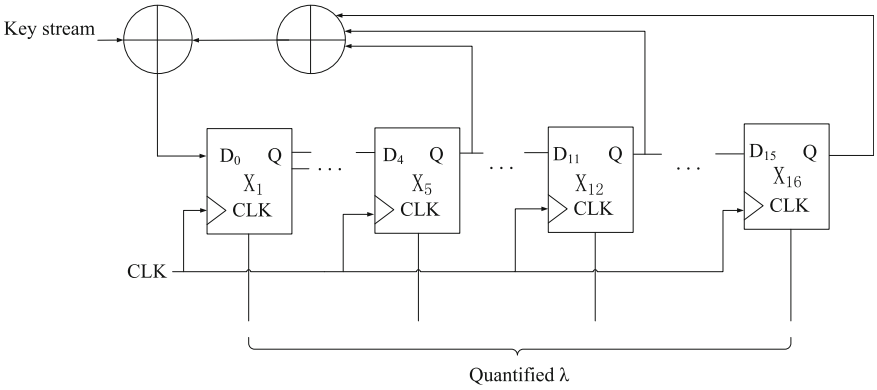



**Fig. 2** The generation of ciphertext base on the LFSR

encrypt the plaintext, whose maximum value is allowed to be $2^N-1$. Moreover, the dynamical updation of the key from the ECG signal every $T$ seconds can realize the real-time encryption, which can extremely lower the possibility of attack.

The polynomial of 16-order LFSR is

$$F(X)=X^{16}+X^{12}+X^5+1. \tag{3}$$

The generation of ciphertext is shown in Fig. 2.

In the BAN initialization, the ECG node sends the preamble sequence and ECG data with a certain length without calculation. The coordinator extracts 4 s ECG data and calculates the key after the reception of the preamble sequence, and then the coordinator broadcasts the key to all nodes.

Receiving the encrypted data generated by the node, the intelligent terminal will calculate the characteristic value according to the ECG signal. In this paper, the amplitude of R-wave and time interval of QS-wave are used, which has large difference among individuals. The flow of the proposed encryption method is shown in Fig. 3.

Since the BAN system requires the node of physiological sign acquisition which should have extremely low consumption, the encryption algorithm should be as

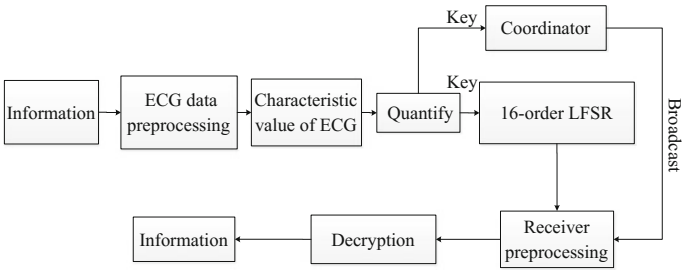



**Fig. 3** The flow of the proposed method

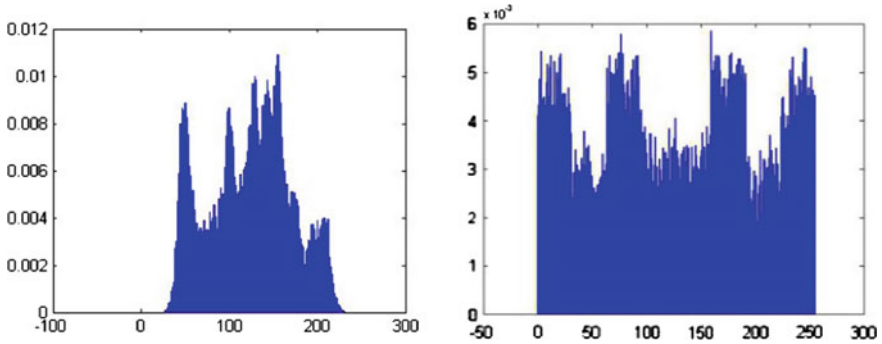
simple as possible. As the method mentioned before, the additional function for the node does not need the extra cost except a simple LFSR circuit. This is very suitable for the implementation of the low-power sensor nodes.

## 3 Experimental Results

The simulation data comes from the MIT-BIH database. Using the 100th data as an example, we can obtain the following results. Figure 4 shows the encryption result of the picture Lena, while Fig. 5 shows the histograms of the original data and encrypted data.

(a) The original image (b) The encrypted image

**Fig. 4** The encryption effect of the proposed algorithm




(a )The original image (b) The encrypted image

**Fig. 5** The original and encrypted statistical histograms

**Table 1** The correlation coefficient of the three directions

| Correlation coefficient | Vertical | Horizontal | Diagonal |
|---|---|---|---|
| Original image | 0.9388 | 0.9633 | 0.9417 |
| Encrypted image | 0.0672 | 0.0218 | 0.0735 |

As shown from Figs. 4 and 5, the original image and encrypted image show obvious change in the domains of time and frequency. In order to analyze the relevance, we calculate the two-dimensional correlation coefficient between the two data matrices defined as follows:

$$\gamma = \frac{\operatorname{cov}(x,y)}{\sqrt{D(x)}\sqrt{D(y)}} \tag{4}$$

$$\operatorname{cov}(x,y) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))(y_i - E(y))^2 \tag{5}$$

$$D(x) = \frac{1}{N}\sum_{i=1}^{N}(x_i - E(x))^2 \tag{6}$$

$$E(x) = \frac{1}{N}\sum_{i=1}^{N} x_i. \tag{7}$$

Here $x_i$ and $y_i$ represent gray values of two adjacent pixels and $N$ is the pixel number. The correlation coefficients of the three directions are shown in Table 1. The correlations in three directions are all beyond 0.9 that denotes adjacent pixels have very similar gray values before encryption, while they decrease to under 0.1 after encryption, which means that the much smaller correlation can make the better resistance to attack.

The entropy of each pixel is calculated as

$$H(M) = \sum_{i=1}^{256} P(m(i))\log_2 P(m(i)). \tag{8}$$

In order to analyze the ability of resisting differential attack, we calculate the *NPCR* (Number of Pixels Change Rate) and *UACI* (Unified Average Changing Intensity):

$$NPCR = \frac{1}{M \times N}\sum_{i=1}^{M}\sum_{j=1}^{N} D(i,j) \times 100\%. \tag{9}$$

$D(i, j)$ is the gray value difference between the original and encrypted images in the pixel $(i, j)$. Gray values of the pixels may change after encryption, so *NPCR* plays a good indicator to reflect the gray value change in number but non-effective

**Table 2** The analysis result of the different data of MIT-BIH

| Data number | 100 | 101 | 102 | 103 | 104 |
|---|---|---|---|---|---|
| Entropy | 6.2967 | 6.0996 | 6.3545 | 6.3912 | 6.3653 |
| NPCR (%) | 99.6025 | 99.4525 | 99.4825 | 99.7 | 99.4935 |
| UACI (%) | 50.1637 | 49.9621 | 49.9241 | 50.142 | 49.9871 |

to express the degree of gray value change. The average gray value change is necessary for evaluation. Therefore, we give the definition of *UACI*:

$$UACI = \frac{1}{M \times N} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{|c_1(i,j) - c_2(i,j)|}{255}. \tag{10}$$

Here $c_1(i, j)$ and $c_2(i, j)$ represent the gray values of the original and encrypted images in the pixel $(i, j)$. Using the 100th–104th data of MIT-BIH,the result is shown in Table 2.

## 4 Conclusions

Body area network is a key technology for solving remote medical problems, where protecting security of vital signs information plays a role of important technique requirement. The encryption method proposed in this paper makes use of the ECG characteristic value as the key and introduces a 16-order LFSR to encrypt the plaintext, which can effectively hide the private information with advantages of dynamic key, low cost, rapid operation, and easy implementation. The investigation of encryption method is very helpful to design low-power sensor nodes, which provides theoretical support and engineering implementation.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China (Grant No. 61301124, 61471075, 61671091), the National Science and Technology Pillar Program (2014BAI11B10), the Natural Science Foundation of CQ CSTC (cstc2016jcyjA0347), the University Innovation Team Construction Plan of Smart Medical System and Core Technology,the Enhancement Plan of Chongqing Key Laboratory of Photoelectronic Information Sensing and Transmitting Technology,Wenfeng Talented Plan of CQUPT.

## References

1. H. Chan, A. Perrig, D. Song, Random key predistribution schemes for sensor networks, in *Symposium on Security and Privacy* (2003), pp. 197–213
2. L. Shi, M. Li, S. Yu, J. Yuan, BANA: body area network authentication exploiting channel characteristics. IEEE J. Sel. Areas Commun. **31**(9), 1803–1816 (2013)

3. S. Bao, Y. Zhang, L. Shen, Physiological signal based entity authentication for body area sensor networks and mobile healthcare systems, in *Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (2005), pp. 2455–2458
4. Z. Zhang, H. Wang, A.V. Vasilakos, H. Fang, ECG-cryptography and authentication in body area networks. IEEE Trans. Inf. Technol. Biomed. **16**(6), 1070–1078 (2012)
5. K. Venkatasubramanian, A. Banerjee, S.K.S. Gupta, Plethysmogram-based secure inter-sensor communication in body area networks, in *Proceedings of the IEEE Military Communications Conference* (2008), pp. 1–7
6. K. Venkatasubramanian, A. Banerjee, S.K.S. Gupta, EKG-based key agreement in body sensor networks, in *Proceedings of the 2nd Workshop Mission Critical Networks* (2008), pp. 1–6
7. K. Venkatasubramanian, A. Banerjee, S. Gupta, PSKA: usable and secure key agreement scheme for body area networks. IEEE Trans. Inf. Technol. Biomed. **14**(1), 60–68 (2010)
8. S. Cherukuri, K. Venkatasubramanian, S. Gupta, Biosec: a biometric based approach for securing communication in wireless networks of biosensors implanted in the human body, in *International Conference on Parallel Processing Workshops* (2003), pp. 432–439
9. H. Wang, H. Fang, L. Xing, M. Chen, An integrated biometric-based security framework using wavelet-domain HMM in wireless body area networks, in *IEEE International Conference on Communications* (2011), pp. 1–5
10. S. Ali, V. Sivaraman, D. Ostry, Secret key generation rate vs. reconciliation cost using wireless channel characteristics in body area networks, in *IEEE/IFIP International Conference on Embedded and Ubiquitous Computing* (2010), pp. 644–650
11. S. Jana, S.N. Premnath, M. Clark, S. Kasera, N. Patwari, S. Krishnamurthy, On the effectiveness of secret key extraction using wireless signal strength in real environments, in *International Conference on Mobile Computing and Networking* (2009), pp. 321–332

# An Improving Fuzzy C-means Algorithm for Concept-Drifting Data Stream

**Baoju Zhang, Lei Xue, Wei Wang, Shan Qin and Dan Wang**

**Abstract** Big Data has been expanding rapidly, concept drift in the data stream receives great attention and has been one of the research focuses. To solve the problems caused by dynamic nature in the data stream, this paper proposed a method of concept drift detection based on fuzzy C-means algorithm and a cumulative update mechanism, a clustering model is established which can not only detect the concept drift in time, but also avoid the problems caused by frequent updates. The results of Experiment on synthetic and real-world data show the efficiency of the proposed algorithm.

**Keywords** Data stream · Concept drift · Fuzzy C-means · Entropy theory

## 1 Introduction

Due to the development of network information technology, the data generated in all application field is characterized by mass, fast and changeable, we call them data stream.

As the dynamic characteristic of data stream, the concept of data is constantly changing over time. This change is called concept drift. The existence of concept drift will cause an immediate impact on the model accuracy. Considering the problems above, data mining algorithms have been asked to digest the data from data stream within limited storage space, and change the model in time to adapt to the data stream environment. The difficulty of addressing the data stream with concept drift, lies in how to ensure the accuracy and applicability of the model and adapt to new concepts as well.

Concept drift detection has been highly valued, and numerous works exist in related research. In [1], Zhou et al. estimated the confidence interval of the error rate

B. Zhang · L. Xue · W. Wang (✉) · S. Qin · D. Wang
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
e-mail: weiwangvip@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_47

of new concepts to detect the occurrence of concept drift, but it is a time delaying system, and this method is restricted to the class of the data samples. In [2], the kNN Model-based algorithm improves the performance, but the model complexity increases greatly. In [3], a dynamic weighted algorithm is proposed to capture the concept drift, in which the entropy theory based on eigenvector distribution is employed to achieve the transition between the old and new concepts, however, this method has a strict limitation on the structure of data stream.

Once the concept drift occurs, the current model is no longer adapted to the new arriving data chunk. CVFDT algorithm update the original subtree adaptively to adapt to new data stream environment [4], as for the change of local nodes which occurs often, INCREDB2 is proposed to make the adjustment to the concept drift occurred in local nodes [5], but the efficiency is reduced when dimension increases to some extent. [6] proposed an incremental classification model algorithm, speeded up the update rate, as a result, the frequent updates cause the deviation from the actual concept. The improved algorithm in [7] proposed a sliding window model which adaptively changes the size of window once the concept drift occurs. Based on sliding window, CFDT algorithm employs the tree index structure to process data by dividing it into different parts [8], and realize classification in real time, but its not applicable for high dimensional data. Chai proposed an online bagging algorithm [9], the system updates or rebuilds the Hoeffding tree adaptively, which the over adaptation would cause the weakening of the performance potentially.

Although there are numerous research on concept drift detection, most of them has the problems on complicated structure and the over adaptation caused by immediate updating. We notice that there are few research on fuzzy clustering, but it is promising and has great values on practice. Therefore, this paper proposed a method of concept drift detection based on fuzzy C-means algorithm and a cumulative update mechanism. In this paper, we use entropy calculated from the membership degree as the measure to detect the concept drift, and then accumulate the entropy to determine when to update system model, eventually, the system avoids the problems caused by frequent updates.

The rest of this paper is organized as follows: In Sect. 2, we proposed a method of concept drift detection based on the fuzzy C-means algorithm, Sect. 3 built a cumulative model updating mechanism, Sect. 4 contained the experimental evaluations of the proposed algorithms. Section 5 concluded the paper.

## 2 Concept Drift Detection of Data Stream

### *2.1 Fuzzy Clustering Theory*

In classical set theory, we can determine clearly whether a member belongs to the set or not, and means that its degree of membership equals to one or zero. Inversely, fuzzy set theory permits that the referenced criterion of membership degree is

subjective and unified, and elements' membership relation is not clear. The fuzziness measure can be denoted by:

$$d(\mu) = F\left[\sum_{i=1}^{n} c_i f_i(\mu_A(x))\right] \tag{1}$$

where, $F \geq 0, and\ F' > 0;\ c_i \in R^+;\ f_i(\mu) = f_i(1-\mu)$.

Given a set $X = \{x_1, x_2, \ldots, x_n\}$, and it can be divided into C categories $C_i (i = 1, 2, \ldots, C)$, whose corresponding probability can be denoted by $\mu_i (i = 1, 2, \ldots, C)$, those probabilities in fuzzy clustering satisfy [10]

(1) $$\sum_{i=1}^{C} \mu_i(x_k) = 1,\ k = 1,\ 2,\ \ldots,\ n;$$

(2) $$0 < \sum_{i=1}^{n} \mu_i(x_k) < n, i = 1, 2,\ \ldots, C.$$

In fuzzy clustering algorithm based on fuzzy mathematics theory, we get the probability in classifying x into different class, this method objectively represents the membership relation, and has been a mainstream solution. And the membership function curve in fuzzy clustering is shown in Fig. 1.

## 2.2 *Fuzzy C-Means Algorithm*

In the domain of clustering, due to the properties of fast convergence and adaptability in massive data environments, fuzzy C-means (FCM) algorithm is one of most widely used algorithms. For any given set $X$, membership function describes a mapping between $X$ and the unit interval [0, 1], usually use $\mu_A$ to represent the membership function of fuzzy set $X$.

The membership degree represents the quantification of membership of the element $x$ to the cluster center. If the value equals to 0, it means that $x$ is not a

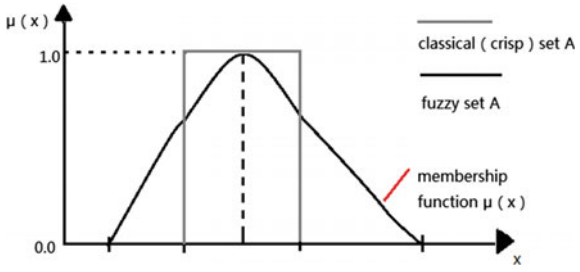



**Fig. 1** The membership function curve

member of this cluster at all; If the value is equal to 1, it means $x$ fully belongs to this cluster; And we consider the values between 0 and 1 as a proportion which belong, and not a truly rigorous probability.

FCM algorithm classify all data into C fuzzy sets, objective function can be expressed as

$$J(U, c_1, \ldots, c_c) = \sum_{i=1}^{n} \sum_{j=1}^{C} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \qquad s.t. \quad \sum_{j=1}^{C} u_{ij} = 1 \tag{2}$$

where, m is a weight value, and $1 < m < \infty$; $u_{ij}$ denotes the membership degree of $x_i \in$ class $j$; $c_j$ is the cluster center of class j.

Through iterative calculation of $u_{ij}$ and $c_j$, objective function can reach the minimum value,

$$u_{ij} = \frac{1}{\sum_{k=1}^{c} \left( \frac{d_{ij}}{d_{kj}} \right)^{2/m-1}} \quad , \quad c_i = \frac{\sum_{j=1}^{n} u_{ij}^{m} x_j}{\sum_{j=1}^{n} u_{ij}^{m}} \tag{3}$$

where k is the number of iterations, when $\left\| U^{(t)} - U^{(t-1)} \right\| < \varepsilon$, then stop the iteration procedure.

In our work, according to the definition of information entropy, we introduce membership degree into entropy theory,

$$E(e_1, e_2, \ldots, e_n) = -\frac{1}{n} \sum_{j=1}^{n} \sum_{i=1}^{C} (u_{ij} \ln(u_{ij}) \tag{4}$$

In (4), when calculating entropy, we take membership degree into account, and draw a diagram consists of entropy values of every corresponding element, and which can be used in the detection of the occurrence of concept drift.

## 2.3 Detection of Concept Drift

Because of the occurrence of concept drift, the system has the low accuracy rate and there has demand to update the model promptly to adapt to new data environment. Meanwhile, frequent updating will cause an adverse effect on the performance of clustering accuracy and system efficiency, even lead to the paralysis of clustering system. In this paper, a cumulative update mechanism is proposed to solve this problem, which requires accumulation of the slight drift, and not updating the model until the variance of entropy is larger than the preseted threshold.

For each instance, FCM algorithm employ the membership to describe the fuzzy characteristics, and we perceive it as the probability in statistics, and plug it into the

entropy calculation formula, then employ it as a measure for overall system stability. The entropy based on membership degree is calculated by the following algorithm:

**Algorithm 1**

**Input:** (1) data set $D$ ; (2) $m$ :exponent/weight value of Partition matrix; (3) $p$ :Maximum number of iterations; (4) $e$ :Minimum error rate; (5) $C$ :the number of clusters。

**Output:** (1) Center point; (2) $U$ : membership matrix; (3) obj_fcn: the value of objective function in iterative process

1. Initialize $U$ ←the number of column in $U$ is 1
2. $i = 1,2\cdots p$ ;

$U_k$ ←calculate membership degree in the iterative equation, where the center point of cluster changes continuously.

**end**

3. **If** $i > 1$

**If** $\left|obj\_fcn(i) - obj\_fcn(i-1)\right| \le e$ ←check the terminating condition;

break;

**end**

**end**

4. $i = 1,2\cdots n$

$u_i \in U$ , $E(e_1, e_2, \ldots, e_n) = -\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{C}(u_{ij}\ln(u_{ij}))$ ←calculate the entropy of data stream

**end**

From the definition of entropy, its apparent that the corresponding entropy value is about 0.1 if the distributions of data stream are stable, conversely, then the entropy value will change proportionally as the result of distributions' change, if the distributions of data stream tend to be stable over time, the entropy also decreases, and is held at around 0.1. Therefore, in this paper we use the change of entropy value to measure the ordering degree in the data stream. By observing the entropy value change of corresponding data chunk, it can be determined whether the concept changes or not over time, such that the concept drift detection can be concluded.

## 3 Updating Mechanism

In the domain of stream environment, we need to update model to adapt to the new data and reduce the influences caused by the occurrence of concept drift. Meanwhile, one-sidedly focusing on model updating will cause excessive occupancy of resource, therefore, avoiding unnecessary updating is the key to improve the efficiency and accuracy of the algorithm.

In this paper, we introduce the classifier pool and propose a cumulative update mechanism, where the system determine whether to update present model by comparing cumulative entropy value with the preset of threshold.

Given a data set $D$, we partition it into N data chunks, $D_1$, $D_2$, $D_3$, …, $D_N$.

The correlation coefficient of two data chunks is given by (X), where x, y denote two different data chunks,

$$\rho_{x,y} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{5}$$

then, plug the value into Eq.(X), and calculate the sum of entropy Es,

$$\mathrm{Es} = -\sum_{i=1}^{n} \rho_{x,y} \ln(\rho_{x,y}) \tag{6}$$

the variance of $E_s$ is then calculated by

$$\sigma = \frac{1}{n}\sum_{i=1}^{n}(E_{sn} - \bar{E}_s)^2 \tag{7}$$

As mentioned before, we preset a threshold, and compare variance $\sigma$ with $\alpha$.

If $\sigma > \partial$,

a. update the model;
b. Es←0 that is to say, we reaccumulate the entropy after the model updating.

If $\sigma > \partial$, do not update.

In our previous study, we use the mean entropy value as the measure to detect the concept drift in the data stream, the entropy calculated based on the correlation coefficient can show the similarity and the change of concepts. The concept drift occurred in the data stream at an unknown time, but in this situation we still accumulate the entropy which is calculated before the concept drift occurred, then we obtained the mean value, which will cause a negative impact on the system. In this paper, we use the variance as a measure to detect the concept drift, a low variance value represents that the data has a slight change. Similarly, by setting a certain threshold, we determine when to update the model. The difference is that we set the sum of entropy as 0 and reaccumulate the entropy after the model updating. We will compare their performance in Sect. 4.

With the deposit of continuous new model update classifier pool. When new data is coming, we use Graham algorithm to extract the shape feature and center point of data chunk. If its an already existing model, then we extract and use it to process data directly; if the models in the pool cannot match the new model, save the new shape feature into the classifier pool.

We calculate the variance of entropy sum, compare it with preset threshold, and update the model until $\sigma > \partial$. Meanwhile, by extracting the shape feature and center point of continuously arrived data chunk, the classifier pool is established. We store the shape feature to improve the efficiency when the new model has similarity with existing models in the classifier pool, that we can reuse the already existing model to train new data chunk.

## 4 Experimental Analysis

We use synthetic data set to validate the proposed algorithm. Since synthetic data is the virtual data, which cannot represent the complexity of the real data, so it is also needed to test on the real-world data set. As a result, the experimental data set is divided into two parts, the synthetic data and the real-world data set.

### *4.1 Synthetic Data*

The synthetic data includes Gaussian data and circular distribution data, and we observe the performance on the two data sets, respectively.

Gaussian data has two attributes: mean and variance. Gaussian data sets with different mean and variance are used to simulate the concept drift. The circular distribution data has two attributes: center and radius, we change the radius to facilitate the observation. The attribute parameter value of two data sets is shown as Table 1.

#### 4.1.1 Gaussian Data

Figure 2 shows the fuzzy clustering result and the entropy curve of Gaussian data, the x-axis denotes the order that the data appears, y-axis denotes its value. Table 2 also lists the result, from those we know that the accuracy rate reaches 90%. When $800 < n < 1000$, there has a junction area, where predictions appear errors, the mean value changes, peak value in the curve appears with corresponding entropy increases, therefore, it suggests that the concept drift has occurred; as the data increasing the data distribution becomes stable, entropy decreases and keep at a low value.

**Table 1** Attribute parameter values

| Data sets | Expression | Variations |
|---|---|---|
| Gaussian data | $f(x) = e^{-(x-\mu)^2/\varepsilon^2}$ | $\mu = 1, 2, 3, 4, 5, 6$ |
| Circular data | $(x-a)^2 + (y-b)^2 \leq r^2$ | a = 1, b = 1, r = −2, −1.5, −1,1, 1.5, 2 |

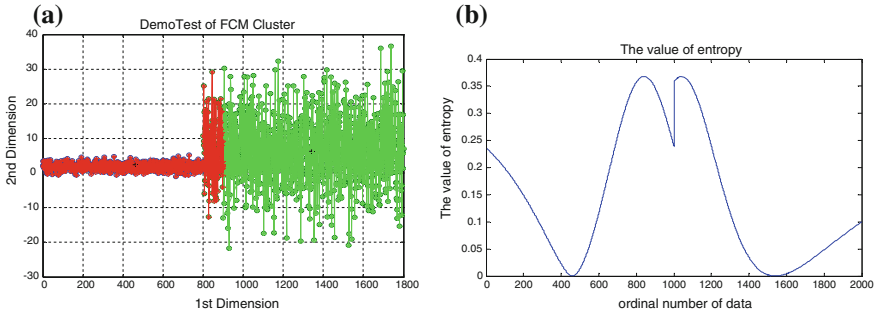



**Fig. 2** Experimental results on Gaussian data. The x-axis denotes the order that the data appears, y-axis denotes its value. **a** The clustering result when $\mu = 1, 6$, *red* area represents the category when $\mu = 1$, and *green* area represents the category when $\mu = 1$. **b** the entropy curve

**Table 2** Experimental accuracy of the Gaussian data

| Feature attribute | Correctly predicted instances | Incorrectly predicted instances | Instances | Accuracy (%) |
|---|---|---|---|---|
| N(1, 1) | 900 | 100 | 1000 | 90 |
| N(6, 9) | 900 | 100 | 1000 | 90 |

### 4.1.2 Circular Distribution Data

As shown in Fig. 3, its the sectional drawing of high dimensional space of the circular distribution data, and circular regions with different colors represent different concept drift caused by the change of radius, the entropy curve is also shown in Fig. 3. In the red, yellow, and green regions, we can clearly see that corresponding data has a larger entropy value due to the rapid concept changes; instead, in blue region, the distribution is stable, so the entropy decreases and keep at a low

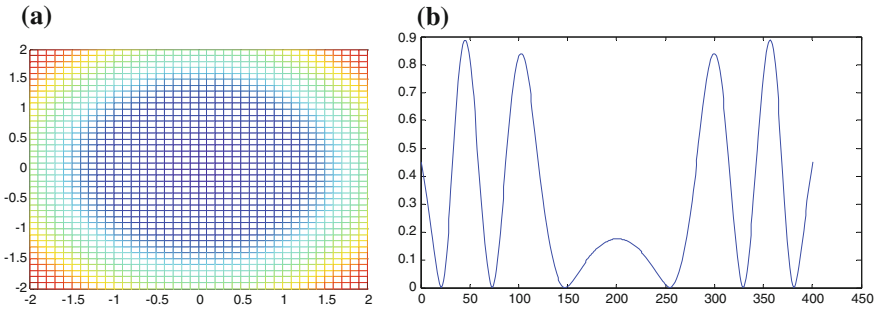



**Fig. 3** Experimental results on circular distribution data. **a** the sectional drawing of circular data, different color represents circular data with different radius. **b** the entropy curve, the x-axis denotes the order that the data appears, y-axis denotes its value

value. The concept changes over time, the entropy increases correspondingly, which suggest that the concept drift has occurred.

## 4.2 *Real-World Data*

Compared with the synthetic data, in real-world data environment, it is more complicate and it has a higher error rate. However, it has a good engineering application value.

We choose the Italy Power Supply Stream from the UCI data repository [11]. This data set contains hourly power supply of an Italy electricity company and its 3-year power supply records, and our learning task is to predict when to update the model. In this paper, we focus on the concept drift caused by hours of a day (e.g., morning and evening).

Figure 4 shows the fuzzy clustering result and the entropy curve of power supply stream. In the fuzzy clustering result, green area represents the concept when the time is 00 am, black area represents the concept when the time is 01 am, and red area represents the concept when its 21 pm. From the correlation coefficient we calculate, we know that the data chunk in 01 am has a strong similarity with 00 am, the data chunk in 21 pm has a weak similarity. In the entropy curve, the x-axis denotes the order that the data appears, y-axis denotes its value. According to the entropy we calculate based on the membership degree of the fuzzy C-means algorithm, as shown in Fig. 4, 01 am has a stable curve, which is coordinated with their similar distribution, while the curve of 21 pm is fluctuate and it suggests that concept drift occurs at 21 pm.

The accuracy of power supply stream is shown in Table 3. Consider the occurence of concept drift, real-world data has a lower accuracy rate.

As shown in Fig. 5a and b, we calculate the correlation coefficient between the value of the each time point of a day and the value in 0 am, and then we obtain the corresponding entropy, and through the calculation of entropy we get its mean entropy and variance, which is shown in Fig. 6a and b, we report the curve of value

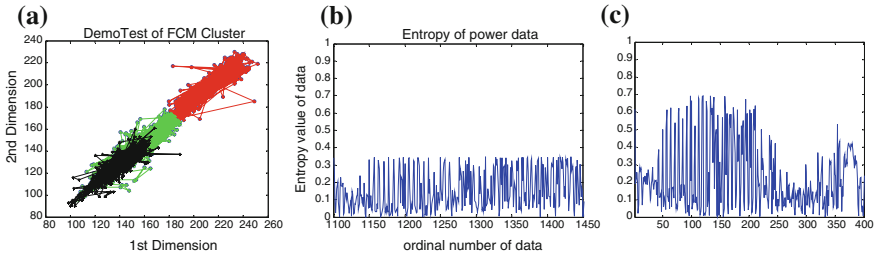



**Fig. 4** Experimental results on power supply stream. **a** The clustering result. **b** the entropy curve of 00 am and 01 am. **c** the entropy curve of 00 am and 21 pm

**Table 3** The accuracy of power supply stream

| Time | Correctly predicted instances | Incorrectly predicted instances | Instances | Accuracy (%) |
|---|---|---|---|---|
| 00 | 762 | 333 | 1095 | 69.6 |
| 21 | 839 | 226 | 1095 | 76.6 |

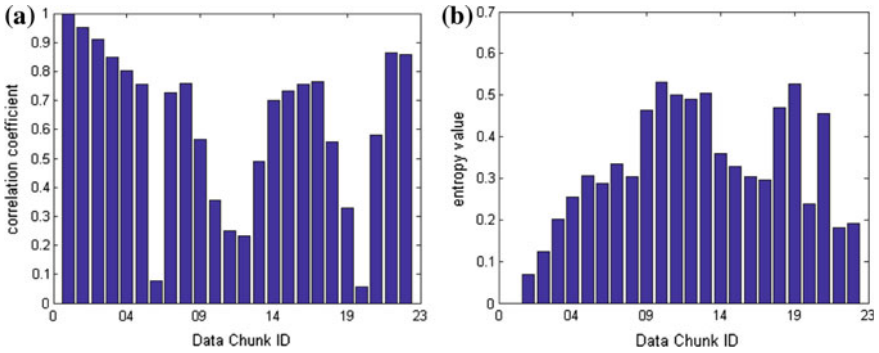



**Fig. 5** The correlation coefficient and the entropy value of the each time point of a day and the value in 0 am

and the variance and compare their performances, which are viewed as the measures to detect the concept drift in the data stream.

It shows that when it's 06 am, the two data chunk has a smaller correlation coefficient value, means that they has a weak similarity and the concept in the data has changed; and in this case, we accumulate the entropy, its the same in both methods.

From Fig. 6a it can be easily found that the mean value has a smoother curve, the mean value of entropy 00–06 am is 0.1592, and the mean value of 00–09 am is

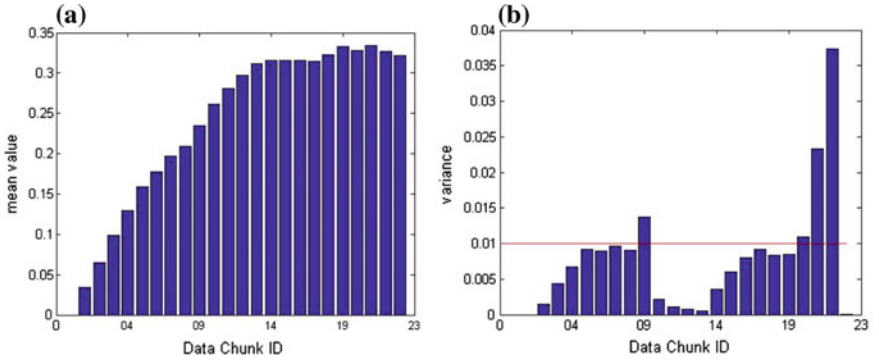



**Fig. 6** The mean value and the variance of entropy

0.2089, we set the threshold to 0.20, and at 09 am the system updates the model; as the accumulation of entropy, the mean value increases, on the one hand, the threshold has to change correspondingly; on the other hand, the curve tends to be stable, the change of mean value becomes smaller, and the system is slow to respond to the change of data, the occurrence of concept drift may not been detected in some cases.

For the situation that variance is used to detect the concept drift, from Fig. 6b we know that the variance of entropy 00–06 am is 0.0089, and the variance of 00–09 am is 0.0138, the increasing of variance means that the increasing extent of drift. When the variance is larger than threshold, we update the model. The threshold represents the tolerance of concept drift of the system, in this paper, the threshold is set to 0.01. The variance is bigger than 0.01 when its 09 am, the model updates.

After the model is updated, we reset the variance value to 0, and calculate the entropy from current data chunk, that is to say, we reaccumulate the entropy since updating the system. The variance of entropy 09 am–17 pm is 0.0092, and variance is 0.0109 until 20 pm, we update the model when its 20 pm.

Compared with the mean value, the system use the variance has a better performance, the choice of threshold is easier, and the detection of concept drift is more accurate and efficient. On Gaussian data, Circular data and power supply data, we detect the occurrence of concept drift, and the entropy curve based on the membership degree demonstrates its occurrence clearly. Meanwhile, the experiments on power supply data show the feasibility of the updating mechanism.

## 5 Conclusions

In this paper, for data stream, where data is massive and concept drift evolved, we proposed a method of concept drift detection based on the fuzzy C-means algorithm, by the calculation of entropy based on membership degree, we use the entropy as the measure to detect the concept drift; then, we build a cumulative model updating mechanism, compare the variance of entropy with the preseted threshold and decide when to update the model, which can decrease of system accuracy caused by frequent updates. Finally, experimental results on synthetic data and power supply data confirmed the validity of the proposed algorithm.

**Acknowledgements** This paper is supported by Natural Science Foundation of China (61271411) and Natural Youth Science Foundation of China (61501326, 61401310). It also supported by Tianjin Research Program of Application Foundation and Advanced Technology (15JCZDJC31500), National Natural Science Foundation of China (Grant No. 61501326), and Tianjin Science Foundation (16JCYBJC16500).

## References


1. Y. Zhang, Y. Chai, L Wang, Method of concept drifting detection based on martingale in data stream. J. Chin. Comput. Syst. **34**(8), 1787–1792 (2013)
2. G. Guo, N. Li, L. Chen, Concept drift detection for data streams based on mixture model. J. Comput. Res. Dev. **51**(4), 731–742 (2014)
3. G. Hulten, L. Spencer, P. Domingos, Mining time-changing data streams, in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. San Francisco, CA, United states, 2001, pp. 97–106
4. L.I. Kuncheva, I. Zliobaite, On the window size for classification in changing environments. Intell. Data Anal. **13**(6), 861–872 (2009)
5. Z. Yin, S. Huang, Adaptive method for handling local concept drift of data streams classification. Comput. Sci. **35**(2), 138–139 (2008)
6. H.A. Aboalsamh, A Boolean algebraic framework for association and pattern mining. Wseas Trans. Comput. 7 (2008)
7. L.I. Kuncheva, Combining pattern classifiers: methods and algorithms: second edition. Technometrics **47**(4), 517–518 (2014)
8. M.R. Henzinger, P. Raghavan, S. Rajagopalan, Computing on data streams, in *External Memory Algorithms* (American Mathematical Society, Washington, 1998), pp. 107–118
9. Y. Chai, C. Zhou, L. Wang, Detecting concept drift and classifying data stream. J. Chin. Comput. Syst. **32**(3), 421–425 (2011)
10. Z. Chen, Clustering and anomaly detection over data stream. Dissertation, Fudan University, 2009
11. Y. Chen, E. Keogh, B.Hu, N. Begum, A. Bagnall, A. Mueen, G. Batista, The UCR Time series classification archive (2015)

# Performance Analysis and Simulation of Turbo Coding System

**Zeng Liu, Jin Chen, Maolin Ji, Ying Tong, Lujia Wang and Hengxin Liu**

**Abstract** Turbo encoding and decoding system plays a key role in improving the reliability of digital communication, and it has been widely used. Based on in-depth analysis of the Turbo code encoding and decoding principle, a comprehensive analysis of the performance of Turbo code system, a simulation scheme and result of Turbo system are given, and based on practical application and simulation results to optimize the performance of Turbo code system.

**Keywords** Channel coding and decoding · Turbo code · System performance · Encoding · Decoding

## 1 Introduction

Since 1990s, with the continuous progress of the information transmission technology and the coding and decoding technology, Turbo code as an iterative decoding algorithm has become an important research focus in the field of channel coding and decoding. The central idea of [1] is dependent on the soft output decision algorithm and encoding technology of component codes to realize the concatenated code distance of pseudorandom encoding and iterative decoding. The advanced decoding technology development thanks to the discrimination algorithm for encoding and decoding technology, combined with the development of information transmission technology, and promote the development of Turbo Technology, further to promote the efficient development of the field of communication, to maximize the use of communication channel capacity, the interleaver in [2] encoding and decoding technology progress plays a key role. The development of

Z. Liu (✉) · J. Chen (✉) · M. Ji · Y. Tong · L. Wang · H. Liu
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
e-mail: 837521973@qq.com

J. Chen
e-mail: cjwoods@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_48

Turbo code through the parallel cascade structure and serial concatenated two stages of [3], the serial concatenated structure is in under the premise of combining the actual demand, for a series of modified parallel cascade structure, and a parallel encoding cascade structure is also the contemporary widely used decoding technology. Considering the importance of each link based on Turbo code design, design process, should be based on the actual demand, combined with the development of technology, comprehensive consideration of various aspects of the impact of link, the establishment of the overall idea, and then achieve the optimization of the design of Turbo codes. In this paper, the key link of the Turbo code analysis system of performance, and the results of the simulation analysis.

## 2 Principle of Encoder and Decoder for Turbo Codes

### *2.1 Encoder Composition*

The coding principle of Turbo code is shown in Fig. 1, the system consists of RSC (recursive system convolutional encoder), delete and reuse unit, and interleaving device. The code word of the encoder is output to the decoder through the channel, and the RSC [4] is a kind of systematic convolutional encoder with feedback for a class of bit rate R = K/N; As an important part of the system, the function of the system is to obtain the same contents as the original information, but the information sequence is arranged in different ways; For the integrity of the optimization of the rate, set up the delete and reuse unit, to obtain a code word of different code rate.

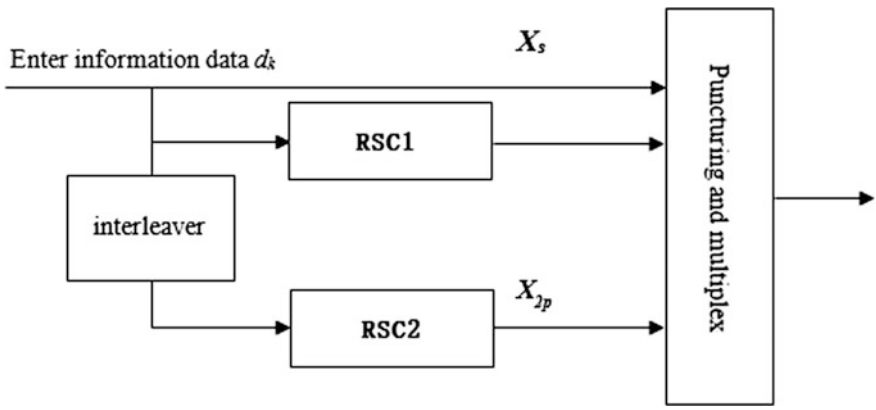



**Fig. 1** Coding principle of Turbo codes

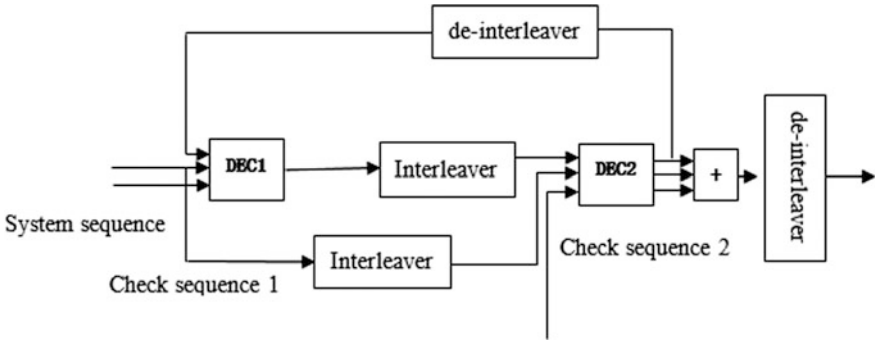



**Fig. 2** Basic structure of Turbo code decoder

## *2.2 Principle of Decoder*

The main components of the Turbo code decoder are shown in Fig. 2. The main components include the soft input/output (SISO) decoder DEC2 and, the interleaving device and the DEC1. Turbo decoder work flow is: the best use of the function of decoding decoder, the likelihood information in RSC1, the information and the information bit sequence information associated, while the transmission function of the interleaver, the likelihood information can be delivered to the outward transmission, DEC2 decoder, DEC2 will receive the likelihood information sequence is processed in the prior information the way of RSC2 component code and interleaver processing function, the information sequence generated for the likelihood information based on bit sequence, after the de-interleaver role transfer to DEC1, so, decoding, information processing cycle. In the iterative function, soft input and soft output (SISO) decoder output information sequence the likelihood ratio asymptotic value approximate maximum likelihood decoding, the best estimate of the information sequence in the optimal likelihood ratio value can be obtained on the basis of the analysis.

# 3 Performance Analysis of Turbo Code System

## *3.1 Influence of Iteration Number*

Taking into account the Turbo code system for iterative decoding algorithm, the number of iterations of the algorithm itself is bound to have a certain impact on the performance of the system. Generally, the number of iterations and the accuracy of the calculated results are proportional to [5]. The iterative algorithm used in the system mainly has two kinds: Log-MAP algorithm and SOVA algorithm, For this two kinds of methods for the number of iterations of bit error rate (BER) and signal-to-noise ratio (Eb/No) analysis of interaction relationship between the

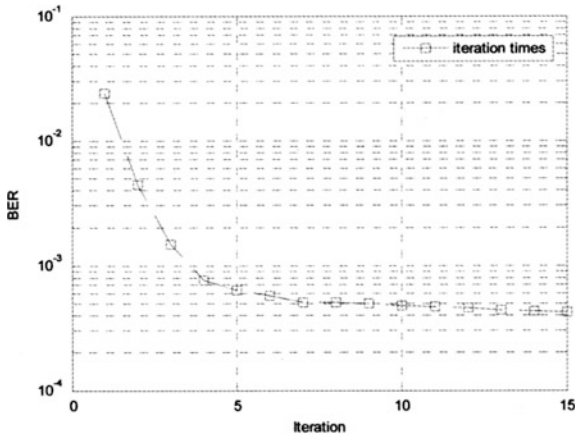



**Fig. 3** The influence of iteration number on performance

**Table 1** Code generation polynomial

| Constraint length K | Feedback generating polynomial f | Forward generating polynomial G |
|---|---|---|
| 2 | [1 1] | [1 0] |
| 3 | [1 1 1] | [1 0 1] |
| 4 | [1 1 0 1] | [1 1 1 1] |
| 5 | [1 1 1 1 1] | [1 0 1 0 1 1] |

number of iterations and show that the decoding rate showed an inverse proportion, and decoding the bit error rate is not unlimited with the increase in the number of iterations without limit decreases. In this paper, it is shown that the decoding performance curve of Log-MAP algorithm is analyzed under the condition of 1–15, the number of iterations is to 2, and the relationship between the number of iterations and the bit error rate can be directly grasped. As shown in Fig. 3, the bit error rate curve is stable after the iteration number is more than 5, which indicates that the error performance of the iterative number is less than after the iteration number is more than 5. On this basis to continue to increase the number of iterations is not only for the error performance of the improvement cannot play a better effect, but also makes the system delay. Therefore there is the best iterative decoding process, for the study of the optimal number of iterations, effectively reduce the computational complexity of iterative decoding cannot affect the bit error rate or the impact of the error rate is very low, has good practical value (Table 1).

## *3.2 Influence of Length of Code Constraint*

Based on the 5 iteration of the Log-MAP algorithm, the length of the constraint length RSC of the K = 2, 3, 4, 5 encoder, the interleaving length of L = 1 024,

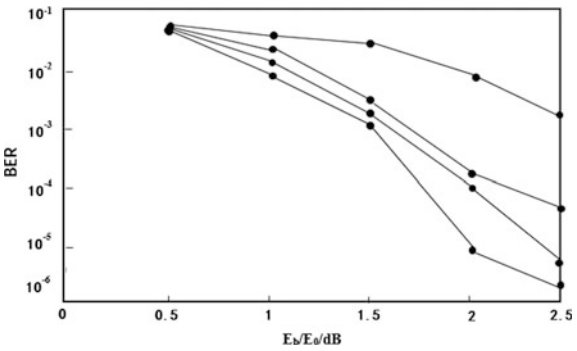



**Fig. 4** The effect of constraint length difference on its performance

Turbo r = 1/2 code corresponding to the code generation polynomial [6]. As shown in Fig. 1, the simulation image is shown in Fig. 4.

Comprehensive analysis of the coding generation polynomial and simulation results, the encoding constraint length and bit error rate presents the inverse proportion of [7]. However, the increase of the length of the encoding constraints will affect the complexity of the encoder's structure, the computation time and the delay performance of the system. Therefore, there is also the best matching between the length of the code constraint and the error rate, and the recursive convolutional code with the length of 4 of the code constraint can obtain the best performance of the system.

## *3.3 The Influence of Different Length of Weaving*

The main effect of the mixture is to go to the relevant. The weaving performance of the interleaving device mainly studies the influence of the interleaving length on the error correcting capability of the code word. In general, the interleaving length and the error correction performance of the code word show a positive proportional relation. With the increase of the interleaving length, the stronger the correlation performance (the larger the bit interleaving distance), however, the increase of the interleaving length will have a negative effect on the performance of the system. Therefore, it is necessary to consider the tradeoff between the error correction performance of the system and the delay performance of the system.

## *3.4 Comparison of Two Kinds of Decoding Algorithms*

In this paper, the decoding performance of SOVA algorithm and MAP-Log algorithm are compared and analyzed. As shown in Fig. 5, to determine the initial conditions of the number of iterations = 5 times, code constraint length = 4,

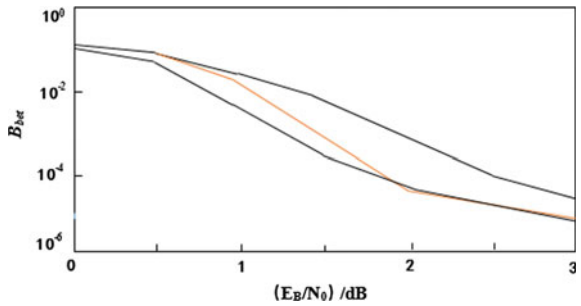



**Fig. 5** The influence of the compilation algorithm on the performance curve

weaving length = 1024, simulation calculation. According to the simulation image, it can be more intuitive to find: The curve of MAP-Log algorithm is lower than the SOVA algorithm, and the MAP-Log algorithm is faster than the SOVA algorithm; Under the same error rate requirements, the MAP-Log algorithm can obtain better coding gain than the SOVA algorithm [8].

## 4 Simulation Scheme and Results of Turbo Code System

In this paper, the system simulation of Turbo code is based on the Matlab language of the variable signal-to-noise ratio of the bit error rate performance analysis. The main components of the system: Recursive system convolutional encoder (2 units), recursive serial concatenated decoder (2 units), random interleaving (size 64 × 64), delete and reuse units. The Gauss white noise model is used in the channel, and the input mode of the [9] is adopted. In different SNR conditions, respectively, using Log-MAP and SOVA algorithm for system analysis. Coded {0/1} sequence BPSK modulation for the 1} {−1/sequence, delete and reuse of the unit under the effect of the source (matrix (07, 05) (2, 1, 2) convolutional code to get the code rate for Turbo 1/2 code. The maximum number of iterative decoding is 8 times, and the signal size of each frame is 1024.

The simulation results of the signal-to-noise ratio (SNR) and bit error rate (BER) performance are known:

(1) the ratio of the signal-to-noise ratio and bit error rate is inversely proportional to the same number of iterations.
(2) the same signal-to-noise ratio, the number of iterations and bit error rate is inversely proportional relationship, Log-MAP algorithm is better than the SOVA algorithm.
(3) the maximum likelihood decoding algorithm for complex and large delay system, the actual application effect is poor, and the SOVA algorithm is easy to realize, and the signal-to-noise ratio is higher than 0.7 dB under the condition of more than 18 times iterative decoding the bit error rate can reach 10–5.

## 5 Conclusion

In this paper, the principle of the encoder and decoder of Turbo code is analyzed in detail. And based on the two types of decoding algorithm for its system performance factors such as the number of iterations, the length of the length of the coding constraints, and so on. Through the Turbo code system simulation program and the results to guide the optimization of the actual performance of the system. Due to the excellent performance, Turbo code has been widely used, but there are still many places to be studied.

**Acknowledgements** Thanks for the supporting by Tianjin Edge Technology and Applied Basic Research Project (14JCYBJC15800), 2015 Google Supported National Undergraduate Innovation and Entrepreneurship Project.

## References

1. B. Vucetic, J. Yuan, *Turbo Codes: Principles and Applications* (Springer Science & Business Media, 2012)
2. C. Heegard, S.B. Wicker, *Turbo Coding*. (Springer Science & Business Media, 2013)
3. G.M. Vitetta, D.P. Taylor, G. Colavolpe, et al., *An Introduction to Channel Coding Techniques. Wireless Communications: Algorithmic Techniques* (2013), pp. 339–347
4. R. Joon, A. Nehra, Performance evaluation of symmetric turbo codes using different decoding algorithms. Imp. J. Interdisc. Res. **2**(8) (2016)
5. C. Berrou, A. Glavieux, Near optimum error correcting coding and decoding: turbo-codes. IEEE Trans. Commun. **44**(10), 1261–1271 (2010)
6. W. Tranter, D. Taylor, R. Ziemer, et al., *Near Optimum Error Correcting Coding and Decoding: TurboCodes* (Wiley-IEEE Press, 2007)
7. C. Sandu, random interleaver design in turbo coding for digital magnetic recording channel, in *information & communication technology electronics & microelectronics (MIPRO), 2013 36th International Convention on (IEEE*, 2013), pp. 137–139
8. J. Chen, J. Hu, High throughput stochastic log-MAP turbo-decoder based on low bits computation. IEEE Signal Process. Lett. **20**(11), 1098–1101 (2013)
9. J.P. Woodard, L. Hanzo, Comparative study of turbo decoding techniques: an overview. IEEE Trans. Veh. Technol. **49**(6), 2208–2233 (2000)

# Part V
# Mobile Communication, Positioning & Tracking

# A Method of Fast Synchronization Based on PSS in TD-LTE Cell Search

**Zengshan Tian, Yujia Yao, Mu Zhou and Yuhang Jiang**

**Abstract** During the first step of cell searching in Time Division Long Term Evolution (TD-LTE) system, the symbol timing synchronization is the most import part, it can not only enhance the performance of anti-frequency-offset but also reduce the complexity. In this paper, we propose a novel Primary Synchronous Signal (PSS) timing synchronization algorithm. Based on the fast convolution and Overlap-save, the proposed algorithm achieves the joint estimation of the symbol timing synchronization and coarse frequency offset in frequency domain. The theoretical analysis and simulation results prove that the proposed algorithm can not only reduce the complexity, but also improve the performance of anti-frequency-offset for PSS timing synchronization in TD-LTE system.

**Keywords** TD-LTE · Timing synchronization · Fast convolution · Overlap-save · Frequency offset compensation

## 1 Introduction

With the development of internet services, the demand for the ubiquitous mobile networks increases remarkably. To satisfy the market demand, the Time Division Long Term Evolution (TD-LTE) system emerges in response to the trend of the time. The TD-LTE system has the main advantages of the high data rate and low time-delay [1, 2]. As one of the key components of TD-LTE system, the PSS timing synchronization algorithms have been significantly studied in recent years.

Up to now, a large number of studies have focused on the development of timing synchronization in TD-LTE system. The authors in [3] introduced a symbol timing

Z. Tian · Y. Yao (✉) · M. Zhou · Y. Jiang
Chongqing Key Lab of Mobile Communications Technology,
Chongqing University of Posts and Telecommunications, Chongqing
400065, People's Republic of China
e-mail: yaoyujia352@163.com

Z. Tian
e-mail: tianzs@cqupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_49

synchronization algorithm by conducting the sliding correlation in time domain. In [4], the authors reduced the complexity based on the down-sampling and autocorrelation feature of PSS. The synchronization algorithms proposed in [3, 4] exhibited well performance of symbol timing synchronization in the scenario of small frequency offset. To enhance the anti-frequency-offset, the authors in [5, 6] put forward a segment correlation-based timing synchronization algorithm, whereas the anti-frequency-offset performance cannot be guaranteed as the frequency offset is significantly large.

This paper is organized as follows. In Sect. 2, we briefly introduce the characteristics of PSS. After that, the fast synchronization algorithm by integrating the fast convolution with Overlap-save is analyzed in Sect. 3. In Sect. 4, the simulations are conducted to examine the performance of complexity and anti-frequency-offset of the proposed algorithm. Finally, we conclude this paper in Sect. 5.

## 2 Primary Synchronous Sequence

In TD-LTE system, the 3GPP agreement requires that the PSS is generated from the Zadoff–Chu (ZC) sequence, which is featured with the constant amplitude, well autocorrelation, and ideal mutual correlation [7, 8], as described in (1) and (2).

$$d_u(n) = \begin{cases} e^{-j\frac{\pi un(n+1)}{63}} & n = 0, 1, ..., 30 \\ e^{-j\frac{\pi u(n+1)(n+2)}{63}} & n = 31, 32, ..., 61 \end{cases} \qquad u = \begin{cases} 25\,, N_{ID}^{(2)} = 0 \\ 29\,, N_{ID}^{(2)} = 1 \\ 34\,, N_{ID}^{(2)} = 2 \end{cases} \tag{1}$$

where $N_{ID}^{(2)}$ is the sector identification and $u$ is the root sequence.

Figure 1 shows the mapping relations of the time-frequency resources of PSS. The PPS is mapped into $31 \times 2$ subcarriers which are centered with the Direct Current (DC). There are 5 guard intervals on each side of the DC.

The length of radio frame which is composed of two half-frame is 10 ms. Each half-frame consists of 5 sub-frames. We map the PSS into the 3rd Orthogonal Frequency Division Multiplexing (OFDM) symbol in the 1st and 6th sub-frames. Due to the fixed location of PSS and periodicity in time domain, the synchronization within the length of half-frame is required to capture the PSS.

# 3 Synchronization Algorithms

## 3.1 Mutual Correlation-Based Synchronization Algorithm

The algorithm addressed in [3] takes the advantage of well mutual correlation of PSS to conduct the symbol timing synchronization based on the sliding correlation of the received radio frame data and local PSS. The PSS, $p_i(n)$, $i = 0, 1, 2$, is generated by (1). We calculate the mutual correlation between the local time-domain PSS, $p_i(n)$, and received radio frame data, $s(n)$, by

$$r_i(n) = \left| \sum_{k=0}^{N-1} s(n+k)\, p_i^*(k) \right|^2, i = 0, 1, 2 \tag{2}$$

where $p_i^*(k)$ is the conjugation of $p_i(n)$ and $N$ is the length of PSS. We calculate 3 groups of correlation results, $r_i(n)$, by using the conventional symbol timing synchronization algorithm. The location of PSS and value of $N_{ID}^{(2)}$ are determined by the maximum of correlation. However, for the $s(n)$ with long length, the time-domain slide relevant algorithm requires a large amount of storage and time cost and is with low robustness.

## 3.2 Segment Correlation-Based Synchronization Algorithm

Although the PSS has the property of well anti-frequency-offset, the performance of synchronization will be greatly affected as the frequency offset is significantly large. In [5, 6], by using the time-domain slide correlation algorithm, the $p_i(n)$ and $s(n)$ are divided into $L$ equal sections to conduct the segment correlation, as shown in (3).

$$r_i(n) = \sum_{l=0}^{L-1} \left| \sum_{k=0}^{N/L-1} s(n+k+Nl/L) \cdot p_i^*(k+Nl/L) \right|^2, i = 0, 1, 2 \tag{3}$$

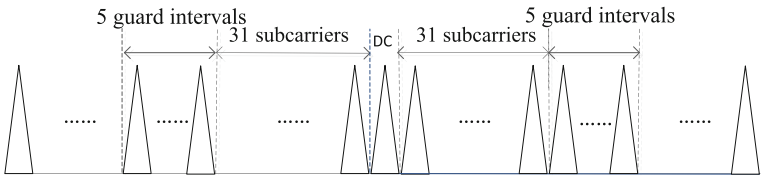



**Fig. 1** Mapping relations of the time-frequency resources of PSS

where $L$ is the number of sections. In general, the larger value of $L$ results in the better anti-frequency-offset performance, whereas the segment correlation-based synchronization algorithm cannot be effective as the frequency offset is significantly large.

## *3.3 The Proposed Fast Synchronization Algorithm*

Based on the pre-frequency offset compensation, we propose the fast synchronization algorithm by integrating the fast convolution and Overlap-save. The fast linear correlation between the long and short sequences is calculated by using the segment FFT transformation to enhance the anti-frequency-offset, as well as reduce the complexity.

### 3.3.1 Pre-frequency Offset Compensation

Different from the existing works which mainly focus on enhancing the anti-frequency-offset by reducing the frequency offset accumulation; we rely on the two-dimensional time-frequency searching algorithm to compensate the frequency offset. The compensated data by using the pre-frequency offset compensation is described as

$$s_v(n) = s(n)\exp\left(2\pi v\Delta f T_s\right), \quad v = 0, \pm 1, ..., \pm n \tag{4}$$

where $\Delta f$ is the interval of the pre-frequency offset compensation and $T_s$ is the sampling interval. When the frequency offset equals to the pre-frequency offset compensation, $v\Delta f$, the maximum correlation peak is reached, and then the coarse frequency offset and location of PSS are obtained.

### 3.3.2 Fast Convolution and Overlap-Save

We calculate the fast correlation between the long and short sequences by integrating the fast convolution and overlap-save. By assuming that the lengths of the received radio frame data, $s(n)$, and local PSS sequence, $p(n)$, are $M$ and $N$, the time-domain sliding cross-correlation is

$$r(n) = \sum_{k=0}^{N-1} s(k+n)p^*(k) \tag{5}$$

We convert (5) into

$$r(n) = \sum_{k=0}^{N-1} s(k)p^*(k-n) = \sum_{k=0}^{N-1} s(k)p^*[-(n-k)] = s(n) * p^*(-n) \tag{6}$$

where the notation "*" indicates the linear convolution operation. By setting $p'(n)$ as the inversion of $p^*(n)$, we have

$$r(n) = s(n) * p'(n) \tag{7}$$

Since the circular convolution outperforms the linear convolution in operating rate, we construct the relationship between the linear convolution and circular convolution based on the zero padding of $s(n)$ and $p'(n)$, as shown in (7).

$$s(n) = \begin{cases} s(n) & n = 0, 1, ..., M-1 \\ 0 & n = M, M+1, ..., L-M \end{cases} \qquad p'(n) = \begin{cases} p'(n) & n = 0, 1, ..., N-1 \\ 0 & n = N, N+1, ..., L-N \end{cases} \tag{8}$$

where $L = 2^{\gamma} \geq M + N - 1$, $\gamma$ is a positive integer. After that, we construct the relationship between the circular convolution and cross-correlation, such that

$$\begin{cases} r(n) = \tilde{r}(n) R_N(n) \\ \tilde{r}(n) = \sum_{k=0}^{N-1} s(k) p'((n-k))_N \end{cases} \tag{9}$$

where $\tilde{r}(n)$ is the result of circular convolution. By setting $R(k) = S(k) P'(k)$, where $S(k)$ and $P'(k)$ stand for the Fourier transform of $s(n)$ and $p'(n)$, respectively, we have

$$r(n) = \text{IDFT}[R(k)] = \left[ \sum_{k=0}^{N-1} s(k) p'((k-n))_N \right] R_N(n) \tag{10}$$

Therefore, the cross-correlation between $s(n)$ and $p(n)$ can be calculated by using the circular convolution. In concrete terms, we first conduct the zero padding of $s(n)$ and $p'(n)$. Second, we calculate $S(k)$ and $P'(k)$ by FFT with the length of $L$. Third, we multiply $S(k)$ and $P'(k)$ in frequency domain to obtain $R(k)$. Finally, we transform $R(k)$ into $r(n)$ by Inverse Discrete Fourier Transform (IDFT), and then select the first $M + N - 1$ numbers in $r(n)$ as the result.

As the length of the received frame data is long, a large number of zeros should be padded in $p'(n)$. In this case, high computation cost is required, and meanwhile the FFT under long length is difficult to be calculated. To solve this problem, the overlap-save is applied to divide $s(n)$ into different short-length segments for the sake of conducting the segmented FFT correlation. First $s(n)$ is divided into $M_1$-length segments, and every segment has points of the $N-1$-previous segment, these two parts compose the $L = M_1 + N - 1$ length sequence of points. and $s_j(n)$ is the $j$-th sequence in $s(n)$, we have

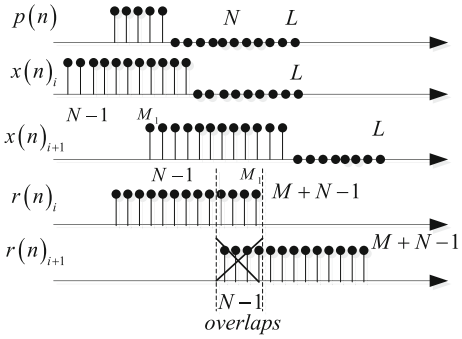



**Fig. 2** Process of overlap-save

$$\begin{cases} s_j(n) = \begin{cases} s(n) \ , \ jM_1 + 1 \le n \le (j+1)M_1 \\ s_{j-1}(n), \ jM_1 + 1 - (N-1) \le n \le (j+1)M_1 - 1 \end{cases} \\ s(n) = \sum_{j=0}^{M/M_1-1} s_j(n) \end{cases} \tag{11}$$

Based on the segmentation, we convert (5) into

$$r(n) = \sum_{k=0}^{N-1} \left[ \sum_{j=0}^{M/M_1-1} s_j(n+k) \right] p^*(k) = \sum_{j=0}^{M/M_1-1} \left[ \sum_{k=0}^{N-1} s_j(n+k) p^*(k) \right] = \sum_{j=0}^{M/M_1-1} r_j(n) \tag{12}$$

In (11), there are $N-1$ overlaps between $r_i(n)$ and $r_{i+1}(n)$ to form $r(n)$, as shown in Fig. 2.

Finally, the fast correlation between the long and short sequences can be calculated by using (9) and (11).

## 4 Simulation Results

In our experiments, the lengths of the local PSS sequence, $N$, received radio frame data, $M$, segment, $M_1$, and FFT, $L$, are set to be 2048, 153600, 2048, and 4096. Considering the anti-frequency-offset capacity of PSS, as well as computation complexity, we set the value $\Delta f$ in the range of $\pm 6$ to $\pm 8$ kHz. To satisfy the $\pm 25$ kHz anti-frequency-offset capacity, the value $v$ is set between $-2$ and 2.

### 4.1 Computation Complexity

The conventional synchronization algorithm which is based on the sliding correlation in time-domain and segmented synchronization algorithm have the same number of complex multiplications and additions, $3NM$. The proposed fast synchronization algorithm in frequency domain divides the received data into $M/M_1$ segments. The correlation calculation for each segment involves 2 FFT and 1 IFFT operations. Since each FFT operation requires $(L/2)\log L$ complex multiplications and $L\log L$ complex additions, the number of complex multiplications and additions involved in the correlation calculation for all the segments equals to $\left(M/M_1\right)\left[(3L/2)\log L + L\right]$ and $\left(M/M_1\right)(3L\log L + N - 1)$. Therefore, after the frequency offset compensation, the number of required complex multiplications and additions in total is $3V\left(M/M_1\right)\left[(3L/2)\log L + L\right]$ and $3V\left(M/M_1\right)(3L\log L + N - 1)$, respectively.

### 4.2 Correlation Calculation

Under the Signal-to-noise Ratio (SNR) = 0 dB and $u = 29$ conditions, Fig. 3 shows the results of timing correlation for three different groups of local PSS sequences and received sequence by using the proposed fast synchronization algorithm with the zero frequency offset. In Fig. 3, p0, p1, and p2 stand for the results of timing correlation between the received sequence and three different groups of local PSS sequences with $u = 25, 29$, and 34, respectively. We observe that there is a significant peak for p1, whereas the peaks for p0 and p2 are difficult to be detected. To be clearer, the amplification of the significant peak for p1 is shown in Fig. 4. Based on this, the timing synchronization of the PSS can be achieved quickly and accurately.

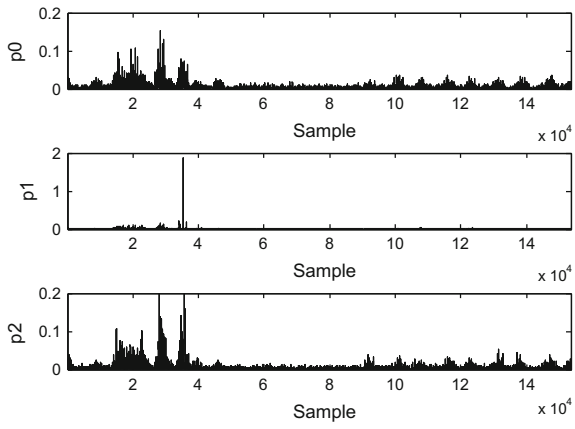



**Fig. 3** Results of timing correlation

### 4.3 Anti-frequency-offset Capacity

Figure 5 compares the anti-frequency-offset capacity under the multi-path channel with (SNR) = 0 dB. As can be seen from this figure, the anti-frequency-offset capacity by using the conventional and segmented synchronization algorithms are 8 kHz and 15 kHz, respectively. As the frequency offset is larger than 18 kHz but smaller than 24 kHz, the correct rates of symbol synchronization by the conventional and segmented algorithms decrease to 0, whereas the one achieved by the proposed algorithm still equals to 100%.

Figure 6 compares the correct rates of symbol synchronization under the 10 kHz frequency offset with different SNRs. From this figure, we can find that when the SNR is higher than −10 dB, the correct rate by the conventional algorithm

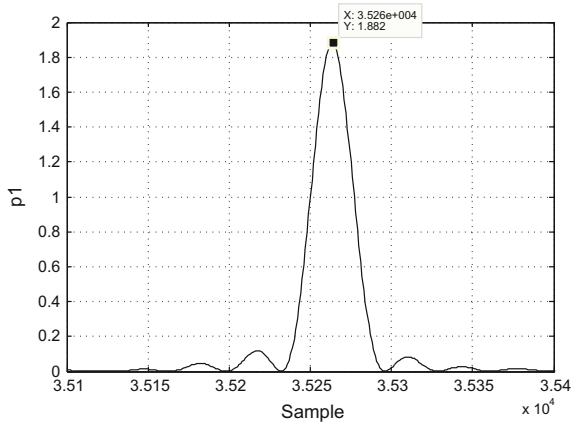



**Fig. 4** Correlation for three different groups of local PSS sequences and received sequence

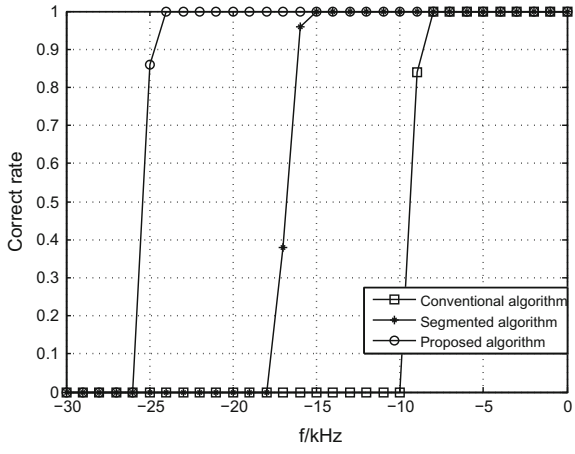



**Fig. 5** Different frequency-offset

approaches zero, whereas the ones achieved by the segmented and proposed algorithms are close to 100%. Furthermore, as the SNR decreases from −15 to −20 dB, the correct rate by the segmented algorithm decreases sharply.

## 5 Conclusion

To enhance the anti-frequency-offset capacity, and solve the high computation complexity problem for the PSS synchronization, a fast synchronization algorithm is proposed based on the fast convolution and overlap-save. The proposed algorithm conducts the time-frequency searching on the received data for the pre-frequency-offset processing, and then achieves the frequency offset compensation to enhance the anti-frequency-offset capacity. The theoretical analysis and simulation results demonstrate that the proposed algorithm is featured with high correct rate of symbol synchronization, well anti-frequency-offset capacity, and low computation complexity.

**Acknowledgements** The authors wish to thank the reviewers for the careful review and valuable suggestions. This work was supported in part by the National Natural Science Foundation of China (61471077), Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), Special Fund of Chongqing Key Laboratory (CSTC), Fundamental and Frontier Research Project of Chongqing (cstc2017jcyjAX0380, cstc2015jcyjBX0065), and University Outstanding Achievement Transformation Project of Chongqing (KJZH17117).

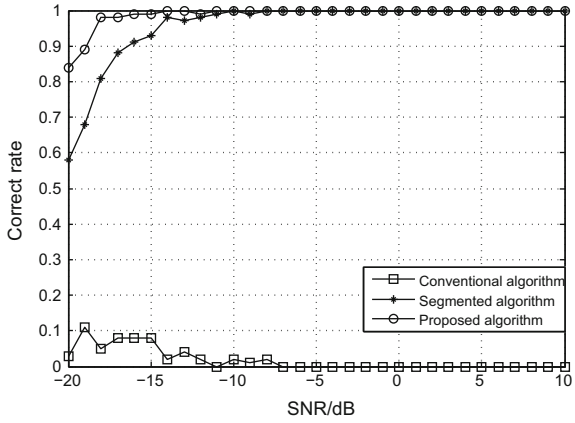



**Fig. 6** Correct rates of symbol synchronization

## References

1. A. Omri, R. Bouallegue, New transmission scheme for MIMO-OFDM system. Int. J. Next-Gener. Netw. **3**(1), 11–19 (2011)
2. Z. Lin, P. Xiao, B. Vucetic, M. Sellathurai, Analysis of receiver algorithms for LTE SC-FDMA based uplink MIMO systems. IEEE Trans. Wirel. Commun. **9**(1), 60–65 (2010)
3. X. Yang, Y. Xiong, G. Jia, W. Fang, X. Zheng, PSS based time synchronization for 3GPP LTE downlink receivers, in *IEEE ICCT* (2011), pp. 930–933
4. Y. Yu, Q. Zhu, A novel time synchronization for 3GPP LTE cell search, in *IEEE International Conference on Wireless Communications, Networking and Mobile Computing* (2013), pp. 328–331
5. Y. Tsai, G. Zhang, Time and frequency synchronization for 3GPP long term evolution systems, in *IEEE Vehicular Technology Conference* (2007) pp. 1727–1731
6. I. Kim, Y. Han, Y.H. Kim, S.C. Bang, Sequence hopping cell search scheme for OFDM cellular systems. IEEE Trans. Wirel. Commun. **7**(5), 1483–1489 (2008)
7. 3GPP TS 36.211(v9.1.0), Evolved Universal Terrestrial Radio Access (E-UTRA), Physical Channels and Modulation (Release 9) (2010)
8. Z. Zhang, J. Liu, K. Long, Low-complexity cell search with fast PSS identification in LTE. IEEE Trans. Vehicular Technology **61**(4), 1719–1729 (2012)

# Vision-Based Positioning Method Based on Landmark Using Multiple Calibration Lines

**Lin Ma, Yingnan Lin, Yang Cui and Yubin Xu**

**Abstract** Currently, with the significant development of the personal terminal's processing rapidly, vision-based indoor positioning has become a hot area of research. Compared with the traditional algorithm, this method is deployed with lower cost. In addition, it provides more robust positioning results and extra visualized services. The positioning results of the traditional method relies heavily on the density of position fingerprinting. Location accuracy could be improved when the fingerprinting is concentrated. However, this causes a greater time delay because of the bigger database and vice versa. This paper proposes a vision-based indoor positioning method based on landmarks to solve above problems. It reduces time complex degree by using SURF-based object image feature matching and improves the location accuracy by adding extra homography matrix and projection constrain information. It leverages several landmarks instead of redundant images in database. Moreover, additional priori information, such as homography matrix constraint and multiple calibration lines' projection relations from landmark to image, could optimize the location results smoothly.

**Keywords** Vision-based indoor positioning · Landmarks · Homography matrix · Multiple calibration lines

This paper is supported by National Natural Science Foundation of China (61571162), Heilongjiang Province Natural Science Foundation (F2016019), Science and Technology Project of Ministry of Public Security Foundation (2015GABJC38).

L. Ma (✉) · Y. Lin · Y. Cui · Y. Xu
Communication Research Center, Harbin Institute of Technology, Harbin, China
e-mail: malin@hit.edu.cn

L. Ma · Y. Lin · Y. Cui · Y. Xu
Key Laboratory of Police Wireless Digital Communication, China Ministry of Public Security, Harbin, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_50

## 1 Introduction

Nowadays, due to fast development of computer vision and the necessity of location-based service in the practical scene, indoor positioning systems have won wide attention, which have opened up a new automatic target detection technology field. In the outdoor environment, the widely diffused Global Navigation Satellite System (GNSS) is one of the most accurate sources of position information when it is available. However, its operation in indoor or obstructed environment is infeasible [1]. At present, the research focus of the indoor positioning system includes WiFi indoor positioning system and Bluetooth indoor positioning system. Among them, WiFi indoor positioning system, with characteristics of extensive deployment of wireless access points and its diffusion, estimates user's location by using mobile terminals detecting received signal strength. However, this technology relies heavily on the number of wireless access points, and the accuracy of positioning drop rapidly when in some certain situation, such as the entrance of a shopping mall. Also, Bluetooth indoor positioning system could achieve the positioning precision of 1 m through Bluetooth signal strength which users get, while it is confined to the high delay when exploring Bluetooth signal [2, 3]. Hence, vision-based indoor positioning system, with deployment of lower cost, smaller delay, higher precision and stronger robust, becomes a new study hotspot. It completes positioning by monitoring objects of indoor scene using mobile terminal and building database without deploying additional equipment of auxiliary positioning. Moreover, vision-based indoor positioning system would provide better visualization services due to the rich scene information in images, which is its incomparable advantage than others.

Generally, vision-based indoor positioning system mainly uses position fingerprint algorithm [4–6]. It builds a database which consists of scene images information acquired by mobile terminals and user's geographical position coordinates information, called Visual Map database. Then it matches images using fast retrieval of scene images algorithm and obtains the user's location as well. In terms of building Visual Map quickly, Walaa Gouda employs SLAM (Simultaneous Localization and Mapping) and SFM (Structure from Motion) to set up a three-dimensional Visual Map in indoor environment [7]. Jason J. Liu and Cody Phillips estimate the camera node location using 2D images obtained by camera sensor in large scale global constraints [8]. In respect of high precision positioning algorithm, Zakhor creates Visual Map with position fingerprint algorithm through multiple cameras and employ SIFT and fast retrieval algorithm of FLANN to estimate user's location [9]. The positioning accuracy comes to 1 m (at 50%), 3 m (at 90%). Rutgers university team also researches the similar positioning system, the difference is that they use color histogram, wavelet decomposition and shape matching to extract image feature points. The room level error probability position precision achieves more than 90% [9]. In [4], geometric constraint is added to the traditional fingerprint positioning algorithm, optimizing positioning accuracy up to 2 m (at 80%).

However, the traditional location fingerprint method based on epipolar geometry has obvious flaw, which is that positioning result depends heavily on intensive degree of fingerprint. When it dense, positioning accuracy improves, and the database storage capacity increases, causing additional retrieval time delay, and vice versa. Therefore, some researchers propose positioning method based on monitoring along-path objects, which are used as landmarks. However, there are few literatures in this field. SURF feature extraction and classification of KNN method is proposed for the recognition of traffic signs, recognition accuracy reaches 97.54% in [10]. From [11], landmark image and its position in the scene are presented in database instead of the traditional method, using SIFT and single calibration line to estimate user's location. Single calibration line method establishes more prior information and geometric constraint in offline phase to ensure that positioning accuracy is not affected when the database capacity reducing in online phase. However, there is a certain problem, which described as the number of a priori information demand is big. Furthermore, the focal length and the tilt angle information are difficult to obtain. Therefore, in this paper we proposed a positioning method based on multiple calibrations line, which provide additional constraints to reduce dependency of priori information, applying in changeful actual scene. Meanwhile, the steering angle computed is helpful for the applications of visual positioning and navigation. The remainder of this paper is organized as follows. In Sect. 2, we will introduce the system model of vision-based indoor positioning system based on landmark. The multiple calibration lines method will be discussed in Sect. 3, which intends to get more constraints to reduce dependency of priori information compared with single calibration line method. Section 4 will provide the implementation and performance analysis. And conclusion will be drawn finally.

## 2 System Model

A. *System Overview*

A typical vision-based indoor positioning system based on landmark involves offline phase and online phase. In offline phase, we build a database, where we selects the landmark image, extracts SURF feature points and corresponding location information, instead of which in the traditional positioning method. Also, set up multiple lines called calibration lines in the landmark image as geometric constraint to position accurately. In online phase, we obtain a database image which is most matched with user's image by SURF feature points. Corresponding calibration lines are described by computing the homography matrix between the two images. Through the parameters of calibration lines and the proposed positioning method, we reduce the delay in ensuring the positioning precision. Figure 1 shows the positioning system based on proposed method.

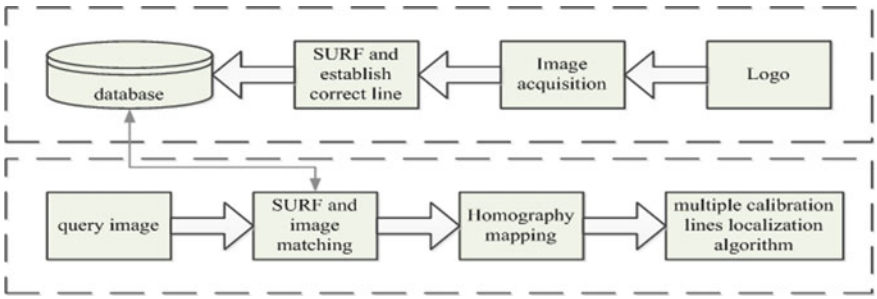



**Fig. 1** The positioning system based on multiple calibration lines

## B. *Database Building in Offline Phase*

To ensure the accuracy of image matching, the selected along-path object, which is regarded as landmark, should contain a sufficient number of SURF feature points. Therefore, the image must have rich content, rich color information and distributed evenly. Furthermore we extract SURF feature points of image $I_l (l = 1, 2, \ldots)$ in the database.

After extracting SURF feature points, we get additional geometric constraint with establishing the calibration lines. Besides, put equations of calibration lines in pixel coordinate system, the coordinates of several points on line and origin in the database as priori information. Single calibration line refers to a line parallel to the ground in the landmark image in world coordinates, which may be naturally occurred or manually tagged. It takes a smaller amount of calculation when selecting a line parallel to the ground as possible. In this paper, we draw a line manually at the top edge using any clue of the image features, obtaining the linear equation artificially. Let pixel coordinates of $n_l$ points on the calibration line describe as the pixel coordinate matrix $[X_n, Y_n]$, of which $X_n$ indicates the coordinate vector of $n$ horizontal ordinate points, $Y_n$ indicates the coordinate vector of $n$ vertical ordinate points. Use the least squares method to solve linear equations, which could be indicated as $y = kx + b$. Assume $\mathbf{W} = [k \quad b]$. Then $\mathbf{W}$ could be derived by

$$\mathbf{W} = \mathbf{Y}^{\mathrm{T}}\mathbf{X}(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1} \tag{1}$$

From the above, we complete t an offline database, with putting parameter matrix $\mathbf{W}$, feature points matrix, the angle of the optical axis of the camera with respect to the RCS as well as the relative height called $Z_0$ between camera and the calibration line as the information of the image $I_l$. When establishing this database, additional prior information allows the system to ensure the accuracy of positioning when reducing time complexity of image retrieval and matching stage.

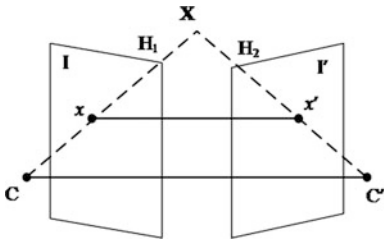



**Fig. 2** Homography relation between two image

C. *Homography Matrix Solving in Online Phase*

The real-world point of a two-dimensional plane projected to the camera CCD imaging sensor is called homography relation. Homography matrix precisely describes point to point corresponding relationship in two images.

In Fig. 2, let $\mathbf{X}(X_0, Y_0, Z_0)$ is a point in the three-dimensional space, $\mathbf{x}(x_0, y_0, 1)$ and $\mathbf{x}(x_0^{'}, y_0^{'}, 1)$ are projected points of $\mathbf{X}$ in imaging planes $I$ and $I^{'}$. Hence, we get the equations of the projection using vector representations as:

$$\begin{aligned} \mathbf{x} &= \mathbf{H}_1 \times \mathbf{X} \\ \mathbf{x}^{'} &= \mathbf{H}_2 \times \mathbf{X} \end{aligned} \tag{2}$$

Therefore, the relation between $\mathbf{x}$ and $\mathbf{x}^{'}$ would be described as follow:

$$\mathbf{x}^{'} = \mathbf{H}_2 \mathbf{H}_1^{-1} \mathbf{x} \tag{3}$$

where $\mathbf{H}_2\mathbf{H}_1^{-1}$ describes the relationship between the corresponding points. Let homography matrix $\mathbf{H} = \mathbf{H}_2\mathbf{H}_1^{-1}$, then (3) would be represented as

$$\mathbf{x}^{'} = \mathbf{H}\mathbf{x} \tag{4}$$

In this paper, we calculate homography matrix of two images using RANSAC, after the introduction of SURF for feature points' extraction and matching.

## 3 Proposed Method

A. *Coordinate Systems of Vision-based Positioning System*

The main idea of using a planar-surfaced object as landmark (such as logo image) for proposed positioning method is to set up a reference coordinate system (RCS) on the calibration lines $L$ and $L^{'}$. Calibration line $L$ is drawn on the landmark, parallel to the environment floors. As is shown in Fig. 3, the left top of the line on the landmark is denoted as the origin of the system $R_0$, with $L$ taken to be the $X$-axis

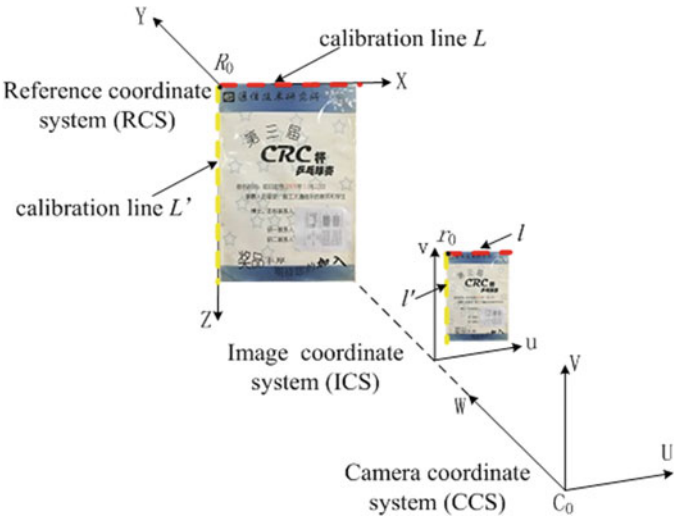



**Fig. 3** Multiple calibration lines positioning model

of RCS, perpendicular to landmark and into it as the *Y*-axis, perpendicular to the ground downward it as the *Z*-axis. Another calibration line $L'$ (e.g., margin-left of the landmark) would be built on the *Z*-axis, passing the origin $R_0$ of RCS. $L'$ is vertical to the ground and the given calibration line *L*.

Figure 4 also reveals a camera coordinate system (CCS), which is set up on the camera with the lens center as the origin and the optical axis as the *W*-axis. The *U*-*V* plane of the CCS is parallel to the imaging plane of the camera, and an image coordinate system (ICS), which is set up in the image taken by the camera.

Let the projection of *L*, $L'$, and $R_0$ in an image be denoted as *l*, $l'$, and $r_0$. In the online phase, the pose of the camera may be represented by coordinates $(X_0, Y_0, Z_0)$

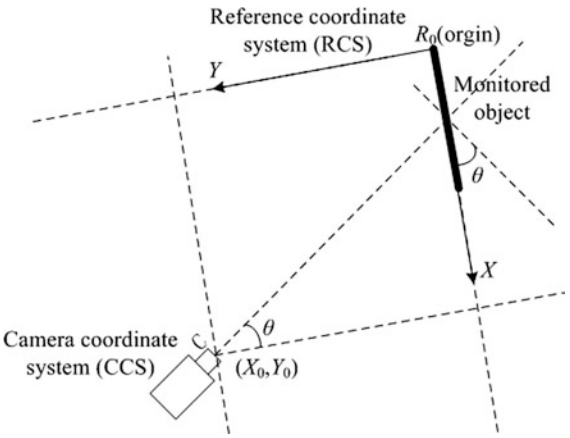



**Fig. 4** Plan view of the relationship between RCS and CCS

in RCS. $Z_0$, which is the vertical distance from the camera to $L$, is assumed to be known. The user takes a landmark image called $I_{on}$, and the best matching image of $I_{on}$ is called $I_{opt}$. After calculating the homography matrix $\mathbf{H}$ between $I_{on}$ and $I_{opt}$, the pixel coordinates of corresponding points on $L$ in $I_{on}$ are described by $[x_i^{'}, y_i^{'}, 1]^{\mathrm{T}} = \mathbf{H} \cdot [x_i, y_i, 1]^{\mathrm{T}} (i = 1, \ldots, n)$, meanwhile the pixel coordinates of corresponding points on $l^{'}$ in $I_{on}$ are described. Afterwards, the pixel coordinate equations are obtained by the least square method. Assume that the equation of $l$ in the ICS is described by $u + bv + c = 0$. The equation of $l^{'}$ in the ICS is described by $u + b_1 v + c_1 = 0$.

Then, coordinates $(X_0, Y_0)$ would be derived in this paper. Figure 3 shows the transformation between the RCS and the CCS, which contains rotation and translation. At first, we translate the origin $R_0$ of the RCS with coordinates $(-X_0, -Y_0, -Z_0)$ to the origin of the CCS. Then, rotate the $X-Y$ plane about the $Z$-axis counterclockwise through the pan angle $\theta$ up to the aim that the $Y-Z$ plane is parallel to the $V-W$ plane of CCS (see Fig. 4). At last, rotate the $Y-Z$ plane about the $X$-axis counterclockwise through the tilt angle $\varphi$ up to the aim that the $X-Y$ plane is parallel to the $U-V$ plane.

B. *Multiple Calibration Lines Positioning Method*

According to the projection in Fig. 3, we propose a positioning method by using multiple calibration lines. Regulate that counterclockwise is the positive direction, RCS accords with left hand coordinate system. As a result, the transformation of the RSC coordinates $(x, y, z)$ of a point in real-world space into the CCS coordinates $(u, v, w)$ may be describes by

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \mathbf{R} \cdot \left( \begin{bmatrix} x \\ y \\ z \end{bmatrix} - \begin{bmatrix} X_0 \\ Y_0 \\ Z_0 \end{bmatrix} \right) = \mathbf{R} \cdot \begin{bmatrix} x \\ y \\ z \end{bmatrix} + \begin{bmatrix} x_0 \\ y_0 \\ z_0 \end{bmatrix} \tag{5}$$

where

$$\mathbf{R} = \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta\cos\varphi & \cos\theta\cos\varphi & -\sin\varphi \\ -\sin\theta\sin\varphi & \cos\theta\sin\varphi & \cos\varphi \end{bmatrix} \tag{6}$$

with

$$\begin{aligned} x_0 &= -X_0\cos\theta - Y_0\sin\theta \\ y_0 &= (X_0\sin\theta - Y_0\cos\theta)\cos\varphi + Z_0\sin\varphi \\ z_0 &= (X_0\sin\theta - Y_0\cos\theta)\sin\varphi - Z_0\cos\varphi \end{aligned} \tag{7}$$

Then, assume that P is a point on the $X$-axis with RCS coordinates $(x, 0, 0)$. Therefore, its CCS coordinates $(u_x, v_x, w_x)$ could be derived to be

$$\begin{bmatrix} u_x \\ v_x \\ w_x \end{bmatrix} = \begin{bmatrix} x\cos\theta + x_0 \\ -x\sin\theta\cos\varphi + y_0 \\ -x\sin\theta\sin\varphi + z_0 \end{bmatrix} \tag{8}$$

After that, assume Q is a point on the $Z$-axis (the calibration line $L'$) with RCS coordinates $(0, 0, z)$. Therefore, its CCS coordinates $(u_y, v_y, w_y)$ could be expressed as:

$$\begin{bmatrix} u_y \\ v_y \\ w_y \end{bmatrix} = \mathbf{R} \cdot \begin{bmatrix} 0 - X_0 \\ 0 - Y_0 \\ z - Z_0 \end{bmatrix} = \begin{bmatrix} x_0 \\ -z\sin\varphi + y_0 \\ z\cos\varphi + z_0 \end{bmatrix} \tag{9}$$

where $\mathbf{R}$ is the same parameter as in (6).

In addition, let $(u_p, v_p)$, $(u_q, v_q)$ be the image coordinates of the projection of P and Q in the ICS. Therefore, according to the camera's optical geometry, we derive that

$$u_p = \frac{u_x \cdot f}{w_x} \quad v_p = \frac{v_x \cdot f}{w_x} \tag{10}$$

$$u_q = \frac{u_y \cdot f}{w_y} \quad v_q = \frac{v_y \cdot f}{w_y} \tag{11}$$

where $f$ is described as the camera focus length. Substituting the three equations of (8) into the two equations of (10) and eliminating the variable x, the equation of $u_p$ and $v_p$ would be derived to be

$$\begin{aligned} u_p - \frac{x_0\sin\theta\sin\varphi + z_0\cos\theta}{y_0\sin\theta\sin\varphi - z_0\sin\theta\cos\varphi} v_p \\ + \frac{f(x_0\sin\theta\cos\varphi + y_0\cos\theta)}{y_0\sin\theta\sin\varphi - z_0\sin\theta\cos\varphi} = 0 \end{aligned} \tag{12}$$

Note that the above Eq. (12) of the calibration line $L$ in the ICS is another representation of $u + bv + c = 0$, which is known from the image. Hence, using the method of undetermined coefficients with (12), we could get the following equalities:

$$\begin{aligned} b &= \frac{Y_0\sin\varphi - Z_c\cos\theta\cos\varphi}{Z_c\sin\theta} \\ c &= -\frac{f(Y_0\cos\varphi + Z_c\cos\theta\sin\varphi)}{Z_c\sin\theta} \end{aligned} \tag{13}$$

where $b$ and $c$ are known from the homography matrix geometric constraint between offline and online phase. $Z_0$ and $\varphi$ are manually measured in advance.

Therefore, we derive the value of $\theta$ by eliminating $Z_0$ and $Y_0$ from the above equalities to get

$$\tan\theta = \frac{f}{fb\ \cos\varphi + c\ \sin\varphi} \tag{14}$$

according to the value of $\theta$ and (13), we could also get the value of $Y_0$ as

$$Y_0 = \frac{Z_0 \cos\theta\ \cos\varphi - bZ_0 \sin\ \theta}{\sin\varphi} \tag{15}$$

By substituting the coordinates $(u_0, v_0)$, which is the projection of the origin of RCS, into (8) and (10), we finally derive $X_0$:

$$X_0 = \frac{u_0(Y_0 \cos\theta\ \cos\varphi + Z_0 \cos\varphi) - fY_0 \sin\ \theta}{u_0 \sin\theta\ \cos\varphi + f\ \cos\ \theta} \tag{16}$$

From above all, when detected single calibration line $L$, we would obtain the RSC coordinates $(X_0, Y_0)$ and the camera pose $\theta$ derived by (14)–(16). Among them, the value of $\theta$, which could not be got in the traditional fingerprint algorithm, is widely used for navigation. Furthermore, by substituting the values of $x_0, y_0, z_0$ got from (7) into (9), the CCS coordinates $(u_y, v_y, w_y)$ in respect of calibration line $L^{'}$ would be derived as follows:

$$\begin{aligned} u_y &= -X_0 \cos\ \theta - Y_0 \sin\ \theta \\ v_y &= -z\ \sin\varphi + (X_0 \sin\ \theta - Y_0 \cos\ \theta)\cos\varphi + Z_0 \sin\varphi \\ w_y &= z\ \cos\varphi + (X_0 \sin\ \theta - Y_0 \cos\ \theta)\sin\varphi - Z_0 \cos\varphi \end{aligned} \tag{17}$$

Next, we get the equation for the projection of $l^{'}$, with computing the coordinates $(u_q, v_q)$ in the ICS by (11) and eliminating variable $z$, as follows:

$$\begin{aligned} u_q + \frac{X_0 \cos\theta\ \cos\varphi + Y_0 \sin\theta\ \cos\varphi}{X_0 \sin\theta - Y_0 \cos\theta} v_q \\ + \frac{f(X_0 \cos\theta\ \sin\varphi + Y_0 \sin\theta\ \sin\varphi)}{X_0 \sin\theta - Y_0 \cos\theta} = 0 \end{aligned} \tag{18}$$

Since the equation of the calibration line $l^{'}$ in the ICS is another representation of $u + b_1 v + c_1 = 0$, which would be known from the image, using the method of undetermined coefficients with (18), we get the following equalities:

$$\begin{aligned} b_1 &= \frac{X_0 \cos\theta\ \cos\varphi + Y_0 \sin\theta\ \cos\varphi}{X_0 \sin\theta - Y_0 \cos\theta} \\ c_1 &= \frac{f(X_0 \cos\theta\ \sin\varphi + Y_0 \sin\theta\ \sin\varphi)}{X_0 \sin\theta - Y_0 \cos\theta} \end{aligned} \tag{19}$$

where $b_1$, $c_1$, $Z_0$, $\varphi$ are known. By putting the two equations in (19) together, we obtain the following equalities:

$$f = \frac{c_1}{b_1 \tan \varphi} \tag{20}$$

By using multiple calibration lines $L$ and $L^{'}$, we conclude that the camera focus length $f$ multiplies the tilt angle $\varphi$ is $c_1/b_1$, which is fixed. Its brief to illustrate the constraint between $f$ and $\varphi$. Therefore we would utilize the proposed method to estimate user's location by knowing $f$ or $\varphi$.

## 4 Implementation and Performance Analysis

The experiment area is our lab, where we set up a logo image as landmark. As is shown in Fig. 5, the first image is the database image and others are obtained in online phase by a smartphone. In real experiment environment, the number of images in inline phase is 41, where it contains the landmark images from different distances and different angles. We regard its tilt angle as $\varphi'$, orientation angle as $\theta'$, and the relative height as $Z_0^{'}$, the focal length of the camera as $f$, coordinates $(X^*, Y^*)$ in RCS for simulation and error analysis.

Figure 6 shows a comparison between single calibration line position method and traditional location fingerprint method based on epipolar geometry, where horizontal axis represents the user's positioning error, vertical axis represents the cumulative distribution function (CDF) of positioning error. Assume that the true coordinates is $(X^*, Y^*)$, Measure that simulation parameters $Z_0 = 82.5$ cm, $\varphi = 98°$, since the result of positioning system is $(X, Y)$, so that the positioning error could be derived by

**Fig. 5** Database image and user input images of simulation scene

$$\Delta d = \left[ (X - X^*)^2 + (Y - Y^*)^2 \right]^{1/2} \tag{21}$$

As could be seen from Fig. 6, the positioning accuracy of single calibration line positioning method is higher than traditional method. Traditional method achieves a $\sigma$ positioning error is 1.2 m, while the single calibration line method achieves a $\sigma$ positioning error of 0.7 m to reach the sub-meter positioning accuracy and the maximum error is less than 1.5 m. The method uses the relative height $Z_0$ and tilt angle $\varphi$ as priori information instead of the fingerprint of the traditional method, realize high precision position.

Tilt angle and the focal length in Fig. 6 are regarded as priori information and known. In practical applicability, the user's arbitrary behaviors, focal length $f$, and tilt angle $\varphi$ often change and could not be read easily. Therefore, multiple calibration lines positioning system creates a constraint between $f$ and $\varphi$, reducing the system of prior information requirements, increasing the actual applicability of the system. Figure 7 gives us the positioning simulation result in two cases of known tilt angle, unknown focal length and unknown tilt angle, known focal length, compared with single calibration line method and traditional method based on epipolar geometry.

Figure 7 shows the CDF probability plot in four methods. As a result, all three methods based on calibration line are better than the traditional location algorithm. However, when position error is less than 50 cm, traditional algorithm performs better, but this is the only part of overall nearly 10% of the weight.

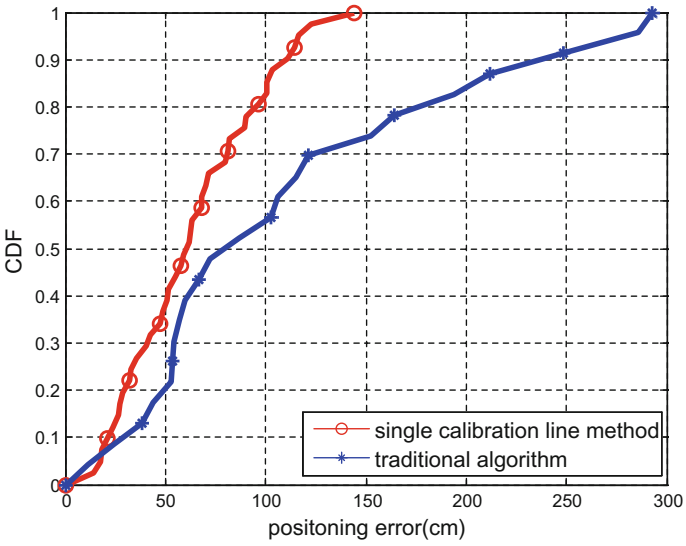



**Fig. 6** CDF of the positioning error for single calibration line method and traditional method based on epipolar geometry

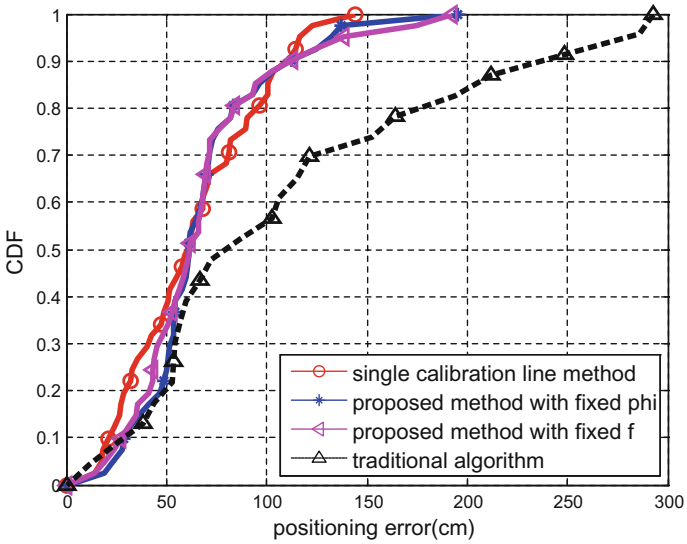



**Fig. 7** CDF of the positioning error for single calibration line method, traditional method and proposed method in two cases

This phenomenon appears because the user happens to stand on the fingerprint or close to the fingerprint point. Nearly 90% of the case, the positioning error of traditional method is bigger, which is because the majority locations of users are not on the fingerprint points. Due to many similar images and existing image noise points in large database, the results of matching images are not accurate, leading to greater positioning error. The proposed method which provides only one similar image with rich SURF descriptors in database does not need to worry about this problem. Since only a stable matching result could be got, and then use the calibration line positioning method to compute user's location, with more robust and higher accuracy.

Figure 8 shows the positioning accuracy comparison of using single calibration line method, multiple calibration lines method when $f$ changing and $\varphi$ known and traditional algorithm. From Sect. 3, we conclude that multiple calibration lines positioning method is not only to reduce the requirement for prior information, but also allows positioning when the focal length $f$ changing. When the image sensor of use's smartphone zooming, the proposed method would achieve that a $\sigma$ corresponds positioning accuracy of 0.8 m, slightly smaller than the single calibration line positioning accuracy corresponding to 0.7 m, higher than the traditional method to 1.2 m. In addition, the maximum positioning error is up to 1.5 m. The positioning result shows better performance than traditional location algorithm, slightly less than single calibration line method. Therefore, in the case of image sensor zooming, multiple calibration lines method further expand the use in the actual scenes ensuring the accuracy of the positioning system.

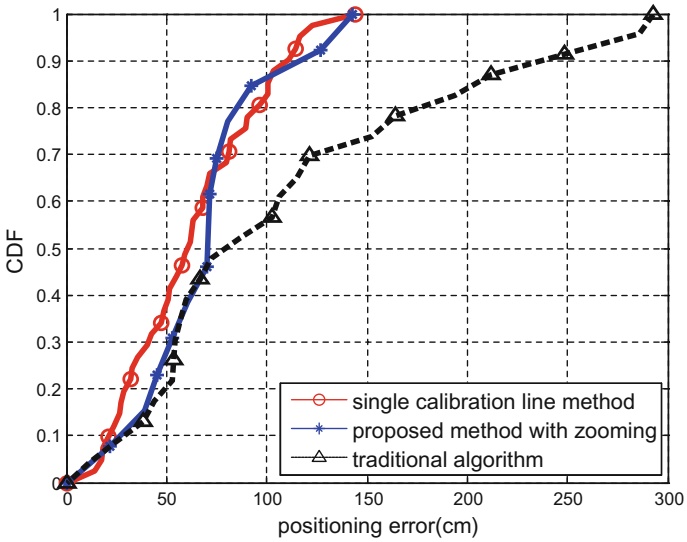



**Fig. 8** Positioning error performance of single line method, traditional method, and proposed method with zooming

## 5 Conclusion

As for the problems of time complexity and positioning accuracy in traditional vision-based positioning system depending on location fingerprint density too much, this paper proposes a method by drawing calibration lines on landmarks to achieve lower time complexity and higher positioning accuracy. In this paper, at first we achieve higher precision position than traditional algorithm based on epipolar geometry using single calibration line positioning method. Then, due to single calibration line positioning method requiring much prior information in offline phase, we propose the method with multiple calibration lines positioning method, which provides additional geometric constraints between the focal length $f$ and the tilt angle $\varphi$ to reduce the demand of the prior information. When only knowing the tilt angle $\varphi$ or the focal length $f$, it could ensure the positioning accuracy. Especially, the proposed method could position accurately in many practical scenes when the image sensor of user terminal zooming, because user's arbitrary behaviors could lead the parameter $f$ change and not be read easily. Furthermore, the value of $\theta$, which could not be got in the traditional fingerprint algorithm, would be obtained for widely using in navigation.

**Acknowledgements** We wish to acknowledge great supports of our team members, who contribute a lot of experiment data for this paper. We also want to express our gratitude to them for their valuable and constructive suggestions.

## References


1. D. Dardari, P. Closas, P.M. Djuric, Indoor tracking: theory, methods, and technologies. IEEE Trans. Vechicular Technol. **64**(4), 1263–1278 (2015)
2. G. Fischer, B. Dietrich, F. Winkler, Bluetooth indoor localization system, in *Proceedings of the 1st Workshop on Positioning, Navigation and Communication* (2004), pp. 147–156
3. S.S. Chawathe, Low-latency indoor localization using bluetooth beacons, in *12th International IEEE Conference on Intelligent Transportation Systems* (2009), pp. 1–7
4. H. Sadeghi, S. Valaee, S. Shirani. A weighted KNN epipolar geometry-based approach for vision-based indoor localization using smartphone cameras, in *IEEE 8th Sensor Array and Multichannel Signal Processing Workshop* (2014), pp. 37–40
5. J.Z. Liang, N. Corso, E. Turner, A. Zakhor, Image based localization in indoor environment, in *2013 Fourth International Conference on Computing for Geo-spatial Research and Application* (2013), pp. 70–75
6. N. Ravi, P. Shankar, A. Frankel, A. Elgammal, L. Iftode, *Indoor Localization using Camera Phones. Mobile Computing Systems and Applications* (2006), pp. 1–7
7. F. Alcantarilla, *Vision Based Localization: From Humanoid Robots To Visually Impaired People*. Electronics (University of Alcala, 2011)
8. R. Tron, R. Vidal, Distributed 3-D localization of camera sensor networks from 2-D image measurements. IEEE Trans. Autom. Control **59**(12), 3325–3340 (2014)
9. N. Corso, A. Zakhor, Indoor localization algorithms for an ambulatory human operated 3D mobile mapping system. Remote Sens. **5**(12), 6611–6646 (2013)
10. Y. Han, K. Virupakshappa, E. Oruklu, Robust traffic sign recognition with feature extraction and k-NN classification methods, in *IEEE International Conference on Electro/Information Technology (EIT)* (2015), pp. 484–488
11. K.C. Chen, W.H. Tsai, Vision-based autonomous vehicle guidance for indoor security patrolling by a SIFT-based vehicle-localization technique. IEEE Trans. Veh. Technol. 3261–3271 (2010)

# Antenna Ports Detection Algorithm in LTE System Using the Repetition of the Reference Signal

**Zeng-Shan Tian, Shan Wei and Mu Zhou**

**Abstract** In the Long Term Evolution (LTE) system, the conventional detection algorithm generally use 1, 2, or 4 transmitting antennas to decode the PBCH. Although the power detection algorithm based on the combination of the PBCH and Secondary Synchronization Sequence (SSS) is featured with low detection complexity, the performance of this algorithm may be much poor under the low Signal Noise Ratio (SNR) condition. In this paper, we perform the repetition of cell reference signal for different antenna ports to detect the number of antenna ports. Simulation results demonstrate that the proposed algorithm can reduce the detection complexity, as well as enhance the anti-noise capacity for antenna ports detection.

**Keywords** LTE · Antenna ports detection · Reference signal

## 1 Introduction

To achieve the integration of mobile communication systems and broadband wireless access techniques, the third Generation Partnership Project (3GPP) launched the Long Term Evolution (LTE) project which is featured with the high communication speed and short transmission delay [1].

In TD-LTE system, the User Equipment (UE) sets up the time-frequency synchronization with the cell during the initial cell searching process by decoding the Physical Broadcast Channel (PBCH) of the cell. The information of the number of antenna ports is contained in the PBCH with the Cyclic Redundancy Check (CRC) mask. The number of antenna ports cannot be determined when the PBCH is not

Z.-S. Tian (✉) · S. Wei (✉) · M. Zhou (✉)
Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, People's Republic of China
e-mail: tianzs@cqupt.edu.cn

S. Wei
e-mail: 1033764762@qq.com

M. Zhou
e-mail: zhoumu@cqupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_51

decoded correctly [2–4]. The multiple antenna port transmission mode supports the transmission with two or four antenna ports. The conventional blind detection algorithm suffers by the high decoding complexity. The authors in [5] put forward a power detection algorithm-based the integration of the PBCH and Secondary Synchronization Sequence (SSS), which is featured with low computation complexity. The performance of this algorithm deteriorates significantly under the low Signal to Noise Ratio (SNR) condition.

To address the problems of the existed works, we propose the antenna ports detection algorithm in LTE system using the repetition of the reference signal which has low computation complexity and well anti-noise capacity.

## 2 Related Works

### 2.1 Generation of Reference Signal

The reference signal [6, 7], $r_{l,n_s}(m)$, is generated as follows.

$$\begin{cases} c(n) = \left(x_1(n+N_C) + x_2(n+N_C)\right) \bmod 2 \\ x_1(n+31) = \left(x_1(n+3) + x_1(n)\right) \bmod 2 \\ x_2(n+31) = \left(x_2(n+3) + x_2(n+2) + x_2(n+1) + x_2(n)\right) \bmod 2 \end{cases} \tag{1}$$

where $N_C = 1600$ and $x_1(n)$ is initialized as $x_1(0) = 1\, and\, x_1(n) = 0, n = 1, 2, ..., 30$ and $x_2(n)$ is denoted by $c_{cnit} = \sum_{n=0}^{30} x_2(n).2^n$ with the value depending on the application of the sequence.

$$c_{init} = 2^{10} * (7 * (n_s + 1) + l + 1) * (2 * N_{ID}^{cell} + 1) + 2 * N_{ID}^{cell} + N_{CP} \tag{2}$$

where $N_{ID}^{cell}$ is the cell ID and $N_{CP}$ =0 or 1.

### 2.2 Mapping of Reference Signal

In LTE system, the reference signal, $r_{l,n_s}(m)$, is mapped into the modulation symbol, $\alpha_{k,l}^{p}$,for the antenna port $p$ and the time slot $n_s$ based on the rule as follows.

$$\alpha_{k,l}^{(p)} = r_{l,n_s}(m') \tag{3}$$

where $k$ is the ID of frequencies and $l$ is defined as the above, which are determined by

$$k = 6s + (v + v_{shift}) \bmod 6 \tag{4}$$

$$l = \begin{cases} 0, N_{symb}^{DL} - 3 & when\ p \in \{0,1\} \\ 1 & when\ p \in \{2,3\} \end{cases} \tag{5}$$

where $s = 0, 1, ..., 2N_{RB}^{DL} - 1$ and $N_{RB}^{DL}$ is the downlink bandwidth configuration obtained from the PBCH. Then, can be calculated by

$$m' = s + N_{RB}^{\max,DL} - N_{RB}^{DL} \tag{6}$$

where $N_{RB}^{\max,DL}$=110,We notate $v$ and $v_{shift}$ as the mapping and offset of reference signal in frequency domain, which are calculated by (7) and (8), respectively.

$$v = \begin{cases} 0 & \text{when } p = 0 \text{ and } l = 0 \\ 3 & \text{when } p = 0 \text{ and } l \neq 0 \\ 3 & \text{when } p = 1 \text{ and } l = 0 \\ 0 & \text{when } p = 1 \text{ and } l \neq 0 \\ 3 & \text{when } p = 2 \\ 3 + 3(n_s mod 2) & \text{when } p = 3 \end{cases} \tag{7}$$

$$v_{shift} = N_{ID}^{cell} \bmod 6 \tag{8}$$

Figure 1 shows the diagram of reference signal transmission for the normal CP by using different antenna ports. We notate $R_p$ as the resource unit used for the reference signal transmission.

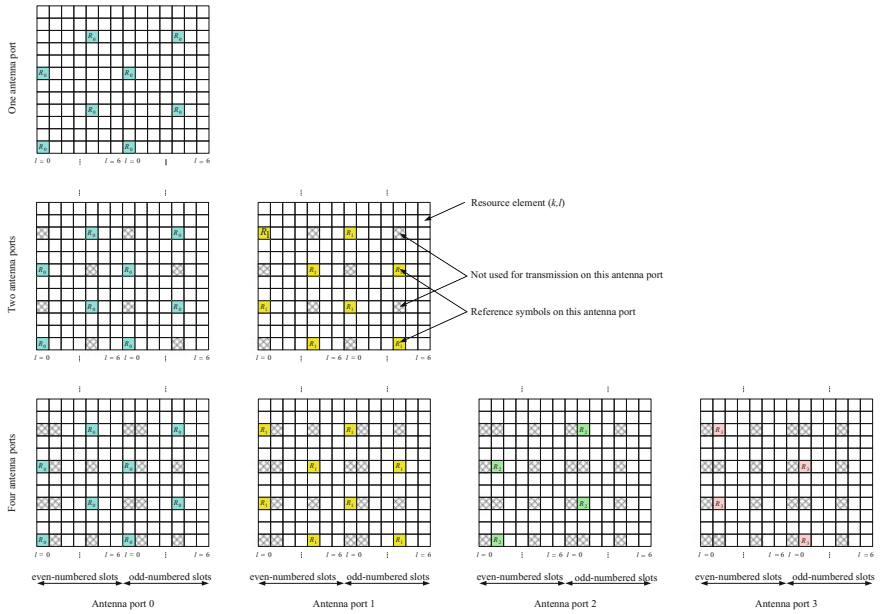



**Fig. 1** Diagram of the reference signal transmission

# 3 Algorithm Description

## 3.1 Blind Detection Algorithm

The conventional algorithms normally rely on configuration of the three available antenna ports to conduct the blind detection. In concrete terms, we first assume that the number of antenna ports is one. Second, after the OFDM decoding and sub carrier mapping, we extract the reference signal from the resource blocks, and then perform the channel estimation, MIMO decoding, demodulation, descrambling, rate matching, and the tail biting convolution. Third, we make CRC check, if it is successful then it means the system is using only one antenna port else assuming two antenna ports are used and then perform the channel estimation repeat all the same processes as one antenna port is used. If CRC is successful then it indicates the system is having two antenna ports else detect the case of four antenna ports. Based on the previous discussion, we can find that the detection of the number of antenna ports requires conducting all the steps involved in PBCH decoding, which significantly increases the computation complexity.

## 3.2 Power Detection Algorithm

The algorithm proposed in [5] is based on the power of reference signal. It extracts the reference signal from the mapping for the 1, 2, and 4 antenna ports, and then calculates the power ratio threshold, S, based on the ratio of the power of SSS in the sub frame with ID 0 and power of protection carrier.

The PBCH is located at the previous four OFDM symbols in the second time slot in the sub frame with ID 0 and occupies the middle 72 sub carriers, while the SSS is located at the last OFDM symbol in the sub frame with ID 0 and occupies the middle 62 sub carriers. The rest 10 sub carriers are selected as the protection interval. Based on this, we name the data occupied by the SSS in frequency domain as the effective signal, while the data transmitted in protection interval as the noise. Then, the ratio of the power of the effective signal and noise equals to

$$S = (((\sum_{k=5}^{66} \alpha(k,l) \cdot \alpha(k,l)^*)/62)/\ ((\sum_{k=0}^{4} \alpha(k,l) \cdot \alpha(k,l)^* + \sum_{k=67}^{71} \alpha(k,l) \cdot \alpha(k,l)^*)/10)) \tag{9}$$

where $\alpha(k,l)$ is the data matrix, which saves the receiving data of SSS.

We set a threshold $T_0 = \rho * S$, where $\rho$ falls into the range of [0.5, 0.8] to calculate the power $P^0_{CRS}$, $P^1_{CRS}$, $P^2_{CRS}$, and $P^3_{CRS}$ of four different reference signals, $r_0(s)$, $r_1(s)$, $r_2(s)$, and $r_3(s)$ by (11). $r_0(s)$, $r_1(s)$, $r_2(s)$, and $r_3(s)$ stand for the reference signal extracted from four different antenna ports.

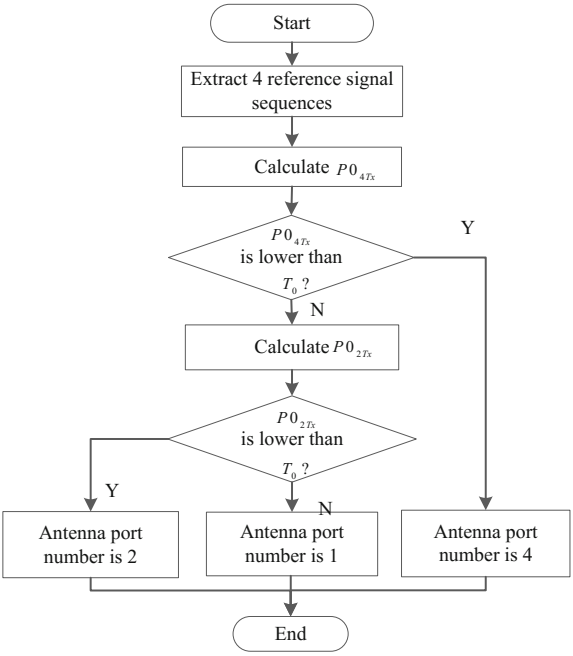



**Fig. 2** Flow chart of the power detection algorithm

$$P = \sum_{n=0}^{11} r(n) \cdot r(n)^* \tag{10}$$

Then, we can calculate the power ratio,$P0_{4Tx}$, by

$$P0_{4Tx} = 2 \cdot P^0_{CRS} / (P^2_{CRS} + P^3_{CRS}) \tag{11}$$

If $P0_{4Tx}$ is lower than $T_0$, the number of antenna ports is set as 4, and otherwise we continue to calculate another power ratio,$P0_{2Tx}$, by

$$P0_{2Tx} = P^0_{CRS} / P^1_{CRS} \tag{12}$$

If $P0_{2Tx}$ is lower than $T_0$, the number of antenna ports is set as 2, and otherwise the number of antenna ports is set as 1. The flow chart is shown in Fig. 2.

## *3.3 The Proposed Algorithm*

To address the problems of the previously discussed blind detection and power detection algorithms, we propose a novel algorithm which can not only reduce the computation complexity, but also enhance the anti-noise capacity.

Based on the mapping of reference signal, we can find that for different antenna ports, only the parameter is different, while the other three parameters, $n_s$, $N_{ID}^{cell}$, and $N_{CP}$ are the same. In addition, the mapping of reference signal is determined by $v$ and $v_{shift}$. Therefore, the reference signal can be extracted based on the mapping of reference signal for each antenna port. Due to the repetition of reference signal for different antenna ports, the number of antenna ports can be estimated by calculating the correlation of reference signal at base station. According to the normalized correlation coefficient equation, we can convert (13) into

$$cor(r,t) = \frac{\left(n\cdot\sum_{i=1}^{n} r_i t_i^* - \sum_{i=1}^{n} r_i\cdot\left(\sum_{i=1}^{n} t_i\right)^*\right)}{\sqrt{n\sum_{i=1}^{n} r_i\cdot r_i^* - \left(\sum_{i=1}^{n} r_i\right)\cdot\left(\sum_{i=1}^{n} r_i\right)^*}\sqrt{n\sum_{i=1}^{n} t_i\cdot t_i^* - \left(\sum_{i=1}^{n} t_i\right)\cdot\left(\sum_{i=1}^{n} t_i\right)^*}} \tag{13}$$

We set a threshold, $T_1$, which falls into the range of [0.6, 0.9]. Then, we record the reference signal from the four different antenna ports, $P_{CRS}^0$, $P_{CRS}^1$, $P_{CRS}^2$, and $P_{CRS}^3$. Then, we calculate the normalized correlation coefficient of $P_{CRS}^2$ and $P_{CRS}^3$ by

$$P1_{4Tx} = cor(r_2(k), r_3(k)) \tag{14}$$

If $P1_{4Tx}$ is higher than $T_1$, the number of antenna ports is set as 4, and otherwise we continue to calculate the correlation coefficient of $P_{CRS}^0$ and $P_{CRS}^1$ by

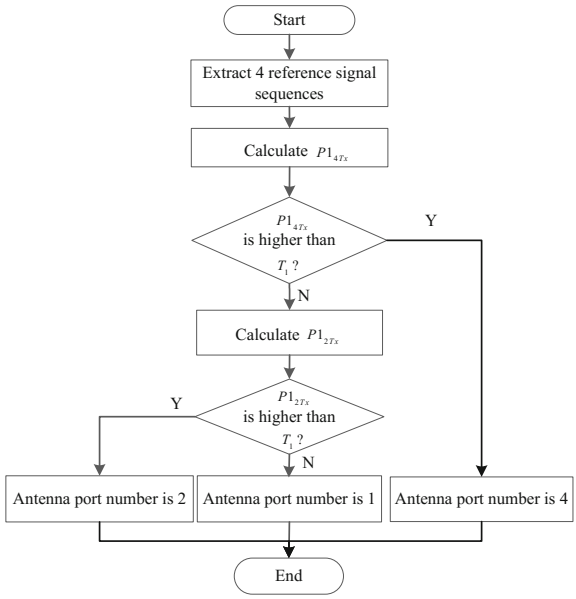



**Fig. 3** Flow chart of the proposed algorithm

$$P1_{2Tx} = cor(r_0(k),\ r_1(k)) \tag{15}$$

If $P1_{2Tx}$ is higher than $T_1$, the number of antenna ports is set as 2, and otherwise the number of antenna ports is set as 1. Figure 3 shows the flow chart.

## 4 Simulation Results

### *4.1 Performance with Different Number of Antenna Ports*

The simulation parameters are as follows. Number of sub carriers = 72, number of OFDM symbols in each sub frame = 14, cell number = 1, number of antenna ports = 1, 2, or 4, and $T_1 = 0.65$. Figure 4 shows the detection success rates with different number of antenna ports.

From Fig. 4, we can find that the proposed algorithm always achieves high detection success rates with the single antenna port under different SNR conditions. However, with two and four antenna ports, the detection success rates significantly decrease when the SNR is lower than 0 dB and 2 dB, respectively. When the SNR is greater than 4 dB, all the detection success rates approach 1.

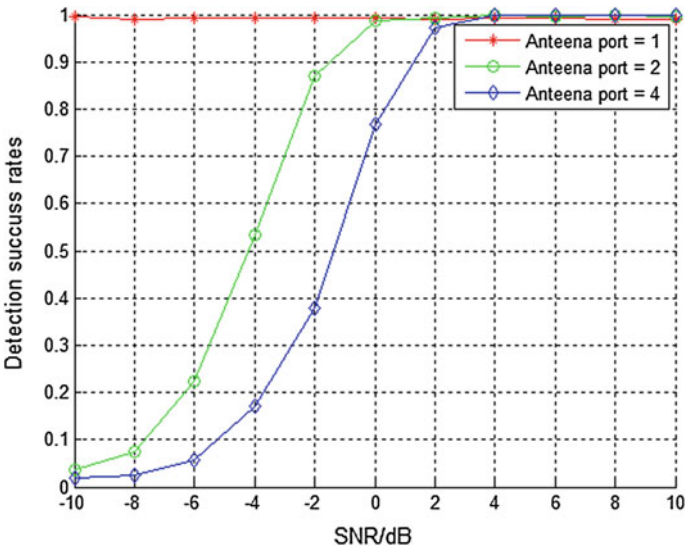



**Fig. 4** Comparison of detection success rates with different number of antenna ports

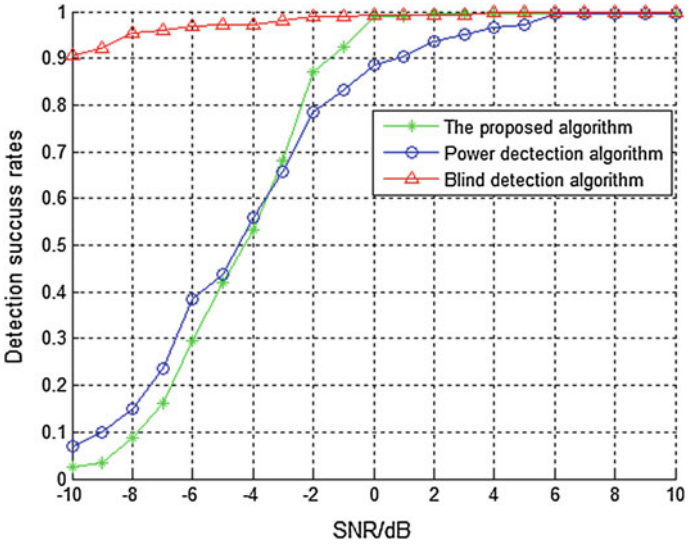



**Fig. 5** Comparison of detection success rates by using different detection algorithms

## *4.2 Performance with Different SNR*

With two transmitting and one receiving antennas, we continue to investigate the impact of Gauss white noise on the performance of antenna ports detection under different SNR conditions. We set the values of $\rho$ and $T_1$ as 0.75 and 0.65, respectively. Figure 5 compares the three algorithms. As can be seen from Fig. 6, all the three algorithms can accurately detect the number of antenna ports when the SNR is larger than 0 dB. The detection success rates by the proposed algorithm are generally higher than the one achieved by the power detection algorithm under the large SNR condition. Considering the larger computational cost of the blind algorithm, if $SNR \geq 0$, we take the proposed method else we take the conventional algorithm. Then the amount of calculation is reduced and the accuracy is also high.

## *4.3 Performance with Different Frequency Offset*

In this section, we focus on the performance of antenna ports detection under different frequency offset conditions. We conduct the simulations with the frequency offset equaling to 50 Hz, 300 Hz, and 1 kHz. The detection success rates achieved by the proposed and existing power detection algorithms are shown in Fig. 6.

From Fig. 6, we can find that as the frequency offset increases, there is no significant variation of detection success rates by the proposed or existing power detection algorithms. In other words, both these two detection algorithms are featured with well anti-frequency offset capacity.

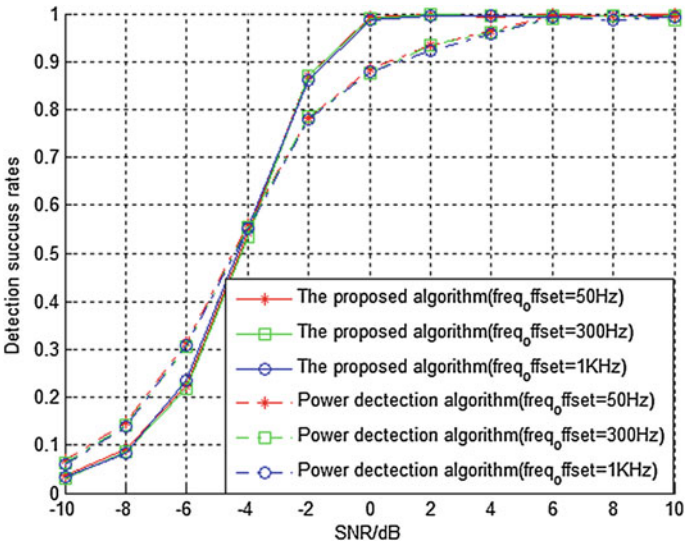



**Fig. 6** Comparison of detection success rates with different frequency offsets

**Table 1** Comparison of time cost by using different detection algorithms

| Algorithms | PBCH decoding (*s*) | Antenna ports detection (*s*) |
|---|---|---|
| Blind detection algorithm | 1.014944 | 1.014944 |
| Power detection algorithm | 0.401926 | 0.000314 |
| Proposed algorithm | 0.401517 | 0.000092 |

### *4.4 Computation Complexity*

From Table 1, we can find that the proposed algorithm requires the least time.

## 5 Conclusion

In this paper, the proposed algorithm has well anti-frequency offset capacity and low computation complexity. The extensive simulation results show that the proposed algorithm generally achieves higher detection success rates compared with the existing power detection algorithm under the large SNR condition.

## References

1. A. Omri, R. Bouallegue, New transmission scheme for MIMO-OFDM system. Int. J. Next-Gener. Netw. **3**(1), 11–19 (2011)
2. Y. Luo, Y. Zhang, Research on TD-LTE downlink packet scheduling. J. Guilin Univ. Electron. Technol. **32**(4), 268–272 (2012)
3. S.B. Lee, S. Choudhury, A. Khoshnevis, et al., Downlink MIMO with frequency-domain packet scheduling for 3GPP LTE, in *Proceeding of IEEE International Conference on Computer Communications* (2009), pp. 1269–1277
4. S. Ahmed, T. Ratnarajah, M. Sellsthurai et al., Iterative receivers for MIMO-OFDM and their convergence behavior. IEEE Trans. Veh. Technol. **8**(1), 1505–1508 (2009)
5. F. Chen, B. Chen, D. Wang, Research on the number of antenna in LTE system, in *IEEE International Conference on Computational and Information Sciences* (2013), pp. 1924–1927
6. 3GPP TS 36.211 v9.1.0 Evolved Universal Terrestrial Radio Access Physical Channels and Modulation (Release 9). 2010-03
7. 3GPP TS 36.212 v9.4.0 Evolved Universal Terrestrial Radio Access Multiplexing and channel coding (Release 9). 2011-09

# TOA Localization in NLOS Environments

**Deliang Liu, Yi Yao and You Zhai**

**Abstract** A new geometric localization approach that is able to locate a stationary tag in non-line-of-sight (NLOS) indoor environment using only the time-of-arrival (TOA) measurements is presented. The novelty of our work resides in converting the NLOS problem into line-of-sight (LOS) problem based on the following three steps, so that the localization algorithm for LOS conditions can be applied in NLOS conditions. First, the geometric TOA model for NLOS is developed, considering the effects of reflection and diffraction. Then, the equivalent anchors can be established based on the geometric features of the real indoor environment. Finally, paths estimation algorithm is derived to estimate the actual signal propagation paths, as well as the position of the tag. Simulation results demonstrate the effectiveness of the proposed method.

**Keywords** Localization · Time-of-arrival · Non-line-of-sight · Equivalent anchor

## 1 Introduction

NLOS and multipath radio propagation, which can cause biased range estimates in Time-of-Arrival (TOA) measurements, will dramatically degrade the performance of conventional TOA localization methods [1]. To mitigate the NLOS impact, a variety of techniques and algorithms have been proposed in the literature. Filtering-based method is presented in [2] which only apply to the tracking of mobile tags. The method in [3] struggles to identify the NLOS paths, from a set of mixed measurements including both LOS and NLOS paths. However, this approach requires LOS measurements involved and cannot be expected to be efficient when signals only propagate through the NLOS paths. In [4], TOA scattering models are adopted to alleviate the NLOS error, whereas the distributions of NLOS error may

D. Liu (✉) · Y. Yao · Y. Zhai
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: liudeliang82@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_52

not be available in the practical environment. A robust scheme in [5] utilizes one-bound scattering and discard multiple-bound scattering to improve the accuracy of NLOS localization. But it needs bidirectional estimation of AOA and TOA, which means more hardware costs.

In this paper, we present a geometric approach for converting the NLOS problem into LOS problem in two-dimensional scenes by using the equivalent anchors. In contrast to [6] which requires both TOA and AOA measurements to reconstruct the LOS path from NLOS measurements, we utilize only TOA measurements and the floor plan information which is easily acquired. Considering the NLOS effects which have not been resolved, we first develop a geometric TOA model for NLOS environments. Then, we establish equivalent anchors and derive the paths estimation algorithm to estimate the position of the tag. Finally, the simulation results confirm the effectiveness of our algorithm.

## 2 Proposed Localization Algorithm

A typical TOA-based localization algorithm for LOS conditions consists of two kinds of parameters: one is distance parameters which contain the measured distances between every anchor and the position-unknown tag; the other is position parameters which include the positions of anchors. In the following analysis, we will take the Least Square (LS) algorithm as an example of localization algorithm for LOS conditions.

To measure the time-of-arrival between an anchor (transmitter) and a tag (receiver), we usually detect the peak of the received signal arriving on the first detectable path, and the time delay$\tau$ of the peak is the measurement of the TOA [7–9]. The estimated distance between the transmitter and the receiver is $\hat{d} = \tau \times c$, in which $c$ is the speed of radio wave propagation.

### *2.1 Geometric TOA Model for NLOS and Establishment of Equivalent Anchors*

In order to estimate the distance between anchor and tag, we measure the length of the first detectable path $\hat{l}_i$, which can be expressed as

$$\hat{l}_i = l_{FPi} + b_m + e \tag{1}$$

where $l_{FPi}$ represents the length of the first detectable path and $i$ is the anchor number, $i = 1, 2, \ldots, M$.

Consider a tag with unknown position at $\boldsymbol{X}_t = [x_t, y_t]^{T,}$ receiving signal from an anchor at a known position, namely, $\boldsymbol{X}_i = [x_i, y_i]^T$, where $^T$ denotes the transpose operator, and $i$ is the anchor number. The TOA measurement between the anchor

and the tag is $\hat{l}_i$. The equivalent anchor is set up at $\mathbf{X}_i^{k_i} = [x_i^{k_i}, y_i^{k_i}]^T$, where $k_i$ is the equivalent anchor number of the $i$th real anchor, and the length of the estimated equivalent path is $\hat{l}_{vi}^{p}$, where $p$ is the path type, and $p = 1, 2, 3$ represents direct path, reflection path and diffraction path respectively.

In indoor environments with obstacles including people, there are mainly four kinds of signal paths: direct paths, reflection paths, diffraction paths, and penetration paths.

### 2.1.1 Direct Paths

Assuming that the signal propagates from the anchor $\boldsymbol{X}_i$ to the tag $\boldsymbol{X}_t$ through a direct path, $l_{FPi}$ in (4) can be expressed as

$$l_{FPi} = \|\mathbf{X}_i - \mathbf{X}_t\|_2 \tag{2}$$

where$\|\cdot\|_2$ represents the 2-norm.

For the consistency of the whole model, we set an equivalent anchor $\boldsymbol{X}_i^1$ at the real anchor position $\boldsymbol{X}_i$, and the length of the estimated equivalent path $\hat{l}_{vi}^{1}$ is equal to the measured path length $\hat{l}_i$

$$\hat{l}_{vi}^{1} = \hat{l}_i \tag{3}$$

### 2.1.2 Penetration Paths

When signal propagates through obstacles, it slows down and gets dispersed. But these effects can be neglected in comparison with other NLOS errors. The only effect of penetration we consider in our work is attenuating the signal strength which may affect the measurement of the first detectable path.

Indoor environments are often populated with people. The experiment results in [12] revealed that no additional ranging error occurs due to the presence of the human body in TOA measurements with the proper choice of the detection threshold. The results also showed that the power of the direct path is strongly attenuated by the presence of the human body. Therefore, we treated human body effects equivalently as penetration effects.

### 2.1.3 Reflection Paths

In our work, we assume that all reflections are specular reflections. Assuming that the signal propagates from the anchor $\boldsymbol{X}_{i,}$ to the tag $\boldsymbol{X}_t$ through a reflection path and the path length we measured is $\hat{l}_i$. As depicted in Fig. 1a, we represent the reflector $A_jB_j$ as a line

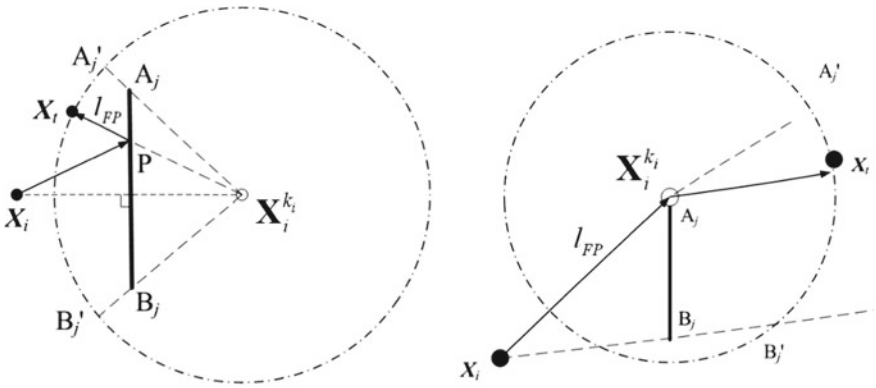



**Fig. 1** TOA model of reflection and diffraction

$$\begin{cases} y = m_j x + c & m_j \neq \infty \\ x = c_j & m_j = \infty \end{cases} \tag{4}$$

where $j$ represents the number of obstacles.

The equivalent anchor $\mathbf{X}_i^{k_i}$ and the real anchor $\boldsymbol{X}_i$ are a pair of symmetry points of line $A_jB_j$, so the position of the equivalent anchor can be expressed as

$$\mathbf{X}_i^{k_i} = \begin{cases} \begin{bmatrix} m_j x_i \\ y_i/m_j - 2c_j \end{bmatrix} & m_j \neq \infty \\ \begin{bmatrix} 2c_j - x_i \\ y_i \end{bmatrix} & m_j = \infty \end{cases} \tag{5}$$

It looks like that the signal transmits from the equivalent anchor $\mathbf{X}_i^{k_i}$ to the tag $\boldsymbol{X}_t$ through a direct path. Therefore, $l_{FPi}$ in (4) can be given as

$$l_{FPi} = \left\| \mathbf{X}_i^{k_i} - \mathbf{X}_t \right\|_2 \tag{6}$$

The equivalent path length can be estimated as

$$\hat{l}_{vi}^2 = \hat{l}_i \tag{7}$$

### 2.1.4 Diffraction Paths

Assuming that the signal propagates from the anchor $\boldsymbol{X}_i$, to the tag $\boldsymbol{X}_t$ through a diffraction path and the path length we measured is $\hat{l}_i$. As depicted in Fig. 1b, the

diffraction happens at point $A_j$ and its position is considered as a known vector. So, $l_{FPi}$ in (4) can be given as

$$l_{FPi} = \left\|\mathbf{X}_i^{k_i} - \mathbf{X}_t\right\|_2 + \left\|\mathbf{X}_i^{k_i} - \mathbf{X}_i\right\|_2 \tag{8}$$

We can get $\left\|\mathbf{X}_i^{k_i} - \mathbf{X}_i\right\|_2$ in (11) from the floor plan information.

We establish the equivalent anchor $\mathbf{X}_i^{k_i}$ at point $A_j$, so that it looks like the signal transmits from the equivalent anchor $\mathbf{X}_i^{k_i}$ to the tag $\boldsymbol{X}_t$ through a direct path and the equivalent path length can be estimated as

$$\hat{l}_{vi}^3 = \hat{l}_i - \left\|\mathbf{X}_i^{k_i} - \mathbf{X}_i\right\|_2 \tag{9}$$

In this work, we assume that the NLOS multipath signals undergo one-bound scattering (reflection or diffraction), because multi-bound scattering can be ignored as they suffer from severe fading and may be too weak to be observed [5].

In NLOS indoor environment, the floor plan can be easily obtained. Then we can develop several equivalent anchors for a real anchor based on the above analysis.

## 2.2 Paths Estimation Algorithm

For one anchor, there may be several equivalent anchors according to the amount of possible signal paths. But when we acquire a TOA measurement from one anchor, only one of the possible paths is the first detectable path. If we can estimate it, the corresponding equivalent anchors can be selected. In order to estimate all of the first detectable paths for all the real anchors, we need to utilize path estimation algorithm. The sequential process of the algorithm is depicted below:

(1) Build a group of equivalent anchors for every real anchor according to the floor plan of the NLOS environment.

We represent obstacles (reflectors) as lines $A_jB_j$, $j = 1, 2, \ldots, J$, where $J$ is the amount of obstacles. The diffraction points are the end points of obstacles $A_j$, $B_j$, at the positions $\mathbf{X}_{Aj}$, $\mathbf{X}_{Bj}$, if they are not the intersection points of lines, the amount of which is $N$.

$$\mathbf{X}_i^{k_i} = \begin{cases} \mathbf{X}_i & \text{if } k_i \text{ is a direct path} \\ \mathbf{X}_{A_j} \text{ or } \mathbf{X}_{B_j} & \text{if } k_i \text{ is a diffraction path} \\ \begin{bmatrix} m_j x_i \\ y_i/m_j - 2c_j \end{bmatrix} (m_j \neq \infty) & \text{if } k_i \text{ is a reflection path} \\ \begin{bmatrix} 2c_j - x_i \\ y_i \end{bmatrix} (m_j = \infty) & \end{cases} \tag{10}$$

$$k_i = 1 + J - N$$

(2) Select one from every group of equivalent anchors $\mathbf{X}_1^{k_1} \sim \mathbf{X}_M^{k_M}$ as the position parameters and use TOA measurements $\hat{l}_1 \sim \hat{l}_M$ as the distance parameters. Then, utilize LS algorithm to estimate a tag position $\hat{\mathbf{X}}_t$

$$\hat{\mathbf{X}}_t = (A^T A)^{-1} A^T b \tag{11}$$

where

$$A = 2\begin{bmatrix} x_1^{k_1} - x_2^{k_2} & y_1^{k_1} - y_2^{k_2} \\ \vdots & \vdots \\ x_1^{k_1} - x_M^{k_M} & y_1^{k_1} - y_M^{k_M} \end{bmatrix}$$

$$b = \begin{bmatrix} (x_1^{k_1})^2 - (x_2^{k_2})^2 + (y_1^{k_1})^2 - (y_2^{k_2})^2 + (\hat{l}_{v1}^{p})^2 - (\hat{l}_{v2}^{p})^2 \\ \vdots \\ (x_1^{k_1})^2 - (x_M^{k_M})^2 + (y_1^{k_1})^2 - (y_M^{k_M})^2 + (\hat{l}_{v1}^{p})^2 - (\hat{l}_{vM}^{p})^2 \end{bmatrix}$$

(3) Decide if the estimated position is the possible tag position by the following rules:

$$\left\|\mathbf{X}_a^{k_a} - \mathbf{X}_b^{k_b}\right\|_2 < \hat{l}_{va}^{p} + \hat{l}_{vb}^{p} + q_1 \quad \forall a, b \in [1, M] \quad a \neq b; \text{if } p = 1 \tag{12}$$

$$\begin{aligned} &\angle A_j \mathbf{X}_i^{k_i} \hat{\mathbf{X}}_t < \angle A_j \mathbf{X}_i^{k_i} B_j + q_2 \cup \angle B_j \mathbf{X}_i^{k_i} \hat{\mathbf{X}}_t < \angle A_j \mathbf{X}_i^{k_i} B_j + q_2 \\ &\cup \left\|\mathbf{X}_i^{k_i} - \mathbf{P}\right\|_2 < \hat{l}_{vi}^{p} + q_3 \qquad \text{if } p = 2 \end{aligned} \tag{13}$$

$$\begin{aligned} &\angle A_j \mathbf{X}_i \hat{\mathbf{X}}_t < \angle A_j \mathbf{X}_i B_j + q_4 \cup \angle B_j \mathbf{X}_i \hat{\mathbf{X}}_t < \angle A_j \mathbf{X}_i B_j + q_4 \\ &\cup \left\|\mathbf{X}_i^{k_i} - \mathbf{X}_i\right\|_2 < \hat{l}_{vi}^{p} + q_5 \quad \text{if } p = 3 \end{aligned} \tag{14}$$

First, as inequality (12) suggests, due to the existence of $b_m$ and $e$, $\hat{l}_i$ are mostly larger than $l_{FPi}$. Therefore all the localization circles should not be far away from one another.

Second, as inequality (13) shows, if the reflection equivalent anchor is selected (i.e. $p = 2$) as depicted in Fig. 1a, the possible tag position $\hat{\mathbf{X}}_t$ must be inside the LOS boundaries $X_i^2 \mathrm{A}_j$ and $X_i^2 \mathrm{B}_j'$. Besides, $\hat{\mathbf{X}}_t$ and $\mathbf{X}_i^{k_i}$ should on the different side of the line $\mathrm{A}_j\mathrm{B}_j$. Point P in (13) is the intersection point of the line $\mathrm{A}_j\mathrm{B}_j$ and the equivalent path. Though diffraction wave is distributed in all directions, direct path signal will play a main role when LOS path exists. So, as inequality (14) shows, if the diffraction equivalent anchor is selected (i.e. $p = 3$) as depicted in Fig. 1b, the possible tag position $\hat{\mathbf{X}}_t$ must be also inside the LOS boundaries $X_i A_j'$ and $X_i B_j'$.

Besides, $\hat{\mathbf{X}}_t$ and $\mathbf{X}_i$ should on the different side of the line $A_jB_j$. We also leave some room for every inequality (represented by $q_1 \sim q_5$) because the estimated tag position has error.

Finally, the possible tag position must be inside the boundaries of the whole indoor environment.

(4) Repeat steps (2)–(3) until all the equivalent anchors have been considered. If the TOA measurements are accurate, there should be only one possible tag position totally, and all the first detectable paths can be estimated exactly. But the TOA errors do exist, so there may be several possible tag positions left after the above filtering steps and the geometric center of them is considered as the final estimation result of the tag position.

## 2.3 Simulation

To evaluate the performance of the proposed method, simulations have been conducted. As shown in Fig. 2, we consider that the simulation environment is a rectangular room with a size of 6 m × 5 m. AB, BC, CD and DA are the 4 walls of the room. There is an obstacle in the room represented as line EF. The 4 anchors are deployed at point A, B, C, and D respectively, and 30 test points are set up at intervals of 1 m as tags positions. For every tag, we presuppose 4 paths as first detectable paths between every anchor and the tag. Under the above environment, there are at least 2 of the presupposed 4 paths in LOS conditions. In order to show the validity of our method, we add one or two people for every test so that at least 3 presupposed paths are NLOS paths. Take test point 9 for example, as depicted in Fig. 2, we presuppose that signal of the first detectable path propagates from anchor A to the tag directly; though the LOS path exists between anchor B and the tag, the signal of the first detectable path which propagates from anchor B to the tag experiences a reflection off the obstacle EF because a human is preset on the LOS

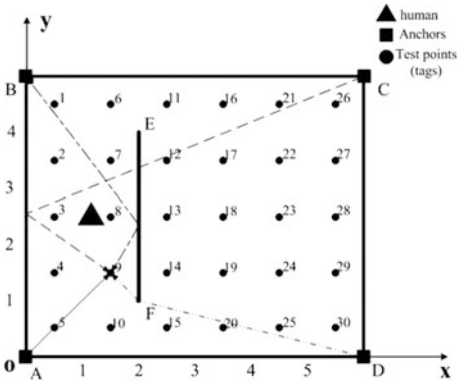
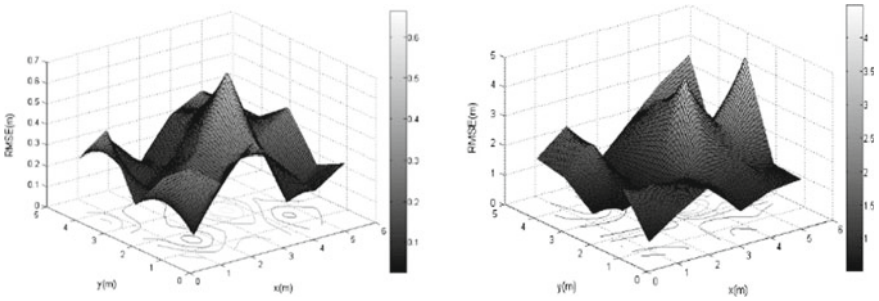



**Fig. 2** Simulation environment




(a) RMSE of proposed method (b) RMSE of conventional LS method

**Fig. 3** Comparison of 30 test points RMSE of proposed method and conventional LS method for $b_m = 0.5$, $\sigma = 0.3$

path and weakens the direct path energy; the signal of the first detectable path which propagates from anchor C to the tag experiences a penetration in the obstacle EF and a reflection off the wall AB; the signal of the first detectable path which propagates from anchor D to the tag experiences a diffraction at the point F. So, there are 3 NLOS paths and 1 direct path when the tag is set on the test point 9.

Figure 3 shows the surfaces representing the root mean square errors (RMSE) of the 30 test points, which are obtained by running simulations of our algorithm and conventional LS algorithm respectively at every test point for $b_m = 0.5$, $\sigma = 0.3$. We can see that the RMSE of our method are below 0.6 m, which has a better performance than the conventional method. The basic localization algorithm of the two methods is the same LS algorithm. But when we derive the equivalent anchors, NLOS problem can be converted into LOS problem, and the LS algorithm can be applied in NLOS conditions with high accuracy.

## 3 Conclusion

In this paper, we have proposed a geometric approach for converting the NLOS problem into LOS problem in two-dimensional scenes, so that the conventional localization algorithm for LOS conditions can be applied to NLOS conditions with high accuracy. We utilized only TOA measurements and the floor plan of the indoor environment. We established TOA model and equivalent anchors considering reflection, diffraction, penetration and human body effect in NLOS environments. The paths estimation algorithm has been derived to search the actual signal propagation paths. Simulation results confirmed the robustness of our method in terms of RMSE.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China under Grant 61601494, 61501493.

## References


1. Z. Li, K. Liu, Y. Zhao et al., MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)
2. K. Yu, E. Dutkiewicz, Improved kalman filtering algorithms for mobile tracking in NLOS scenarios. In *Proc. IEEE WCNC*, Apr. 2012; 2390–2394
3. K. Yu, Y.J. Guo, Non-line-of-sight detection based on TOA and signal strength, in *Proceedings of the 19th IEEE International Symposium* PIMRC, Sept 2008, pp. 1–5
4. S. Al-Jazzar, J. Caffery, You Heung-Ryeol, Scattering-model-based methods for TOA location in NLOS environments. IEEE Trans. Veh. Technol. **56**(2), 583–593 (2007)
5. C.K. Seow, S.Y. Tan, Non-line-of-sight localization in multipath environments. IEEE Trans. Mob. Comput. **7**(5), 647–660 (2008)
6. M.A. Soleimani, A.R. Sharafat, A novel geometric approach for mitigating NLOS effects in wireless location estimation, in *Proceedings of the 19th Iranian Conf. Electrical Engineering*, May 2011, pp. 1–6
7. Z. Li, Y. Chen, H. Shi et al., NDFB-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. **48,** 1–5 (2016)
8. X. Liu, Z. Li, P. Yang et al., Information-centric mobile ad hoc networks and content routing: a survey. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.04.005
9. Z. Li, L. Song, H. Shi et al., Approaching the capacity of k-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.02.009
10. B. Alavi, K. Pahlavan, Modeling of the TOA based distance measurement error using UWB indoor radio measurements. IEEE Commun. Lett. **10**(4), 275–277 (2006)
11. N.A. Alsindi, B. Alavi, K. Pahlavan, Measurement and modeling of ultrawideband TOA-based ranging in indoor multipath environments. IEEE Trans. Veh. Technol. **58**(3), 1046–1058 (2009)
12. Y. Kilic, A.J. Ali, A. Meijerink, M.J. Bentum, W. Scanlon, The effect of human-body shadowing on indoor UWB TOA-based ranging systems, in *9th Workshop on Positioning Navigation and Communication (WPNC)*, Mar 2012, pp. 126–130

# Complex Networks Analysis Based on IP Data of Mobile Communication System

**Bilun Wu, Zhuo Sun, Qingyi Quan and Ruixue Zhang**

**Abstract** In order to analyze the IP label data from mobile communication gateway, this paper creates the network based on IP label data. By using the degree analysis method, the paper analyses the data and gets IP network degree distribution condition, fits deviation analysis and extracts the import node to obtain the size of the IP network and the network boundary. Then, the Eigenvector centrality analysis method is used to calculate the core of interconnected nodes. The IP label data network is partitioned and three pure data set is achieved according to the important nodes in combination with label propagation algorithm. Each divided data set can be used for further data analysis, including analyzing the user behaviour through a purer user behaviour data extracted from the complex network data and stripping the invalid data from the data set.

## 1 Introduction

In mobile communication network, mobile communication gateway is in charge of providing routing in the packet data service of mobile communication network [1]. The correlation analysis of IP label data is the basis of the research on equipment of mobile communication network and objects of public Internet. With some or all of the properties of self-organization, self-similar, attractor, small world in scale-free network known as complex networks [2].

B. Wu (✉) · Z. Sun (✉) · Q. Quan (✉) · R. Zhang (✉)
Key Laboratory of Universal Wireless Communications,
Beijing University of Posts and Telecommunications, Beijing, China
e-mail: eternity@bupt.edu.cn

Z. Sun
e-mail: zhuosun@bupt.edu.cn

Q. Quan
e-mail: qyquan@bupt.edu.cn

R. Zhang
e-mail: zhangruixue@bupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_53

In this paper, the second chapter models the IP data of the mobile communication network. The third chapter analyzes the complex networks nature of mobile communication network IP data, and studies the network topology in detail. This paper analyzes the topology in aspect of the degree distribution, average path length, diameter and eigenvector centrality of mobile communication network. The fourth chapter of this paper filters out the irrelevant data according to the internal relationships of network data, and uses the label propagation algorithm to realize the division of the network. The paper gives a summary in the fifth chapter. The method provided by this paper can be used for separation of IP label data of mobile communication network gateway.

## 2 The Data Pre-processing and Network Building

The mobile communication network IP label data is collected from operators mobile communication network gateway and storage in IP packet form. This paper chooses to use analysis method of complex network to build and analyse the IP network and intercepts the data of three groups of different sizes in accordance with the original data acquisition time. To study the IP label data, the paper build a data network first. This paper uses the concept of graph theory to define the elements in the complex network. And then the paper uses IP label data and communication condition between the data to structure the complex network, model the IP data and use IP address as the network nodes. Two nodes have a common edge if and only if there is data communication between them.

## 3 Analysis of Network Topology

### *3.1 Degree Analysis*

The degree of a node is defined as the number of edges connected to that node [3]. The degree distribution in a completely random network is often approximated by the Poisson distribution, and the actual network has another description of the power law distribution. In this paper, the data set is fitted by the two forms namely power law distribution and Poisson distribution. The fitting formula of power law distribution is (1), and the fitting formula of Poisson distribution is (2).

$$P(k) = ak^b \tag{1}$$

$$P(k) = ae^{bk} \tag{2}$$

The paper analyses the degree and fits the data of the three data sets of the mobile communication network which includes 5000 sides, 50000 sides and 500000 sides network.Tables 1 and 2 are the results of the power law distribution and the exponential distribution.

**Table 1** Power law function fitting effect

| Set | Number of connections | Number of nodes | Standard error | The sum of squared residuals | $P(k) = ae^{bk}$ | |
|---|---|---|---|---|---|---|
| | | | | | a | b |
| 1 | 5,000 | 2232 | 0.0005708 | 1 | 6.487 | −3.071 |
| 2 | 50,000 | 15996 | 0.0001766 | 1 | 82460 | −4.863 |
| 3 | 500,000 | 60688 | 0.00005827 | 1 | 3.13 * 1011 | −5.912 |

**Table 2** The exponential function fitting effect

| Set | Number of connections | Number of nodes | Standard error | Adjusted sum of squared residuals | $P(k) = ae^{bk}$ | |
|---|---|---|---|---|---|---|
| | | | | | a | b |
| 1 | 5k | 2232 | 0.001978 | 0.9989 | 6.825 | −1.08 |
| 2 | 50k | 15996 | 0.0002071 | 1 | 16.22 | −0.272 |
| 3 | 500k | 60688 | 0.000062 | 1 | 27.14 | −0.038 |

From Table 1, power law function fitting effect can be seen on three groups of different size of network, the data of standard errors of the power law function fitting is respectively 0.0005708, 0.0001766 0.00005827 and it decreases along with the increase of the amount of data; adjusted sum of squares of deviations were 0.9999 and 1, it increases along with the increase of data and after retained four decimal places it can be thought that the value is approximately to 1.

From Table 2 exponential function fitting effect, the result of exponential function fitting to the three set of network data are 0.001978, 0.000207160688; result of adjusted sum of squares of deviations are 0.9989 and 1 respectively.

Combining the results of two table, the degree distribution of mobile communication network is closer to the power law distribution, and the distribution form of the power law distribution network satisfies in the probability distribution of the scale-free condition. Figure 1 is the fitting function of the 500,000 edges.

From connection degree distribution of the double log distribution figure of Fig. 1, we can see that there is a small number of node degree more than 400 points, which have more number than the distribution curve of fitting function. There are a few relatively high degree nodes called the network's "hubs" in the complex network [4]. This paper adds the "hub" tag to the node with more than 400 degrees to facilitate the analysis below. Table 3 are the nodes with more than 1500 degrees.

From Table 3, except some network hubs, network contains a special point connected with most of the other point, namely IP x.x.0.200 which is connect nearly half of the node refer to 34830 nodes, this point in IP data network has special significance. The point on the connected with x.x.0.200 is counted and analyzed in this paper. By information inquiry the paper found x.x.0.200 corresponds to the node for the WAP proxy server of operators so it can be identified x.x.0.200 communication IP nodes represent mobile devices of mobile communication network.

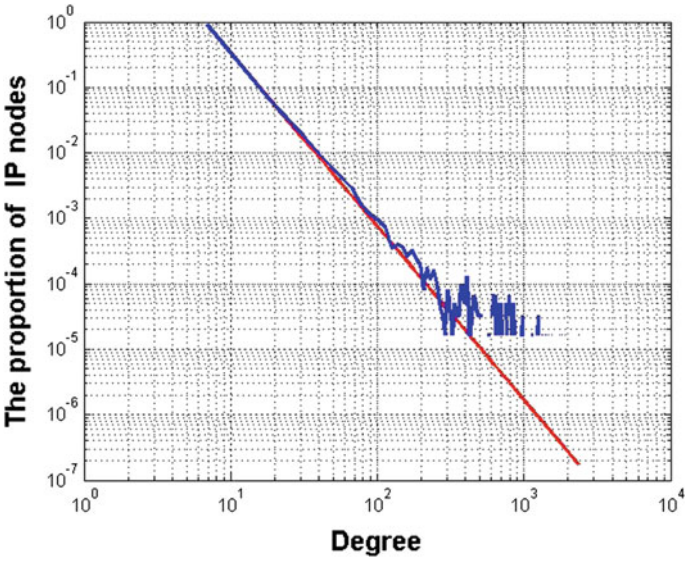



**Fig. 1** 500,000 connections degree distribution (*blue line*) and the fitting curve (*red line*)

**Table 3** IP nodes with more than 2000 degrees of 500,000 connection data

| Serial number | IP of nodes | Degree |
|---|---|---|
| 60683 | x.x.220.150 | 2360 |
| 60684 | x.x.223.27 | 2700 |
| 60685 | x.x.66.12 | 5212 |
| 60686 | x.x.66.14 | 5606 |
| 60687 | x.x.66.10 | 5940 |
| 60688 | x.x.0.200 | 34830 |

## *3.2 The Statistical Property of Network*

In network data, the average of degree $k_i$ of all the nodes $v_i$ in the network is called the average degree [5], remember to $< k >$, namely:

$$< k >= \frac{1}{N} \sum_{i=1}^{N} k_i \tag{3}$$

The maximum value of the distance between any two nodes of the network calls diameter for the network [6], remember to D, namely:

$$D = \max_{1 \le i \le j \le N} d_{ij} \tag{4}$$

Average path length L of network is defined as the average distance between any two nodes [7], namely:

**Table 4** Network statistical properties

| Set number | Number of connections | Number of nodes | The average degree | The average path length | The network diameter |
|---|---|---|---|---|---|
| 1 | 5,000 | 2232 | 1.516 | 2.64423 | 12 |
| 2 | 50,000 | 15996 | 2.174 | 2.63940 | 15 |
| 3 | 500,000 | 60688 | 3.516 | 5.83743 | 22 |

$$L = \frac{1}{C_N^2} \sum_{1 \le i \le j \le N} d_{ij} \tag{5}$$

This section calculates statistical properties of three groups network data with different number of connections and gets the results in Table 4.

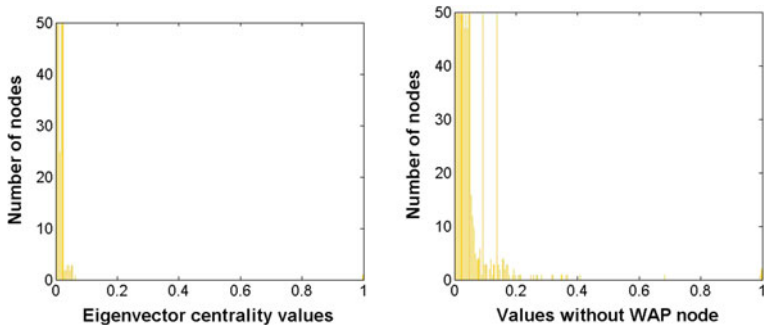
The data in Table 4 represent the mobile communication network IP data average degree, average path length and diameter of the three data set with 5000, 50000 and 500000 connections. The average degrees are respectively 0.758, 0.758 and 1.087. Because of three groups of data is selected in accordance with the time, the first group of data is the first $\frac{1}{10}$ data of the second set, the second group of data is the first $\frac{1}{10}$ data of the third set. It can be seen that with the increase of time, the relevance between the different nodes is increase, so as to traffic between nodes. With data recording time increase 100 times, the average degree only increase 3 times, indicating that as the change of amount of data, connection types keeps stable. Average path length are respectively 2.64423, 2.63940 and 5.83743. Average path length is not as big as the scale of node number, and it reach 5.83 when the number of connections reach 500,000.

The network diameter of three set of network data are 12, 15, 22, respectively. It can be seen that the diameter increases with the increase of the number of the nodes. The mobile communications network contains both the LAN IP and public IP, so the physical gateway of mobile communication network draws the edge for the Intranet and we can use the network diameter data to determine the network boundaries of the real mobile communication network.

## *3.3 Eigenvector Centrality Analysis*

Eigenvector centrality is an important parameter of node importance measure, it will give a relative score to each node according to the importance of the connection the node it connected to, the node connects to the high score node has more score than the node connection to the low score node [4]. For node $v_i$, make its centrality score $x_i$ be proportional to the sum total of all nodes score connect to his centrality, then

$$x_i = \frac{1}{\lambda} \sum_{j=1}^{N} a_{ij} x_j \tag{6}$$




(a) Autocorrelation matrix (b) Cyclic autocorrelation

**Fig. 2** Eigenvector centrality normalized histogram

This paper calculates characteristics vector centricity of IP data of the 50,000 connections set and obtains that the node X.X.0.200 which has the maximum centricity value that has the normalized centricity score 1, this result gets the same conclusion with the degree analysis. Figure 2 is the distribution of eigenvector centrality value.

It can be seen from the Fig. 2 that exception the IP node X.X.0.200 whose normalized eigenvector set to 1, namely the WAP node, all distribution is in the scope of 0–0.1. This article cuts out the effect WAP involved in the connection and rebuilds the model, recalculates the eigenvector centrality values. Figure 3 shows the new centricity distribution. It can be seen from the Fig. 3, the most value of the normalized eigenvector centre distributes in 0–0.1 range. Through comparing the IP value of the network node, it can be found that these data give priorities to LAN IP begin with 10 as well as the public IP of each provinces and cities. A few of nodes have score within the scope of 0.1–0.3, these IP data are given priority to the public IP

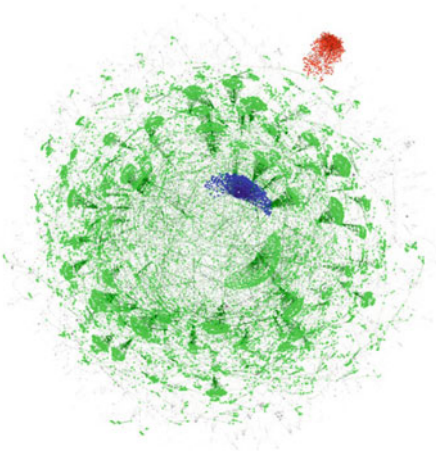


**Fig. 3** Label propagation algorithm network diagram

of operators; A handful of IP have score over 0.24. It can be seen from the data that three IP nodes of operators have the largest eigenvector normalized values, means that the three nodes in the current network have an important position and they are likely to the key nodes of the network. The paper adds special mark to such data to convenient for later data analysis.

## 4 The Network Structure Analysis

### *4.1 The Application of Label Propagation in the Network Structure Decomposition*

Label propagation algorithm is a kind of a semi-supervised learning algorithm which uses the tag node information to predict tag-less nodes [8]. This paper improves the label propagation algorithm according to the label of nodes isolated from the third chapter to separate the mobile communication network gateway IP label data, algorithm of this paper is as follows:

1. Mark $(id_1, label_1) \ldots (id_i, label_i)$ as the labelled data, let $id_1, id_2 ..., id_i \in \Omega_0$ , $\Omega_0$ is the tag node set. $Label = \{label_1, label_2, \ldots, label_i\}$ is category labels. The category is the key nodes of the paper above indicates. Keep $(id_{i+1}, label_t) \ldots (id_{i+j}, label_t)$ for not labelled data, make $id_{i+1}, id_{i+2} \ldots, id_{i+j} \in \Omega_1$, $\Omega_1$ is not tag node set and $label_t$ un-labelled tags.
2. For the edge set $\Omega_2$ of the network, $(id_i, id_j) \in \Omega_2$ $(0 < i \leq N, 0 < j \leq N)$, when one side of the two nodes $id_a$ and $id_b$ belong to the collection $\Omega_0$ at the same time, the algorithm change the label of $id_a$ and $id_b$ to the same value which is the minimum one of the two nodes and then remove the element from $\Omega_2$.
3. For the edge set $\Omega_2$ of the network, $(id_i, id_j) \in \Omega_2$ $(0 < i \leq N, 0 < j \leq N)$, when one side of the two nodes $id_a$ and $id_b$ respectively belong to the collection $\Omega_0$ and $\Omega_1$, change the $label_t$ of the $id_b$ to the $id_a$s tag from the set *Label* and remove the element from set $\Omega_2$. The algorithm uses iteration method to spread the *Label* to $label_t$.
4. Repeat step 2, 3
5. The set of two nodes $id_a$ and $id_b$ belong to the collection $\Omega_2$, does not exist $id_a$ and $id_b$ respectively belong to the collection $\Omega_0$ at the same time. Then, change the $label_t$ of nodes belong to set $\Omega_1$ to new labels with different value in random which are larger than the value of set *Label*.
6. For the edge set $\Omega_2$ of the network, $(id_i, id_j) \in \Omega_2$ $(0 < i \leq N, 0 < j \leq N)$, change the label of $id_a$ and $id_b$ to the same value which is the minimum one of the two nodes and then remove the element from $\Omega_2$.

### *4.2 The Statistical Property of Network*

This paper uses the mark results of the Chap. 3 to process the network data based on previous mark label propagation algorithm, each node has its own label. This article uses Hadoop to process and sort the label data and removes the discrete points (the point does not communication with other nodes) from the set. Then the paper visualizes the formatted result after visualization processing by using the GEPHI complex network visualization software. The paper uses Fruchterman Reingold layout algorithm [9] and [10] ForceAtlas2 layout algorithm for visual processing, Fig. 4 is the result.

It can be seen from the Fig. 3, the network nodes start from X.X.0.200 constitute an independent network (marked by green dot, as a group 1) and the network nodes start from X.X.66.12, X.X.66.14 and X.X.66.10 with special labels constitute an independent network (marked by the blue nodes, as the group 2) independently. In addition, in the data network also includes another kind of independent network with different data types (marked by the red node, as the group 3).

This article compares and analyzes the IP protocol type data of three groups of different labels. The three sets of network distinguished by the method of this paper can be analyzed by different groups for further research. Group 1 is simple IP network, contains most of the network nodes, and the data from which can reveal the regional access tendency of the IP equipment which can be used to excavate the content of the packets and analyses the user behaviour combined with depth of mining algorithm. Group 2 centres on operators IP and involves with other IP data from different provinces which can be used to analyses the network information and data communication structure of the real network. Group 3 is similar with group 2, but the core node of the operator is the IP data without province label. The actual network topology and the working mode can be explored by comparing the data of two different network.

## 5 Conclusions

This paper models the mobile communication network IP data, analyzes the complex networks nature of the mobile communication network and studies the network topology in detail. Article analyses mobile communication network topology from the degree distribution, average path length, diameter and eigenvector centrality point of view. Finally, the paper filters out the irrelevant data and realizes the division of the network according to the internal association of network data by using the label propagation algorithm. Then the paper comes to the following conclusions:

1. Mobile network node degree distribution follows power law distribution; the number of nodes with degree of more than 400 is less but slightly higher than the fitted curve, a small amount of nodes has characteristic of network hub.
2. IP data network has small average path length and diameter network.

3. IP data network has a few key nodes which are of great importance in the data analysis, the nodes can be achieved through the eigenvector centrality analysis.
4. The three un-associated networks divided by this paper have a different protocol type, IP encapsulation mode, and home ownership.

Results of this paper are important to parse mobile communication gateway IP label data. And, the method provided by this paper can be used for separation of mobile communication network gateway IP label data to acquire data sets with different protocol type and encapsulation mode. Each data set can be used for more in-depth analysis of data, such as the analysis of user behaviour through purer data.

## References

1. IP multimedia services: analysis of mobile IP and SIP interactions in 3G networks
2. G.-R. Chen, Introduction to the complex network and its recent research progress. Adv. Mech. **38**(6), 653–662 (2008)
3. R. Albert, A.L. Barabsi, Statistical mechanics of complex networks. Rev. Mod. Phys. **74**(1), 47 (2002)
4. X.-Q. Sun, S.-K. Si, Complex network algorithm and application
5. National Defense Industry Press (2015)
6. W. Ye, X. Jing-Hua, W. Zhi-Yuan, et al., Research on the growing process of short message networks (2007)
7. Xu Ye, Su Hai Zhao, Wei-Ji, et al., Analysis of the Internet network's access diameter. J. Comput. Sci. **29**(5), 690–698 (2006)
8. T. Liu, Z. Chen, X.-R. Chen, An overview of complex network theory and its application. Syst. Eng. **23**(6), 1–7 (2005)
9. J.-L. Zhang, Y.-L. Chang, S. Wen, Research on the theory and application of label propagation algorithm. Comput. Appl. Res. **30**(1), 21–25 (2013)
10. T.M.J. Fruchterman, E.M. Reingold, Graph drawing by force directed placement (1991)
11. M. Jacomy, T. Venturini, S. Heymann et al., ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. Plos One **9**(6), e98679 (2014)

# Indoor WLAN Collaborative Localization Algorithm Based on Geometric Figure Overlap

**Xiaolong Geng, Mu Zhou, Yacong Wei and Yunxia Tang**

**Abstract** Most indoor localization methods only focus on the relationship between the user locations and environmental layout, while ignoring the relations among different user locations. Thus, we come up with an idea of collaboration to reduce the impact of noise on localization performance. First of all, according to mutual information between the target user and collaborative ones, we construct the geometric figure for different user locations. Second, the candidate marker reference points with maximum overlap are selected for solving a specific localization problem. Finally, the extensive experiments conducted in both the indoor straight corridor and lab demonstrate the effectiveness of the proposed approach with the average localization error within 2 m.

**Keywords** WLAN · Location fingerprint · Geometric figure · Collaborative localization

## 1 Introduction

IEEE 802.11b/g wireless network has been widely deployed in major public places, such as schools, office buildings, airports, train stations, and large shopping malls. The majority of mobile phones, laptops, tablets, and many other mobile terminals are equipped with Wi-Fi module, which accelerates the development of indoor localization and navigation based on Wireless Local Area Networks (WLAN) technology

X. Geng (✉) · M. Zhou · Y. Wei · Y. Tang
Chongqing Key Lab of Mobile Communications Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, People's Republic of China
e-mail: 343097884@qq.com

M. Zhou
e-mail: zhoumu@cqupt.edu.cn

Y. Wei
e-mail: 2540462563@qq.com

Y. Tang
e-mail: tang013@qq.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_54

[1]. In the year 2000, the Microsoft Research Institute proposed the famous Radar indoor localization system by using the WLAN location fingerprinting technique [2]. After that, the University of Maryland proposed the Horus indoor localization system based on probability matching [3] and University of California at Los Angeles proposed the Nibble indoor localization system based on Signal-to-Noise Ratio (SNR) matching [4].

The indoor localization methods using the sensors are based on the smart terminals with motion sensors, such as accelerometer, gyroscope, and magnetometer. The pedestrian dead reckoning (PDR) method is generally used for detecting the number of walking steps, heading direction, and step length, and consequently estimate the locations of the target [5]. However, the PDR is always suffered with significant accumulative error in long time duration, which leads to the high localization error.

In this paper, to further improve the indoor localization accuracy, we propose a new method by constructing the geometric figure for different user locations based on their angles and distances. Afterward, the candidate marker reference points are selected by figure overlapping to accomplish the estimation of the target locations.

The rest of this paper is organized as follows. In Sect. 2, we give some related work on collaborative localization. Section 3 describes the proposed method in detail. The experimental results are presented in Sect. 4. Finally, Sect. 5 concludes the paper and provides an outlook of future work. Different from the works mentioned before, we first select the collaborative nodes which are associated with the target one. Second, different geometric figures are constructed based on the variation of the received signal strength and angle of the collaborative nodes. Finally, we rely on the figure overlap to correct the location estimation for the target node.

## 2 Related Work

The collaborative localization technique combines the collaboration and localization methods for the sake of achieving the higher localization accuracy. At present, in wireless sensor network, the mutual communication between the collaborative and anchor nodes enhances the localization performance to some extent. In cellular network, the mobile terminals always share their coordinates with each other for the collaboration to improve the localization accuracy. In [6], the authors proposed a new algorithm for indoor collaborative localization in wireless network by dividing the coverage area into grids with different weights to achieve the optimal collaborative combination. Based on the time of arrival and angle of arrival of the signal measured by the base station, the locations of the target can be estimated with the help of the collaborative information from the other terminals [7]. Duke University [8] proposed the iSee system to spot the virtual events in the real world. The user only needs to swipe on the touchscreen of the smartphone in the direction of the virtual event for the collaborative localization.

Different from the works mentioned before, we first select the collaborative nodes which are associated with the target one. Second, different geometric figures are

constructed based on the variation of the received signal strength and angle of the collaborative nodes. Finally, we rely on the figure overlap to correct the location estimation for the target node.

# 3 System Overview

## *3.1 Algorithm Description*

As shown in Fig. 1, the proposed algorithm consists of two phases, namely offline phase and online phase. In offline phase, the coordinates of the $i$-th Reference Point (RP) $(x_i, y_i)$ and the corresponding Received Signal Strength (RSS) vector $(rss_{i1}, rss_{i2}, \ldots, rss_{in})$ forms a location fingerprint $(x_i, y_i, rss_{i1}, rss_{i2}, \ldots, rss_{in})$, where n is the number of APs. Then, the location fingerprint database is constructed by traversing all the RPs.

The online phase is divided into the rough and precision localization steps. In rough localization step, the K-nearest neighbor (KNN) algorithm is used to estimate the initial locations of the target node and collaborative ones, as well as select collaborative nodes. In precision localization step, the direction of the collaborative nodes with respect to the target one is obtained by the magnetometer. Then, based on the

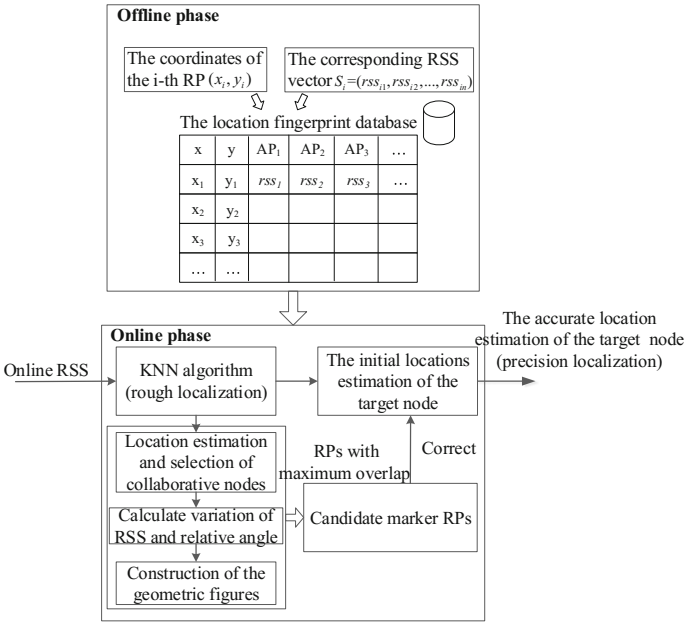



**Fig. 1** Algorithm process

variation of the received signal strength and angle of the collaborative nodes, we rely on construction of the geometric figures to correct the initial location estimation according to the candidate marker reference points with the maximum overlap, and consequently obtain the accurate location estimation of the target node $(x_u, y_u)$.

## 3.2 Geometric Figure Construction

The key factors affecting the localization accuracy for the target node are the errors of localization and angle of the collaborative ones. Based on this, we set the variance thresholds for the RSS ${\sigma_l}^2$ and angle ${\sigma_a}^2$. In view of each collaborative node, we focus on three cases for the geometric figure construction as follows.

- In the case that the variation of the received signal strength and angle of the collaborative nodes is smaller than ${\sigma_l}^2$ and greater than ${\sigma_a}^2$ respectively, we only focus on the error of angle. Figure 2a shows that the geometric figure is simplified into a triangle model.
- In the case that the variation of the received signal strength and angle of the collaborative nodes is greater than ${\sigma_l}^2$ and smaller than ${\sigma_a}^2$ respectively, we only focus on the error of localization. Figure 2b shows that the geometric figure is simplified into a rectangle model.
- In the case that the variation of the received signal strength and angle of the collaborative nodes is greater than ${\sigma_l}^2$ and ${\sigma_a}^2$ respectively, we focus on the errors of both the localization and angle. Figure 2c shows that the geometric figure is simplified into a trapezium model.

Since the triangle and rectangle models can be recognized as the special cases of trapezium model, we only focus on the trapezium model in the results that follow. In Fig. 2c, $\theta$ is the angle of the collaborative node with respect to the target one, r and $\alpha$ are the errors of localization and angle of the collaborative node with respect to

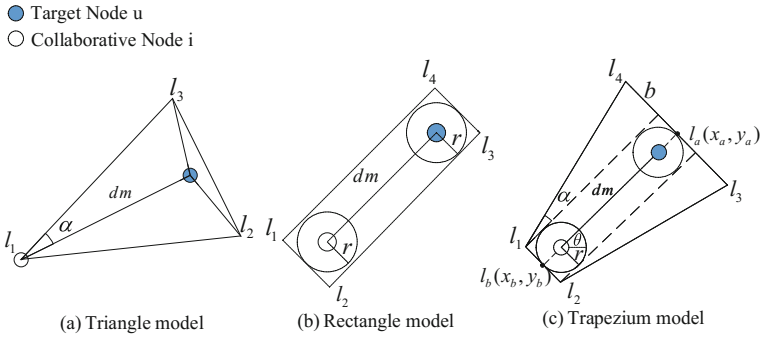



**Fig. 2** Geometric figures under different cases

the target one, and *dm* is the estimated distance between the collaborative and target ones, such that

$$dm_i = \sqrt{(x_i - x_u')^2 + (y_i - y_u')^2} \quad i = 1, \ldots v \tag{1}$$

In the trapezium model, by assuming that the coordinates of the four vertices are $l_1(x_1, y_1)$, $l_2(x_2, y_2)$, $l_3(x_3, y_3)$ and $l_4(x_4, y_4)$, and the center point of the upper and lower lines are $l_a(x_a, y_a)$ and $l_b(x_b, y_b)$ respectively, we have

$$\begin{cases} b = r + (dm + 2r) \times \tan(\theta) \\ x_b = x + r \times \cos(\pi + \theta) \quad y_b = y + r \times \sin(\pi + \theta) \\ x_a = x + (r + dm) \times \cos\theta \quad y_a = y + (r + dm) \times \sin\theta \\ x_1 = x_b + r \times \cos(\theta + \frac{\pi}{2}) \quad y_1 = y_b + r \times \sin(\theta + \frac{\pi}{2}) \\ x_2 = x_b + r \times \cos(\theta + \frac{3\pi}{2}) \quad y_2 = y_b + r \times \sin(\theta + \frac{3\pi}{2}) \\ x_3 = x_a + b \times \cos(\theta + \frac{3\pi}{2}) \quad y_3 = y_a + b \times \sin(\theta + \frac{3\pi}{2}) \\ x_4 = x_a + b \times \cos(\theta + \frac{\pi}{2}) \quad y_4 = y_a + b \times \sin(\theta + \frac{\pi}{2}) \end{cases} \tag{2}$$

## 4 Experimental Results

### *4.1 Environmental Layout*

As shown in Fig. 3, the experimental environment is selected on the fifth floor in a building with the dimensions of 57 m × 25 m. There are 9 D-Link2310 APs fixed in target environment which includes a straight corridor and a lab. The 34 RPs (with black points) are uniformly distributed and there are 200 RSS measurements collected at each of them. The 100 RSS measurements are collected at each of the randomly selected 17 collaborative and 19 target nodes for the testing.

### *4.2 Parameters Discussion*

The three main parameters involved in the proposed algorithm are localization errors of the collaborative nodes r, heading angle errors of the collaborative nodes with respect to the target one $\alpha$, and number of collaborative nodes $v$.

- From Fig. 4, since localization errors of the most collaborative nodes are within 4 m, we set r equal to 4 m in the results that follow.
- From Fig. 5, since heading angle errors of the most collaborative nodes with respect to the target one are within 10 degree, we set $\alpha$ equal to 10 degree in the results that follow.

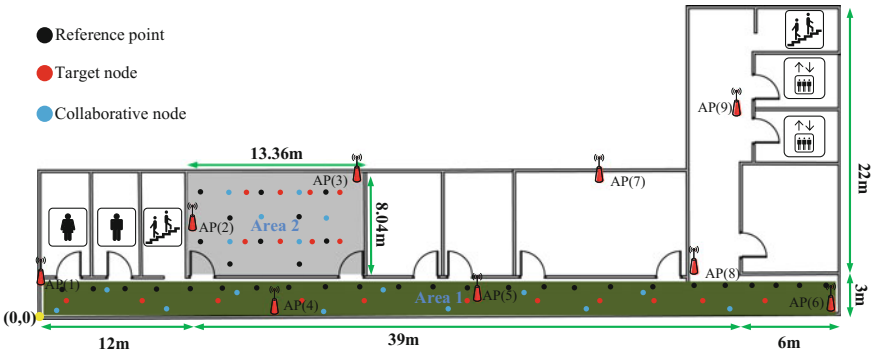



**Fig. 3** Environmental layout

**Fig. 4** Localization errors of the collaborative nodes

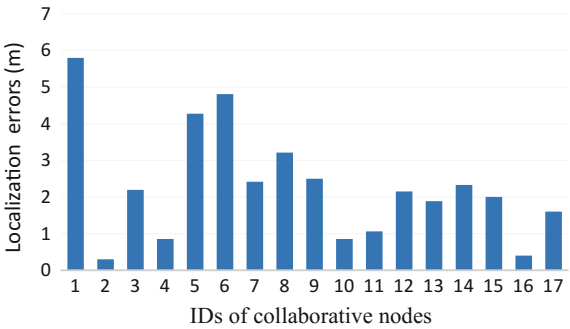
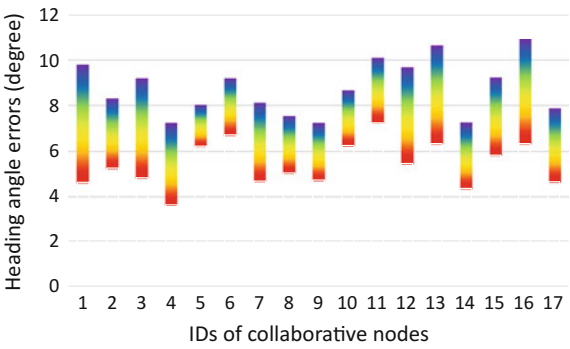



**Fig. 5** Heading angle errors of the collaborative nodes




- From Table 1, a small amount of collaborative nodes generate large overlapping area that contains a lot of RPs with maximum overlap recorded as candidate marker reference points, which makes the localization accuracy decreased owing to bringing some little value RPs. However, a large amount of collaborative nodes will generate small overlapping area that contains a few of RPs with maximum

**Table 1** The number of candidate marker RPs under different number of collaborative nodes

| IDs of TNs | 2 CNs | 3 CNs | 4 CNs | 5 CNs |
|---|---|---|---|---|
| 1 | 4 | 4 | 4 | 4 |
| 3 | 9 | 7 | 7 | 6 |
| 5 | 8 | 6 | 6 | 2 |
| 7 | 5 | 5 | 4 | 1 |
| 9 | 6 | 6 | 3 | 0 |
| 11 | 8 | 7 | 6 | 2 |
| 13 | 5 | 5 | 5 | 4 |
| 15 | 7 | 6 | 5 | 1 |
| 17 | 5 | 4 | 3 | 3 |

overlap, which may lead to 0 candidate marker reference point, affecting the localization correction of target nodes, thus we set $v$ equal to 3 or 4 in the results that follow.

where TNs and CNs are the abbreviation of target nodes and collaborative nodes, respectively.

## *4.3 Localization Performance*

In the straight corridor, we compare the cumulative distribution functions (CDFs) of errors and average localization errors achieved by the proposed geometric figure overlap, single figure overlap, and traditional KNN algorithms in Figs. 6 and 7 respectively. From these figures, we can find that the proposed algorithm is supe-

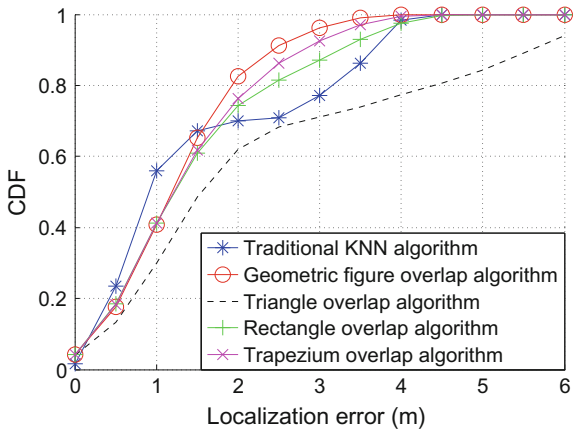



**Fig. 6** CDFs of errors in straight corridor

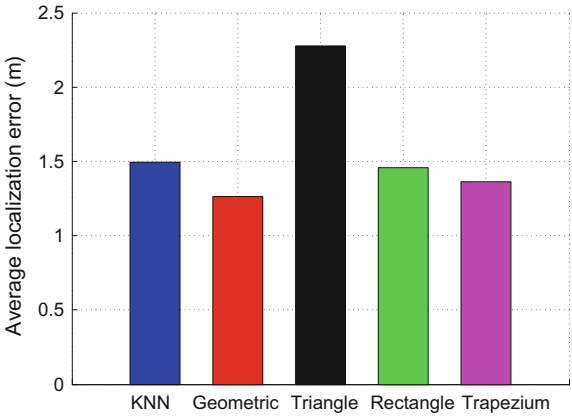



**Fig. 7** Average localization errors in straight corridor

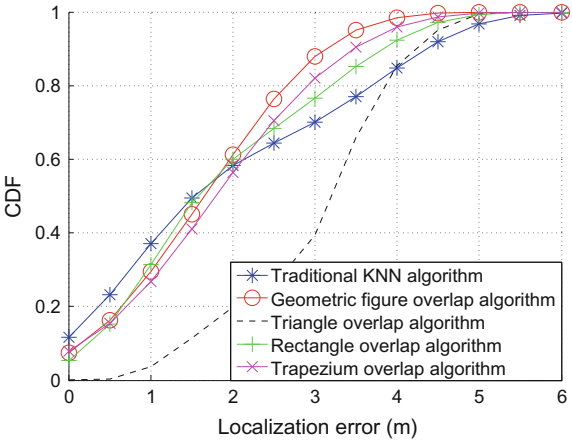



**Fig. 8** CDFs of errors in straight corridor

rior to both the single figure overlap and traditional KNN algorithms in localization accuracy. For example, compared with the traditional KNN algorithm, the probabilities of errors within 2 m, 2.5 m, and 3 m by the proposed algorithm are 83%, 91%, and 96% respectively, which are about 13, 20, and 19% higher than the ones by the traditional KNN algorithm.

In the lab, we similarly compare the CDFs of errors and average localization errors achieved by the proposed geometric figure overlap, single figure overlap, and traditional KNN algorithms in Figs. 8 and 9 respectively. We can find geometric figure overlap algorithm is also superior to single figure overlap algorithm. Compared with traditional KNN algorithm, the probabilities of errors within 2.5 m, 3 m, and 3.5 m by the proposed algorithm are 76%, 87%, and 95% respectively, which are about 12, 17, and 18% higher than the ones by the traditional KNN algorithm.

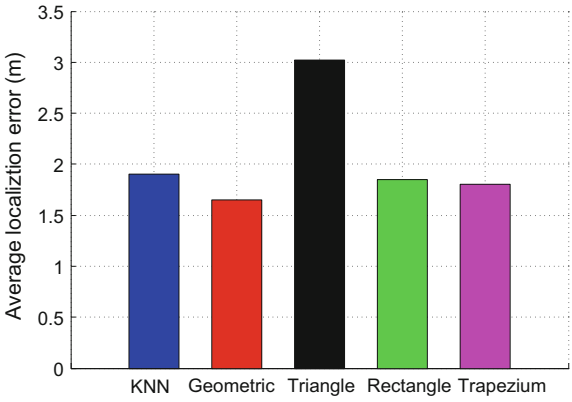



**Fig. 9** Average localization errors in straight corridor

## 5 Conclusion

We propose a new indoor WLAN collaborative localization algorithm based on the geometric figure overlap, as well as demonstrate the effectiveness of the construction of geometric figure. By using the mutual information between the collaborative and target nodes, the accuracy of location estimation for the target can be well improved for both the indoor straight corridor and lab environments. In future, the selection of the optimal parameters involved in the proposed algorithm for the further improvement of localization accuracy in large-scale indoor environment forms an interesting topic.

**Acknowledgements** The authors wish to thank the reviewers for the careful review and valuable suggestions. This work was supported in part by the National Natural Science Foundation of China (61471077), Program for Changjiang Scholars and Innovative Research Team in University (IRT1299), Special Fund of Chongqing Key Laboratory (CSTC), Fundamental and Frontier Research Project of Chongqing (cstc2017jcyjAX0380, cstc2015jcyjBX0065), and University Outstanding Achievement Transformation Project of Chongqing (KJZH17117).

## References

1. R.D. Xu, J. Liu, J.S. Zhu, X.F. Jiang, The design and construction of WLAN-based indoor navigation system. Commun. Comput. Inf. Sci. **482**, 437–446 (2015)
2. P. Bahl, V.N. Padmanabhan, RADAR: an in-building RF-based user location and tracking system. Proc.—IEEE INFOCOM **2**, 775–784 (2000)
3. M. Youssef, A. Agrawala, The horus WLAN location determination system, in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services* (2005), pp. 205–218
4. E. Coca, V. Popa, G. Buta, An indoor location system performance evaluation and electromagnetic measurements, in *IEEE 16th International Symposium for Design and Technology of Electronics Packages* (2010), pp. 69–72

5. G.L. Chen, X.L. Meng, Y.J. Wang, Y.Z. Zhang, P. Tian, H.C. Yang, Integrated WiFi/PDR/smartphone using an unscented kalman filter algorithm for 3D indoor localization. Sens. (Switzerland) **15**(9), 24595–24614 (2015)
6. J. Luo, Z.Y. He, Y.X. Zhang, Y.C. Song, Indoor cooperation localization algorithm based grid in wireless networks. J. Huazhong Univer. Sci. Technol. (Natural Science Edition) **43**(5), 114–118 (2015)
7. J.X. Wang, Z.H. Liu, J. Li, Cooperative positioning algorithm with single station. J. Xian Univer. Posts Telecommun. **18**(2), 32–35 (2013)
8. R.W. Ouyang, A. Srivastava, P. Prabahar, R.R. Choudhury, M. Addicott, F. McClernon, If you see something, swipe towards it: crowdsourced event localization using smartphones, in *UbiComp—Proceedings ACM International Joint Conference Pervasive Ubiquitous Computing* (2013), pp. 23–32

# Particle Swarm Optimized Indoor Localization for Tracking of Moving Target

**Chunyue Li, Xiao Peng and Chenglin Zhao**

**Abstract** Dynamic tracing is a continuous process to estimate and predict the motion state of moving target using observable data. Probabilistic methods such as Bayesian filtering, Monte Carlo box (MCB) and particle filtering (PF) make use of historical data and state transition distribution as the posterior distribution function of sampling particles. In this paper, particle swarm method is added to sampling process to move particle to regions with higher posterior distribution. This optimization significantly enhances positioning accuracy and shortens localization time compared with conventional MCB and PF methods.

**Keywords** Monte Carlo box · Particle filtering · Particle swarm · Indoor localization

## 1 Introduction

Confronted with the increasingly demand of smart phone applications for awareness of users' location [1, 2], localization capability in those devices is equipped to provide location-based services which is called LBS. Global position system (GPS) is a kind of outdoors location service. But non-line-of-sight as well as attenuation and scattering of satellite signal make GPS not appropriate for indoor localization. While indoor location services, such as WIFI localization, Bluetooth localization and ZigBee, RFID, are frequently used in many scenarios. Meanwhile, Internet of Things has been a hot topic since the beginning of this decade in China, and the system cannot be constructed without indoor localization.

C. Li (✉) · C. Zhao
School of Information and Communication Engineering,
Beijing University of Post and telecommunication, Beijing 100876, China
e-mail: 929147299@qq.com

X. Peng
The State Radio Product Quality Supervision and Inspection Center,
Beijing 100041, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_55

Static position locates the target every second independently. If the target is moving, previous positions are relevant to the present position, which provides more information to the system. Probabilistic methods have been used in a lot of applications to track the dynamic system's state. Arnaud [3] presents prediction, smoothing, and evaluation of the likelihood for Monte Carlo Sampling method in dynamic models. Hu [4] first brings sequential Monte Carlo Localization (MCL) method and exploits mobility to enhance the accuracy and precision of localization. Baggio [5] builds a Monte Carlo box (MCB) based on MCL and makes better use of information a sensor node gathers and drawing the necessary location sample faster. Xu [6] comes up with wavelet transform prediction to enhance the accuracy in wireless sensor network (WSN). Hu and Xu [4] propose a multi-feature selection algorithm based on the support vector machine (SVM) to track the target with the maximum confidence probability.

## 2 Monte Carlo Localization

### A. *Basic idea*

Monte Carlo localization [4] can be concluded as two steps: prediction and filtering. After initialization, which is to select a random set of samples $L_0 = \{l_0^0, l_0^1, \ldots l_0^{n-1}\}$, comes the iteration of prediction and filtering. In prediction period, a node forecasts a new bunch of samples $L_t$ based on the previous samples $L_{t-1}$ and posterior distribution; during filtering period, all impossible positions are eliminated while possible ones are given normalized weights.

$$\pi(l_t|o_0, o_1, \ldots o_t) = p(l_0) \prod_{k=1}^{t} p(l_k|l_{k-1}) \tag{1}$$

$$\widetilde{w}_t^i = \widetilde{w}_{t-1}^i p(o_t|l_t^i), \quad \widetilde{w}_t^i = \frac{\widetilde{w}^i}{\sum_{k=1}^{n} \widetilde{w}_t^k} \tag{2}$$

Equation (1) is the construction of posterior distribution, (2) is the update period in which new samples use weighted set $(l_t^i, w_t^i)$ to simulate the posterior distribution.

B. *Monte Carlo box*

MCB algorithm [5] introduce anchor box and sample box into prediction period. Anchor box can be built while $(x_j, y_j)$ being the coordinates of the anchor $j$ and $n$ being the total number of anchor heard, as for one-hop anchor:

$$\begin{aligned} x_{\min} &= \max_{j=1}^{n}(x_j - r), x_{\max} = \min_{j=1}^{n}(x_j + r) \\ y_{\min} &= \max_{j=1}^{n}(y_j - r), y_{\max} = \min_{j=1}^{n}(y_j + r) \end{aligned} \tag{3}$$

Replace $r$ by $2r$ in the form (3) when using two-hop anchors. Sample box $Box_t^i = \{(x_{\min}^i, x_{\max}^i), (y_{\min}^i, y_{\max}^i)\}$ is built after the anchor box, where $(x_{t-1}^i, y_{t-1}^i)$ is the coordinates of the old sample $l_{t-1}^i$ and $v_{\max}$ is the maximum moving speed.

$$\begin{aligned} x_{\min}^i &= \max(x_{\min}, x_{t-1}^i - v_{\max}), \quad y_{\min}^i = \max(y_{\min}, y_{t-1}^i - v_{\max}) \\ x_{\max}^i &= \min(x_{\max}, x_{t-1}^i + v_{\max}), \quad y_{\max}^i = \min(y_{\max}, y_{t-1}^i + v_{\max}) \end{aligned} \tag{4}$$

Prediction process is to construct the posterior distribution

$$P(l_t | l_{t-1}^i) = \begin{cases} 1, x_{\min}^i \le x_t^i \le x_{\max}^i \& y_{\min}^i \le y_t^i \le y_{\max}^i \\ 0, otherwise \end{cases} \tag{5}$$

And then abandon the samples beyond the communication coverage of one-hop and two-hop anchor box. Denoted by

$$P(o_t | l_t^i) = \begin{cases} 1, & \forall s \in S, d(l_t, s) \le r \wedge \forall s \in T, r < d(l_t, s) \le 2r \\ 0, & otherwise \end{cases} \tag{6}$$

S is the set of one-hop anchors and T is the set of two-hop anchors. The building process of anchor box and sample box narrows the sampling area, which improves sampling rate and reduces the energy consumption of positioning. However, MCB method has disadvantages. First, sampling area is pretty large and positioning error is big when anchor node is not dense enough; Second, the maximum sampling points are fixed, which leads to a waste of memory and computational overhead when sampling area is small. The last, only one-hop and two-hop anchor node are heard during prediction and filtering process rather than all historical motion trail that can improve the positioning accuracy.

## 3 Optimized Particle Filtering

### A. *Particle Filtering*

Particle filtering uses approximate method to realize recursive Bayesian filtering. It basically represents the current state of the posterior distribution with weighted random calculation sample, and replaces the integral operation with sample mean when estimating the target state parameters. Particle filtering has very strong practicability in dealing with nonlinear and non-Gaussian problems.

Particle filtering uses $n$ weighted particle collection $\{x_{0:k}^i, w_k\}_{i=1}^n$ to illustrate the posterior probability distribution of target motion state at $k$ time interval.

$$p(x_k|z_{1:k}) \approx \sum_{i=1}^{n} w_k^i \delta(x_k - x_k^i), \tag{8}$$

where $\delta(\bullet)$ is the Dirac-delta function. Weights are updated by formula (9):

$$w_k^i = w_{k-1}^i \frac{p(z_k|x_k^i)p(x_k^i|x_{k-1}^i)}{q(x_k^i|x_{k-1}^i, z_k)}, \tag{9}$$

where $q(\bullet)$ is the importance distribution. In this way, the state of the moving target can be estimated by

$$x_k^{'} = E(x_k|z_{1:k}) \approx \sum_{i=1}^{n} w_k^i x_k^i (k = 1, 2, \ldots, n) \tag{10}$$

### B. *Particle Swarm Theory*

Conventional particle filter algorithm chooses subprime importance function, therefore, the particle swarm optimization algorithm is introduced to bring optimization to the sampling process of the algorithm.

First of all, introduce the latest observations to sampling process and define the fitness function:

$$F_f = \exp\left[ -\frac{1}{2\sigma}(a_{lat} - a_{pred})^2 \right], \tag{11}$$

where $a_{lat}$ is the latest observation and $a_{pred}$ is the result of prediction. The optimization process [7] moves those particles far away from the true state to area with large probability of showing true state, and improves the estimation precision of

moving object. When the initial state is unknown, particle filter needs a vast number of particles to accurately estimate the state of a moving object. In this way, the optimized moving target tracking based on particle filtering can be realized with following steps:

(a) Initialization: get samples $\{S_0^i\}_{i=1}^N$ with equal weight $\frac{1}{N}$ from $p(S_0)$
(b) calculate the fitness value of each sample, then use particle swarm optimization algorithm to make particles move to high likelihood area to get a new sample set $\{Q_0^i\}_{i=1}^N$
(c) normalized significant priority:

$$w_k^i = p(z_1(k), \ldots z_{N_k}(k)|Q_k^i) = \coprod_{j=1}^{N_k} p(z_j(k)|Q_k^i) \tag{12}$$

(d) Re-sampling to cope with particle degeneracy. After re-sample, particles near the true state will increase weights.

$$[Q_k^i, \frac{1}{N}]_{i=1}^N = [Q_k^i, w_k^i]_{i=1}^N \tag{13}$$

(e) State estimation:

$$S_k^{'} = \sum_{i=1}^{N} w_k^{i'} S_k^i \tag{14}$$

## 4 Target Tracking

Suppose the target moving within CT/CV model [8]. $S_k$ is the location state of the target at k time interval.$S_k = [x_k, y_k, v_{xk}, v_{yk}]^T$. Where $(x_k, y_k)$ is the coordinate of the target in time $k$ and $v_{xk}, v_{yk}$ is the velocity in x and y direction r.

$$F_{CV} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad F_{CT} = \begin{bmatrix} 1 & 0 & \frac{\sin(\omega\Delta t)}{\omega} & \frac{\cos(\omega\Delta t)-1}{\omega} \\ 0 & 1 & \frac{1-\cos(\omega\Delta t)}{\omega} & \frac{\sin(\omega\Delta t)}{\omega} \\ 0 & 0 & \cos(\omega\Delta t) & -\sin(\omega\Delta t) \\ 0 & 0 & \sin(\omega\Delta t) & \cos(\omega\Delta t) \end{bmatrix} \tag{15}$$

m $\omega$ is the turn rate, $\Delta t$ is the time interval of positioning system, motion noise is:

$$u = \begin{bmatrix} 0.5\Delta t^2 & 0 \\ 0 & 0.5\Delta t^2 \\ \Delta t & 0 \\ 0 & \Delta t \end{bmatrix} \begin{bmatrix} \sigma_x \\ \sigma_y \end{bmatrix}, \tag{16}$$

where $\sigma_x, \sigma_y$ are the motion noise standard deviation in x and y direction. Suppose there are $N$ mutually independent access points(AP) deploying in test area, $a_j$ is the signal strength received from the $j$th AP, so the particle likelihood $p_j^{(i)}$ can be calculated by the $j$th AP:

$$p_j^{(i)} = \frac{1}{\sqrt{2\pi}\sigma} \exp\left( - \frac{(a_j - a_j^{(i)})^2}{2\sigma^2} \right) \tag{17}$$

$$p_j = \prod_{j=1}^{N} p_j^{(i)} \tag{18}$$

## 5 Simulation Result

Suppose signals transmit accords with logarithm model, so the signal strength can be calculated by the distance between the sample and AP. Time interval $\Delta t = 1s$ and the total simulation time is 30 s. turn rate $\omega = 0.5\ rad/s$, the standard deviation of motion noises are all 0.1. Initial state is $S_0 = [7, 5, 0.5, 1.5]^T$. The maximum particle filtering sample size = 100 and the threshold value of re-sampling is 50.

60 times of simulation are conducted independently. And we use the average value of all 60 times as the root-mean-square errors (RMSE) (Fig. 1).

$$RMSE = \sqrt{(x_k^{'} - x_k)^2 + (y_k^{'} - y_k)^2} \tag{19}$$

By comparing the tracking trails of three algorithms, it is noticeable that the last trajectory is closest to the target's actual path as expected. In order to verify PSOF outperformance in detail, maximum error (m/s), average RMSE (m/s), standard deviation of RMSE (m/s) are listed in the following chart and '<0.5 m' means the percentage of RMSE smaller than 0.5 m (Table 1).

As we can see, PSOF algorithm has the least RMSE both in maximum and average value, which indicates the best accuracy in localization. The maximum error of MCB localization is the biggest and PF and PSOF has better performance.

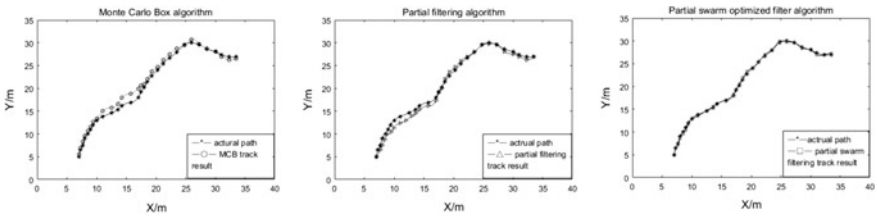



**Fig. 1** Monte Carlo box versus particle filtering versus particle swarm optimized filter

**Table 1** RMSE comparison of 3 positioning methods

| RMSE | Average | Stdev | Max | <0.5 m (%) |
|---|---|---|---|---|
| MCB | 0.831 | 0.893 | 2.00 | 47 |
| PF | 0.560 | 0.527 | 1.63 | 70 |
| PSOF | 0.488 | 0.436 | 1.08 | 92 |

**Table 2** RMSE of several methods when is $\sigma$ different

| RMSE | $\sigma = 1$ | $\sigma = 2$ | $\sigma = 3$ | $\sigma = 4$ |
|---|---|---|---|---|
| MCB | 1.203 | 1.525 | 2.171 | 2.964 |
| PF | 0.781 | 0.939 | 1.039 | 1.350 |
| PSF | 0.490 | 0.682 | 0.908 | 1.055 |

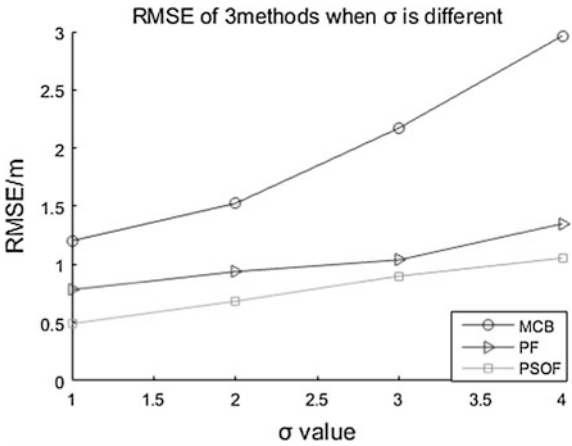



**Fig. 2** RMSE of several methods when is $\sigma$ different

After state estimation, 92% of the distances between estimated and true coordinate are less than 0.5 m, it meets the requirements of positioning accuracy basically. Meanwhile, PSOF is more stable in positioning.

As is noted at the beginning of simulation, noises are added to simulate all unstable factors caused by multi-path effect of signal or obstacles of the environment. So the adjustment of the standard deviation of motion noises will bring influential changes to the accuracy of tracking. Details are listed in the following chart (Table 2, Fig. 2).

## 6 Conclusion

Particle swarm optimization filtering algorithm introduces the latest observation to the sampling process, aiming at the existing problem of PF positioning algorithm, making sample particle set moving toward probabilistic higher region, which avoids the problem of poor particles, and improve the positioning precision of the target.

The simulation results show the feasibility and effectiveness of the improved algorithm. Compared with Monte Carlo box and basic particle filtering, PSOF has improved accuracy by 41.2 and 12.9% approximately.

**Acknowledgement** The work has been supported by the National Science Fund No. 61271180.

## References

1. D. Munoz, F. Bouchereau, C. Vargas, et al., *Position Location Techniques and Applications*. Academic Press (2009)
2. S. Zafer, G. Sinan, G. Ismail, Ultra-wideband positioning systems: theoretical limits, in *Ranging Algorithms, and Protocols*. Cambridge (2008)
3. A. Doucet, S. Godsill, C. Andrieu. On sequential Monte Carlo sampling method for Bayesian filtering. Stat. Comput. **10**, 197–208 (2000)
4. L. Hu, D. Evans, Localization for mobile sensor network, in *Wireless Communication*, pp. 45–57 (2004)
5. A. Baggio, K. Langendoen, Monte Carlo localization for mobile wireless sensor networks. Ad Hoc Netw. **6**(5), 718–733 (2008)
6. Z. Xu, Z. Shan, A localization algorithm based on wavelet transform prediction for mobile nodes. Chin. J. Sens. Actuators **4**(4) (2016)
7. H. Shaorong, Survey of particle swarm optimization algorithm. Comput. Eng. Des. **30**(8), 1977–1980 (2009)
8. X. Yuan, C. Han, Z. Duan, et al., Comparison and choice of models in tracking target with coordinated turn motion (IEEE, 2005)
9. Xu Ma Lin, Wu Di Yubin, A novel two-step WLAN indoor positioning method. J. Commun. Inf. Syst. **6**(14), 4627–4636 (2010)
10. Y. Ning, Q. Feng, Z. Rui, Improved particle filter based on genetic algorithm. J. Shanghai Jiaotong Univer. **10**, 1526–1530 (2011)

# Application of the EKF Algorithm in the DTMB Positioning System

**Chengbiao Fu, Zengshan Tian and Anhong Tian**

**Abstract** Because single GPS technology cannot realize localization in a dense urban environment, and in view of Chinese digital television terrestrial broadcasting (DTMB) signals using with fixed transmitters, wide coverages, high positioning accuracy, and low cost, this paper suggests a location method based on DTMB. Because the observation model of the DTMB's location system is gauss non-linear, this article contrasts and analyzes the tracking effect of two kinds of environment: a line-of-sight (LOS) environment and a non-line-of-sight (NLOS) environment. Simulation results show that: the tracking trajectory is in line with the true trace, due to the interference of the NLOS environment; the error of NLOS is bigger than LOS; and the tracking accuracy of NLOS is better than LOS. This all proves the feasibility and superiority of a location system based on DTMB.

**Keywords** DTMB positioning system · Filter algorithm · Simulation result

## 1 Introduction

GPS has been widely used in the outdoor environment, and its precision is very high. However, the positioning accuracy, availability, and reliability of GPS are related to the number of satellites and geometrical structure of the satellite constellation. In an urban environment, there is a location blind area. Due to the advantages of digital television signals [1], such as fixed transmitters, wide coverages, and low cost, this chapter puts forward a positioning method based on a digital television terrestrial broadcasting (DTMB) signal [2]. At present, there are four digital television signal standards worldwide, such as the US ATSC standard,

C. Fu (✉) · A. Tian
Department of Information Engineering, Qujing Normal College, Qujing 655000, China
e-mail: tianfucb@163.com

Z. Tian
College of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_56

the European DVB standard, and China's DTMB standard [3, 4]. China's DTMB standard was formulated on August 18, 2006, and was promulgated later, so research of DTMB has varying significance.

## 2 The DTMB Positioning System

In this system a key point that the time between DTMB stations must be strictly synchronized. The TOA value can be calculated through pseudorange measurement, and the trilateration method can be used to calculate the localization of receivers (Fig. 1).

Assuming that the coordinates of stations are respectively (x1, y1), (x2, y2), and (x3, y3), the distance between a base station and receiver is respectively d1, d2, and d3. Assuming that the coordinates of MS are (x, y), the distances can be described as follows:

$$d_1 = \sqrt{(x_1 - x)^2 + (y_1 - y)^2} \tag{1}$$

$$d_2 = \sqrt{(x_2 - x)^2 + (y_2 - y)^2} \tag{2}$$

$$d_3 = \sqrt{(x_3 - x)^2 + (y_3 - y)^2} \tag{3}$$

Then the coordinates of MS can be written as:

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 2(x_1 - x_3)2(y_1 - y_3) \\ 2(x_2 - x_3)2(y_2 - y_3) \end{bmatrix}^{-1} \begin{bmatrix} x_1^2 - x_3^2 + y_1^2 - y_3^2 + d_3^2 - d_1^2 \\ x_2^2 - x_3^2 + y_2^2 - y_3^3 + d_3^2 - d_2^2 \end{bmatrix} \tag{4}$$

The coordinates of a user terminal can be calculated using a trilateral localization algorithm, at the same time, the geometric distance $d_i$, clock offset $\Delta t$, and pseudorange can be written as:

$$\rho_i = d_i + c \cdot \Delta t \tag{5}$$

The location method based on DTMB is shown in Fig. 2.

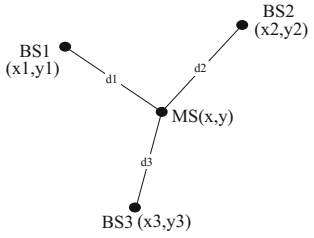



**Fig. 1** Trilateral positioning principle

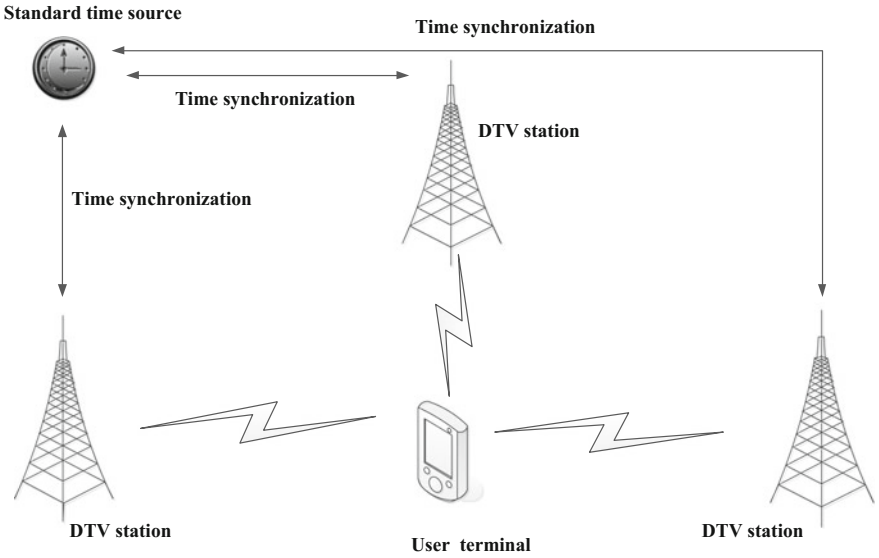



**Fig. 2** Positioning principle of DTMB

## 3 The EKF Non-linear Filtering Algorithm

The state equation and observation equation for flying targets are non-linear. According to the characteristics of an integrated positioning system its equation is non-linear, [5–7] therefore, an EKF algorithm is more suitable for use as a tracking filter. In the process of calculating an EKF algorithm, the state equation and observation equation of a non-linear system needs to be converted to the state equation and observation equation of a linear system, [8, 9] namely, it needs to expand its estimated value via Taylor series, remove the secondary and higher items, only keep one item, and reuse the KF filtering algorithm to generate an estimate.

The assumed state equation and observation equation model of a non-linear system can be written as follows:

$$\begin{cases} X(k) = f(X(k-1)) + W(k-1) \\ Y(k) = h(X(k)) + V(k) \end{cases} \tag{6}$$

where, the state equation and observation equation's noise is $W(k)$ and $V(k)$, being unrelated white noise sequences. The covariance matrix is respectively $Q(k)$ and $R(k)$.

The EKF algorithm should calculate the optimal linear estimate $\hat{X}(k|j)$ of state $X(k)$, using observations of $\{Y(1), Y(2), \ldots, Y(j)\}$ at moment j. The linearization process of non-linear Eq. (15) can be written as:

$$\begin{aligned} X(k) &\approx f(\hat{X}(k-1)) + \frac{\partial f}{\partial X}|_{X(k-1)=\hat{X}(k-1)}[X(k-1)-\hat{X}(k-1)] + W(k-1) \\ &= AX(k-1) + W(k-1) \end{aligned} \tag{7}$$

$$\begin{aligned} Y(k) &\approx h(\hat{X}(k)) + \frac{\partial h}{\partial X}|_{X(k)=\hat{X}(k)}[X(k)-\hat{X}(k)] + V(k) \\ &= HX(k) + V(k) \end{aligned} \tag{8}$$

The estimation error is:

$$\tilde{X}(k|j) = X(k) - \hat{X}(k|j) \tag{9}$$

The state error covariance matrix is:

$$P(k|j) = E[\tilde{X}(k|j)\tilde{X}^T(k|j)] \tag{10}$$

The optimal estimate of the EKF algorithm is confirmed, when $X$ is minimum. If $k=j$, there is filtering problem: $\hat{X}(k)$, $\tilde{X}(k)$, $P(k)$ are respectively $\hat{X}(k|j)$, $\tilde{X}(k|j)$, $P(k|j)$. The step state prediction is $\hat{X}(k|k-1)$, the error matrix of the step state prediction is $P(k|k-1)$, and the recursive process of state estimation can be shown below.

The state estimation update equation is:

$$\hat{X}(k) = A\hat{X}(k-1) + K(k)[Y(k) - HA\hat{X}(k-1)] \tag{11}$$

The filter gain equations are:

$$K(k) = P(k|k-1)H^T[HP(k|k-1)H^T + R(k)]^{-1} \tag{12}$$

$$P(k|k-1) = AP(k-1)A^T + Q(k-1) \tag{13}$$

The filtering covariance update equation is:

$$P(k) = P(k|k-1) - K(k)HP(k|k-1) \tag{14}$$

If the initial value of the state vector is $X(0)$, and the initial value of the state estimation error covariance matrix is $P(0)$, the EKF algorithm can calculate the optimal linear estimate $\hat{X}(k|j)$ of state $X(k)$, according to Eqs. (20)–(23), thus completing the state estimation of the flight target.

# 4 Algorithm Simulation

## 4.1 Setting the Simulation Parameters

Assuming that the model of the target user is plane turning speed movement, the simulation parameters are: the initial position of the target is (5, 25, 15) m, the acceleration vector is $0.5, -0.5, -0.5\, m/s^2$, the initial velocity vector is $4, -4, 4\ m/s$, the process noise obeys a Gaussian distribution, the observation time of the simulation is 80 s 110 s, the step is 1 s, there are three Beidou satellites, three DTMB signals, and the number of particles is 400.

## 4.2 Simulation Results

In order to verify the feasibility of such a DTMB positioning system, in this chapter, two kinds of environment have been considered: a LOS environment and a NLOS environment. The contrast effect of location tracking is shown in Fig. 3.

The simulation result in Fig. 3 shows that the trajectory of an EKF filtering algorithm is essentially coincident with the true trajectory. The tracking accuracy of

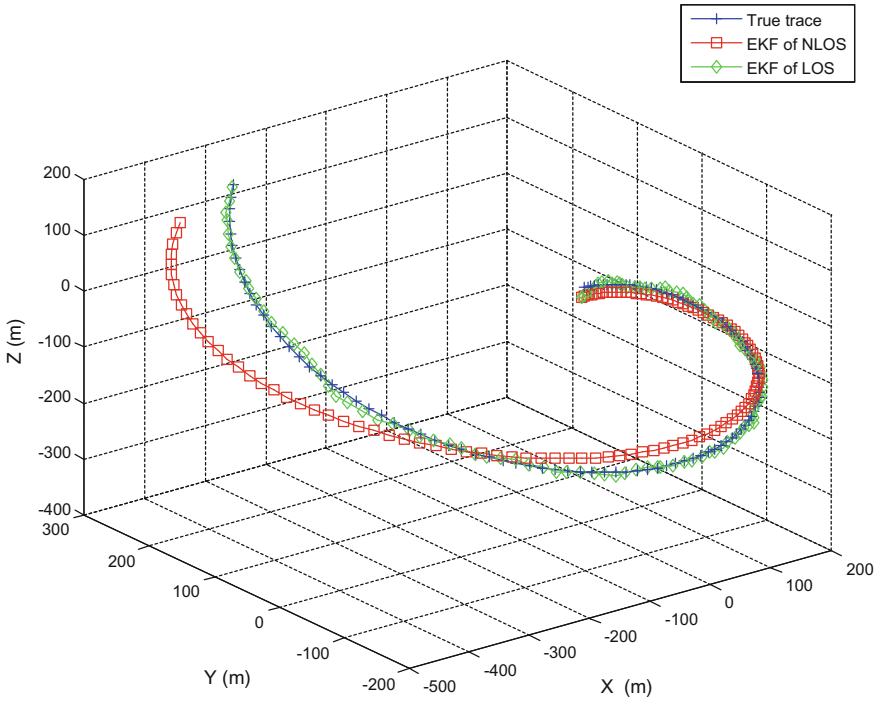



**Fig. 3** Comparative tracking trajectory of NLOS and LOS

the EKF of NLOS is better than the EKF of LOS. Due to the interference of the NLOS environment, the error of NLOS is bigger than that of the error in LOS, however, the tracking effect proves that the filtering trace is in line with the true trace.

## 5 Conclusion

GPS cannot achieve the demand for accurate locations in an urban environment. This chapter puts forwards a novel navigation positioning system, which is based on DTMB. Due to the characteristics of the non-linear observation equation, an EKF algorithm is applied to the positioning system. Simulation results show that the tracking trajectory of two environments is essentially coincident with true trace, proving the feasibility and superiority of a location system based on DTMB.

**Acknowledgements** This work was supported by the Youth Project Foundation of Science and Technology Department of Yunnan Province under Grant 2014FD048 and the Nature Science Foundation of Yunnan Province Education Department under Grant (2016ZDX127).

## References

1. F. Zhang, Xi. Wu, S. Zhang, GPS location accuracy improvement by WLAN. Comput. Modell. New Technol. **17**(4), 224–228 (2013)
2. X. Wang, M.J. Rahman, S.I. Park, H.M. Kim, Y. Wu, A new positioning system using DVB-T2 transmitter signature waveforms in single frequency networks. IEEE Trans. Broadcast. **58**(3), 347–359 (2012)
3. P. Thevenon, O. Julien, C. Macabiau, D. Serant, L. Ries, S. Corazza, M. Bousquet, Positioning principles with a mobile TV system using DVB-SH signals and a single frequency network, in *l6th International Conference on Digital Signal Processing* (2009), pp. 1–8
4. X. Jinsong, L. Xiaochun, W. Haitao, et al., Seamless Indoor/Outdoor Positioning Using Digital Television Terrestrial Broadcasting (DTMB) Technology Integrated with GNSS, in *27th International Technical Meeting of the Satellite-Division of the Institute of Navigation, ION, GNSS*, vol. 2 (2014), pp. 1646–1651
5. H. Xiang, Y. Wuwei, L. Bin, Fast RCS modeling for dynamic target tracking. Int. J. Smart Sens. Intell. Syst. **8**(4), 1956–1976 (2015)
6. S. Zhengyi, C. Changxiao, W. Honglin, Y. Hairong, Application of improved EKF algorithm in fixed single observer passive localization. Mod. Def. Technol. **41**(6), 122–127 (2013)
7. J.M. Pak, C.K. Ahn, P. Shi, Self-recovering extended Kalman filtering algorithm based on model-based diagnosis and resetting using an assisting FIR filter. Neurocomputing **173**, 645–658 (2016)
8. G. Bourmaud, R. Megret, A. Giremus, et al., From intrinsic optimization to iterated extended kalman filtering on lie groups. J. Math. Imag. Vis. **55**(3), 284–303 (2016)
9. G. Battistelli, L. Chisci, Stability of a consensus extended Kalman filter for distributed state estimation. AUTOMATICA **68**, 169–178 (2016)

10. W. Jun, J. Xiaoyuan, Z. Yu, D. Linglong, H. Kunling, Positioning based on the Chinese digital terrestrial television broadcasting system. J. Tsinghua Univer. (Science and Technology), **51**(6), 729–733 (2011)
11. Q. Wang, S. Li, Y. Zhao, H. Wu, et al., Indoor location fingerprinting system using DTMB signal. Lecture Notes in Electrical Engineering, vol. 342 (2015), pp. 439–447

## Author Biography

**Chengbiao Fu (1982)** Male,of han nationality,educated to Master degree, working as a lecturer, mainly engaged in wireless positioning navigation and Internet technology.

# DV-Hop Node Localization Algorithm Based on Improved Particle Swarm Optimization

**Fei Zhou and Shu Chen**

**Abstract** In order to further improve the positioning precision of the DV-Hop node localization algorithm in wireless sensor network (WSN), this paper proposes a modified particle swarm algorithm (MPSO) to optimize the result. The algorithm is based on particle swarm optimization algorithm, meanwhile overcomes the disadvantage of the PSO easily falling in local optimum. To adaptively balance local search capability and global search capability, MPSO updates each particle's inertia weight and acceleration coefficients self-adaptive, and sorts the particles. The simulation results show that the average localization error is lower than the standard DV-Hop algorithm and the DV-Hop algorithm based on PSO algorithm (PDV-Hop).

**Keywords** Wireless sensor network · DV-Hop · Particle swarm optimization · Node localization

## 1 Introduction

WSN is a self-organizing network, which consists of a large number of sensor nodes [1]. Integrating every sensor node with global position system (GPS) receiver or equivalent technology is impossible to predetermine the position of sensor nodes. Because the process of localization is restricted by WSN application, cost and energy consumption, and so on. Thus, the self-positioning of sensors is necessary for WSN [2]. Localization algorithm can be divided into two categories, one is range based, and another is range free. Range-based localization algorithm uses the distance, angle, and received signal strength information between nodes to estimate the unknown node. Well known techniques are time of arrival (TOA) [3], Time difference of arrival (TDOA), angle of arrival (AOA) [4], and received signal strength (RSS) [5]. Range-free algorithms use communication between nodes in locating the unknown nodes. Typical range-free localization algorithms include

F. Zhou · S. Chen (✉)
Fuzhou, China
e-mail: chenshu929@foxmail.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_57

centroid [6], APIT [7] and DV-Hop [8]. Although range-based techniques can achieve precise position, they require costly devices and consume high energy. So the range-free localization is pursued as a cost-effective alternative, which does not rely on any additional hardware.

DV-Hop is one of the most popular range-free localization algorithms, but the position accuracy is not precise. The localization error is largely affected by the network topology and the density of WSN node. Based on properties of the DV-Hop, two main improved schemes are proposed to reduce the localization error. In [9], it uses weighted node distances to calculate the unknown node's coordinate. In [10, 11], particle swarm optimization (PSO) is used to improve the traditional DV-Hop algorithm. But PSO algorithm has some disadvantages such as slow convergence, low precision in the late evolutionary. In this paper, a modified adaptive particle swarm optimization (MPSO) is presented. MPSO updates each particle's inertia weight and acceleration coefficients self-adaptive, and the particles are sorted. The new algorithm significantly improves the positioning accuracy of the DV-Hop algorithm.

The rest of this paper is organized as follows. Section 2 describes the original DV-Hop algorithm. In Sect. 3, particle swarm optimization algorithm is introduced. Sections 4 and 5 expound the improved algorithm and process. In Sect. 6, simulation results are shown and localization performances are discussed.

## 2 DV-Hop Algorithm

### *2.1 DV-Hop Algorithm*

The algorithm has three steps:

Step1: Calculating the minimum hop count from the unknown node to anchor nodes.

Anchor node broadcast information to the adjacent nodes through distance vector protocols, the information include anchor node identifier, the location information, hop, which is expressed as $\{x_i, y_i, h_i\}$. $h_i$ initialized to zero, adjacent nodes record the information and $h_i$ plus one when they receive the message. By these steps, they get the minimum hop count between nodes.

Step 2: Computing the distance from the unknown node to anchor nodes.

The number of hops and distances between anchor nodes is known, then the average hop distance is calculated.

$$HopSize_i = \frac{\sum_{i \neq j} \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}}{\sum_{i \neq j} h_{ij}}, \tag{1}$$

where $HopSize_i$ is the average jump distance, $h_{i,j}$ is the minimum number of hops from $i$ to $j$, $\{x_i, y_i\}$ is the coordinate of anchor node $i$, $\{x_j, y_j\}$ is the coordinate of anchor node $j$.

Step 3: Locating the unknown node by maximum likelihood estimation method or trilateral method after the distance information from the unknown nodes to at least three anchor nodes is obtained.

### 2.2 The Drawbacks of DV-Hop

The positioning accuracy is affected by many factors. On the one hand, the sparseness of local area nodes will produce bad nodes, which are easy to ignore and further affect the accuracy of positioning. On the other hand, the triangulation method is simple to calculate, but its positioning error is large. Because the selected three reference nodes are not evenly distributed and the average distance per hop is large, the positioning error of the unknown coordinates is large.

## 3 PSO Algorithm

Particle swarm optimization (PSO) is proposed by Dr. Eberhart and Dr. Kenney in 1995 [12].

The algorithm needs to initialize a population randomly. Every particle in the population is likely to be one of the solutions. The speed and position of these particles rely on its "individual optimal value" and "global optimal value" to update. Updating its speed and location can make particles move at the optimum place. The updating equation for the particles can be described as:

$$v_i(t+1) = \mathrm{w} \cdot v_i(t) + c_1 \cdot r_1(pBest_i - x_i(t)) + c_2 \cdot r_2(gBest - x_i(t)) \tag{2}$$

$$x_i(t+1) = x_i(t) + v_i(t+1), \tag{3}$$

where $v_i$ is particle's velocity; $x_i$ is particle's position; $c_1$ and $c_2$ are called learning factors or acceleration coefficients; $r_1$ and $r_2$ are random positive numbers between 0 and 1; $pBest_i$ is individual optimal solution position; $gBest$ is group optimal solution position.

## 4 Modified Adaptive Particle Swarm Optimization (MPSO)

The standard particle swarm optimization (PSO) has some defects, such as suffering from the premature convergence problem, which is easy to fall into local optima and lead to the premature convergence accuracy. In these cases, a modified adaptive particle swarm optimization (MPSO) is presented.

### 4.1 Updating Each Particle's Inertia Weight

Particle's inertia weigh is an important parameter to adjust local search capability and global search capability. Large inertia weight gives rise to strong global search capability, then small inertia weight leads to strong local search capability. So the inertia weight can be expressed as:

$$w(i) = \frac{w_{start} - w_{end}}{Iter}(Iter - i) + w_{end} rand, \tag{4}$$

where *Iter* is the maximum number of iterations, $i$ is the current number of iterations. In the algorithm implementation process, the overall decline of $w$ endures that the algorithm has a larger search in the early stage and finds the global optimal point in the later. Random number rand can take the opportunity to make the particles smaller or larger weights in the whole process of evolution, this let particles obtain the appropriate weights and prevent algorithm fitness value of stagnation phenomenon.

### 4.2 Adapting Learning Factor

Learning factor $c_1$ and $c_2$ respectively determine the particles themselves "self-awareness" and "social cognition" of the group's influence on the particles' trajectories. By adaptively adjusting the size of the learning factor, so that at the beginning of the algorithm it can keep the diversity of particles and enhance search capability, meantime it can strengthen the ability of particles to move to the optimal position to get high quality particles. The proposed acceleration coefficient can be respectively expressed as:

$$c_1 = 0.5 + 1.55\frac{\exp\frac{10(Iter - t + 1)}{Iter - 1} - 1}{\exp(10) - 1} \tag{5}$$

$$c_2 = 0.5 + 1.55\frac{\exp\frac{10(t - 1)}{Iter - 1} - 1}{\exp(10) - 1}, \tag{6}$$

where $t$ is the current number of iterations. $c_1$ is larger in the early stage, so the algorithm has a strong ability of "self-awareness" in the evolutionary process for a long time, this enhances the search area of the particles, avoid falling into local optimum; $c_2$ is larger in the late stage, so the algorithm obtain a strong ability of "social cognition", it is beneficial to local fine search, So that the algorithm converges to the global optimal solution quickly.

### 4.3 Sorting Particles

Due to particle swarm optimization (PSO) algorithm is based on simple random global optimization of intelligent algorithm, Particle evolved to a certain extent can't well keep the diversity of population, easily get into convergence. At the same time, the algorithm of each particle only depends on individual extremum *pBest* and global extremum *gBest* to update itself, and not make full use of the individual extremum of other particles. This paper proposed an improved method, after each particle updates their own position, particles sort (fitness value from small to large) according to the size of the population fitness value of particles. Thus, each particle can have a high probability to update their location by using individual optimal position of other particles, so that it obtains more useful information and information sharing between the particles of the population are more frequent and full.

## 5 DV-Hop Localization Algorithm Based on MPSO Optimization (MPDV-Hop)

The estimate location coordinate $(x, y)$ of the unknown node is obtained through the traditional DV-Hop algorithm. Then the improved algorithm uses the modified particle swarm algorithm to optimize positioning results, which can be described by six steps as follows:

Step 1: Randomly generate N flying particles, assign random initial value to each particle's speed, the location, the current individual optimal solution and group optimal solution.

Step 2: Get the unknown node to anchor node distance according to the original DV-Hop algorithm in the first two steps, nonlinear equations constructed as shown in the formula:

$$\begin{cases} \sqrt{(x_1 - x)^2 + (y_1 - y)^2} = d_1 \\ \vdots \\ \sqrt{(x_n - x)^2 + (y_n - y)^2} = d_n \end{cases} \tag{7}$$

Assuming $f_i = \sqrt{(x_i - x)^2 + (y_i - y)^2}$, so the fitness function of particles can be expressed as:

$$F = \sum_{i=1}^{n} (\alpha_i (f_i - d_i))^2, \tag{8}$$

where $\alpha_i$ is the reciprocal of $h_i$, $h_i$ is the hops between the unknown node and anchor node $i$ which was obtained in first step of DV-Hop algorithm. In the DV-Hop, the more hops unknown nodes and anchor nodes, the greater the

estimated distance of the error accumulated. So each node join a weight, the farther the distance of the anchor nodes added the smaller the weight.

Step 3: According to the value of the fitness function of each particle, we sort all the particle's position from small to large.

Step 4: Particles respectively update individual optimal value and groups optimal value of particle according to Eqs. (9) and (10).

$$pBest^{t+1} = \begin{cases} x^{t+1}, & if \quad \mathrm{F}(X^{t+1}) \le F(pBest^{t}) \\ pBest^{t}, & if \quad F(X^{t+1}) > F(pBest^{t}) \end{cases} \tag{9}$$

$$gBest^{t+1} = \begin{cases} pBest^{t+1}, & if \quad \mathrm{F}(pBest^{t+1}) \le F(gBest^{t}) \\ gBest^{t}, & if \quad F(pBest^{t+1}) > F(gBest^{t}) \end{cases} \tag{10}$$

Step 5: Particles continually update the position and speed according to the improved equation. Formula (4), (5) and (6) to (2) and (3).

Step 6: The iteration is terminated when the search value is not good enough or has already reached the maximum number of iterations.

## 6 Algorithm Simulation and Analysis of Effects

The MATLAB software is used to simulate. In the simulation, the topology of the network settings is as follows: the simulation region set in the two-dimensional of 100 m * 100 m. PSO algorithm in the parameters setting are as follows: the number of particles is 30; the number of iterations is 100; $w_{start} = 0.9$, $w_{end} = 0.4$.

The average localization error is selected as a measure of accuracy, the average localization error is defined as follows:

$$AverageError = \frac{\sum_{i=1}^{N} \sqrt{(\widehat{x_i} - x_i)^2 + (\widehat{y_i} - y_i)^2}}{N*R} *100\%, \tag{11}$$

where N is the number of all nodes; $R$ is communication radius; $(\widehat{x_i}, \widehat{y_i})$ is the estimated coordinates of the unknown node $i$; $(x_i, y_i)$ is the real coordinates of the unknown node $i$. Lower localization error of the algorithm shows better performance. The simulation results are shown in Figs. 2, 3, 4, and 5.

In simulation environment both anchor and unknown nodes are distributed randomly as shown in Fig. 1. The black dot indicates the unknown node, the red star point represents the anchor node.

MPDV-Hop localization algorithm is used to locate the unknown nodes 100. Figures 2 and 3 is respectively localization error and average localization error of each unknown node, which reflects the positioning effect.

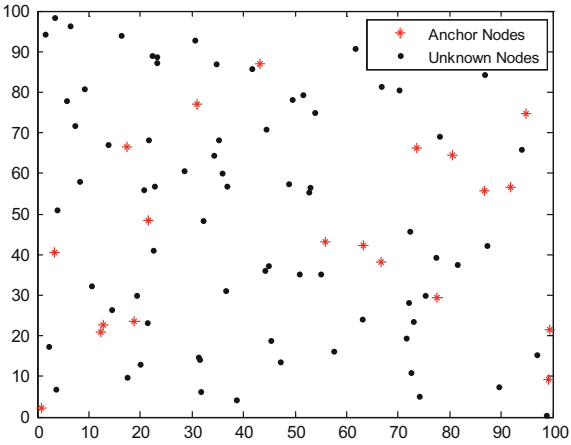



**Fig. 1** Distribution of sensor nodes in the monitoring area

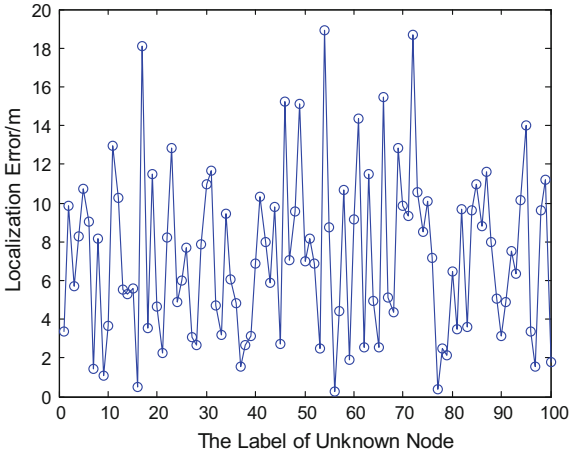



**Fig. 2** Localization error with different unknown node. (N = 120, R = 20 m)

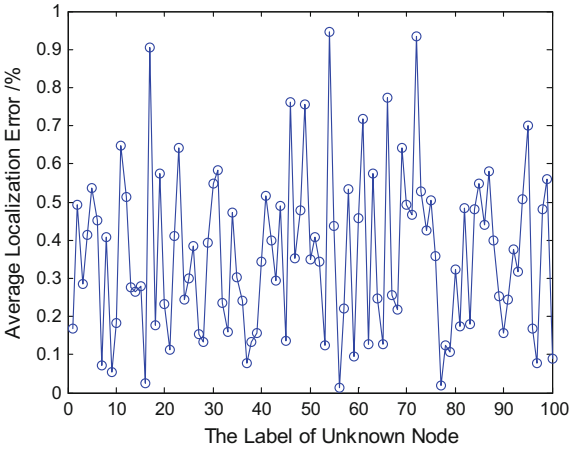



**Fig. 3** Average localization error with different unknown node. (N = 120, R = 20 m)

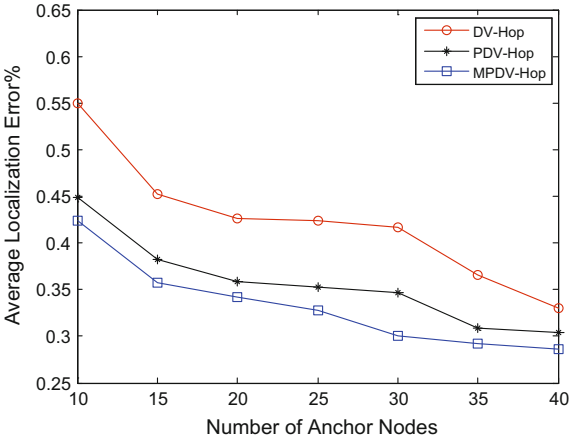



**Fig. 4** Average localization error with different number of anchor nodes. (N = 150, R = 20 m)

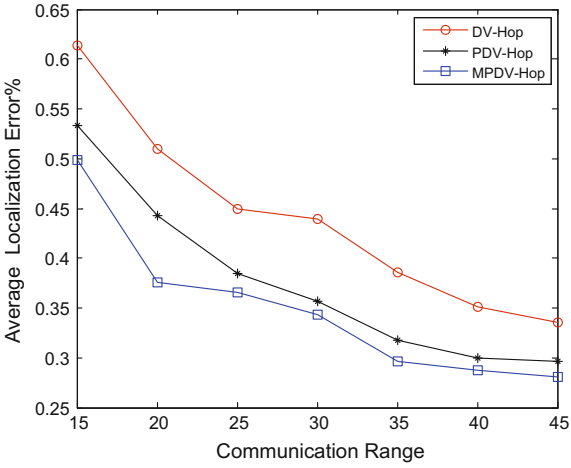



**Fig. 5** Average localization error with different communication range. (N = 150, 10% anchor nodes)

Percentage localization error has been plotted with different number of anchor nodes when the total number of nodes is 150 and communication range is set to 20 m in Fig. 4.

It can be known from Fig. 4 that with the increase of anchor nodes, the average localization error of DV-Hop, PDV-Hop and MPDV-Hop is gradually reduced, and the average localization error of MPDV-Hop is significantly less than that of DV-Hop and PDV-Hop, which means the improved algorithm has better localization accuracy.

As depicted in Fig. 5, with the increase of communication radius, the average localization error of all algorithms shows a downward trend. Compared to DV-Hop, MPDV-Hop has higher localization accuracy.

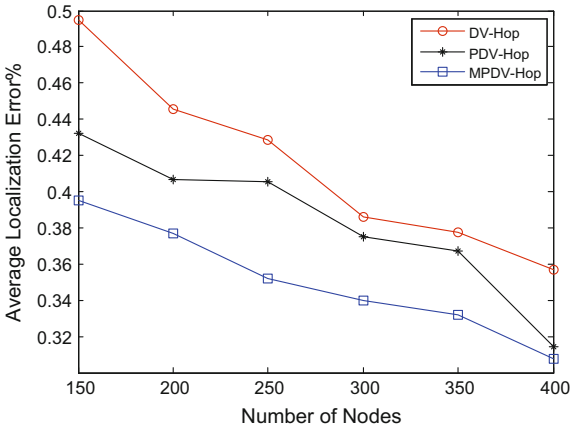



**Fig. 6** Average localization error with different number of nodes. (R = 20 m, 10% anchor nodes)

As shown in Fig. 6 number of nodes is varied from 150 to 400, and communication range is 20 m, while number of anchor nodes is 10% of total nodes. It can be seen that with the increase of nodes, the average localization error of all algorithms shows a downward trend. The average localization error of MPDV-Hop is less than that of DV-Hop, which indicates that the localization of MPDV-Hop is better than that of DV-Hop.

## 7 Conclusions

In this paper, we proposed an advanced DV-Hop algorithm by using the PSO algorithm. But PSO algorithm has some disadvantages such as slow convergence, low precision in the late evolutionary. According to these problems, a modified adaptive particle swarm optimization (MPSO) is proposed. MPSO updates each particle's inertia weight and acceleration coefficients self-adaptive, and the particles are sorted. These measures can balance local search capability and global search capability. Through simulation experiments, the improved DV-Hop algorithm based on the modified adaptive PSO algorithm for the localization coverage rate is obviously better than the traditional DV-Hop algorithm.

## References

1. A.A. Agashe, A.A. Agashe, R.S. Patil, Evaluation of DV hop localization algorithm in wireless sensor networks, in 2012 *International Conference on Advances in Mobile Network, Communication and its Applications (MNCAPPS)* (IEEE, 2012), pp. 79–82
2. X. Li, L. Yan, W. Pan et al., Optimization of DV-hop localization algorithm in hybrid optical wireless sensor networks. J. Heuristics **21**(2), 177–195 (2015)

3. S. Kumar, D.K. Lobiyal, An advanced DV-Hop localization algorithm for wireless sensor networks. Wirel. Pers. Commun. **71**(2), 1365–1385 (2013)
4. X. Chen, B. Zhang, Improved DV-Hop node localization algorithm in wireless sensor networks. Int. J. Distrib. Sens. Netw. (2012)
5. N.A.M. Maung, M. Kawai, Experimental evaluations of RSS threshold-based optimised DV-HOP localisation for wireless ad-hoc networks. Electron. Lett. **50**(17), 1246–1248 (2014)
6. S. Gayan, D.H. Dias, Improved DV-hop algorithm through anchor position re-estimation, in *2014 IEEE Asia Pacific Conference on Wireless and Mobile* (IEEE, 2014), pp. 126–131
7. W. Dargie C. Poellabauer, *Fundamentals of Wireless Sensor Networks* (Wiley, 2010)
8. Z. Zhou, M. Xiao, L. Liu et al., An improved DV-HOP localization algorithm, in *2009 Second International Symposium on Information Science and Engineering (ISISE)* (IEEE, 2009), pp. 598–602
9. H. Chen, K. Sezaki, P. Deng et al., An improved DV-Hop localization algorithm for wireless sensor networks, in *3rd IEEE Conference on Industrial Electronics and Applications, 2008. ICIEA 2008*, (IEEE, 2008), pp. 1557–1561
10. F. Zhang, Positioning research for wireless sensor networks based on PSO algorithm. Elektronika ir Elektrotechnika **19**(9), 7–10 (2013)
11. Q. Zhang, M. Cheng, A node localization algorithm for wireless sensor network based on improved particle swarm optimization, in *Mechatronics and Automatic Control Systems* (Springer International Publishing, 2014), pp. 135–144
12. R.C. Eberhart, J. Kennedy, A new optimizer using particle swarm theory, in *Proceedings of the Sixth International Symposium on Micro Machine and Human Science*, vol. 1 (1995), pp. 39–43

# Part VI
# Optical Communication

# Inductorless SiGe BiCMOS Optical Receiver Front End for 25 Gb/s Optical Links

**Jingqiu Wang, Fujiang Lin, Liang Chen and Qiwei Song**

**Abstract** A novel inductorless optical receiver analog front end (AFE) design is demonstrated to require less chip area and is suitable for both low cost and high-speed optical communication applications. The optimized transimpedance amplifier (TIA) has a differential regulated cascode (RGC) topology, with a novel zero-pole canceling technique. The proposed limiting amplifier (LA) using Cherry–Hooper topology and negative Miller capacitance broaden the bandwidth. Based on the IBM 7WL 0.18 μm SiGe BiCMOS process, the post-simulation results show a total transimpedance gain of 107.1 dBΩ and −3 dB bandwidth of 17 GHz. The chip consumes 132 mW power dissipation from a single 3.3 V supply and occupies the core area of only 110 × 340 $\mu m^2$.

**Keywords** SiGe BiCMOS · Optical receiver · Inductorless · Zero-pole canceling · Negative miller capacitance

## 1 Introduction

The required data traffic is rapidly increasing in the Internet, supercomputing, and data centers. For these applications, existing electrical communications face severe performance limitation due to signal integrity and jitter, density, cross-talk noises, and power dissipation. Fiber-optic techniques are gaining popularity as they can

J. Wang · F. Lin
School of Information Science and Technology, University of Science and Technology of China, Hefei 230026, China

J. Wang · L. Chen · Q. Song (✉)
National ASIC Design Engineering Center, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
e-mail: sqw6307@tju.edu.cn

Q. Song
School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_58

solve above problems. Furthermore, silicon photonics interconnect techniques are emerging as a major driving force because they can fabricated on bulk silicon platforms [1, 2].

An essential electronic building block is the optical receiver in optical links, and its performance can affect the whole optical interconnect systems. Transimpedance amplifier (TIA) and limiting amplifier (LA) are the two critical blocks in the optical receiver. Parasitic capacitances of transistors greatly reduce the bandwidth [3], several hybrid optical integration technologies such as inductive peaking and capacitive degeneration have been widely explored to demonstrate coherent wideband optical receivers [4]. Furthermore, we have realized SiGe BiCMOS RGC TIA with one on-chip spiral inductor, which achieves 20 Gb/s [5]. However, using inductors result in the chip size dramatically increases, also the magnetic cross talk increases through the inductors resulting in performance degradation of the clock and data recovery (CDR) [6]. Therefore, inductive peaking technology can be realized with active devices if the area is critical.

This letter present a fully inductorless differential SiGe BiCMOS optical receiver AFE tailored to IEEE 100 GBASE-LR4 standard for mid-to-long-range transmissions at a channel speed of 25 Gb/s. The IC comprises a TIA with a novel capacitive emitter degeneration technology, an equalizer, two-stage LA with negative Miller compensation and a $f_T$ doubler buffer.

## 2 Proposed Circuit Implementation

### *2.1 Transimpedance Amplifier*

The TIA is the first critical block in the optical receiver AFE. The main bandwidth restriction of a conventional TIA is usually at the input node due to the large parasitic photodiode capacitance. By modifying conventional common gate (CG) input stage to regulated cascade (RGC) or common gate feedforward topology containing negative feedback, very small input impedance can be obtained to relax the gain-bandwidth tradeoff at the input node [7]. The schematic of the proposed TIA circuit is shown in Fig. 1a. Obviously, this is a fully differential RGC input stage. However, unlike traditional structure, resistor $R_{E1}$, and capacitor $C_{E1}$ can provide an extra zero for zero-pole canceling. There are two dominant poles in this fully differential, $V_{out1}$ and $V_{out2}$, which will seriously reduce the stability and bandwidth of the amplifier. Fortunately, the two dominant poles can be canceled with zero by adjusting $R_{E1}$ and $C_{E1}$. The new dominant pole is pushed to $V_{in1}$ and $V_{in2}$. Because of this low input resistance characteristic of RGC topology, the bandwidth of the modified fully differential TIA will be higher. The detailed analysis of this TIA topology can be found in [5].

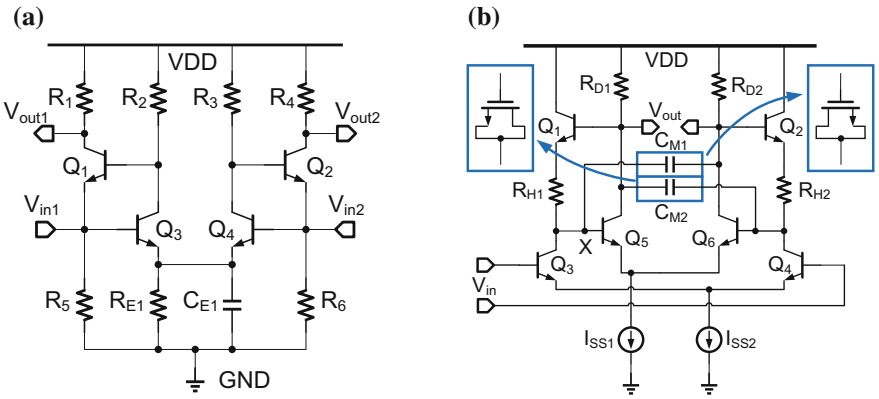



**Fig. 1** Schematic of the proposed **a** TIA and **b** LA

## 2.2 *Limiting Amplifier*

A traditional LA is difficult to achieve high gain-bandwidth product (GBW) due to the identical cascaded gain cells topology [8]. Several circuit techniques have been proposed to enhance the bandwidth of LA in the past. Cherry–Hooper amplifier with emitter–follower feedback is widely used in LA and decision circuits in photoelectronic integrated receivers. Operation in wideband communication systems requires high slew rates and thus high tail current, so the power dissipation is rather high [9]. In our design, a modified Cherry–Hooper gain cells is introduced to achieve high GBW performance [10]. The LA is a combination of a modified Cherry–Hooper amplifier with negative Miller compensation. Figure 1b shows the schematic of the proposed LA.

The $Q_1$ ($Q_2$) and $R_{D1}$ ($R_{D2}$) provide equivalent active inductor load to improve bandwidth, and the $R_{H1}$ ($R_{H2}$) can achieve higher gain performance. It is assumed that the circuit is symmetrical, so that the small-signal parameters of $Q_3$ and $Q_4$ are equal. The base–collector capacitance of the transistor $Q_3$, $C_{\mu 3}$, was reflected to the base and collector of $Q_3$ using the Miller effect. Hence, the equivalent capacitance of node X, $C_X$, are the addition of $C_{\mu 3}$ reflected to the collector of $Q_3$, the base–emitter capacitance of $Q_5$, $C_{\pi 5}$ and the collector-buck capacitance of $Q_3$, $C_{sub1}$.

Obviously, $C_X$ will create the dominate pole if the pole due to $C_{in}$ and $C_{out}$ is at a relatively high frequency. Thus, the conventional Cherry–Hooper LA topology cannot work in high speed communication systems. If we can weaken the effect of this pole, the bandwidth of the LA will be limited at higher frequencies. Fortunately, this supposition can be realized using negative Miller capacitance method. As shown in Fig. 1b, two capacitors are connected across the non-inverting nodes of the amplifier to cancel the some part of the amplifier's input capacitance. If a capacitance $C_{M1}$ ($C_{M2}$) is connected across the non-inverting nodes of an amplifier, effective miller capacitance at the input of the amplifier becomes $-C_M$ (1−A),

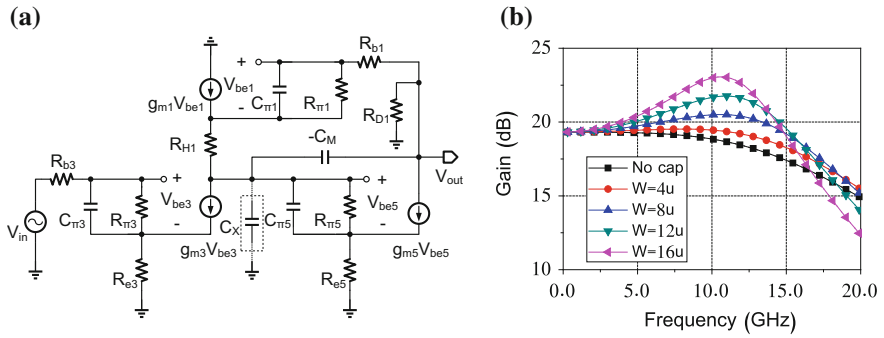



**Fig. 2** **a** Small-signal differential-mode half circuit of the proposed LA. **b** Simulated frequency response with different values of W

where A is the gain between these non-inverting nodes. Figure 2a is the small-signal differential-mode half circuit of proposed LA. In order to save the area, $C_{M1}$ and $C_{M2}$ are realized by placing NMOS devices inside an N-well, thus providing a greater fraction of the gate-oxide capacitance. Efficiency of this method depends on the cutoff frequency of capacitors and their series resistances. Therefore, the new equivalent capacitance of node X, $C_X$, new, is given

By

$$\begin{aligned} C_{X,new} &= C_{\pi 5} + \left(1 - \frac{1}{A_3}\right) C_{\mu 3} + C_{sub3} + C_{\mu 5}(1 - A_5) + [-C_M(1 - A_5)], \\ &= C_X - C_M(1 - A_5) \end{aligned} \tag{1}$$

where $A_3$ and $A_5$ are the gains across $C_{\mu 3}$ and $C_{\mu 5}$, respectively, and are given by

$$A_3 = -\frac{g_{m3}(1 + R_H \cdot g_{m1})}{g_{m1}(1 + R_D \cdot g_{m5})} \tag{2}$$

$$A_5 = g_{m5} R_D \tag{3}$$

where $g_{mn}$ is the transconductance of transistor n, $R_{\pi n}$ is the base–emitter resistance for a transistor n. Figure 2b shows the simulated frequency response of proposed LA with different values of channel width (W) of NMOS capacitor (the channel length of NMOS is fixed). It can be found that the −3 dB bandwidth of the proposed topology is 1.3 times than that without negative Miller NMOS capacitor.

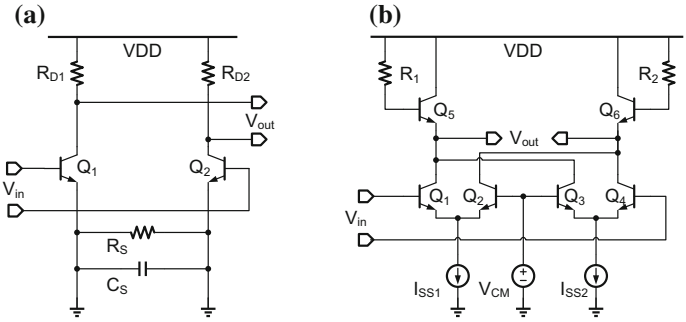



**Fig. 3** **a** Schematic of the proposed EQ and **b** Buffer

## *2.3 Equalizer and Buffer*

Figure 3a shows the schematic diagram of the equalizer (EQ) circuit includes the emitter degeneration resistor and capacitor to extend the bandwidth. The equivalent transconductance of the EQ is given by

$$G_m = \frac{g_m(R_s C_s + 1)}{R_s C_s + 1 + g_m R_s/2} \tag{4}$$

If the zero ($1/R_sC_s$) cancels the pole ($1/R_DC_L$) at the collector, the bandwidth of the EQ is extended. Where $C_L$ represents the load capacitance of EQ's output. However, emitter degeneration may cause undesired gain peaking in the frequency response. Thus, the tradeoff between gain flat degree and bandwidth should be considered.

To drive the testing instruments with an input impedance of 50 Ω, an output buffer is included in the design. To deliver a single-ended voltage swing of 0.5 V, the buffer must steer 10 mA, which requires a tail current of 20 mA. This work employs a $f_T$ doubler as the output buffer with active inductive peaking. Depicted in Fig. 5, the circuit exhibits an input capacitance roughly equal to half the base–emitter capacitance of $Q_1$ while provide the same transconductance as that of $Q_1$.

## 3 Layout and Simulation Results

In order to verify the effectiveness and feasibility of the proposed topology, we performed cadence simulation using IBM7WL 0.18 μm SiGe BiCMOS technology, which offers HBTs with a maximum transit frequency of 60 GHz. The chip layout is depicted in Fig. 4. Two on-chip 300 fF MIM capacitors are used to mimic the effect of the photodiode parasitic capacitance, and together with the parasitic capacitance of the input pad, the total input parasitic capacitance is about 0.35 pF.

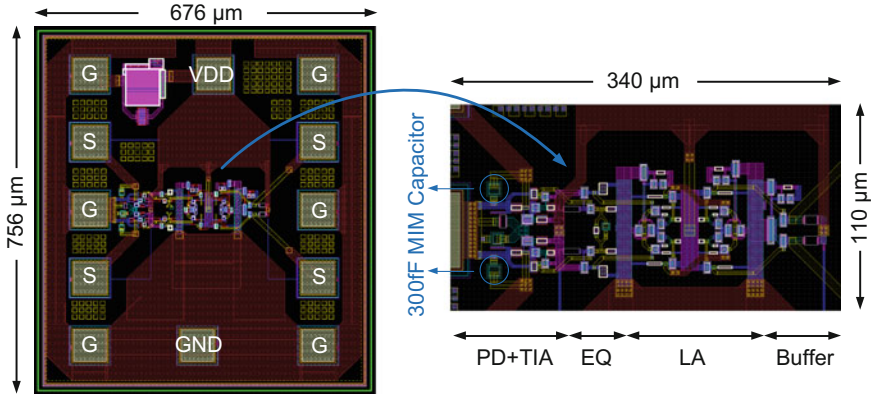



**Fig. 4** Layout of the proposed optical receiver AFE

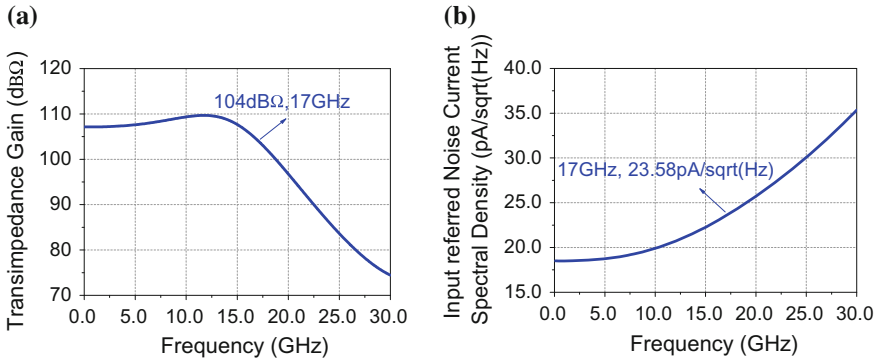



**Fig. 5** Post-layout simulated **a** differential frequency response and **b** noise response

The size of the IC is 756 × 676 µm$^2$ due to the GSGSG differential probe, but the active area occupies only 340 × 110 µm$^2$. Figure 5a shows the post-layout simulated frequency response for the proposed optical receiver AFE exhibits a transimpedance gain of 107.1 dBΩ and bandwidth of 17.03 GHz. The receiver consumes 132 mW from 3.3 V supply. Figure 5b illustrates the post-layout simulated input noise current spectral density. The simulation results show an equivalent input noise current spectral density below 23.6 pA$\sqrt{\text{Hz}}$ up to 17.03 GHz. Figure 6 shows the post-layout simulated output eye diagram with 10 µApp input current as well as $2^{31}-1$ PRBS (pseudorandom binary sequence) NRZ (non-return-to-zero) input data.

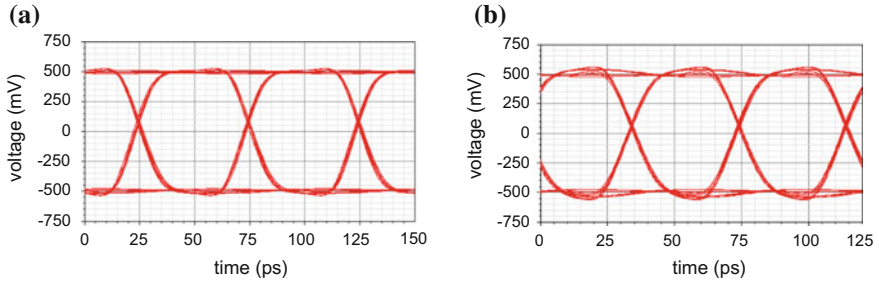



**Fig. 6** Post-layout simulated eye diagram of the differential output with 10 μApp input current at **a** 20 Gb/s and **b** 25 Gb/s

## 4 Conclusion

This paper presents a novel inductorless fully differential optical receiver AFE. It is implemented in 0.18 μm SiGe BiCMOS technology which offers HBTs with a maximum transit frequency of 60 GHz. One of the innovations is the Cherry–Hopper LA with negative NMOS Miller capacitance significantly increases the bandwidth. It is observed that optical receiver AFE achieves a high transimpedance gain of 107 dBΩ over a high bandwidth of 17 GHz in the presence of a 300 fF photodiode capacitance which is sufficient for 25 Gb/s. The optical receiver AFE occupies 340 × 110 $\mu m^2$ (without PAD) and the power dissipation is 132 mW at a 3.3 V supply. Simulation results show that the optical receiver AFE is very proficient for applications in high speed optical transceivers.

## References

1. J.-S. Youn, M.-J. Lee, K.-Y. Park, H. Rücker, W.-Y. Choi, A bandwidth adjustable integrated optical receiver with an on-chip silicon avalanche photodetector. IEICE Electron. Express. **8** (7), 404–409 (2011)
2. D. Li, G. Minoia, M. Repossi, D. Baldi, E. Temporiti, A. Mazzanti, F. Svelto, A low-noise design technique for high-speed CMOS optical receivers. IEEE J. Solid-State Circuits **49**(6), 1437–1447 (2014)
3. C. Li, S. Palermo, A low-power 26-GHz transformer-based regulated cascode SiGe BiCMOS transimpedance amplifier. IEEE J. Solid-State Circuits **48**(5), 1264–1275 (2013)
4. K. Park, B.C. Kim, B. Jung, W.-S. Oh, A 1–13 Gbps tunable optical receiver with supply voltage scaling. IEICE Electron. Express. **11**(18), 1–6 (2014)
5. Q. Song, L. Mao, S. Xie, Wideband SiGe BiCMOS transimpedance amplifier for 20 Gb/s optical links. IEICE Electron. Express. **12**(13), 1–8 (2015)
6. O. Momeni, H. Hashemi, E. Afshari, A 10-Gb/s Inductorless transimpedance amplifier. IEEE Trans. Circuits Syst.-II, Exp. Briefs **57**(12), 926–930 (2010)
7. Z. Lu, K.S. Yeo, J. Ma, M.A. Do, W.M. Lim, X. Chen, Broad-band design techniques for transimpedance amplifiers. IEEE Trans. Circuits Syst.-I, Reg. Papers **54**(3), 590–600 (2007)

8. F. Jiang, J. Chen, D. Wu, J. Wu, Z. Jin, X. Liu, A wide dynamic range transimpedance amplifier with high gain-bandwidth product for 10 Gb/s optical links, in *International Workshop on Microwave and Millimeter Wave Circuits and System Technology (MMWCST)*, vol. 1 (2012)
9. M. Maadani, M. Atarodi, A Low-area, 0.18 μm CMOS, 10 Gb/s optical receiver analog front end, in *IEEE International Symposium on Circuits and Systems* (2007), pp. 3904–3907
10. C.D. Holdenried, J.W. Haslett, M.W. Lynch, Analysis and design of HBT Cherry-Hooper amplifiers with emitter-follower feedback for optical communications. IEEE J. Solid-State Circuits **39**(11), 1959–1967 (2004)

# Nonlocal Means Denoising Based on LJS for Optical Sensing Signal

**Han Zhang, Xianyang Qian, Ling Wang, Nuoya Long, Bin Zhang and Wei Sun**

**Abstract** A robust nonlocal means with local James-Stein center pixel weight (LJSCPW) filtering algorithm has been proposed to enhance the performance of distributed optical sensing signal, and the simulation results show the proposed method has no signal distortion even if the input parameter noise level is inaccurate.

**Keywords** Distributed optical fiber sensing · Nonlocal means deionising · James-Stein estimator

## 1 Introduction

In the past decades, distributed optical fiber sensing (DOFS) has attracted the interest of both the academic and the industrial sectors for its significant advantages, such as large number of monitored points, simple deployment, over thousand sensing range, etc. [1, 2]. Brillouin optical time domain analyzer (BOTDA), which is based on Brillouin gain spectrum (BGS) measurement, has been increasing for the accurate estimation of the static strain/temperature over tens of km measurement range with metric spatial resolution [3–6].

Various technologies have been investigated to enhance the performance of DOFS, such as distributed Raman amplification [4], optical pulse coding [5], balanced detector [6], etc. Recently, nonlocal means (NLM) image denoising is an impactful and cost-effective way to intensify the response of sensing system [7, 8]. NLM processes the noisy sensing signal by averaging data from different regions in the image that has similar structures, depending on the fact that sensing signal often contains repeated patterns from in the time or frequency domain.

H. Zhang · L. Wang · N. Long
Guizhou Power Grid Information and Communication Company,
Guiyang 550003, Guizhou, China

X. Qian (✉) · B. Zhang · W. Sun
Key Laboratory of Optical Fiber Sensing & Communications (Ministry of Education),
University of Electronic Science and Technology of China, Chengdu 611731, China
e-mail: qianxianyang@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_59

However, the value of the selecting center pixel weights (CPW) in NLM is an important factor affecting the performance of signal denoising [8], and the image denoising algorithm should be optimal for DOFS. In this paper, we study the different designed CPWs for NLM denoising by BOTDA sensing system simulation. The results show that the NLM with James-Stein type CPW (LJSCPW) has better performance than other methods, and the noise is effectively removed while without signal distortion. Moreover, the NLM with the LJSCPW is more robust for the inaccurate parameter input than other methods.

## 2 Principle

Nonlocal means algorithm is essentially spatial filtering of Gaussian model, and it exploits the self-predictions and similarity of noisy signal to average data by different weights. Given a discrete noisy image, these weights are defined as [9]

$$\omega(k,l) = \frac{1}{Z(k)} \exp\left(-\frac{\|I(N_k - I(N_l))\|^2_{2,Ga}}{h^2}\right), \tag{1}$$

where $I(N_k)$ and $I(N_l)$ are the intensities of the local neighborhoods (patches), $G_a$ is a Gaussian kernel function, $Z(k) = \sum_{l \in P} \exp(\|I(N_k - I(N_l))\|^2_{2,Ga}/h^2)$ is the normalizing constant, and $h$ acts as smoothing kernel width.

We can see from Eq. (1) that when the two local patches overlap completely, the CPW in the classic NLM is unitary. But this unitary CPW is unreasonable, because the unitary value means the contribution of noisy center pixel is dominated in pixel denoising. Several other CPWs have been proposed, such as the zero CPW, the Stein CPW, the max CPW, and the LJSCPW [8].

$$cpw_l^{one} = 1 \tag{2}$$

$$cpw_l^{zero} = 0 \tag{3}$$

$$cpw_l^{max} = \max_{k \in S \setminus \{l\}} (w_{l,k}) \tag{4}$$

$$cpw_l^{LJS} = 1 - (|B| - 2)\sigma^2 / \|yb_l - z\hat{b}_l\| \tag{5}$$

where $s$ is the search window, $y\hat{b}_l$ and $z\hat{b}_l$ are two local image blocks around the $l$th pixel, and $B$ is local block regions.

These CPWs can be classified as two groups: global CPWs (2), (3) and local CPWs (4), (5). In the global case, researches use a constant CPW for all pixels, while the local CPWs vary for pixels. To further discuss the CPW problem, we separate the contributions of the non-center and of the center pixels in the considered pixel $\hat{x}_l$,

$$\hat{x}_l = \frac{W_l}{W_l + cpw}\hat{z}_l + \frac{cpw}{W_l + cpw} y_l, \tag{6}$$

where the $W_l$ is the summation of all weights except the center pixel, and $\hat{z}_l$ is the denoised pixel by using all non-center weights.

Given a prescribed local patch, the local James-Stein (LJS) CPW can be expressed as

$$p_l^{LJS} = 1 - (|B| - 2)\sigma^2 / ||yb_l - z\hat{b}_l||^2 \tag{7}$$

where $|B|$ is the local patch, is $|B|$ the number of pixels in the region, and $z\hat{b}_l = \{z\hat{b}_{l+b} | \forall b \in |B|\}$ and $yb_l = \{yb_{l+b} | \forall b \in |B|\}$ are the two local patch around the $l$th pixel.

So the considered pixel $\hat{x}_l$ can be expressed as

$$\hat{x}_l^{LJS} = (1 - p_l^{LJS})\hat{z}_l + p_l^{LJS} y_l. \tag{8}$$

## 3 Simulation and Results

We simulate the Brillouin interaction along 20 Km single-mode fiber in the ideal condition, where the Brillouin gain is unchanged along the whole fiber. And the frequency offset range is 100 MHz with 1 MHz step, while the full-widths at half-maximum (FWHM) of BGSs is set to 30 MHz. Figure 1 shows the noisy BGS, where the added noise is a zero mean Gaussian white noise, with 0.01 noise level. Every position has typical Brillouin interaction intensity for different frequency offsets, and every frequency-position point can be treated as pixel. So the 2D BGS image has $20000 \times 100$ pixels.

Then, we use NLM filtering algorithm to processing the BGS image. The similarity window is $5 \times 5$, and the search window is $15 \times 15$, while the smoothing kernel width is the 10–15 times noise level. Figure 2 shows the denoising performance with different CPWs, and the noise level is equal to the added noise level. When we used the accurate noise level as the input parameter, NLM processing can effectively remove the noise and significantly enhance the SNR of sensing signal. But in

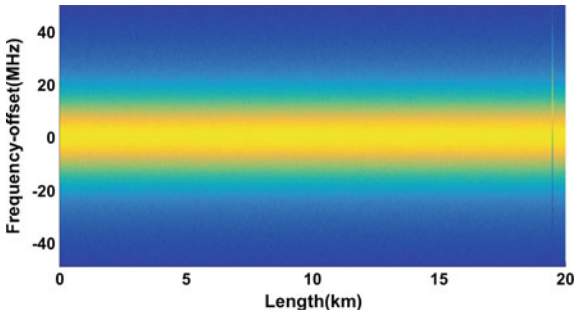



**Fig. 1** Simulation of a noisy BGS

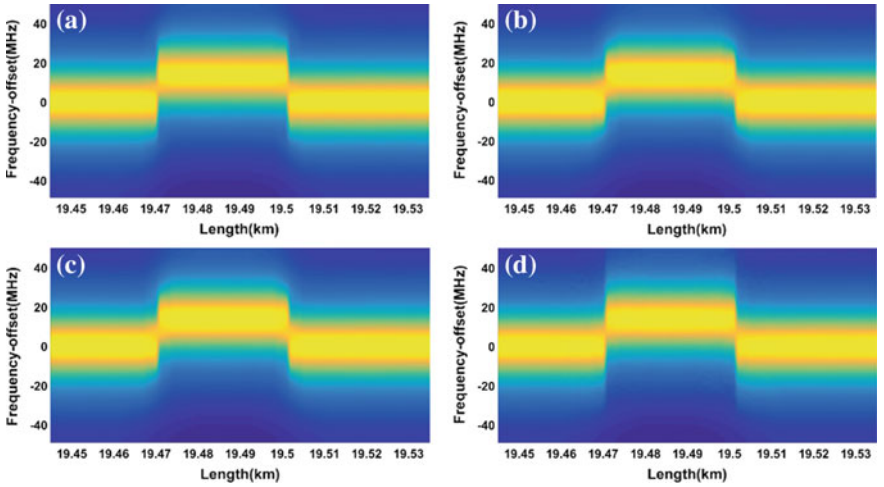



**Fig. 2** NLM denoising with different CPWs for accurate noise level input: **a** CPW is one; **b** CPW is zero; **c** CPW is max; **d** CPW is LJS

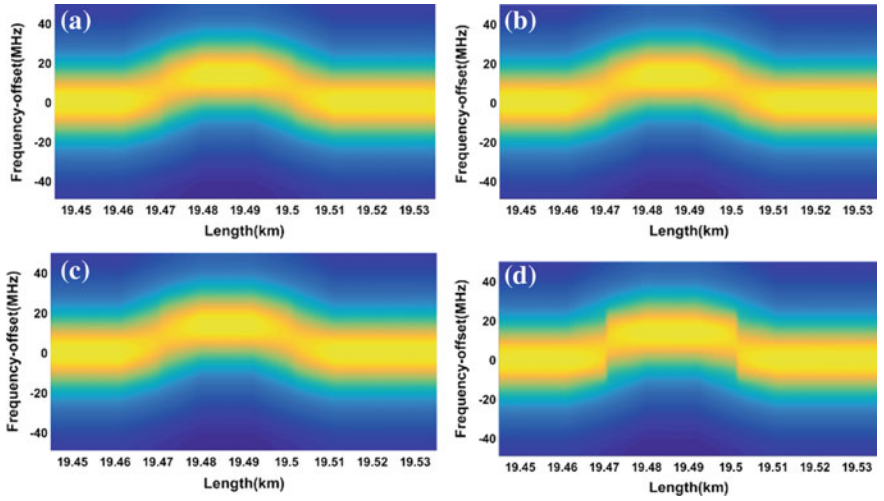



**Fig. 3** NLM denoising with different CPWs for inaccurate noise level input: **a** CPW is one; **b** CPW is zero; **c** CPW is max; **d** CPW is LJS

the case CPW is zero and max, and NLM filter induces a little signal distortion, as shown in Fig. 2b, c.

NLM is a blind image denoising method, and the noise level is unknown in practical application [10]. So we may use an inaccurate noise level as the input parameter in many conditions. Figure 3 shows the case, where the noise level is 10 times as large as the added noise level. When the noise level input is inaccurate, NLM can

also enhance the performance of BOTDA, but the filtered signal has obvious distortion except NLM with LJSCPW. As shown in Fig. 3d, the NLM with LJSCPW suppresses effectively the noise of signal while keeping the detail of perturbation.

## 4 Conclusion

In conclusion, we proposed a robust NLM filter algorithm to denoise the distributed optical sensing signal, and the NLM with LJSCPW can significantly enhance the SNR of optical sensing signal without signal distortion even when the noise level input is 10 times larger than accurate noise level.

**Acknowledgements** This work is supported by the Fundamental Research Funds for the Central Universities under grants ZYGX2015J021.

## References

1. Y. Lu, T. Zhu, L. Chen et al., Distributed vibration sensor based on coherent detection of phase-OTDR. J. Lightwave Technol. **28**(22), 3243–3249 (2010)
2. Z. Wang, J. Zeng, J. Li et al., Ultra-long phase-sensitive OTDR with hybrid distributed amplification. Opt. Lett. **39**(20), 5866–5869 (2014)
3. A. Motil, A. Bergman, M. Tur, State of the art of Brillouin fiber-optic distributed sensing. Opt. Laser Technol. **78**, 81–103 (2016)
4. M.A. Soto, X. Angulo-Vinuesa, S. Martin-Lopez et al., Extending the real remoteness of long-range Brillouin optical time-domain fiber analyzers. J. Lightwave Technol. **32**(1), 152–162 (2014)
5. X.H. Jia, Y.J. Rao, K. Deng et al., Experimental demonstration on 2.5-m spatial resolution and 1 C temperature uncertainty over long-distance BOTDA with combined Raman amplification and optical pulse coding. IEEE Photonics Technol. Lett. **23**(7), 435–437 (2011)
6. L. Zhang, Z. Wang, J. Li et al., Ultra-long dual-sideband BOTDA with balanced detection. Opt. Laser Technol. **68**, 206–210 (2015)
7. M.A. Soto, J.A. Ramrez, L. Thevenaz, Intensifying the response of distributed optical fibre sensors using 2D and 3D image restoration. Nat. Commun. **7** (2016)
8. X. Qian, Z. Wang, S. Wang et al., 157km BOTDA with pulse coding and image processing, in *Sixth European Workshop on Optical Fibre Sensors (EWOFS'2016), International Society for Optics and Photonics* (2016), pp. 99162S–99162S–4
9. Y. Wu, B. Tracey, P. Natarajan et al., James-Stein type center pixel weights for non-local means image denoising. IEEE Signal Process. Lett. **20**(4), 411–414 (2013)
10. A. Buades, B. Coll, J.M. Morel, A review of image denoising algorithms, with a new one. Multiscale Model. Simul. **4**(2), 490–530 (2005)
11. X. Liu, M. Tanaka, M. Okutomi, Single-image noise level estimation for blind denoising. IEEE Trans. Image Process. **22**(12), 5226–5237 (2013)

# Performance of Probabilistic Non-local Means on the Brillouin Optical Time Domain Analysis

**Xingjie Sa, Xianyang Qian, Baisen Li, Cheng Xiong, Bin Zhang and Wei Sun**

**Abstract** In this paper we proposed using a robust method, probabilistic non-local means (PNLM), to enhance the performance of optical fiber sensing system. The simulation results show PNLM significantly improves the signal SNR. Comparing with the traditional Non-Local Means (NLM), the PNLM filtering can effectively remove system noise without signal details distortion.

**Keywords** Probabilistic non-local means · Distributed optical fiber sensing · Non-local means

## 1 Introduction

Distributed optical fiber sensing system (DOFS) is a well-developed technology for long-range parameter monitoring, showing a promising prospect in pipeline, large-scale structures, and electric cable safety application [1, 2]. Brillouin optical time domain analysis is typical optical fiber sensing system, and it has significant advantage in long-range industrial application [3–5]. Recently, image processing methods, such as non-local means, are put forward to improve the signal-to-noise ratio (SNR) of DOFS [3, 4].But there are many works should be done to achieve optimal image denoising performance, taking the characteristic of distributed optical fiber sensing signal into consideration.

Non-local means has attracted significant attention since it was introduced by Buade et al. in the 2005 [6]. This method is proven to be effective in image denoising on a patch-by-patch basics. The NLM processes the noisy pixel by calculating the weight of their neighboring pixels, and the weight reflects the similarity between

X. Sa · B. Li · C. Xiong
Guizhou Power Grid Information and Communication Company, Guiyang 550003, Guizhou, China

X. Qian (✉) · B. Zhang · W. Sun
Key Laboratory of Optical Fiber Sensing & Communications (Ministry of Education), University of Electronic Science and Technology of China, Chengdu 611731, China
e-mail: qianxianyang@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_60

neighboring pixels and the central pixel, then these weights are utilized to estimate the value of the central pixel.

However, the NLM does not consider the impact of the noise. Probabilistic non-local means (PNLM), which takes the noise distribution into consideration, achieves better performance than classical NLM [7]. In this letter, we focus on the denoising performance of the PNLM for DOFS. As the validation of the proposed method, we simulate the Brillouin interaction along the single mode fiber. The denoising results show PNLM is more effective filter than NLM, and the proposed significantly suppress the noise while keeping the signal details.

## 2 Principle

For simple consideration, we define a clean 2D image $x = \{x(i)|i \in I\}$ on the spatial domain $I$. After the image being contaminated by zero mean Gaussian noise $n(i)$ which is identically and independently distributed (i.i.d) with variance $\sigma^2$, the image can be molded [6]:

$$y(i) = x(i) + n(i), n(i) \sim N(0, \sigma^2) \tag{1}$$

where $y(i)$ is the noisy pixel.

The NLM estimate the clean pixel $x(i)$ by using weighted average of pixels in the search area $S$, typically a square region:

$$\hat{x(i)} = \sum_{j \in S} w(i,j)y(j)/W_i \tag{2}$$

where the $W_i$ is the summation of all weights within the search area denoted as:

$$W_i = \sum_{j \in S} w(i,j) \tag{3}$$

Weight $w(i,j)$ is determined by the similarity between pixels of two similar windows, these two windows centered on $y(i)$ and $y(j)$ respectively. The function is shown as follow [6]:

$$w(i,j) = exp(-\sum_{k \in P}(y(i+k) - y(j+k))^2/)h \tag{4}$$

where the $h$ determines the rate of decay of the exponential function, and the parameter $h$ $10\sigma^2 \sim 15\sigma^2$ to achieve better performance.

However, traditional NLM denoising method does not consider the influence of noise. From now on we assume the two local patches centered on $x(i)$ and $x(j)$ match perfectly to analyze the drawback of NLM. In theory the weight $w(i,j)$ should be 1, but after corrupted by noise the weight $w(i,j)$ will be less than 1 by the function (4). So it is necessary to take account into the influence of the noise when calculating

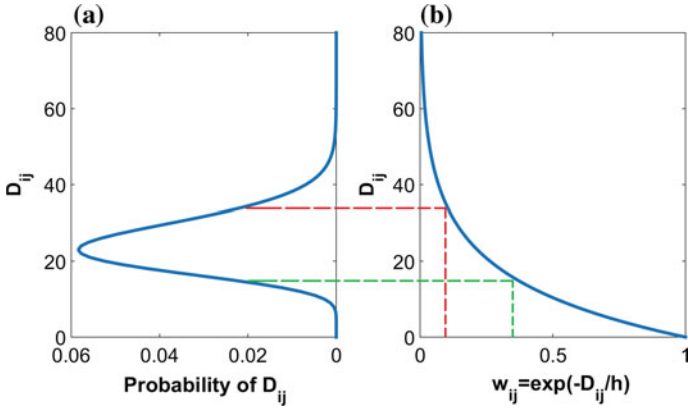



**Fig. 1** NLM weight function and the patch differences distribution of $5 \times 5$ similar window: **a** the probability distributions of $D_{i,j}$ **b** the function of NLM weight

the weight. Because the value $(y_{i+k} - y_{j+k})$ agree with Gaussian distribution with variance $2\sigma^2$ and mean 0. So we have the equation:

$$(y_{j+k} - y_{i+k})^2 / 2\sigma^2 \sim \chi^2(1) \tag{5}$$

where the $\chi(1)$ is chi-square distribution with 1° of freedom.

Figure 1 shows NLM weight function gives two different weight for the equal probable $D_{i,j}$, and it did not give largest weight for the most probable. We can see from above analysis that the NLM gives a biased weight when the image is contaminated by noise. Taking the distribution of noise into consideration, the probabilistic weight function is as follow [6]:

$$w_{i,j} = f_n(D_{i,j}) = \frac{1/2^{n/2}}{\Gamma(n/2)} D_{i,j}^{\frac{n}{2}-1} e^{-aD_{i,j}/2} \tag{6}$$

where the $f_n(D_{i,j})$ is the probability density function of chi-square distribution.

## 3 Simulation Result

We simulated BOTDA sensing system with the ideal condition, assuming the Brillouin gain is unchanged along the whole fiber. The fiber length is 100 km, and a 10 MHz Brillouin frequency shift occurred at the location 95 km while the length of the change section is 30 m. And the frequency scanning range is 100 MHz with 1 MHz step, while the full-widths at half-maximum (FWHM) of BGS is set to 30 MHz. We compare the denoising results of NLM and PNLM with the same para-

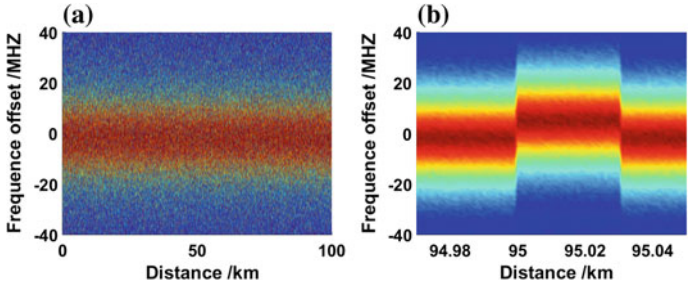



**Fig. 2** Raw data with noise level 0.02: **a** along the whole fiber; **b** around the change section

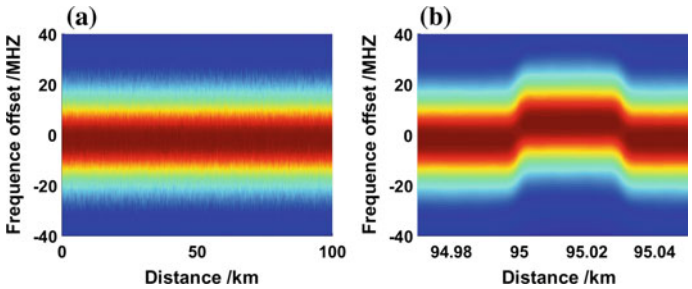



**Fig. 3** Denoising result with NLM algorithm: **a** along the whole fiber; **b** around the change section

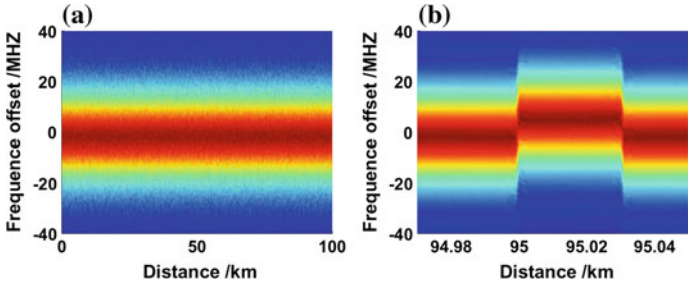



**Fig. 4** Denoising result with PNLM algorithm: **a** along the whole fiber; **b** around the change section

meters, where the search window is $21 \times 21$, the similar window is $5 \times 5$ and $h$ is $10 \times \sigma^2$. Figure 2 shows the clean BGS contaminated by zero mean Gaussian while the noise level is 0.02, and the BGS image consists of $100 \times 100000$ pixels, assuming every data-point is treated as pixel.

NLM and PNLM can both significantly remove the noise of optical sensing signal, as show in Figs. 3 and 4 respectively. Comparing the noisy sensing signal in Fig. 2, both of the two denoising methods can smooth the image, intensifying the response of signal. But the denoising detail shows the NLM processing induces sig-

**Table 1** PSNR comparisons

| PSNR\$\sigma$ | 0.01 | 0.02 | 0.04 | 0.06 | 0.08 | 0.10 |
|---|---|---|---|---|---|---|
| NLM | 32.21 | 31.03 | 29.47 | 25.48 | 23.67 | 23.08 |
| PNLM | 32.75 | 31.48 | 29.86 | 25.96 | 24.02 | 23.32 |

nal distortion around the change section, as shown in Fig. 3b. And PNLM denoising keeps a steeper edge of temperature raised than NLM, as show in Fig. 4b.

The denoising performance is then evaluated by computing the peak signal and noise ratio (PSNR). We corrupt the simulated BGS image with different noise levels, from 0.01 to 0.1, then the noisy signal is processed by NLM and PNLM. The simulation results are summarized in Table 1. It can be seen from the table that the PNLM method has higher PSNR than NLM approach in all case, and the proposed PNLM method outperforms the NLM method.

## 4 Conclusion

In this paper, we proposed using PNLM denoising algorithm to enhance the SNR of optical sensing signal. The typical DOFS, BOTDA, is implemented by simulation. The denoising performance of noisy signal shows PNLM method can effectively remove the noise without signal distortion, comparing with NLM approach.

**Acknowledgements** This work is supported by the Fundamental Research Funds for the Central Universities under grants ZYGX2015J021.

## References

1. Y. Lu, T. Zhu, L. Chen et al., Distributed vibration sensor based on coherent detection of phase-OTDR. J. Lightwave Technol. **28**(22), 3243–3249 (2010)
2. Z. Wang, J. Zeng, J. Li et al., Ultra-long phase-sensitive OTDR with hybrid distributed amplification. Opt. Lett. **39**(20), 5866–5869 (2014)
3. L. Zhang et al., Ultra-long dual-sideband BOTDA with balanced detection. Opt. Laser Technol. **68**, 206–210 (2015)
4. M.A. Soto, J.A. Ramrez, L. Thevenaz, Intensifying the response of distributed optical fibre sensors using 2D and 3D image restoration. Nat. Commun. **7** (2016)
5. X. Qian, Z. Wang, S. Wang et al., 157 km BOTDA with pulse coding and image processing, in *Sixth European Workshop on Optical Fibre Sensors (EWOFS'2016), International Society for Optics and Photonics* (2016), pp. 99162S–99162S–4
6. A. Buades, B. Coll, J.M. Morel, A review of image denoising algorithms with a new one. Multiscale Model. Simul. **4**(2), 490–530 (2005)
7. Y. Wu, B. Tracey, P. Natarajan et al., Probabilistic non-local means. IEEE Sig. Process. Lett. **20**(8), 763–766 (2013)

# Performance of Coherent FSO System Operating over Terrestrial Link

**Ming Li**

**Abstract** We studied the performance of coherent FSO system. The results showed that the terrestrial air turbulence increases the symbol-error-rate, eventually degrade the performance drastically.

**Keywords** Free-space optical communication · Coherent communication · Symbol-error-rate

## 1 Introduction

Transferring huge information data over the atmospheric media has attracted numerous attentions in both scientific and industrial communities. It is well known that the traditional radio frequency (RF) communications have been well-exploited. As a most promising alternative and complement of the high-speed RF, the free-space optical communication (FSO) has been emerged. Compared to the traditional RF, FSO system is able to offer the following certain advantages. Light wave has higher frequency than radio wave so that FSO system could provide larger information capacity; FSO system has the nature of immunity against eavesdropping because the laser beam has extremely narrow divergence in the FSO link; optical spectrum is unregulated while radio spectrum needs license; smaller size of terminals and larger gain of optical antenna make the FSO system consume less power. On the other hand, the FSO system is also highly desirable to establish the high-speed communication links for some important application scenarios. For instance, communication during the disaster recovery; information transfer in the remote areas, links between the ground station and low-orbit satellite; the interconnection of ship-to-ship, ship/ground-to-aircraft where data transfer with extremely high security is required; establishing high-speed data links in urban areas, etc.

M. Li (✉)
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
e-mail: mliece@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_61

Advances in manufacturing and exploitation of high-level modulation formats activate the development of coherent detection scheme. The coherent detection enables to suppress the impact of background noises effectively, for example, the sun, star, light from indoor and outdoor, scattering light and thermal noise. As a result, the sensitivity of the coherent detector is remarkably higher than the traditional intensity modulation direct detection scheme with the additional complex of system design. Furthermore, the coherent detection scheme uses multiple-level modulation formats so that the spectrum efficiency increases significantly. Consequently, the FSO system with coherent detect scheme has potential to realize high-speed information data transfer over the atmospheric media.

The atmosphere channel is one of the most key challenges to the coherent FSO system. The existence of temperature gradient and wide leads to air turbulence, which resulting in the fluctuations of air refractive index randomly. When the optical signals propagate in the turbulent air, their wavefront phase and amplitude vary randomly. Therefore, the phase distortions and amplitude fluctuations would be significant, which degrades the performance of coherent FSO system. Up to now, researchers have reported the influence of air turbulence on the performance of coherent FSO system. Belmonte et al. [1] developed an asymptotic expression of symbol-error-rate (SER) for the terrestrial FSO links by employing some assumptions. Based on [1], a modified SER expression has been proposed to assess the performance of coherent FSO system by taking the spatial diversity technique into account [2]. Combining the misalignment and air turbulence, a comprehensive analysis of coherent FSO system has been presented in [3]. However, accurate evaluation with respect to performance of coherent FSO system is highly limited. In this paper, the deleterious effects resulting from air turbulence are characterized by phase distortions and amplitude fluctuations. The former is produced by the use of Monte-Carlo phase screen method; while the latter is calculated according to log-amplitude fluctuation variance. We consider the coherent FSO system with QPSK (i.e. $M = 4$). Subsequently, the performance is evaluated with calculated phase and log-amplitude fluctuation variances.

## 2 Deleterious Effects Through Terrestrial FSO Link

As we mentioned, the air turbulence causes the random fluctuations of air refractive index. After optical signals propagate through the FSO link, their wavefront phase distortions as well as amplitude fluctuations occur, thus leading to degradation of coherent FSO performance. Fortunately, these random fluctuations obey certain statistics. The modified Von Karman spectrum model is widely used to reflect the terrestrial air turbulence. With this model, the power spectrum density (PSD) function of air refractive index fluctuations is written as

$$\Phi_n = 0.033 C_n^2 \frac{\exp(-\kappa^2/\kappa_m^2)}{\left(\kappa^2 + \kappa_0^2\right)^{11/6}}, \quad \text{for } 0 \le \kappa < \infty, \tag{1}$$

where $\kappa$ is spatial frequency; $\kappa_m = 5.92/l_{\text{inner}}$ and $\kappa_0 = 2\pi/L_{\text{outer}}$; $l_{\text{inner}}$ and $L_{\text{outer}}$ represent the inner and outer scale of turbulence, respectively; $C_n^2$ is crucial for the FSO link, and referred as atmospheric refractive index structure parameter with unit of $\text{m}^{-2/3}$. Based on Eq. (1), the phase PSD function can be obtained simply by multiplying the factor $2\pi k^2 L$, thus

$$\Phi_\phi(\kappa) = 0.49 r_0^{-5/3} \frac{\exp(-\kappa^2/\kappa_m^2)}{\left(\kappa^2 + \kappa_0^2\right)^{11/6}}, \quad \text{for } 0 \le \kappa < \infty, \tag{2}$$

where $r_0 = (0.423 k^2 C_n^2 L)^{-3/5}$ is the Fried parameter; $L$ is the FSO link distance.

Now, we employ the well-known Monte-Carlo phase screen method to generate the phase distortions resulting from the air turbulence. The generated phase screen is actually an array of random numbers created by computer program, which has the same statistics with the Eq. (2) above.

Next, we calculate the log-amplitude fluctuation variance caused by the air turbulence. In terms of Eq. (1), its specific expression is given by

$$\delta_\chi^2 = 2\pi^2 k^2 L \int_0^1 \int_0^\infty \kappa \Phi_n(\kappa) \left[1 - \cos\left(\frac{L\kappa^2 \eta}{k}\right)\right] d\kappa d\eta, \tag{3}$$

where $\eta = z/L$. By substituting Eqs. (1) and (3) becomes

$$\delta_\chi^2(L) = 0.97 \sigma_R^2 \left[\left(1 + Q_m^{-2}\right)^{11/12} \sin\left(\frac{11}{6} \tan^{-1} Q_m\right) - \frac{11}{6} Q_m^{-5/6}\right], \tag{4}$$

where $Q_m = 35.05 L/k l_{\text{inner}}^2$, $\sigma_R^2 = 1.23 C_n^2 k^{7/6} L^{11/6}$.

## 3 Symbol-Error-Rate of Coherent FSO System

Turbulence-induced phase and amplitude fluctuations degrade the performance of coherent FSO system. We use the symbol-error-rate to measure the performance, which has the expression [4]

$$P_s = \int_0^{\infty} \int_{-\pi}^{\pi} \int_{-\pi/2}^{\pi/2-\pi/M} \frac{1}{4\pi^2\gamma_0\delta_r\delta_i} \times \exp\left[ -\frac{\left(\sqrt{\frac{\gamma}{\gamma_0}}\cos\beta - \zeta\right)^2}{2\delta_r^2} \right] \times \exp\left( -\frac{\gamma \sin^2\beta}{2\delta_i^2\gamma_0} \right) \exp\left( -\gamma \frac{\sin^2\frac{\pi}{M}}{\sin^2\theta} \right) \mathrm{d}\gamma \mathrm{d}\beta \mathrm{d}\theta, \tag{5}$$

where $\gamma$ and $\gamma_0$ are the SNRs with and without turbulence effect; $\zeta = \exp\left(-\delta_\chi^2/2\right)\exp\left(-\delta_\phi^2/2\right)$, $\delta_r^2 = \frac{1}{2N}\left[1 + \exp(-2\delta_\phi^2) - 2\exp(-\delta_\chi^2)\exp(-\delta_\phi^2)\right]$, $\delta_i^2 = \left[1 - \exp(-2\delta_\phi^2)\right]/2N$ with $\delta_\chi^2$ and $\delta_\phi^2$ denote the variances of phase and log-amplitude distortions; parameter $N$ represents the number of statistical cell with respect to the air turbulence over the aperture.

One can see from Eq. (5) that the variances $\delta_\chi^2$ and $\delta_\phi^2$ would be the essential parameters to determine the value of SER in light of the Monte-Carlo phase screen method and expression of Eq. (4) described in Sect. 2, $\delta_\chi^2$ and $\delta_\phi^2$ enable to be calculated. Subsequently, we evaluate the performance of coherent FSO system measured by SER by using Eq. (5). Note that here we pay our attention to the QPSK modulation format as an example, but the results are also appropriate to other multiple-level modulation format.

## 4 Results and Discussions

To perform the simulations, without any loss of generality, we assume that the FSO link path is 10 km; air turbulence parameters are $C_n^2 = 10^{-15}$ m$^{-2/3}$ with outer and inner scale of turbulence $L_{\text{outer}} = 5$ m, $l_{\text{inner}} = 6$ mm; the wavelength of optical signals $\lambda = 1550$ nm is adopted. In order to achieve the ensemble average of phase distortion variance, we produced 500 phase screens based on the same FSO link by using the Monte-Carlo phase screen method. Eventually, we calculate the values of SER within the concerned SNR $\gamma 0$ range [0, 30] dB, which is presented in Fig. 1. As one can see, compared to the case of turbulence-free, the values of SER are higher remarkably in the presence of air turbulence. Also, the gap of curves with and without air turbulence becomes larger as SNR $\gamma 0$ increases. It reveals that the deleterious air turbulence impacts including the phase and amplitude variances increase the SER significantly, and deteriorate the performance of coherent FSO system drastically. Therefore, effective schemes need to be employed to suppress the turbulence effects. The potential techniques have been reported. The adaptive

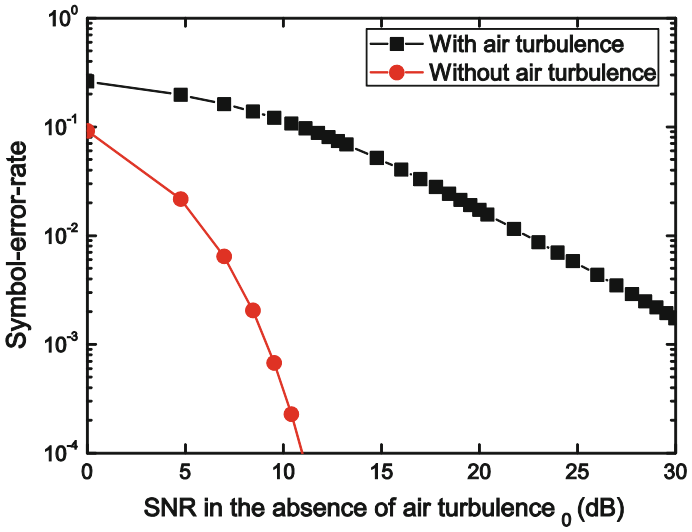



**Fig. 1** SER of coherent FSO system within the concerned $\gamma 0$ range [0, 30] dB

optics system as a formidable tool enables to correct the phase distortions [4]. The spatial diversity technique can mitigate amplitude fluctuations [5]. Gerchberg–Saxton algorithm is an effective method to compensate for turbulence effects [6], etc.

## 5 Conclusions

We have studied the performance of coherent FSO system measured by the symbol-error rate. The terrestrial air turbulence effects are characterized by phase distortions and log-amplitude fluctuations. The numerical results showed that the deleterious terrestrial air turbulence effects increase the value of SER, and deteriorate the performance of coherent FSO system significantly.

**Acknowledgements** The author acknowledges the support from the doctorial funding at Tianjin Normal University (No. 52XB1506).

## References

1. A. Belmonte, J. Khan, Performance of synchronous optical receivers using atmospheric compensation techniques. Opt. Express **16**, 14151–14162 (2008)
2. L. Zuo, Y. Ren, A. Dang, and G. Hong, Performance of coherent BPSK systems using phase compensation and diversity techniques, in *Global Telecommunications Conference* (IEEE, 2010), pp. 1–5

3. J. Park, E. Lee, C.B. Chae, G. Yoon, Impact of pointing errors on the performance of coherent free-space optical systems. IEEE Photonics Technol. Lett. **28**, 181–184 (2016)
4. M. Li, M. Cvijetic, Coherent free space optics communications over the maritime atmosphere with use of adaptive optics for beam wavefront correction. Appl. Opt. **54**, 1453–1462 (2015)
5. T. Xuan, X. Zhengyuan, Z. Ghassemlooy, Coherent polarization modulated transmission through MIMO atmospheric optical turbulence channel. J. Lightwave Technol. **31**, 3221–3228 (2013)
6. A. Jesacher, A. Schwaighofer, S. Fürhapter, C. Maurer, S. Bernet, M. Ritsch-Marte, Wavefront correction of spatial light modulators using an optical vortex image. Opt. Express **15**, 5801–5808 (2007)

# Part VII
# Digital Signal Processing

# Active Band-Stop Filter Synthesis Based on Nodal Admittance Matrix Expansion

**Lingling Tan, Yunpeng Wang and Guizhen Yu**

**Abstract** Active network synthesis is important for circuit designer to find new circuits with desired performance. In this paper, a synthesis method for synthesizing active band-stop filters is presented, which starts from voltage transfer function and linked infinity variables to describe nullors in both nodal admittance matrix (NAM) and port admittance matrix of the circuit to be synthesized. Then circuit topology is derived by nodal admittance matrix expansion. The Tow-Thomas band-stop filter circuit and Åkerberg-Mossberg band-stop filter circuit are synthesized by nodal admittance matrix expansion on the same port admittance matrix. A design example of band-stop filter verifies the effectiveness of the circuit design method.

**Keywords** Active network synthesis · Nullor · Nodal admittance matrix (NAM) expansion · Band-stop filter

## 1 Introduction

Alternative circuits are expected to be available in electronic circuit design. For the passive circuit synthesis, based on the given transfer function, design impedance or admittance can be derived and then the circuit topology as well as component values can be synthesized by expanding the input impedance or admittance, with

L. Tan · Y. Wang · G. Yu (✉)
School of Transportation Science and Engineering, Beihang University, Beijing 100191, China
e-mail: yugz@buaa.edu.cn

L. Tan
e-mail: tanlingling@buaa.edu.cn

Y. Wang
e-mail: ypwang@buaa.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_62

consideration of the transmission zeros of the transfer impedance or transfer admittance. Unfortunately, for the active circuits, systematic active circuit synthesis theory needs to be further developed. D.G. Haigh has put forward a systematic method of active circuit synthesis after years of intensive research [1–4]. Research of network synthesis can be divided into two phases. One is the approximation to find a transfer function satisfying the specifications. The other is the realization to find a circuit to realize the transfer function. Electronic circuits synthesized by the latter method, such as filters, are derived from several given requirements without considering the final topology structures of the final synthesized circuits.

In the method proposed in this paper, a $p \times p$ port admittance matrix derived by a certain transfer function is expanded to a port-equivalent $n \times n$ nodal admittance matrix, such that $n > p$. The matrix expansion process starts from introducing blank rows and columns which represent internal nodes of the nodal admittance matrix, and then nullors are added. Nodal admittance matrix expansion [5] is applied thereafter to all matrix terms until every term in the admittance matrix becomes a single admittance term. Then the added nullors are used to move the resulting admittance matrix terms to their final locations, which properly describing either floating or grounded passive terms [6]. Thus, the final nodal admittance matrix is obtained, which includes finite terms and unbounded terms representing passive circuit components and active ones, respectively.

Band-stop filters are used to filter out a single frequency or a band of frequencies, which are widely used in wireless communication system. It can restrain the stray output of high power transmitter and parasitic band of the nonlinear power amplifier or band-pass filter [7] and so on. In this paper, active band-stop filter topology structures of The Tow-Thomas band-stop filter circuit and Åkerberg-Mossberg band-stop filter circuit are synthesized using nodal admittance matrix expansion and a design example verify the effectiveness of the circuit design method.

The purpose of the present paper is to extend previous work to more complicated circuits with more than one operational amplifier and verify the feasibility of the method of the symbolic circuit design using nodal admittance matrix expansion. As extra nullors are introduced, nullators, and norators may be paired as operational amplifiers in alternative ways, then different active filter circuits can be derived from the same circuit network with nullors.

## 2 Basic Theories of Active Network Synthesis

### *2.1 Nullor*

Nullor is a universal active two-port network composed of nullator and norator, which are specified according to the constraints imposed on their terminal voltages and currents, that is:

$$\begin{bmatrix} V_1(S) \\ I_1(S) \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_2(S) \\ -I_2(S) \end{bmatrix} \tag{1}$$

Its admittance matrix is [8]

$$\begin{array}{c} \\ j \\ k \end{array} \begin{array}{c} \begin{array}{cc} g & h \end{array} \\ \begin{bmatrix} \infty_i & -\infty_i \\ -\infty_i & \infty_i \end{bmatrix} \end{array} \tag{2}$$

The matrix (2) indicates that there is a nullator between nodes $g$ and $h$ and a norator between nodes $j$ and $k$. Where $\infty_i$ is linked infinity parameter and $i$ refers to the $i$th nullor. In the case of two-port circuits with prescribed voltage or current transfer function, $\infty_i$ is also included in the port admittance matrix, thus $\infty_i$ creates the possibility of a transformation from port admittance matrix to nodal admittance matrix.

The widely application of the pathological element of nullor is for its capability to model active blocks [9–11], such as the ideal operational amplifier as shown in Fig. 1.

## 2.2 *Admittance Matrix Descriptions for Circuits with Prescribed Voltage and Current Transfer Functions*

For the 2 × 2 port admittance matrix of VCVS and CCCS, which may be expanded to the 3 × 3 or 4 × 4 matrix as shown in Table 1 [1], where $N$ and $D$, respectively, represents the numerator and denominator of the transfer function, $Q$ represents an arbitrary admittance function parameter.

Table 1 describes four different types of port admittance matrix of the voltage control voltage source(VCVS) and current control current source (CCCS), the rows and the columns of the matrices are numbered in sequence starting from 1 and we assume that nodes 1 and 2 are input and output node, respectively. The following criterion explains which type of matrix we can choose to start the expansion: for

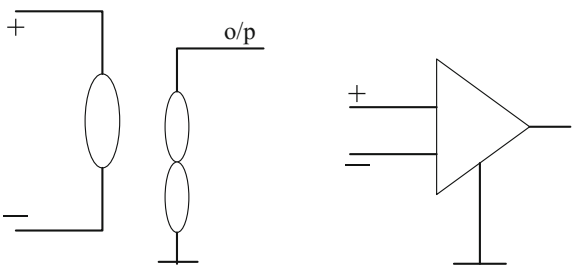



**Fig. 1** Nullor equivalent for ideal operational amplifier

**Table 1** Port admittance matrix descriptions of VCVS and CCCS

| | Type I | Type II | Type III | Type IV |
|---|---|---|---|---|
| VCVS<br>$A = A_v = N/D$ | $\begin{bmatrix} 0 & 0 & 0 \\ 0 & \infty_1 & -\infty_1 \\ -N & 0 & D \end{bmatrix}$ | $\begin{bmatrix} 0 & 0 & 0 \\ \infty_1 & 0 & -\infty_1 \\ 0 & -D & N \end{bmatrix}$ | $\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \infty_1 \\ -N & -D & Q \end{bmatrix}$ | $\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \infty_1 & -\infty_1 \\ -N_1 & -D_1 & P_1 & 0 \\ -N_2 & -D_2 & 0 & P_2 \end{bmatrix}$ |
| CCCS<br>$A = A_i = N/D$ | $\begin{bmatrix} \infty_1 & 0 & 0 \\ 0 & 0 & -N \\ -\infty_1 & 0 & D \end{bmatrix}$ | $\begin{bmatrix} 0 & 0 & -D \\ \infty_1 & 0 & 0 \\ -\infty_1 & 0 & N \end{bmatrix}$ | $\begin{bmatrix} 0 & 0 & -D \\ 0 & 0 & -N \\ \infty_1 & 0 & Q \end{bmatrix}$ | $\begin{bmatrix} 0 & 0 & -D_1 & -D_2 \\ 0 & 0 & -N_1 & -N_2 \\ \infty_1 & 0 & P_1 & 0 \\ -\infty_1 & 0 & 0 & P_2 \end{bmatrix}$ |
| $A$ | $N/D$ | $N/D$ | $-N/D$ | $\dfrac{N_1P_2 - N_2P_1}{D_2P_1 - D_1P_2}$ |

type I, the terms in the numerator of the transfer function are included in the denominator and vice avers; for type II, the numerator and the denominator terms are positive; for type III, the numerator and denominator terms are independent and the transfer function is negative; for type IV, the numerator and denominator have a combination of positive and negative terms. Through adding nullors and applying the pivotal expansion in the port admittance matrix, circuits with more operational amplifiers can be obtained [12].

### *2.3 Matrix Port-Equivalence Transformations [13]*

Arbitrary element theorem: for single instance of a $\infty_i$, arbitrary finite elements may be added to any position in the row and column which the $\infty_i$ exists; for the instance of $\pm\ \infty_i$ in the same row or column, arbitrary finite elements may be added anywhere in the row or column that $\pm\ \infty_i$ exists.

Element shift theorem: for a pair of $\pm\ \infty_i$ occurs in a row or column, other elements may be moved between the corresponding columns or rows. In addition, elements in a row or column of an admittance matrix which are corresponding to that of an internal node of an admittance matrix may be scaled by any factor and added to any other row or column, respectively.

## 3 The Method of Active Network Synthesis

The synthesis of passive-RC circuits can be expanded to that of active RC circuits, the process of which consists of the following six steps [14]:

(1) Choose a suitable matrix from Table 1 according to the given transfer function;
(2) Determine each element of the starting matrix referring to the transfer function;
(3) Carry out expansion of *N, D, P* or *Q* terms until all terms become 1st-order admittance functions;
(4) Introducing missing terms or shift original matrix terms according to the arbitrary element theorem and element shift theorem;
(5) If all matrix elements now correctly describe passive elements, the circuit is a single nullor circuit. Otherwise:
(6) Add extra nullors to introduce missing matrix terms.

## 4 Active Band-Stop Filter Synthesis Using Nodal Admittance Matrix Expansion

The transfer function of a band-stop filter may be in the form

$$A_v = -\frac{k(s^2+\omega_z^2)}{s^2+\frac{\omega_0}{Q_0}s+\omega_0^2} \tag{3}$$

Assuming the elements $C_1$, $C_2$, $C_3$, $g_1$, $g_2$, $g_3$, $g_4$, $g_5$, $g_6$ are used to build the band-stop filter, where $g_1$ is used for tuning of $\omega_z$, $g_2$, $g_3$, and $g_4$ may be used to form the gain of the amplifier, $g_6$ is used for tuning of $\omega_0$, $g_5$ is used for tuning of $Q_0$. The transfer function is then in the form

$$\begin{aligned} A_v &= -\frac{C_1C_3s^2+g_1g_4\frac{g_2}{g_3}}{C_1C_2s^2+C_1g_5s+g_4g_6\frac{g_2}{g_3}} \\ &= -\frac{g_3C_1C_3s^2+g_1g_4g_2}{g_3C_1C_2s^2+g_3C_1g_5s+g_4g_6g_2} \\ &= -\frac{g_3C_1C_3s^2+g_1g_4g_2}{g_3C_1s(C_2s+g_5)+g_4g_6g_2} \end{aligned} \tag{4}$$

For simplicity, let $a$ stand for $g_1$, $b$ stand for $g_2$, $c$ stand for $g_3$, $d$ stand for $g_4$, $e$ stand for $C_2s + g_5$, $g$ stand for $g_6$, $h$ stand for $C_1s$, $i$ stand for $C_3s$, yields

$$A_v = -\frac{ihc+abd}{ehc+gbd} \tag{5}$$

From the given transfer function, type III in Table 1 is selected according to the content description in Sect. 2.2, that is,

$$A_v = -\frac{N}{D} \tag{6}$$

Allocate every terms of the matrix after choosing transfer function

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \infty_1 \\ -N & -D & Q \end{bmatrix} \longrightarrow \begin{matrix} N = ihc+abd \\ D = ehc+gbd \\ Q = (d+e+\mathrm{i})hc \end{matrix}$$
$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \infty_1 \\ -\frac{ihc+abd}{Q_1} & -\frac{ehc+gbd}{Q_1} & \frac{(d+e+i)hc}{Q_1} \end{bmatrix} \tag{7}$$

An arbitrary 1st-order function $Q_1$ is introduced to Eq. (7) in order to carry out pivotal expansion, where $Q_1 = hc$. After pivotal expansion, another function $Q_2 = b$ is introduced to the right side matrix of Eq. (8).

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & \infty_1 \\ -i-\frac{abd}{hc} & -e-\frac{gbd}{hc} & d+e+i \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \infty_1 & 0 \\ -i & -e & d+e+i & -d \\ -\frac{ab}{Q_2} & -\frac{gb}{Q_2} & 0 & \frac{hc}{Q_2} \end{bmatrix} \tag{8}$$

Carry pivotal expansion on the terms in the fourth row of the right side matrix of Eq. (8), we get Eq. (9).

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & \infty_1 & 0 \\ -i & -e & d+e+i & -d \\ -a & -g & 0 & \frac{hc}{b} \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \infty_1 & 0 & 0 \\ -i & -e & d+e+i & -d & 0 \\ -a & -g & 0 & 0 & -h \\ 0 & 0 & 0 & -c & -b \end{bmatrix} \tag{9}$$

According to the arbitrary element theorem, $\pm e$ is introduces to the row where $\infty_1$ exists. The $\pm i$ elements are moved to the first row from the zeroth row. Then the equivalent matrix of (9) is obtained as Eq. (10).

$$\begin{bmatrix} i & 0 & -i & 0 & 0 \\ 0 & e & \infty_1 - e & 0 & 0 \\ -i & -e & d+e+i & -d & 0 \\ -a & -g & 0 & 0 & -h \\ 0 & 0 & 0 & -c & -b \end{bmatrix} \tag{10}$$

Extra nullors need to be introduced since the elements $-a$, $-b$, $-c$, $-d$, $-h$, $-g$ cannot be represented by floating terms. The matrix of Eq. (11) can be obtained by introducing all zeroes in the 4th, 6th row and all zeroes in the 5th, 7th column; adding a nullator between node 5 and the grounded node, and another one between node 7 and the grounded node; adding a norator between node 4 and the grounded node and another one between node 6 and the grounded node.

$$\begin{bmatrix} i & 0 & -i & 0 & 0 & 0 & 0 \\ 0 & e & \infty_1 - e & 0 & 0 & 0 & 0 \\ -i & -e & d+e+i & -d & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \infty_3 & 0 & 0 \\ -a & -g & 0 & 0 & 0 & -h & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \infty_2 \\ 0 & 0 & 0 & -c & 0 & -b & 0 \end{bmatrix} \tag{11}$$

According to the arbitrary element theorem, Eq. (12) is obtained.

$$\begin{bmatrix} a+i & 0 & -i & 0 & -a & 0 & 0 \\ 0 & e+g & \infty_1 - e & 0 & -g & 0 & 0 \\ -i & -e & d+e+i & -d & 0 & 0 & 0 \\ 0 & 0 & -d & c+d & \infty_3 & 0 & -c \\ -a & -g & 0 & 0 & a+g+h & -h & 0 \\ 0 & 0 & 0 & 0 & -h & b+h & \infty_2 - b \\ 0 & 0 & 0 & -c & 0 & -b & b+c \end{bmatrix} \tag{12}$$

It can be seen that the circuit includes three pairs of nullors from Eq. (12). For one pair, the nullator is connected between node 3 and the grounded node, and the norator between node 2 and the grounded node. For another pair, the nullator is connected between node 5 and the grounded node, and the norator between node 6 and the grounded node. And for the third one, the nullator is connected between node 7 and the grounded node, and the norator between node 4 and the grounded node. Therefore, the new active circuit topology is obtained as shown in Fig. 2.

As assumed before, the elements $a$, $b$, $c$, $d$, and $g$ are selected as the resistor, $h$ and $i$ as the capacitor, $e$ as the combination of a resistor and a capacitor and nullors are in pair in succession from the input node to the output node, then the Tow-Thomas band-stop filter circuit shown in Fig. 3 is obtained.

Alternatively, if we reorder the nullors in Fig. 2 as the following sequence shown in Fig. 4, then the reconstructed circuit topology with nullors replaced by

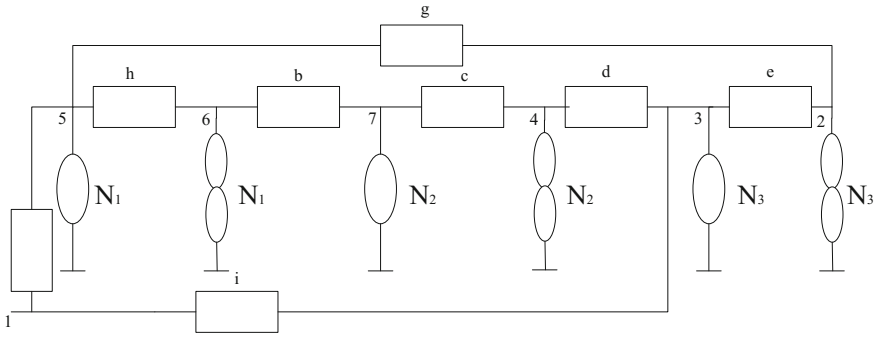



**Fig. 2** The circuit topology synthesized from Eq. (5)

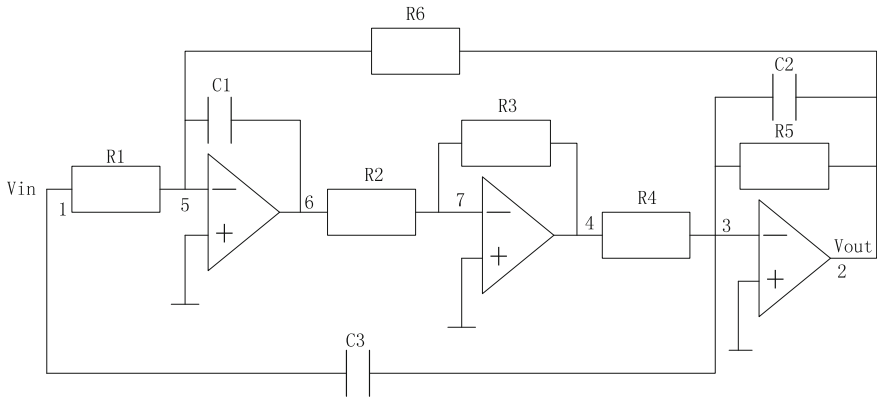



**Fig. 3** Tow-Thomas band-stop filter circuit

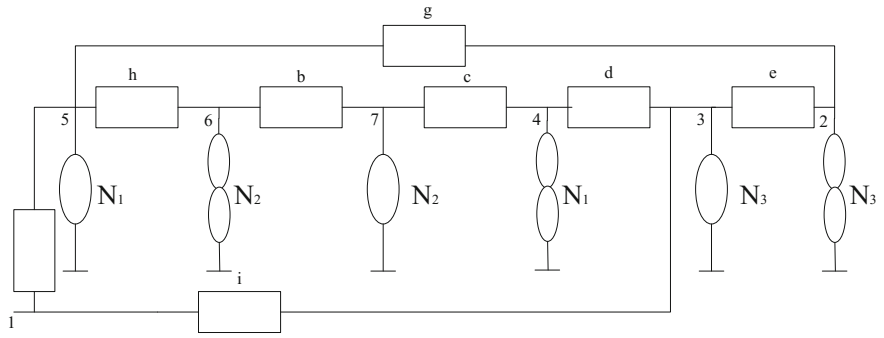



**Fig. 4** The reconstructed circuit topology

operational amplifiers is synthesized to the Åkerberg-Mossberg band-stop filter, which is shown in Fig. 5.

Band-stop filter can be obtained by Eq. (13), and a design example is provided to illustrate the design procedure. Design a band-stop filter according to the parameters below:

$$fp = 1\,\text{kHz}, Qp = 2, \quad K = 2$$

where $fp$ represents the stop-band central frequency, $Qp$ represents the quality factor, $K$ represents the passband voltage magnification.

$$A_V(S) = \frac{K(S^2 + \omega_p^2)}{S^2 + \frac{\omega_p}{Q_p}S + \omega_p^2} \tag{13}$$

In this design example, for simplicity, capacitor $C_2$ is chosen to 1 μF, and $R_2 = R_3 = R_4 = 1$ KΩ. Component values which are calculated according to Eq. (13) are listed in Table 2.

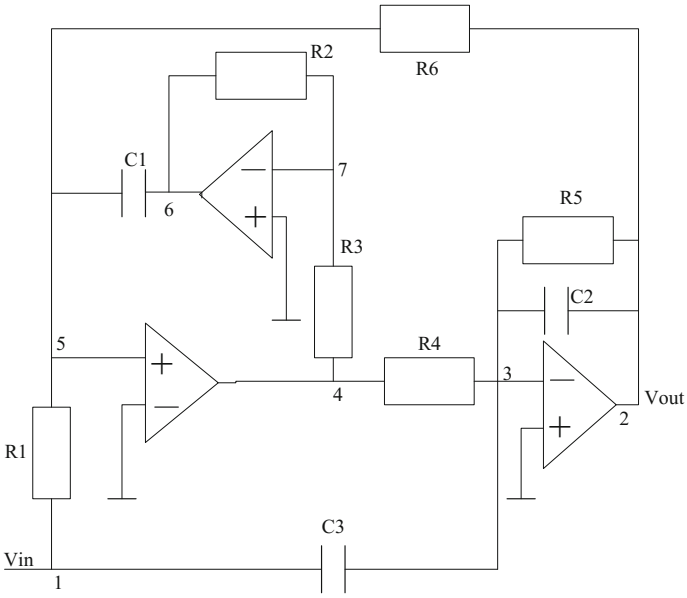



**Fig. 5** Åkerberg-Mossberg band-stop filter circuit

**Table 2** Component values of the design band-stop filter

| | | |
|---|---|---|
| $C_1 = 12.7$ nF | $R_1 = R_2 = R_3 = R_4 = 1$ kΩ | $R_5 = 318.47$ Ω |
| $C_2 = 1$ μF | $C_3 = 2$ μF | $R_6 = 2$ KΩ |

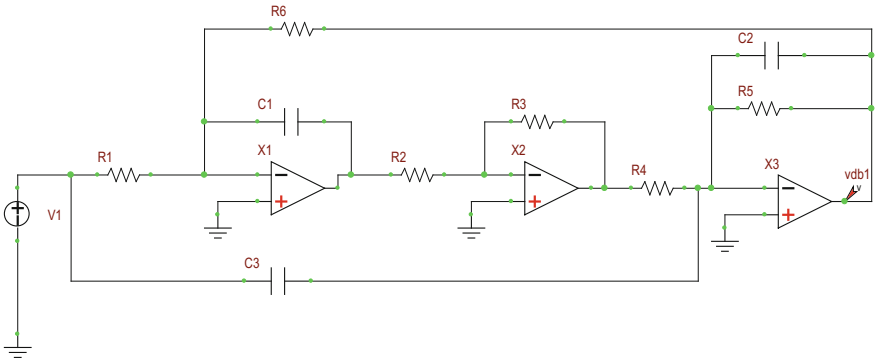



**Fig. 6** The simulation of Tow-Thomas band-stop filter

The Tow-Thomas band-stop filter and the Åkerberg-Mossberg band-stop filter are simulated according to the component values presented above, respectively, which are shown in Figs. 6 and 7.

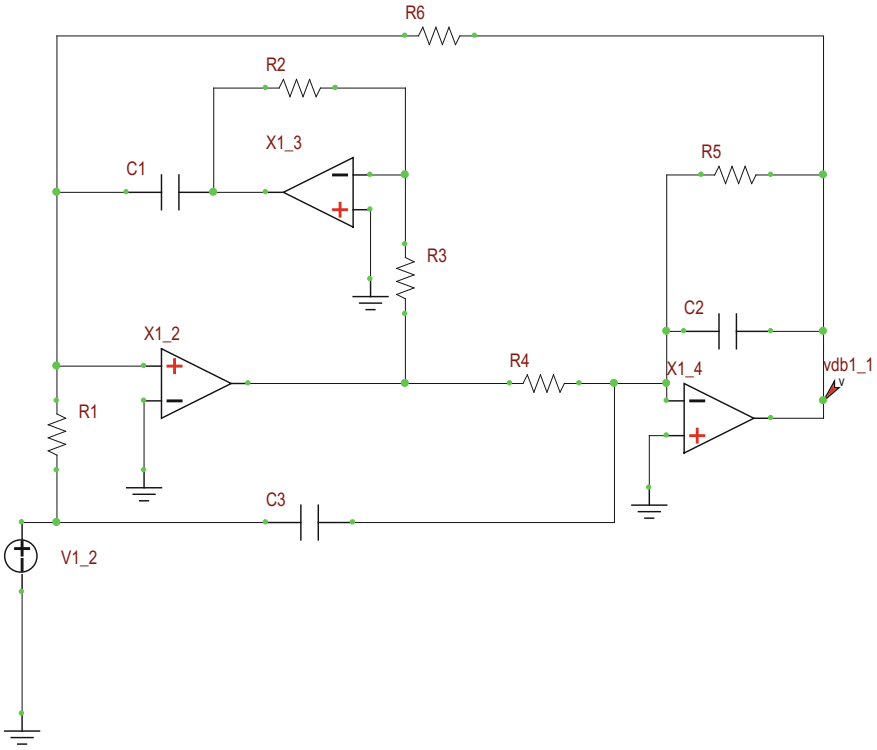



**Fig. 7** The simulation of Åkerberg-Mossberg band-stop filter

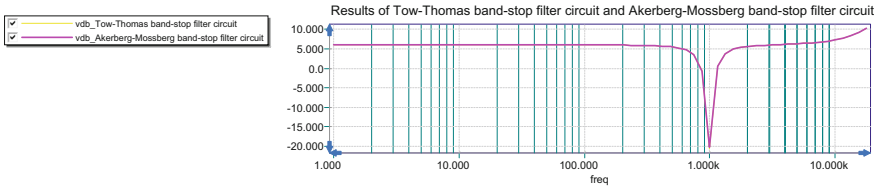



**Fig. 8** The simulation results of Tow-Thomas band-stop filter circuit and Åkerberg-Mossberg band-stop filter circuit

The simulation results are shown in Fig. 8, which explains that the results of the two band-stop filter are consistent.

## 5 Conclusions

This paper demonstrates a systematic synthesized method based on nodal admittance matrix expansion using nullors. The active circuit topologies of the Tow-Thomas band-stop filter and the Åkerberg-Mossberg band-stop filter are synthesized from the same transfer function. Unlike traditional methods of circuits design, the circuit topology is derived by the original transfer function without assumptions about the synthesized circuit topology, the method of which verify the feasibility of systematic circuit synthesis method based on nodal admittance matrix expansion. Finally, a practical band-stop filter is provided, and the simulation results verify the correctness of the proposed design method.

**Acknowledgements** This research was funded partially by the National Science Foundation of China under Grant #61371076.

## References

1. D.G. Haigh, P. Radmore et al., Systematic synthesis method for analogue circuits - Part I Notation and synthesis toolbox, in *International symposium of Circuit and System, ISCAS*, vol. 1, 23–26 May 2004, pp. 701–704
2. D.G. Haigh, T.J.W. Clarke, P.M. Radmore, Symbolic framework for linear active circuits based on port equivalence using variables. Int. Symp. Circuit Syst. **53**, 2011–2024 (2006)
3. D.G. Haigh, F.Q. Tan, C. Papavassiliou, Systematic synthesis method for analogue circuits – Part III All-transistor synthesis, in *International symposium of Circuit and System, ISCAS*, vol. 1, 23–26 May 2004, pp. 701–712
4. D.G. Haigh, F.Q. Tan, C. Papavassiliou, Systematic synthesis of active-RC circuit building-blocks. Analog Integr. Cir. Sig. Process. **43**(3), 297–315
5. R.A. Saad, A.M. Soliman, Use of mirror elements in the active device synthesis by admittance matrix expansion. IEEE Trans. Circuits Syst. I Reg. Papers **55**(9), 2726–2735 (2008)
6. A.M. Soliman, Adjoint network theorem and floating elements in the NAM. J. Circuits Syst. Comput. **18**(3), 597–616 (2009)
7. W.D. Yan, R.R. Mansour, Compact tunable bandstop filter integrated with large deflected actuators, in *IEEE/MTT-S International Microwave Symposium*, 3–8 June 2007, pp. 1611–1614
8. D.G. Haigh, P.M. Radmore, Admittance matrix models for the nullor using limit variables and their application to circuit design. IEEE Trans. Circuits Syst. I Reg. Papers **53**(10), 2214–2223 (2006)
9. L. Tan, K. Liu, Y. Bai, J. Teng, Construction of CDBA and CDTA behavioral models and the applications in current-mode symbolic circuits analysis. Analog Integr. Circuit Signal Process. (2013). doi:10.1007/s10470-013-0065-3
10. H.Y. Wang, W.C. Huang, N.H. Chiang, Symbolic nodal analysis of circuits using pathological elements. IEEE Trans. Circuits Syst. II **57**(11), 874–877 (2010)
11. Sanchez-Lopez, Fernandez, Tlelo-Cuautle, Tan, Pathological element-based active device models and their application to symbolic analysis. IEEE Trans. Circuits Syst. **58**, 1382–1395 (2011)
12. L. Tan, Y. Bai, J. Teng, K. Liu, W. Meng, Trans-impedance filter synthesis based on nodal admittance matrix expansion. Circuits Syst. Signal Process. (2012). doi:10.1007/s00034-012-9514-y

13. D.G. Haigh, Symbolic active-RC circuit synthesis by admittance matrix expansion. IEEE Trans. Circuits Syst. I Reg. Papers **1**, 248–251 (2005)
14. D.G. Haigh, A method of transformation from symbolic transfer function to active-RC circuit by admittance matrix expansion. IEEE Trans. Circuits Syst. I Reg. Papers **53**(12), 2715–2728 (2006)

# Matrix Reconstruction-Based Algorithm for Two-Dimensional Coherent DOA Estimation

**Heping Shi, Jihua Cao, Dun Liu and Hua Chen**

**Abstract** In this paper, an effective matrix reconstruction-based two-dimensional (2-D) direction-of-arrival (DOA) estimation algorithm is addressed. In the proposed algorithm, the coherency of incident signals is decorrelated through two equivalent covariance matrices, which are constructed by utilizing cross-correlation information of received data between the two parallel ULAs and the changing reference element. Then, the 2-D DOA estimation can be estimated by using eigenvalue decomposition (EVD) of the new constructed matrix. Compared with the previous works, the proposed algorithm can offer remarkably good estimation performance. In addition, the proposed algorithm can achieve automatic parameter pair-matching without additional computation. The theoretical analysis and simulation results demonstrate the effectiveness and efficiency of the proposed algorithm.

**Keywords** Matrices reconstruction · 2-D DOA estimation · Coherent signals · Decoupled estimation

## 1 Introduction

Two-dimensional (2-D) direction-of-arrival (DOA) estimation of incident coherent signals has received increasing attention in radar, sonar, and wireless communication [1–3]. Many DOA estimation methods, such as multiple signal classification (MUSIC) [4] and estimation of signal parameter via rotation invariance techniques

H. Shi (✉)
School of Automotion and Transportation, Tianjin University of Technology and Education (TUTE), Tianjin 300222, China
e-mail: shiheping@tju.edu.cn

J. Cao · D. Liu
School of Electronic Engineering, Tianjin University of Technology and Education (TUTE), Tianjin 300222, China

H. Chen
School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_63

(ESPRIT) [5], have achieved good estimation performance. However, the aforementioned methods assume the incident signals are independent, and thus these methods would encounter performance degradation due to the rank deficiency of signal covariance matrix when coherent signals exist. In order to decorrelate coherent incident signals, the spatial smoothing (SS) [6] is especially noteworthy. However, this technique generally reduces the effective array aperture and the maximum number of resolvable signals cannot exceed the number of array sensors. In [7], X. Nie et al. have introduced a computationally efficient subspace algorithm for 2-D DOA estimation with L-shaped array. In [8], an effective 2-D DOA estimation method using a sparse L-shaped array is proposed to obtain good estimation performance and less computational complexity. Some decorrelation algorithms are proposed in [9, 10] to achieve 2-D DOA estimation by utilizing two parallel ULAs. However, the limitation of the abovementioned algorithms is that the estimation performance cannot be satisfactory due to the fact that the structure of the array is not being fully exploited.

For the purpose of description, the following notations are used. Boldface italic lower/upper case letters denote vectors/matrices. $(\cdot)^*$, $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\dagger}$, and $(\cdot)^{\mathrm{H}}$ stand for the conjugation, transpose, Moore–Penrose pseudo-inverse and conjugate transpose of a vector/matrix, respectively. The notation $E(\mathrm{x})$ and $diag\ (\cdot)$ separately denote the expectation operator and the diagonal matrix, respectively.

## 2 Data Model

As illustrated in Fig. 1, the antenna array consists of two parallel ULAs ($\mathrm{X}_a$ and $\mathrm{Y}_a$) in the $x-y$ plane. Each ULA has $N$ omnidirectional sensors with spacing $d_x$, and the interelement spacing between the two ULAs is $d_y$. Suppose that $M$ far-field narrowband coherent signals impinge on the two parallel ULAs from 2-D distinct directions $(\alpha_i, \beta_i)(1 \le i \le M)$, where $\alpha_i$ and $\beta_i$ are measured relatively to the $x$ axis and to the $y$ axis, respectively.

Let the $k$th element of the subarray $\mathrm{X}_a$ be the phase reference, and then the observed signals $x_m^k(t)$ at the $m$th element can be expressed as

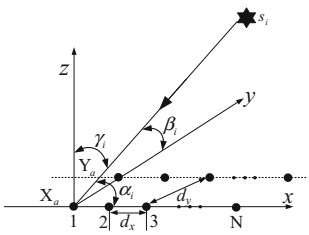



**Fig. 1** Parallel array configuration for 2-D DOA estimation

$$x_m^k(t) = \sum_{i=1}^{M} e^{-j(2\pi/\lambda)(m-k)d_x \cos\alpha_i} s_i(t) + n_{x,m}(t), \tag{1}$$

where $s_i(t)$ denotes the complex envelope of the $i$th coherent signal, $\lambda$ is the signal wavelength, and $d_x$ represents the spacing between two adjacent sensors. The superscript $k(k = 1, 2, \ldots, N)$ of the $x_m^k(t)$ stands for the number of the reference element in subarray $X_a$, and the subscript $m(m = 1, 2, \ldots, N)$ of the $x_m^k(t)$ denotes the number of the element along the $x$ positive axis in subarray $X_a$. $n_{x,m}(t)$ is the additive white Gaussian noise of the $m$th element in subarray $X_a$.

With a similar processing, employing the $k$th element of the subarray $Y_a$ as the phase reference, and then the observed signals $y_m^k(t)$ at the $m$th element can be expressed as

$$y_m^k(t) = \sum_{i=1}^{M} e^{-j(2\pi/\lambda)(m-k)d_y \cos\alpha_i} e^{j(2\pi/\lambda)d_y \cos\beta_i} s_i(t) + n_{y,m}(t). \tag{2}$$

The observed vectors $X_a$ and $Y_a$ can be written as

$$\mathbf{X}^k(t) = [x_1^k(t), x_2^k(t), \ldots, x_N^k(t)]^T \tag{3}$$

$$\mathbf{Y}^k(t) = [y_1^k(t), y_2^k(t), \ldots, y_N^k(t)]^T. \tag{4}$$

## 3 The Proposed Algorithm

For the subarray $X_a$, the autocorrelation calculation is defined as follows:

$$\begin{aligned} r_{x_m^k (x_k^k)^*}^k &= E[x_m^k(t)(x_k^k(t))^*] \\ &= \sum_{i=1}^{M} g_i(t) e^{-j(2\pi/\lambda)(m-k)d_x \cos\alpha_i} + \sigma^2 \delta(m,k), \end{aligned} \tag{5}$$

where

$$g_i(t) = \sum_{j=1}^{M} s_i(t) s_j^*(t) \tag{6}$$

$$\delta(m,k) = \begin{cases} 1, & m = k \\ 0, & m \neq k \end{cases}. \tag{7}$$

Assume that the $k$th element of the subarray $X_a$ is the phase reference. Thus, the autocorrelation vectors $\mathbf{r}^k_{\mathbf{X}^k(x_k^k)^*}$ between $\mathbf{X}^k(t)$ and the corresponding reference element $x_k^k(t)$ can be defined as follows:

$$\begin{aligned}\mathbf{r}^k_{\mathbf{X}^k(x_k^k)^*} &= E\left[\mathbf{X}^k(t)(x_k^k(t))^*\right] \\ &= \left[r^k_{x_1^k(x_k^k)^*}, r^k_{x_2^k(x_k^k)^*}, \ldots, r^k_{x_N^k(x_k^k)^*}\right]^{\mathrm{T}}.\end{aligned} \tag{8}$$

It is obvious that $N$ column vectors can be achieved as the superscript $k$ of the $\mathbf{r}^k_{\mathbf{X}^k(x_k^k)^*}$ is changed from 1 to $N$. Thus, we construct an equivalent auto-covariance matrix $\mathbf{R}_{xx}$ as follows:

$$\begin{aligned}\mathbf{R}_{xx} &= \left[\mathbf{r}^1_{\mathbf{X}^1(x_1^1)^*}, \mathbf{r}^2_{\mathbf{X}^2(x_2^2)^*}, \ldots, \mathbf{r}^N_{\mathbf{X}^N(x_N^N)^*}\right] \\ &= \begin{bmatrix} r^1_{x_1^1(x_1^1)^*} & r^2_{x_1^2(x_2^2)^*} & \cdots & r^N_{x_1^N(x_N^N)^*} \\ r^1_{x_2^1(x_1^1)^*} & r^2_{x_2^2(x_2^2)^*} & \cdots & r^N_{x_2^N(x_N^N)^*} \\ \vdots & \vdots & \ddots & \vdots \\ r^1_{x_N^1(x_1^1)^*} & r^2_{x_N^2(x_2^2)^*} & \cdots & r^N_{x_N^N(x_N^N)^*} \end{bmatrix}.\end{aligned} \tag{9}$$

Similarly as in (5), for the subarray $Y_a$, the cross-correlation calculation $\tilde{r}^k_{y_m^k(x_k^k)^*}$ can be written as

$$\begin{aligned}\tilde{r}^k_{y_m^k(x_k^k)^*} &= E[y_m^k(t)(x_k^k(t))^*] \\ &= \sum_{i=1}^{M} g_i(t) e^{-j(2\pi/\lambda)(m-k)d_x \cos\alpha_i} e^{j(2\pi/\lambda)d_y \cos\beta_i}.\end{aligned} \tag{10}$$

Then, the cross-correlation vectors $\tilde{\mathbf{r}}^k_{\mathbf{Y}^k(x_k^k)^*}$ between $\mathbf{Y}^k(t)$ and the reference element $x_k^k(t)$ in subarray $X_a$ can be expressed as

$$\begin{aligned}\tilde{\mathbf{r}}^k_{\mathbf{Y}^k(x_k^k)^*} &= E\left[\mathbf{Y}^k(t)(x_k^k(t))^*\right] \\ &= \left[\tilde{r}^k_{y_1^k(x_k^k)^*}, \tilde{r}^k_{y_2^k(x_k^k)^*}, \ldots, \tilde{r}^k_{y_N^k(x_k^k)^*}\right]^T.\end{aligned} \tag{11}$$

Obviously, we can obtain another $N$ column vectors when the superscript $k$ of the $\tilde{\mathbf{r}}^k_{\mathbf{Y}^k(x_k^k)^*}$ is varied from 1 to $N$. Based on the $N$ column vectors, an equivalent cross-covariance matrix $\mathbf{R}_{yx}$ can be given by

$$\begin{aligned}\mathbf{R}_{yx} &= \left[\tilde{\mathbf{r}}^1_{\mathbf{Y}^1(x_1^1)^*}, \tilde{\mathbf{r}}^2_{\mathbf{Y}^2(x_2^2)^*}, \ldots, \tilde{\mathbf{r}}^N_{\mathbf{Y}^N(x_N^N)^*}\right] \\ &= \begin{bmatrix} \tilde{r}^1_{y_1^1(x_1^1)^*} & \tilde{r}^2_{y_1^2(x_2^2)^*} & \cdots & \tilde{r}^N_{y_1^N(x_N^N)^*} \\ \tilde{r}^1_{y_2^1(x_1^1)^*} & \tilde{r}^2_{y_2^2(x_2^2)^*} & \cdots & \tilde{r}^N_{y_2^N(x_N^N)^*} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{r}^1_{y_N^1(x_1^1)^*} & \tilde{r}^2_{y_N^2(x_2^2)^*} & \cdots & \tilde{r}^N_{y_N^N(x_N^N)^*} \end{bmatrix}.\end{aligned} \tag{12}$$

In order to obtain the final matrix form of the equivalent auto-covariance matrix $\mathbf{R}_{xx}$ as in (9), we need to further investigate the autocorrelation calculation $r^k_{x_m^k(x_k^k)^*}$ in (5):

$$\begin{aligned} r^k_{x_m^k(x_k^k)^*} &= E[x_m^k(t)(x_k^k(t))^*] \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M} s_i(t)s_j^*(t)e^{-j(2\pi/\lambda)(m-k)d_x\cos\alpha_i} + \sigma^2\delta(m,k) \\ &= \sum_{i=1}^{M}\sum_{j=1}^{M} s_i(t)s_j^*(t)e^{-j(2\pi/\lambda)(m-1)d_x\cos\alpha_i}\cdot e^{j(2\pi/\lambda)(k-1)d_x\cos\alpha_i} + \sigma^2\delta(m,k) \\ &= \boldsymbol{a}_m(\alpha)\mathbf{G}\boldsymbol{a}_k^H(\alpha) + \sigma^2\delta(m,k), \end{aligned} \tag{13}$$

where

$$\mathbf{G} = diag\begin{bmatrix} g_1(t) & g_2(t) & \cdots & g_M(t) \end{bmatrix} \tag{14}$$

$$\boldsymbol{a}_m(\alpha) = \begin{bmatrix} e^{-j(2\pi/\lambda)(m-1)d_x\cos\alpha_1} & \ldots & e^{-j(2\pi/\lambda)(m-1)d_x\cos\alpha_M} \end{bmatrix}. \tag{15}$$

It can be seen from (15) that $a_m(\alpha)$ is the $m$th row of the steering matrix in covariance matrix with the scenario when the first element of the subarray $\mathrm{X}_a$ is set to be the reference element. According to (13), (14), and (15), Eq. (8) can be rewritten as

$$\begin{aligned}\mathbf{r}^k_{\mathbf{X}^k(x_k^k)^*} &= [r^k_{x_1^k(x_k^k)^*}, r^k_{x_2^k(x_k^k)^*}, \ldots, r^k_{x_N^k(x_k^k)^*}]^T \\ &= \mathbf{A}(\alpha)\mathbf{G}a_k^H(\alpha) + \sigma^2\delta(m,k),\end{aligned} \tag{16}$$

where $\mathbf{A}(\alpha) = [a(\alpha_1) \quad a(\alpha_2) \quad \cdots \quad a(\alpha_M)]$ is the steering matrix of the covariance matrix along the subarray $\mathrm{X}_a$, and $a(\alpha_i) = \begin{bmatrix} 1 & e^{-j(2\pi/\lambda)d_x\cos\alpha_i} & \ldots & e^{-j(2\pi/\lambda)(N-1)d_x\cos\alpha_i} \end{bmatrix}^T$.

Based on (16), the equivalent auto-covariance matrix $\mathbf{R}_{xx}$ in (9) can be rewritten as

$$\begin{aligned}\mathbf{R}_{xx} &= \left[\mathbf{r}^1_{\mathbf{X}^1(x_1^1)^*}, \mathbf{r}^2_{\mathbf{X}^2(x_2^2)^*}, \ldots, \mathbf{r}^N_{\mathbf{X}^N(x_N^N)^*}\right] \\ &= \mathbf{A}(\alpha)\mathbf{G}\mathbf{A}^H(\alpha) + diag[\sigma_1^2, \sigma_2^2, \ldots, \sigma_N^2],\end{aligned} \tag{17}$$

where $\sigma_i^2$ is the noise power on the $i$th element of the subarray $\mathrm{X}_a$.

Similar to the equivalent auto-covariance matrix $\mathbf{R}_{xx}$ in (17), the equivalent cross-covariance matrix $\mathbf{R}_{yx}$ in (12) can be rewritten as

$$\begin{aligned}\mathbf{R}_{yx} &= \left[\tilde{\mathbf{r}}^1_{\mathbf{Y}^1(x_1^1)^*}, \tilde{\mathbf{r}}^2_{\mathbf{Y}^2(x_2^2)^*}, \ldots, \tilde{\mathbf{r}}^N_{\mathbf{Y}^N(x_N^N)^*}\right] \\ &= \mathbf{A}(\alpha)\mathbf{\Psi}(\beta)\mathbf{G}\mathbf{A}^H(\alpha),\end{aligned} \tag{18}$$

where

$$\mathbf{\Psi}(\beta) = \begin{bmatrix} \upsilon(\beta_1) & 0 & \cdots & 0 \\ 0 & \upsilon(\beta_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \upsilon(\beta_M) \end{bmatrix} \tag{19}$$

$$\mathbf{G} = diag[\, g_1(t) \quad g_2(t) \quad \cdots \quad g_M(t) \,]. \tag{20}$$

From (17) and (18), it is easy to see that since $\alpha_i \neq \alpha_j$, $(i \neq j)$, $\mathbf{A}(\alpha)$ is a full column rank matrix with $\mathrm{rank}(\mathbf{A}(\alpha)) = M$. Similarly, since $\beta_i \neq \beta_j$, $(i \neq j)$, $\mathbf{\Psi}(\beta)$ is a full-rank diagonal matrix with $\mathrm{rank}(\mathbf{\Psi}(\beta)) = M$. According to (6) and (14), note that the incident source signals $s_i(t) \neq 0$, $(i = 1, 2 \cdots M)$, so $g_i(t) \neq 0$. As a result, $\mathbf{G}$ is a full-rank diagonal matrix, namely, $\mathrm{rank}(\mathbf{G}) = M$. That is, the coherency of incident signals is decorrelated through matrices constructing no matter whether the signals are uncorrelated, coherent or partially correlated.

From (17), we can obtain the noiseless auto-covariance matrix $\hat{\mathbf{R}}_{xx}$

$$\hat{\mathbf{R}}_{xx} = \mathbf{A}(\alpha)\mathbf{G}\mathbf{A}^H(\alpha). \tag{21}$$

The eigen-decomposition of $\hat{\mathbf{R}}_{xx}$ can be written in the form

$$\hat{\mathbf{R}}_{xx} = \sum_{i=1}^{M} \lambda_i \mathbf{U}_i \mathbf{U}_i^H, \tag{22}$$

where $\{\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M\}$ and $\{\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_M\}$ are the nonzero eigenvalues and eigenvector of the noiseless auto-covariance matrix $\hat{\mathbf{R}}_{xx}$, respectively. Then, the pseudo-inverse of $\hat{\mathbf{R}}_{xx}$ is

$$\mathbf{R}_{xx}^{\dagger} = \sum_{i=1}^{M} \lambda_i^{-1} \mathbf{U}_i \mathbf{U}_i^H. \tag{23}$$

Since $\mathbf{A}(\alpha)$ is a column full-rank matrix, Eq. (21) can be expressed as

$$\begin{aligned} &\mathbf{G}\mathbf{A}^H(\alpha)\mathbf{A}^{-1}(\alpha)\hat{\mathbf{R}}_{xx} \\ &\quad = \left(\mathbf{A}^H(\alpha)\mathbf{A}(\alpha)\right)^{-1}\mathbf{A}^H(\alpha)\hat{\mathbf{R}}_{xx}. \end{aligned} \tag{24}$$

According to (18) and (24), the equivalent cross-covariance matrix $\mathbf{R}_{yx}$ can be rewritten as

$$\begin{aligned} \mathbf{R}_{yx} &= \mathbf{A}(\alpha)\mathbf{\Psi}(\beta)\mathbf{G}\mathbf{A}^H(\alpha) \\ &= \mathbf{A}(\alpha)\mathbf{\Psi}(\beta)\left(\mathbf{A}^H(\alpha)\mathbf{A}(\alpha)\right)^{-1}\mathbf{A}^H(\alpha)\hat{\mathbf{R}}_{xx}. \end{aligned} \tag{25}$$

Then, right-multiplying both sides of (25) by $\mathbf{R}_{xx}^{\dagger}\mathbf{A}(\alpha)$, we have

$$\begin{aligned} \mathbf{R}_{yx}\mathbf{R}_{xx}^{\dagger}\mathbf{A}(\alpha) = &\mathbf{A}(\alpha)\mathbf{\Psi}(\beta)\left(\mathbf{A}^H(\alpha)\mathbf{A}(\alpha)\right)^{-1}\mathbf{A}^H(\alpha) \\ &\hat{\mathbf{R}}_{xx}\mathbf{R}_{xx}^{\dagger}\mathbf{A}(\alpha). \end{aligned} \tag{26}$$

Substituting (22) and (23) into (26) yields

$$\begin{aligned} \mathbf{R}_{yx}\mathbf{R}_{xx}^{\dagger}\mathbf{A}(\alpha) &= \mathbf{A}(\alpha)\mathbf{\Psi}(\beta)\left(\mathbf{A}^H(\alpha)\mathbf{A}(\alpha)\right)^{-1}\mathbf{A}^H(\alpha) \\ &\quad \left(\sum_{i=1}^{M}\lambda_i\mathbf{U}_i\mathbf{U}_i^H\right)\left(\sum_{i=1}^{M}\lambda_i^{-1}\mathbf{U}_i\mathbf{U}_i^H\right)\mathbf{A}(\alpha) \\ &= \mathbf{A}(\alpha)\mathbf{\Psi}(\beta)\left(\mathbf{A}^H(\alpha)\mathbf{A}(\alpha)\right)^{-1}\mathbf{A}^H(\alpha) \\ &\quad \left(\sum_{i=1}^{M}\mathbf{U}_i\mathbf{U}_i^H\right)\mathbf{A}(\alpha) \\ &= \mathbf{A}(\alpha)\mathbf{\Psi}(\beta)\left(\mathbf{A}^H(\alpha)\mathbf{A}(\alpha)\right)^{-1}\left(\mathbf{A}^H(\alpha)\mathbf{A}(\alpha)\right) \\ &= \mathbf{A}(\alpha)\mathbf{\Psi}(\beta). \end{aligned} \tag{27}$$

Notice that $\sum_{i=1}^{M}\mathbf{U}_i\mathbf{U}_i^H = \mathbf{I}$. Based on (23) and (25), a new matrix $\mathbf{R}$ can be defined as follows:

$$\mathbf{R} = \mathbf{R}_{yx}\mathbf{R}_{xx}^{\dagger}. \tag{28}$$

From (28), (27) can be further rewritten as

$$\mathbf{R}\mathbf{A}(\alpha) = \mathbf{A}(\alpha)\boldsymbol{\Psi}(\beta). \tag{29}$$

Obviously, the columns of $\mathbf{A}(\alpha)$ are the eigenvectors corresponding to the major diagonal elements of diagonal matrix $\boldsymbol{\Psi}(\beta)$. Therefore, by performing an eigen-decomposition of $\mathbf{R}$, the $\mathbf{A}(\alpha)$ and $\boldsymbol{\Psi}(\beta)$ can be obtain. Then, the DOA estimation of the coherent signals $\boldsymbol{\theta}_i = (\alpha_i, \beta_i)$ can be achieved according to $\upsilon(\beta_i) = e^{j(2\pi/\lambda)d_y \cos\beta_i}$ and $\boldsymbol{a}(\alpha_i) = [1, e^{-j(2\pi/\lambda)d_x \cos\alpha_i}, \ldots, e^{-j(2\pi/\lambda)(N-1)d_x \cos\alpha_i}]^T$ without additional computations for parameter pair-matching and 2-D peak searching.

## 4 Simulation Results

In this section, simulations are performed to ascertain the performance of the proposed algorithm. The proposed method is compared with another efficient algorithm (DMR-DOAM) in [10]. The number of sensors in each subarray is $N = 7$ with $d_x = d_y = \lambda/2$. Consider $M = 4$ coherent signals with carrier frequency $f = 900$MHz coming from $\boldsymbol{\alpha} = (75^o, 100^o, 120^o, 60^o)$ and $\boldsymbol{\beta} = (65^o, 75^o, 90^o, 50^o)$. The phases of coherent signals are $[\pi/5, \pi/3, \pi/3, \pi/3]$. Results on each of the simulations are analyzed by 1000 Monte Carlo trials. The root-mean-square error (RMSE) is defined to evaluate the performance of the proposed algorithm and DMR-DOAM algorithm:

$$\begin{aligned} \mathrm{RMSE}(\alpha) &= \sqrt{\frac{1}{1000K}\sum_{i=1}^{1000}\sum_{n=1}^{M}(\hat{\alpha}_n(i) - \alpha_n)^2} \\ \mathrm{RMSE}(\beta) &= \sqrt{\frac{1}{1000K}\sum_{i=1}^{1000}\sum_{n=1}^{M}(\beta_n(i) - \beta_n)^2}, \end{aligned} \tag{30}$$

where $\hat{\alpha}_n(i)$ and $\beta_n(i)$ are the estimate of $\alpha_n$ and $\beta_n$ for the $i$th Monte Carlo trial, respectively, and $K$ is the source number.

In the first simulation, we evaluate the performance of the two algorithms with respect to the input SNR. The number of snapshots is fixed at 1000, and the SNR varies from −10 to 10 dB. The RMSE of the DOAs versus the SNR is shown in Fig. 2. It can be seen from Fig. 2 that the proposed algorithm can provide better DOA estimation than the DMR-DOAM algorithm no matter whether the RMSE curve of the $\alpha$ or the RMSE curve of the $\beta$. The reason is that the proposed algorithm takes full advantage of all the received data of the two parallel ULAs to construct the equivalent auto-covariance matrix $\mathbf{R}_{xx}$ and cross-covariance matrix $\mathbf{R}_{yx}$, which can improve the estimation precision.

In the second simulation, we investigate the performance of the two algorithms versus the number of snapshots. The simulation conditions are similar to those in

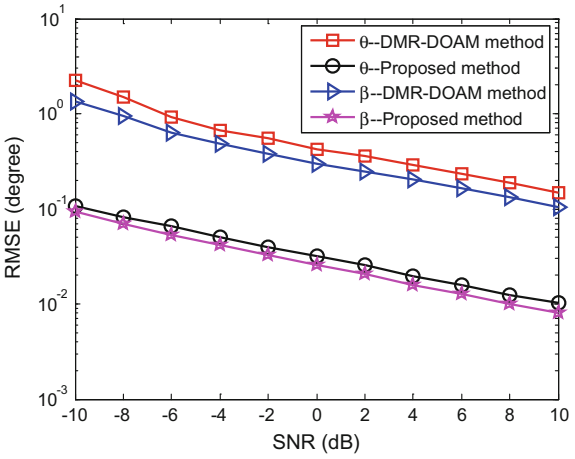



**Fig. 2** The RMSE of the DOA estimates versus input SNR

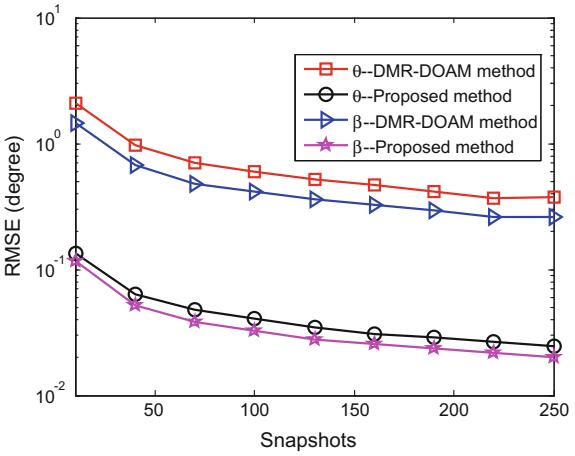



**Fig. 3** The RMSE of the DOA estimates versus input snapshots

the first simulation, except that the SNR is set at 5 dB, and the number of snapshots is varied from 10 to 250. The RMSE of the DOAs versus the number of snapshots is depicted in Fig. 3. As shown in Fig. 3, the proposed algorithm behaves better performance than the DMR-DOAM algorithm.

## 5 Conclusion

A novel decoupling algorithm for 2-D DOA estimation with two parallel ULAs has been presented. It has been shown that the proposed algorithm yields remarkably better estimation performance than the DMR-DOAM algorithm, especially in low SNR and small number of snapshots.

**Acknowledgements** This work was supported by the Key Talents Project for Tianjin University of Technology and Education (TUTE) (KYQD16001).

## References

1. H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach. IEEE Signal Process. Mag. **13**(4), 67–94 (1996)
2. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. 1–6 (2016)
3. Z. Li, Y. Chen, H. Shi, K. Liu, NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wireless Commun. Networking **2016**(48), 1–5 (2016)
4. R.O. Schmidt, Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag. **34**(3), 276–280 (1986)
5. R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Trans. Acoust. Speech Signal Process. **37**(7), 984–995 (1989)
6. N. Tayem, H.M. Kwon, L-shape 2-dimensional arrival angle estimation with propagator method. IEEE Trans. Antennas Propag. **53**(5), 1622–1630 (2005)
7. X. Nie, L.P. Li, A Computationally efficient subspace algorithm for 2-D DOA estimation with L-shaped array. IEEE Signal Process. Lett. **21**(8), 971–974 (2014)
8. J.F. Gu, W.P. Zhu, M.N.S. Swamy, Joint 2-D DOA estimation via sparse L-shaped array. IEEE Trans. Signal Process. **31**(5), 1171–1182 (2015)
9. T.Q. Xia, Y. Zheng, Q. Wan, X.G. Wang, 2-D angle of arrival estimation with two parallel uniform linear arrays for coherent signals. IEEE Radar Conf. **55**(9), 244–247 (2007)
10. L. Wang, G.L. Li, W.P. Mao, New method for estimating 2-D DOA in coherent source environment based on data matrix reconstruction data matrix reconstruction. J. Xidian Univ. **40**(2), 159–168 (2013)

# Sub-Nyquist Sampling Based on Exponential Reproducing Gabor Windows

**Wang Cheng, Chen Peng, Meng Chen and Luo Jin**

**Abstract** For analog signals comprised of several, possibly overlapping, finite duration pulses with unknown shapes and time positions, an efficient Sub-Nyquist sampling system is based on Gabor frames. To improve the realizability of this sampling system, we present alternative method for the case that the windows are exponential reproducing Gabor windows. Then, the time translation element could be realized with exponential filters. In this paper, we also construct the measurement matrix and prove that it has better coherence than Fourier matrix. Simulations prove that the sampling proposed in this paper, the sampling system require quite low sampling frequency and has nice reconstruction performance.

**Keywords** Compressed sensing · Gabor frames · Exponential reproducing windows · Sub-Nyquist sampling

## 1 Introduction

Sub-Nyquist sampling, which can acquire signals even at very low sampling rate and maintain high approximation precision, has been developed over the past years to process certain signal models. Unlike the traditional Sub-Nyquist method, Xampling originates from CS theory and ends up in some less sophisticated front-end hardware [1]. Until now, a variety of different applications of Xampling is developed for multipulse signals, such as radar signals, ultrasound signals, etc. However, all the applications are based on finite rate of innovation (FRI) signals sampling [2, 3], and the pulse shape should be a priori known. Gabor frame Sub-Nyquist sampling is proposed for making up for the weakness of FRI signals

W. Cheng · M. Chen · L. Jin
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: 32626364@qq.com

C. Peng (✉)
Institution of Mechanical Engineering, Shijiazhuang, China
e-mail: beimingke@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_64

sampling and can acquire both location and shape information for multipulse signals [4]. Unlike the FRI signals sampling, which just collects the signal Fourier coefficients and reconstruct original signal with kernel functions, it operates short-time Fourier transform on signal and collects the Gabor time-frequency coefficients.

Theoretically, all square-integrable time limited signals can be well approximated by truncated Gabor series. Under Gabor sampling scheme, the sampling rate equals $1/T$ and the sampling number is about only $4\mu^{-1}\Omega' WN_p$, where $\Omega'$ is related to the essential bandwidth of the signal and $\mu \in (0, 1)$ is the redundancy of the Gabor frames used for processing. The sampling scheme in [4] is demonstrated to possess well reconstruction performance with time domain modulation measurement functions constructed by Bernoulli random matrix and Gabor windows, such as piecewise spline or B-spline window sequences. Unfortunately, there still exists a gap between the theory and practice, because the shifting Gabor windows modulated by random measurement matrix is hard to be realized with simple circuit, and its complexity and synchronization precision also greatly affect the reconstruction performance.

The main contribution of this paper is introducing the exponential reproducing windows into Gabor sampling scheme and reducing the time domain modulation functions, constructed by appropriately weighted window sequences, to simple exponential functions. The time domain modulation functions of this study can avoid holding complicated function structures and intricate system functions, which are difficult to realize in real world. Any time domain response function from E-splines system could be defined as exponential reproducing function, the simplest one being E-spline itself [5]. For this study, we deduced the representation of the filter form of the time domain modulation functions in Gabor sampling scheme and chose the appropriate E-Splines parameter vector $\boldsymbol{\alpha}$ for guaranteeing the windows positive real functions. Then we constructed the measurement matrix for CS recovery and calculated each entry and proved that such measurement matrix has better coherence than that of DFT matrix. Next, we study the effect of frame windows width $W_g$ on the signal reconstruction performance. In [4], the frame windows width is the same as the pulse width of the signal to be measured and the Gabor frames is not quite redundant. So the sparsity $S$ of the Gabor coefficients recovered for signal representation is very small and the columns dimension to row numbers ratio $r$ of the measurement matrix is large enough to result in good RIP. Simulations prove that the sampling proposed in this paper, the sampling system requires quite low sampling frequency and has nice reconstruction performance.

## 2 Sampling System with Exponential Reproducing Windows

### 2.1 Basic Sampling System

We focus here on a class of multipulse signals, which have limited time duration and short pulses with arbitrary shapes and positions, and may even overlap. The only parameters assumed are the duration $T$, the number $N_p$ of pulses and the maximal width $W$. The multipulse signal $x(t)$, supported on $[-T/2, T/2]$, can be expressed as $x(t) = \sum_{n=1}^{N} h_n(t)$, where $\max_n |\text{supp} h_n| \le W$.

In the sampling scheme, the signal $x(t)$ enters $J \times M$ channels simultaneously. In the $(j,m)th$ channel, $x(t)$ was multiplied by function $q_{j,m}(t)$ and processed by an integrator. The structure of the scheme is shown in Fig. 1.

Matusiak and Eldar [4] set $W_g = W$, and $g(t)$ essentially bandlimited to $\left[-B_g/2, B_g/2\right]$. For some $\mu \in (0,1)$, $q_{j,m}(t) = w_j(t)s_m(t)$ is defined, where $w_j(t) = \sum_{l=-L_0}^{L_0} d_{jl} e^{-2\pi i b l t}$, $s_m(t) = \sum_{k=-K_0}^{K_0} c_{mk}\overline{g(t-ak)}$. Here, $K_0 = \lceil (T+W)/2W\mu \rceil - 1$, and $L_0 = \lceil (\Omega + B_g)W/2 \rceil - 1$, and then the numbers of time and frequency shifts are, respectively $K = 2K_0 + 1$ and $L = 2L_0 + 1$. The output of the $(j,m)th$ channel is

$$y_{j,m} = \int_{-T/2}^{T/2} x(t) q_{j,m}(t) dt = \sum_{l=-L_0}^{L_0} d_{jl} \sum_{k=-K_0}^{K_0} c_{mk} z_{kl} \tag{1}$$

Equation (1) can be written in the matrix equation as $\mathbf{Y} = \mathbf{D}\mathbf{U}^T$, $\mathbf{U} = \mathbf{C}\mathbf{Z}$ For general multipulse signals, matrix **D,** was used only to simplify hardware

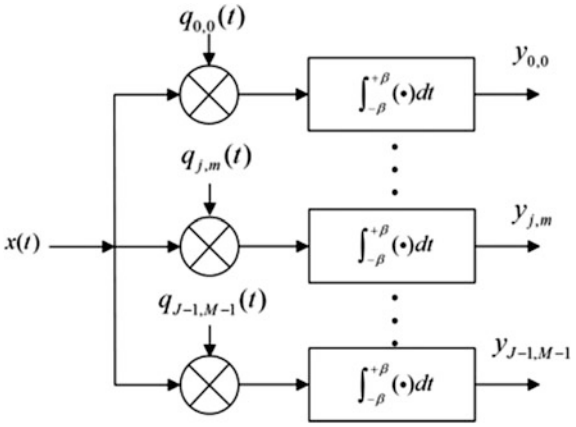



**Fig. 1** Gabor sampling scheme for multipulse signals

implementation, but not to reduce sampling rate. Matusiak and Eldar [4] chose matrix $\mathbf{D}$ as $\mathbf{D}=\mathbf{I}$ or other full bank matrix with $J<L$, reducing the primary mission of (1) to $\mathbf{U}=\mathbf{CZ}$.

Relying on ideas of CS, the signal can be recovered from a small number of samples by exploiting its sparsity. Then, the multipulse signal $x(t)$ can be recovered according to equation $x(t)=\sum_{k,l\in\mathbb{Z}} z_{kl}e^{-2\pi ibl}g(t-ak)$.

Recovering $\mathbf{Z}$ is referred to as a multiple measurement vector (MMV) problem [6]. In [4], matrix $\mathbf{C}$ was chosen as Gaussian or Bernoulli random matrices, which have RIP of the order $S$, if $M\geq 2\lceil 2\mu^{-1}\rceil Nlog(K/2\lceil 2\mu^{-1}\rceil N)$.

## 2.2 Exponential Reproducing Windows Design

We now present a sampling scheme with exponential reproducing windows that can greatly simplify the time domain transform function $s_m(t)$. Primarily, compared with the time duration $[-T/2,T/2]$ set in [4], we shift to a complete positive scope $[0,T]$. For short pulses stream, it satisfies that $WN_p\ll T$ and matrix $\mathbf{Z}$ is sparse.

An exponential reproducing window is any function $g(t)$ that, together with its shifted versions, can generate complex exponentials of the form $e^{\alpha_n t}$, such as

$$\sum_{k\in\mathbb{Z}} v_{n,k}g(t-k)=e^{\alpha_n t} \tag{2}$$

where $n=0,1,\ldots,N$, and $\alpha^n\in\mathbb{C}$.

The theory relating to the reproduction of exponentials derives from the concept of E-splines. A function with the time domain representation $\beta_\alpha=e^{\alpha}rect(t-1/2)$ is called cardinal E-spline of first order. Through convolution of $\beta_\alpha$, $N$th order E-splines can be obtained, e.g., $\beta_{\boldsymbol{\alpha}}(t)=(\beta_{\alpha_1}*\beta_{\alpha_2}*\cdots*\beta_{\alpha_N})(t)$, where $\boldsymbol{\alpha}=(\alpha_1,\alpha_2,\ldots,\alpha_N)$, and it can be written in the Fourier domain as $\beta_{\boldsymbol{\alpha}}(\omega)=\prod_{n=1}^{N}\frac{1-e^{\alpha n-j\omega}}{j\omega-a_n}$.

Exponential reproducing windows $g(t)=\psi(t)*\beta_{\boldsymbol{\alpha}}(t)$ is considered as response to sampling kernel function $\psi(t)$ filtered with E-splines $\beta_{\boldsymbol{\alpha}}(t)$. Under ideal conditions, we let $\psi(t)=\delta(t)$, the windows is reduced to $g(t)=\beta_{\boldsymbol{\alpha}}(t)$. Operating a time scaling on $N$th order cardinal E-spline function $\beta_{\boldsymbol{\alpha}}(t)$ with factor $N/W_g$, the Gabor window and its dual window can be expressed as $g(t)=\beta_{\boldsymbol{\alpha}}(tN/W_g)$.

Then Gabor window $g(t)$ is compactly supported in $[0,W_g]$. In this scenario, the lattice parameters are $\mu W_g$, and $b=1/W_g$. If $\mu=1/N$, Eq. (3) is obtained

$$\sum_{k\in\mathbb{Z}} v_{n,k}g(t-k)=\sum_{k\in\mathbb{Z}} v_{n,k}\beta_{\boldsymbol{\alpha}}(tN/W_g-k)=e^{\alpha_n Nt/W_g} \tag{3}$$

In fact, the time duration of the signal is [0, $T$] and the sampling rate is restricted to $1/T$, so that the window sequence can be truncated. To ensure that the exponential

functions constructed by the shift windows can cover [0, $T$] in time domain, they were calculated by assuming the lower and upper shift count limits as $K_1$ and $K_2$:

$$\begin{array}{l} K_1 a \geq -W_g \\ K_2 a \leq T + W_g \end{array} \rightarrow \begin{array}{l} K_1 = -N \\ K_2 = \lceil (T + W_g)N/W_g \rceil - 1 \end{array} \tag{4}$$

Equation (7) shows that the time domain is divided into $K = K_2 - K_1$ pieces. Then according to (3), the time domain transform function is

$$s_m(t) = e^{\overline{\alpha_n} Nt/W_g} \mathrm{rect}(t - T/2) \tag{5}$$

## 2.3 Filter Form

Equation (5) has an exceedingly simple form, as compared with that of the general settings discussed in [4]. Nonetheless, $s_m(t)$ is implemented with a filter.

First of all, construct function $\chi_m^*(t) = \sum_{p=K_1}^{K_2} v_{m,p} \overline{\beta_{\boldsymbol{\alpha}}(-tN/W_g + \mu N(p-1))}$ here. Given that $\mu = 1/N$, $p, q \in \{K_1, K_1 + 1, \ldots, K_2\}$, $p + q = K - 2N$, and the integral interval is restricted to [0,$T$], let $\chi_m(t) = \chi_m^*(t)\mathrm{rect}(t - T/2)$. Then

$$\begin{aligned} y_{j,m} &= \int_0^T x(t) w_j(t) s_m(t) dt = \int_0^T x(t) w_j(t) \sum_{q=K_1}^{K_2} c_{m,q} \overline{\beta_{\boldsymbol{\alpha}}(tN/W_g - q)} dt \\ &= \int_0^T x(t) w_j(t) \sum_{p=K_1}^{K_2} c_{m,p} \overline{\beta_{\boldsymbol{\alpha}}(tN/W_g - \mu N(K - 2N - p))} dt \\ &= \int_0^T x(t) w_j(t) \sum_{p=K_1}^{K_2} c_{m,p} \overline{\beta_{\boldsymbol{\alpha}}\big(-(\tau - t)N/W_g + \mu N(p-1)\big)} dt \\ &= \big(x(t) w_j(t) * \chi_m(t)\big)[\tau] \end{aligned} \tag{6}$$

Evidently, $\chi_m(t)$ represents the unit impulse response of the filter in the $(j, m)$th channel. What is notable is that $y_{j,m}$ is the integral result in time $\tau = W_g(K - 2N + 1)/N$. According to (5), $\tau = W_g(K_2 - K_1 - 2N + 1)/N = \lceil T \rceil$. Consequently, if the sampling action occurs in $t_s$, it suggests that the sample $y_{j,m}$ is acquired from the $(j, m)$th channel. In addition, index $m$ corresponds to $n$ and the total number of time domain transform functions equals the order of E-splines, namely $M = N$. Figure 2 shows the transformation of Gabor sampling scheme with $s_m(t)$, carried out by filters other than waveforms.

According to (3) into the expression, $\chi_m(t)$ can be reduced to a simple form as $\chi_m(t) = e^{-\overline{\alpha_m} Nt/W_g} \mathrm{rect}(t - T/2)$, which is an exponential function truncated by a rectangle window supported in [0, $T$]. Then, the Laplace transform of $\chi_m(t)$ is expressed as

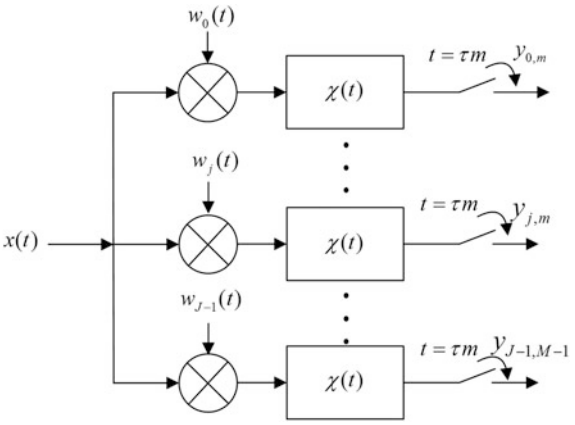



**Fig. 2** Filter form of Gabor sampling scheme

$$X_m(s) = \frac{1 - e^{-T(s+\overline{\alpha_m}N/W_g)}}{s + \overline{\alpha_m}N/W_g} \tag{7}$$

Besides the order of the E-splines and window width, the parameter vector $\boldsymbol{\alpha}$ is the adjustable variable that decides the filter characteristics. Each entry $\alpha_m$ consists of real and imaginary parts. For the exponent item of (7), the real part was used to simplify the expression. Choose $\alpha_m = \alpha_0 + im\lambda$, where $m = 1, 2, \ldots M$. To fulfill the convergence requirement of Laplace transform, it is necessary that $\mathrm{Re}[s] > -\alpha_0$. The optimal setting being $\alpha_0 = W_g/WN$, if $s = i\omega$, for $T \gg W$ and $W_g \geq W$, it follows that $e^{-T(s+\overline{\alpha_m}N/W)} = e^{-TW_g/W^2} e^{iT(m\lambda N/W - \omega)} \to 0$. Particularly, if $W_g = W$, $\alpha_0$ will have simpler form as $\alpha_0 = 1/N$. In this case, the utility of $s_m(t)$ becomes a first order RC filter $X_m(s) = \frac{1}{s+\overline{\alpha_m}N/W}$. At the same time, for $\tau = \lceil T \rceil \geq T$, the sample acquired in $t = t_s$, after passing the filter, does not lose any signal information in interval [0, $T$].

## 2.4 Signal Reconstruction and Measurement Matrix

For this study, we focus mainly on the construction of measurement matrix **C**. To bring out the concrete measurement matrix entries $c_{m,k}$, the first and foremost task is to fix $\alpha_m$. For the above analysis, we have chosen $\alpha_m = \alpha_0 + i\lambda\xi(m)$, with $\alpha_0 = W_g/NW$, and here we just need to decide $\lambda$. According to [5], $c_{m,k} = e^{\alpha_m k} c_{m,0}$. If there is a component such as $2\pi im$ in $\lambda$, it is possible to construct a measurement matrix that has the same properties as those of Fourier matrix. So, we fix $\lambda$ as $\lambda = 2\pi/K$, and then $\alpha_m = \frac{W_g}{NW} + i\frac{2\pi\xi(m)}{K}$.

According to Sect. 2.3, and given that $p, q \in \{K_1, K_1+1, \ldots, K_2\}$, $p+q=K-2N$, we can compute $c_{m,k}$ as follows:

$$\begin{aligned} c_{m,k} = v_{m,q} = v_{m,(K-2N-p)} &= \int_{-\infty}^{+\infty} e^{-\overline{\alpha_m} Nt/W_g} \overline{\beta_{\boldsymbol{\alpha}}(-tN/W_g + (K-2N-p-1))} dt \\ &= e^{\overline{\alpha_m}(K-2N-p)} \int_{-\infty}^{+\infty} e^{-\overline{\alpha_m} Nt/W_g} \overline{\beta_{\boldsymbol{\alpha}}(-tN/W_g - 1)} dt \\ &= e^{\overline{\alpha_m} q} \int_{-\infty}^{+\infty} e^{-\overline{\alpha_m} Nt/W_g} \overline{\beta_{-\boldsymbol{\alpha}}(tN/W_g)} dt \\ &= e^{\overline{\alpha_m} k} \frac{\Xi(m)}{|\Xi(m)|^2} \end{aligned} \tag{8}$$

where $\Xi(m) = \frac{W_g}{N} \prod_{n=1}^{N} \frac{1 - e^{\overline{\alpha_n} + \overline{\alpha_m}}}{\overline{\alpha_n} + \overline{\alpha_m}}$, $k \in \{K_1, K_1+1, \ldots, K_2\}$.

However, for perfect CS recovery, $\mathbf{C}$ should satisfy RIP. We can see that $\mathbf{C}$ is a weighted DFT matrix. As there is still no effective and easy method to compute the RIC of such kind of matrices except ergodic calculating, we detect the RIP by comparing it with DFT matrix, which was proved to satisfy RIP for sparse signal reconstruction [7]. First, we need to explore the matrix coherence $\theta$.

Then by applying Gershgorin's circle theorem, we can conclude that $\delta_S \le (S-1)\theta(\boldsymbol{\Phi})$. According to the definition, matrix coherence of $\mathbf{C}$ satisfy

$$\theta(\mathbf{C}) = \frac{|<c_k, c_l>|}{||c_k||_2 ||c_l||_2} = \left| \sum_{m=0}^{M-1} \frac{e^{i\frac{2\pi m}{K}(l-k)}}{|\Xi(m)|^2} \right| \Big/ \sum_{m=0}^{M-1} \frac{1}{|\Xi(m)|^2} \le \sum_{m=0}^{M-1} \left| e^{i\frac{2\pi m}{K}(l-k)} \right| / M = \theta(\mathbf{DFT}_Q), \tag{9}$$

Where $\theta(\mathbf{DFT}_Q)$ is a class of submatrices of Fourier matrices **DFT**. In [8], Theorem 3.3 shows that for satisfying RIP the best known bound on the number of Fourier measurements $M = O(S \log^4 K)$. So, for perfectly reconstructing the original signal, the critical problem is how to construct the sampling scheme with an appropriate relation between $S$ and $K$.

In this study, we can solve the problem from the two entry points: the windows width $W_g$ and the index set of measurement matrix $\mathbf{C}$. On the one hand, in a fixed time duration, $K$ is determined by the time shifting parameter a. With $W_g$ increasing, $K$ gets smaller. On the other hand, $W_g$ is intimately involved with sparsity $S$. If the windows are wide enough, the sparsity $S$ can be reduced to half of that proposed in [4]. What is more, E-splines have a special property: the higher the order $N$, the more the energy is centralized to a shorter time domain support [5]. Then the essential window's width gets short and the ratio $M/S = N/S$ brought down.

## 3 Simulation and Discussion

The sampling scheme was tested on multipulse signals of duration $T = 20$ ms and the pulses making up the signals were randomly chosen as a set of three different windows: cosine, Gaussian, B-spline of three orders. The number of pulses was varied between $N_p = 1, 3, 5$, the maximum pulse width being $W = 0.5$ ms. The locations of the pulses were also random. Monte Carlo method was used for simulations averaged over 500 trials. Throughout the experiments we chose $\mathbf{D} = \mathbf{I}$ and measured the relative error $||x - \hat{x}||_2 / ||x||_2$.

First, we studied the effects of stretching Gabor frame window width on the reconstruction error. The smoothness order of the E-spline windows was chosen as $N = M$=100. $\mathbf{Z}$ was acquired by solving the MMV problem with SOMP.

Figure 3 shows the relative error between the reconstructed signals and the original signals with the increase in windows stretching factor $\zeta$ and the pulse number $N_p$.

The two curved surfaces represent, respectively, the error variation trends under the conditions of $b = 1/W$ and $b = 1/\zeta W$. It can be seen that when $b = 1/\zeta W$ and $\zeta \geq 7$, the signals can be reconstructed with minimum error, and the variation of $N_p$ had little effect on the error. With $\zeta$ increasing, the dimension $K$ of $\mathbf{Z}$ significantly decreased, which means that the measurement matrix has correspondingly fewer columns, and can hence be recovered with higher accuracy.

Then we compared the reconstruction error of the sampling scheme proposed in this paper with that of a specific example in [4]. Under appropriate parameter setting, the reconstruction performance of both schemes was similar. Therefore, the sampling scheme proposed here can be considered effective and feasible.

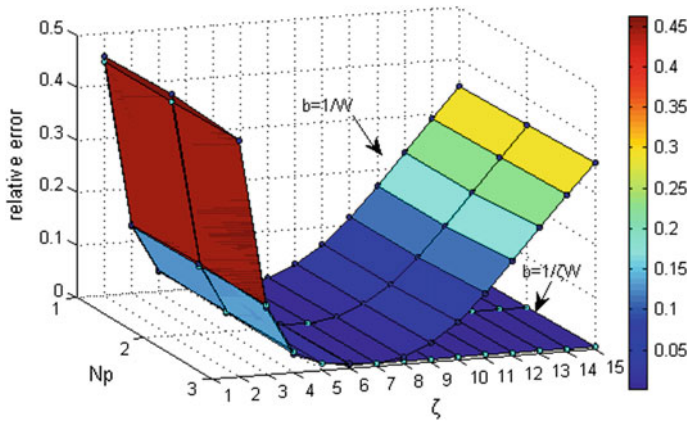



**Fig. 3** Relative reconstruction error variation curve with the increasing of windows stretching factor $\zeta$ and pulse number $N_p$

## 4 Conclusion

We introduced exponential reproducing windows into Gabor sampling scheme for Sub-Nyquist sampling of short pulses, which notably simplified the filter structure of the sampling system and enabled it to be realized by exponential filters. Subject to choosing an appropriate E-splines parameter vector $\boldsymbol{\alpha}$, the Gabor windows could be a positive real function and it is possible to construct the measurement matrix as a weighted DFT matrix for Gabor coefficients recovery, which is proved to have good coherence. Simulations prove that the sampling proposed in this paper the sampling system require quite low sampling frequency and has nice reconstruction performance.

## References

1. M. Mishali, Y.C. Eldar, A.J. Elron, Xampling: signal acquisition and processing in union of subspaces. IEEE Trans. Signal Process. **59**(10), 4719–4734 (2011)
2. G. Itzhak, E. Baransky, N. Wagner, I. Shmuel, E. Shoshan, Y.C. Eldar, A hardware prototype for Sub-Nyquist radar sensing, in *ITGFachbericht-SCC 2013*, 2013
3. T. Michaeli, Y.C. Eldar, Xampling at the rate of innovation. IEEE Trans. Signal Process **60**(3), 1121–1133 (2012)
4. E. Matusiak, Y.C. Eldar, Sub-Nyquist sampling of short pulses. IEEE Trans. Signal Process **60** (3), 1134–1148 (2012)
5. M. Unser, T. Blu, Cardinal exponential splines: part i-theory and filtering algorithms. IEEE Trans. Signal Process. **53**(4), 1425–1438 (2005)
6. J.D. Blanchard, M. Cermak, D. Hanle, Y. Jing, Greedy algorithms for joint sparse recovery. IEEE Trans. Signal Process. **62**(7), 1694–1699 (2013)
7. M. Rudelson, R. Vershynin, On sparse reconstruction from fourier and gaussian measurements. Commun. Pure Appl. Math. **61**(8), 1025–1045 (2008)
8. H. Olkkonen, J.T. Olkkonen, Sampling and reconstruction of transient signals by parallel exponential filters. IEEE Trans. Circuits Syst. II: Express Briefs **57**(6), 426–429 (2010)
9. Z. Li, K. Liu, Y. Zhao, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)

# DOA Estimation for Wideband Chirp Signals

**Deliang Liu, Xiwei Guo, Peng He and Shen Zhao**

**Abstract** Conventional DOA estimation approaches suffer from low-angular resolution or relying on a large number of snapshots which are unavailable in numerous practical applications such as underwater array processing. The sparsity-based IAA can work with a few snapshots and has high resolution and low sidelobe levels, but it is only applied to narrowband signals. To solve the above problem, a new FrFT-IAA method was proposed to estimate the DOA of wideband chirp signals with high resolution based on a few snapshots. First, the wideband chirp signal was taken on the Fractional Fourier Transform (FrFT) under a specific order so that the chirp wave in time domain could be converted into sine wave with a single frequency in FrFT domain. Then the steering vector of the received signal can be obtained in FrFT domain. Finally, IAA algorithm was utilized with the obtained steering vector to estimate the DOA of the wideband chirp with a few snapshots. The simulation results demonstrate the effectiveness of the proposed method.

**Keywords** Direction of arrival · Fractional fourier transform · Iterative adaptive approach · Chirp

## 1 Introduction

DOA estimation has many applications in radar, sonar, and communications systems [1–4], and has been extensively studied. The classical delay-and-sum (DAS) method suffers from low resolution and high sidelobe levels, while the

D. Liu · X. Guo (✉) · P. He · S. Zhao
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: xixinting@sohu.com

D. Liu
e-mail: liudeliang82@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_65

well-known Multiple Signal Classification (MUSIC) method provides super-resolution DOA estimation for narrowband uncorrelated signals but requires a large amount of snapshots [5]. Wideband chirp signals are widely used in radar and sonar systems in recent years, and many DOA estimation approaches have been proposed for them, such as MUSIC method in FrFT domain [6], Estimating Signal Parameters via Rotational Invariance Techniques (ESPRIT) in FrFT domain [7] and Ambiguity-Function based techniques [8], etc. However, none of these methods is able to provide high angular resolution with very low snapshots which is the case when the environment being sensed by the array is stationary for a short duration of time. The Iterative Adaptive Approach (IAA) [5] is a sparsity-based technique that can estimate the DOA with a few snapshots with high resolution and low sidelobes. But it does not applied to wideband signals, because the steering vector of the received signal model is time-variant. So some extensions to this method have been proposed to deal with the wideband chirp signals, such as utilizing all-phase fast Fourier transform (apFFT) algorithm [9] for multiband radar system.

In this paper, we propose a novel FrFT-IAA method to estimate DOA for wideband chirp signals with only a few snapshots based on the IAA approach and FrFT algorithm which is a generalization of the Fourier transform. First, FrFT algorithm with a specific transform order is adopted to obtain the representation of wideband chirp signal which is a sine wave with a single frequency so that we can obtain a time-invariant steering vector which can be used for IAA. Then, we utilize the original IAA to estimate the DOA values in FrFT domain. Finally, the simulation results confirm the effectiveness of our algorithm.

## 2 System Model and Problem Statement

Without loss of generality, we show the fundamental of the proposed method based on active radar system with a linear array with $M$ sensors uniformly placed along the $x$ axis with an interelement spacing $d$, as shown in Fig. 1. Far-field targets are

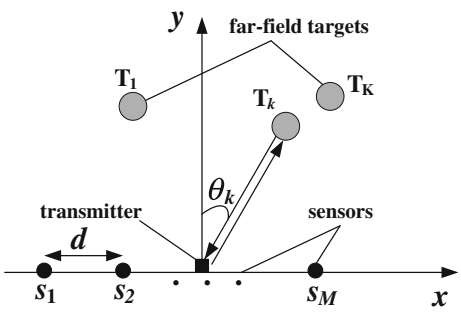



**Fig. 1** Active radar system

located at $\boldsymbol{\theta}$, where $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_K]$. The number of targets, $K$, is usually unknown, so, here, $K$ is considered to be the number of potential targets (scanning points) in the region and it is much larger than the number of actual targets. Only a few signal power estimates of the potential targets will be nonzero, so sparsity-based algorithm can be used in array processing applications.

The transmitter emits a chirp signal

$$x(t) = a \exp(j2\pi f_0 t + j\pi\mu t^2) \tag{1}$$

where $a$ is the signal amplitude, $f_0$ is the center frequency and $\mu$ is the chirp rate.

The received signal at the $m$th sensor can be expressed as the sum of $K$ delayed versions of $x(t)$, given by

$$x^m(t) = \sum_{k=1}^{K} \rho_k x_k(t - \tau_k^m) + e^m(t) \tag{2}$$

where $e^m(t)$ for $m = 1, 2, \ldots, M$ is the additive Gauss white noise at the $m$th sensor, $\rho_k$ is the backscattering coefficient of target $k$ for $k = 1, 2, \ldots, K$. $\tau_k^m$ is the time delay of the $k$th signal traveling to the $m$th sensor relative to the reference sensor (the first sensor), which can be expressed as

$$\tau_k^m = (m-1)d \sin\theta_k / c \tag{3}$$

where $c$ is the wave speed.

The received signals at the sensors can be written in matrix form as

$$\mathbf{x}(t) = \mathbf{a}\mathbf{s}(t) + \mathbf{e}(t) \tag{4}$$

where $\boldsymbol{s}(t) = [\, s_1(t), s_2(t), \ldots, s_K(t)]^T$ is the waveform vector, $\boldsymbol{a} = [\boldsymbol{a}(\theta_1), \boldsymbol{a}(\theta_2), \ldots, \boldsymbol{a}(\theta_K)]^T$ is the steering vector with

$$\begin{aligned} \mathbf{a}(\theta_k) = [\, &\exp(-j2\pi f_0\tau_1 + j\pi\mu\tau_1^2) \exp(-j2\pi\mu\tau_1 t), \ldots, \\ &\exp(-j2\pi f_0\tau_M + j\pi\mu\tau_M^2) \exp(-j2\pi\mu\tau_M t)]^T \end{aligned} \tag{5}$$

If the bandwidth is small compared to the carrier frequency, $\boldsymbol{a}(\theta_k)$ can be considered as time-invariant (i.e., the term $\exp(-j2\pi\mu\tau_m t)$ in (5) can be neglected), but for wideband chirp, it cannot. So, the steering vector $\boldsymbol{a}(\theta_k)$ of wideband chirp depends on the time $t$, and the narrowband IAA algorithm cannot be applied directly to wideband signal. To solve this problem, we derive FrFT method.

## 3 Proposed FrFT-IAA Method

The FrFT of signal $x$(t) is represented as [7, 10]

$$X(\alpha,u)=F^{p}[x(t)]=\int_{-\infty}^{+\infty}x(t)K_{\alpha}(t,u)dt \tag{6}$$

where

$$K_{\alpha}(t,u)=\begin{cases}\sqrt{1-j\cot\alpha}\,\exp\{j\pi[t^{2}\cot\alpha-2tu\csc\alpha+u^{2}\cot\alpha]\}, & \alpha\neq n\pi\\ \delta(t-u), & \alpha=2n\pi\\ \delta(t+u), & \alpha=(2n\pm1)\pi\end{cases}$$

with $p$ is the transform order, $F^{p}$ is the FrFT operator, $K_{\alpha}(t, u)$ is the kernel function, $\alpha$ is the rotation angle, $\alpha = p\pi/2$.

As a generalization of the standard Fourier transform, the FrFT can be regarded as a counterclockwise rotation of the signal coordinates around the origin in the time-frequency plane and the rotation angle is $\alpha$. When $\alpha = 2n\pi + \pi/2$, FrFT is equal to Fourier transform.

The FrFT of $x(t)$ in (1) about angle $\alpha$ can be represented as

$$X(\alpha,u)=a\sqrt{\frac{1+j\tan\alpha}{1+\mu\tan\alpha}}\times\exp\left[j\pi\frac{u^{2}(\mu-\tan\alpha)+2uf_{0}\sec\alpha-f_{0}^{2}\tan\alpha}{1+\mu\tan\alpha}\right] \tag{7}$$

When $\alpha=\alpha_{d}=\arctan(\mu s_{c}^{2})$, we can get

$$X(\alpha_{d},u)=C\exp[j2\pi f_{0}\cos\alpha_{d}u] \tag{8}$$

where $s_c$ is the scale factor to normalize signals. In this paper, we use Ozaktas's fast sampling-type discrete FrFT method [11] to compute the digital values for FrFT, so, here, $s_{c}=\sqrt{N}/f_{s}$ with $N$ denotes the number of snapshots, $f_s$ is the sampling frequency.$C=a\cos\alpha_{d}\sqrt{1+j\tan\alpha_{d}}\exp[-j\pi f_{0}^{2}\sin\alpha_{d}\cos\alpha_{d}]$ is a constant. Therefore, after the Frft, the chirp signal becomes a sine wave with a single frequency $f=f_{0}\cos\alpha_{d}$.

According to the properties of FrFT, we can get

$$F^{p}[x(t-\tau)]=\exp(j\pi\tau^{2}\sin\alpha\cos\alpha-j2\pi u\tau\sin\alpha)\times X(\alpha_{d},u-\tau\cos\alpha) \tag{9}$$

Therefore, the received signal, reflected by the $k$th target, at the $m$th sensor in FrFT domain can be given by

$$X_k^m(\alpha_d,u)=F^p[x(t-\tau_k^m)]=\rho_k aC\exp[j2\pi(f_0\cos\alpha_d-\tau_k^m\sin\alpha_d)u] \times \exp\{j\pi[(\tau_k^m)^2\sin\alpha_d\cos\alpha_d-2f\cos^2\alpha_d\tau_k^m]\} \tag{10}$$

Because $\tau_k^m\sin\alpha_d$ is very small, we can get

$$X_k^m(\alpha_d,u)\approx\rho_k aC\exp(j2\pi f_0\cos\alpha_d u)\times\exp\{j\pi[(\tau_k^m)^2\sin\alpha_d\cos\alpha_d-2f\cos^2\alpha_d\tau_k^m]\} \tag{11}$$

Therefore, (10) can be rewritten as

$$X_k^m(\alpha_d,u)=A_k^m S_k(u) \tag{12}$$

where

$$A_k^m=\exp\{j\pi[(\tau_k^m)^2\sin\alpha_d\cos\alpha_d-2f\cos^2\alpha_d\tau_k^m]\} \tag{13}$$

$$S_k(u)=\rho_k aC\exp(j2\pi f_0\cos\alpha_d u) \tag{14}$$

Put (3) into (13), we can get

$$A_k^m(\theta_k)=A_k^m=\exp\{j\pi[(m-1)^2d^2\sin^2\theta_k/c^2\sin\alpha_d\cos\alpha_d - 2f\cos^2\alpha_d(m-1)d\sin\theta_k/c]\} \tag{15}$$

Take FrFT on (2), according to the property of FrFT, we can get

$$X^m(\alpha_d,u)=\sum_{k=1}^{K}X_k^m(\alpha_d,u)+E^m(\alpha_d,u) \tag{16}$$

Therefore the FrFT of (4) can be expressed as

$$\mathbf{X}=\mathbf{AS}+\mathbf{E} \tag{17}$$

where

$$\mathbf{S}=diag\{S_1(u),S_2(u),\ldots,S_K(u)\},\mathbf{X}=[\mathbf{X}^1,\mathbf{X}^2,\ldots,\mathbf{X}^M]^T,\mathbf{A}=[\mathbf{A}_1,\mathbf{A}_2,\ldots,\mathbf{A}_K]^T$$

with

$$\mathbf{X}^m=[X_1^m(\alpha_d,u),X_2^m(\alpha_d,u),\ldots,X_K^m(\alpha_d,u)]^T,\mathbf{A}_k=[A_k^1(\theta_k),A_k^2(\theta_k),\ldots,A_k^M(\theta_k)]^T$$

After FrFT, the steering vector will not depend on $u$, so we can use the IAA algorithm to estimate the DOA values $\theta_k$.

The discrete form of (14) can be given as

$$\mathbf{X}(n) = \mathbf{AS}(n) + \mathbf{E}(n) \tag{18}$$

where $n = 1, 2, \ldots, N$ is the number of snapshots.

Let $\mathbf{P}$ be a $K \times K$ diagonal matrix, whose diagonal contains the power at each angle on the scanning grid, given by

$$\mathbf{P}_k = \frac{1}{N} \sum_{n=1}^{N} |s_k(n)|^2, \quad k = 1, 2, \ldots, K \tag{19}$$

where $s_k(n)$ represents the signal waveform at angle $\theta_k$ at the $n$th snapshot.

The initial estimates $\hat{\mathbf{P}}_k^{(0)}$ can be obtained using the SFLS method [12]

$$\hat{\mathbf{P}}_k^{(0)} = \frac{\sum_{n=1}^{N} \left| \mathbf{A}_k^H \mathbf{X}(n) \right|^2}{(\mathbf{A}_k^H \mathbf{A}_k)^2 N} \tag{20}$$

Define the interference and noise covariance matrix $\mathbf{Q}_k$ to be

$$\mathbf{Q}_k = \mathbf{R} - \mathbf{P}_k \mathbf{A}_k \mathbf{A}_k^H \tag{21}$$

where $\mathbf{R} = \mathbf{APA}^H$. Then the weighted least squares (WLS) cost function is given by

$$\sum_{n=1}^{N} ||\mathbf{X}(n) - s_k(n)\mathbf{A}_K||^2_{\mathbf{Q}_k^{-1}} \tag{22}$$

where $||\mathbf{x}||^2_{\mathbf{Q}_k^{-1}} = \mathbf{x}^H \mathbf{Q}_k^{-1} \mathbf{x}$.

Minimizing (22) with respect to $s_k(n)$, we can get

$$\hat{s}_k(n) = \frac{\mathbf{A}_k^H \mathbf{Q}_k^{-1} \mathbf{X}(n)}{\mathbf{A}_k^H \mathbf{Q}_k^{-1} \mathbf{A}_k} \tag{23}$$

Using the definition of $\mathbf{Q}_k$ and the matrix inversion lemma, (23) can be written as

$$\hat{s}_k(n) = \frac{\mathbf{A}_k^H \mathbf{R}^{-1} \mathbf{X}(n)}{\mathbf{A}_k^H \mathbf{R}^{-1} \mathbf{A}_k} \tag{24}$$

**Table 1** IAA algorithm

| |
|---|
| **Initialization** |
| calulate $\hat{\mathbf{P}}_k^{(0)}$ by (20) |
| **Iteration** |
| $\mathbf{R}^{(i)} = \mathbf{A}\hat{\mathbf{P}}^{(i-1)}\mathbf{A}^H$ |
| for $k$=1,2,…,$K$ |
| calculate $\hat{s}_k^{(i)}(n)$ by (23), $n$=1, 2,…, $N$ |
| calculate $\hat{\mathbf{P}}_k^{(i)}$ by (19) |
| end for |
| **Termination** |
| $\lvert \hat{s}_k^{(i)} - \hat{s}_k^{(i-1)} \rvert^2$ is less than a specified tolerance |

The iteration procedure can be summarized in Table 1. After the above procedures, we search the positions of the spectral peak of $\hat{\mathbf{P}}_k$ which are the final DOA estimates.

## 4 Simulation Results

We investigate the performance of the proposed FrFT-IAA about angular resolution, sidelobe level and accuracy in the angular spectrum. We also compare it with FrFT-MUSIC algorithm [6] which utilizes MUSIC algorithm combined with FrFT. But in order to apply for active radar, the rotation angle $\alpha$ has been changed from -arccot$\mu$ to arctan$\mu$.

The chirp signal with center frequency of 12 MHz and a bandwidth of 6 MHz is used in the simulation. The array contains 32 sensors which are uniformly spaced at half wavelength. The sampling frequency is 60 MHz and the number of snapshots is 2. We consider three targets at 70°, 72°, and 110°. The noise is assumed to be Gaussian random processes with zero mean and the SNR is 20 dB. The scanning grid for the both 3 algorithms is uniform in the range from 1° to 180°, with 1° increment between adjacent grid points, so $K = 180$. The number of iterations used in IAA is 10. 100 independent runs are simulated to obtain the angular spectrum. Furthermore, each power value is normalized as $10\log 10(P_k/P_{\max})$, where $P_{\max}$ is the maximum value of $P_k$, $k = 1, 2, \ldots, 180$.

We can see from Fig. 2a that the FrFT-MUSIC method can exactly estimate the DOA of the target at 110°, but it suffers from low-angular resolution so that it cannot separate the two targets at 70° and 72° and consider them as one target at 71°. From Fig. 2b, our FRFT-IAA algorithm produces the angular spectrum with lower sidelobe levels, higher angular resolution and higher accuracy in low snapshots conditions.

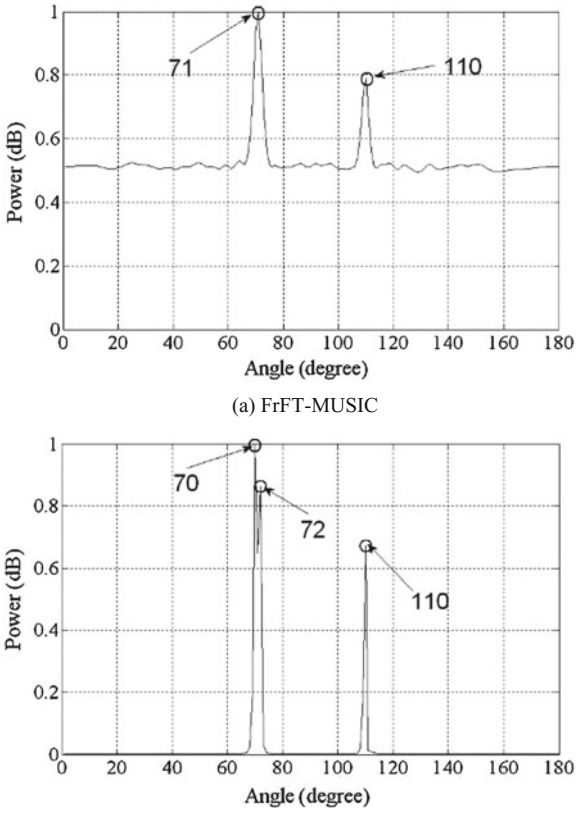



**Fig. 2** Comparison of the angular spectrums

(a) FrFT-MUSIC

(b) FrFT-IAA

## 5 Conclusion

This paper has presented a FrFT-IAA method for the DOA estimation of wideband chirp signals with a few snapshots. We extend the IAA algorithm in FrFT domain so that the DOA of wideband chirp signals can be estimated with a few snapshots. The proposed method has high angular resolution and low sidelobe levels. The simulation results have demonstrated the effectiveness of the proposed method.

**Acknowledgements** This work is supported by the National Natural Science Foundation of China under Grant 61601494 and 61501493.

## References


1. Z. Li, K. Liu, Y. Zhao et al., MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)
2. Z. Li, Y. Chen, H. Shi et al., NDFB-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. **48**, 1–5 (2016)
3. X. Liu, Z. Li, P. Yang et al., Information-centric mobile ad hoc networks and content routing: a survey. *Ad Hoc Networks* (2016). doi:10.1016/j.adhoc.2016.04.005
4. Z. Li, L. Song, H. Shi et al., Approaching the capacity of k-user MIMO interference channel with interference counteraction scheme. *Ad Hoc Networks* (2016). doi:10.1016/j.adhoc.2016.02.009
5. T. Yardibi, J. Li, P. Stoica, M. Xue, Source localization and sensing: a nonparametric iterative adaptive approach based on weighted least squares. IEEE Trans. Aerosp. Electron. Syst. **46** (1), 425–443 (2010)
6. F. Ye, Y. Li, S. Yang, 2-D DOA estimation of LFM signal based on FRFT, in *Proceedings of the 1st ISSCAA*, Harbin, China, Jan. 2006, pp. 1080–1083
7. P. Luo, K. Liu, W. Shi, G. Yan, 2-D DOA estimation of wideband LFM signals for arbitrary planar array, in *Proceedings of the 10th ICSP*, Oct 2010, pp. 307–310
8. N. Ma, J.T. Goh, Ambiguity-function-based techniques to estimate DOA of broadband chirp signals. IEEE Trans. Signal Process. **54**(5), 1826–1839 (2006)
9. J. Tian, J. Sun, G. Wang, Multiband radar signal coherent fusion processing with IAA and apFFT. IEEE Signal Process. Lett. **20**(5), pp. 463–466 (2013)
10. R. Tao, F. Zhang, Y. Wang, Research progress on discretization of fractional Fourier transform. Sci. China series F: Inf. Sci. **51**(7), 859–880 (2008)
11. M.F. Erden, M.A. Kutay, H.M. Ozaktas, Repeated filtering in consecutive fractional Fourier domains and its application to signal restoration. IEEE Trans. Signal Process. **47**(5), 1458–1462 (1999)
12. H.M. Ozaktas, O. Arikan, M.A. Kutay, G. Bozdagt, Digital computation of the fractional Fourier transform. IEEE Trans. Signal Process. **44**(9), 2141–2150 (1996)

# Electrocardiogram Signal De-noising and Reconstruction Based on Compressed Sensing

**Jinchao Sun**

**Abstract** The electrocardiogram signal consists in a character of smaller amplitude together with a larger interference range and the reconstructed signal, according to the classical compressed sensing theory, cannot be accurately conveyed by the signal. To solve this problem, compressed sensing based on the wavelet transform was stressed on. We carry out a compressed sensing algorithm based on wavelet transform, thus is to use the wavelet decomposition to separate the electrocardiogram, to reduce the noise pollution, to compress and reconstruct the high-frequency coefficient and to recover the signal by inversing the wavelet transform. Meanwhile, analysis on the data effect was also made. The result of the simulation shows that it obviously proves the noise suppressing effect on combining wavelet transform with compressed sensing to recover the signal. The integrity of useful information is enhanced, as well as obtaining a higher signal-to-noise ratio.

**Keywords** Electrocardiogram (ECG) · Compressed sensing (CS) · Wavelet transform · Noise reduction · Signal reconstruction

## 1 The Introduction

Compressed sensing theory (Compressed sensing, CS) is a kind of a new signal acquisition technology which breaks through the traditional sampling rate [1, 2]. This theory has attracted more scholars' attention ever since it is put forward. So far it has been widely used in image processing, radar imaging, wireless sensor networks, and many other fields [3, 4]. Peyre implementations the image grid points to adapt section image by using CS theory and which meanwhile cares not only the detection time but also the detection efficiency. LiuYubo, aiming at both the specific process and the sparseness for the target of radar's imaging space while

J. Sun (✉)
College of Information and Communication, Tianjin Sino-German University of Applied Sciences, Tianjin 300350, China
e-mail: polarissjc@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_66

UWB radar goes through the wall, presents an imaging method based on compressed sensing. Tang Yajuan, with her partners, uses the CS technology to improve the utilization rate of the spectrum of digital TV signal. Although the research on CS theory has made some encouraging achievements, we can still further develop its applications, we are sure that CS will have an extremely broad prospect.

Compressed sensing is applied to ECG signals' (electrocardiogram, ECG) acquisition and restoration, which greatly reduces the measurement sequence in doing the analysis and processing on the usage of CS. It not only can reduce the amount of calculation and complexity during the treatment process, but also can reduce the cost of ECG signal acquisition. However, in fact, in the process of ECG signal acquisition [5], due to the influence caused by instruments, human beings, etc., the collected ECG signal has always interfered in noises, such as Electrical noise, baseline drift, power frequency disturbance, and random noise [6]. Thus, it has a decisive impact on how to effectively reduce the interference caused by signal preprocessing in order to improve the accuracy of medical diagnosis. Thanks to the theory of wavelet which provides us with good time-frequency resolution characteristics, wavelet has played a crucial role for the noise reduction processing of ECG signal [7, 8]. Based on the above theory, both the compressed sensing and wavelet filtering are combined in this paper and a new method is proposed for noise reduction and reconstruction of the ECG signal. Experimental results also show that this method has added a very important value to the theory of reducing noise interference, recognizing features of ECG signal clarity and extracting the integrity of the useful information.

## 2 Introduction to Compressed Sensing and Wavelet Transforming Theory

### *2.1 Compressed Sensing*

The traditional signal acquisition and processing includes four parts: sampling, transforming, compression and reconstruction, as shown in Fig. 1. The raw data in this processing method has huge and hence effects in both high demands of transferring speed and storage space. Meanwhile, some important information may also occur during the transferring period. The compressed sensing theory, however, is different from others. The basic principle is among the information collections, together with the compression being executed at the same time [9], as shown in


**Fig. 1** Traditional information acquisition and processing procedure

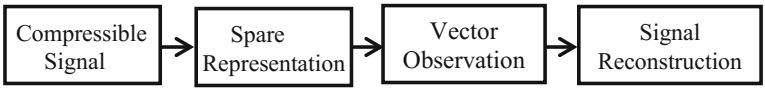



**Fig. 2** Frame of compressed sensing theory

Fig. 2. The compressed sensing theory mainly includes three parts: spare representation, coding measurement, and algorithm reconstruction [10].

The mathematical model could be understood as N-dimensional signal $x$, after a linear transform:

$$y = \boldsymbol{\Phi} x \tag{1}$$

$x$ could be transformed to $M$ sets of linear measuring signal $y$, while $\boldsymbol{\Phi}$ is $M \times N$ dimensional matrix, and $M \gg N$ It's apparently shown that the dimension of $y$ is much smaller than that of $x$, hence Eq. (1) has infinite solutions. It is very difficult to reconstruct $y$ from $x$ as there is much uncertainty. The compressed sensing theory could execute sparse representation in a known transforming domain $\boldsymbol{\Psi}$ and this is also the prerequisite of the compressed sensing theory:

$$x = \boldsymbol{\Psi} \boldsymbol{\alpha} \tag{2}$$

where $\boldsymbol{\Psi} = [\boldsymbol{\psi}_1, \boldsymbol{\psi}_2, \ \ldots, \boldsymbol{\psi}_N]$, $\boldsymbol{\psi}_i$ is a column vector and column vector $\boldsymbol{\alpha}$ is the weighted coefficient sequence. $\boldsymbol{\alpha}_i = \langle x, \boldsymbol{\psi}_i \rangle = \boldsymbol{\psi}_i^T x$. Hence $\boldsymbol{\alpha}$ is an equivalent representation of signal $x$. When $\boldsymbol{\alpha}$ has $K$ numbers of nonezero elements, it is $K$ sparse representation of signal $x$, and $\mathrm{K} < M \ll N$. Considering Eq. (1), there is

$$y = \boldsymbol{\Phi} x = \boldsymbol{\Phi} \boldsymbol{\Psi} \boldsymbol{\alpha} = \boldsymbol{\Theta} \boldsymbol{\alpha} \tag{3}$$

where $\boldsymbol{\Theta} = \boldsymbol{\Phi} \boldsymbol{\Psi}$ is a $M \times N$ dimensional matrix, and is also called sensing matrix. According to Eq. (1), $y$ could be viewed as value of measuring matrix $\boldsymbol{\Theta}$ for sparse signal $\boldsymbol{\alpha}$. In the literature [1, 2], it is proved that while matrix $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are irrelevant, and measuring matrix $\boldsymbol{\Theta}$ satisfies the constraint equidistant condition, we could restore signal $\boldsymbol{\alpha}$ very well from $M = O(KlnN)$ times measurement through nonlinear optimization. In other words, sparse signal $\boldsymbol{\alpha}$ could be reconstructed by solving an optimization equation f the minimum of normal $l_0$.

$$\hat{\boldsymbol{\alpha}} = argmin \|\boldsymbol{\alpha}\|_0 \quad s.t. \quad \boldsymbol{\Theta} \boldsymbol{\alpha} = y \tag{4}$$

For one-dimensional signal, $\boldsymbol{\Phi}$ normally uses Gauss independent equal-distributed matrix.

### 2.2 Wavelet Transform

According to the non-stable feature of ECG, and the noise-signal during the ECG sampling, the most popular filtering treatment is using wavelet transform. Compared to Fourier transform, it has better local features both in time and frequency domain.

Wavelet transform based on threshold treatment is an effective and intuitive algorithm. The frequently used threshold functions includes hard threshold function and soft threshold function, while the hard threshold function is

$$\tilde{w}_{j,k} = \begin{cases} w_{j,k} & |w_{j,k}| \geq \lambda \\ 0 & |w_{j,k}| < \lambda \end{cases} \tag{5}$$

And the soft threshold function is

$$\tilde{w}_{j,k} = \begin{cases} sgn(w_{j,k})(|w_{j,k}| - \lambda) & |w_{j,k}| \geq \lambda \\ 0 & |w_{j,k}| < \lambda \end{cases} \tag{6}$$

In the above equations, $w_{j,k}$ is the wavelet transform coefficient, and $\tilde{w}_{j,k}$ is the wavelet transform coefficient after threshold treatment. Consider the non-stability of ECG, soft threshold function is used in the article.

## 3 The De-noising and Reconstruction of ECG Based on Wavelet Compressed Sensing

Based on the advantages of the wavelet filtering and the excellent performance of compressed sensing on the sampling and reconstruction, this paper puts forward an electrocardiogram signal de-noise and reconstruction algorithm based on compressed sensing. The main characteristics of ECG signal are as followings: weaker signal, lower noise–signal ratio, stronger noise interference. The superiority of the compressed sensing theory is applied to ECG signal monitoring measurement. Using the principles of sparse signal, the sampling values can be less restored into the original signal. Namely, the reconstruction algorithm can be utilized to recover the original ECG signal out of M measurement noise. According to the CS theory, when both the measurement matrix $\boldsymbol{\Phi}$ and matrix $\boldsymbol{\Psi}$ meet the conditions of RIP, use minimum norm $L_1$, orthogonal matching pursuit method or other optimization methods to restore the ECG signal. Because of the recursive orthogonalization of selected collection of atoms, make sure per iteration in the result of its experiments conducts the condition of optimization, at the same time guarantees the signal integrity of useful information, reduces the energy of the transmission equipment, while being used to obtain a stable ECG signal. And the pacific steps are as followings:

**Table 1** The SNR gain η comparison of the de-noising ECG by Daubechies, Coiflets and Symmlet

| Wavelet | η | Wavelet | η |
|---|---|---|---|
| db2 | 1.3144 | coif4 | 1.3148 |
| db3 | 1.3259 | coif5 | 1.3431 |
| db4 | 1.3559 | sym2 | 1.3357 |
| db5 | 1.3318 | sym3 | 1.3499 |
| coif1 | 1.2721 | sym4 | 1.3082 |
| coif2 | 1.3046 | sym5 | 1.3226 |
| coif3 | 1.3238 | sym6 | 1.3422 |

Step 1: In order to carry out the de-noise processing for the original noise ECG signal $s$, we have several steps. First of all, track the signal-to-noise ratio (SNR) gain η of each wavelet basis function by using a series of simulation, thus to select the most suitable wavelet base. During the simulation, the above noise ECG signal $x$ expands three-level wavelet decomposition, and then the decomposed wavelet coefficients are quantified with soft threshold method. In this characteristics analysis of the wavelet basis function, can we get the three wavelet bases: Daubechies, Coiflets, and Symmlet, of which are more suitable for reducing the noise ECG signal. Therefore, this experiment only compares the SNR gains of the three kinds of wavelet basis functions.

As it is seen from Table 1, it is appropriate to use db4 and sym6 as the de-noising wavelet basis functions for the noise ECG signal, while the relative SNR gain of db4 is much higher than the two wavelet basis functions. Therefore, it is suggested that db4 should be chosen as a better de-noising wavelet based to decompose the noise ECG signal into three-level wavelet decomposition rather than others. The purpose of this step is to acquire the low-frequency wavelet coefficient $d$, the first level high- frequency wavelet coefficient $g_1$, the second-level high-frequency wavelet coefficient $g_2$ and the third-level high-frequency wavelet coefficient $g_3$. The layers of signal waveform are shown in Fig. 3.

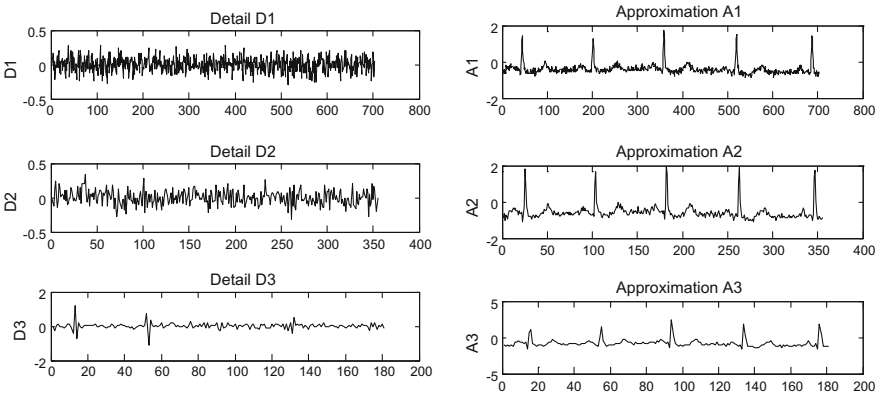



**Fig. 3** Three-level wavelet transform of the ECG signal

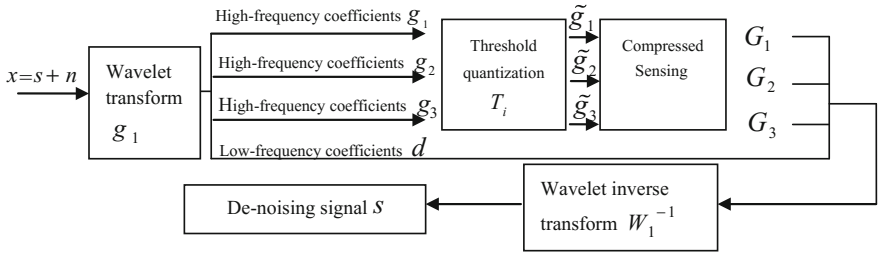



**Fig. 4** The flowchart of de-noising and reconstruction of ECG based on wavelet compressed sensing

Step 2: Since the ECG signal is decomposed by wavelet, the important useful information is mainly concentrated in the low-frequency part, and the noise-signal is mainly contained in the details of the higher frequency, so we only set each high-frequency coefficient into the threshold quantization treatment as an example. According to the amplitude of all high-frequency coefficients, by selecting the adaptive thresholds $T_i$ corresponds to each level, the treated high-frequency coefficients will be approved, they are $\tilde{g}_1$, $\tilde{g}_2$, $\tilde{g}_3$, respectively.

Step 3: As can be seen from Fig. 3, though the low-frequency coefficients do not have a good sparsity, it still contains a lot of useful information, while comparatively the de-noised high-frequency coefficient after its thresholder does much better in both. In order to reserve the pure ECG signal and reduce the noise-signal, if the compressed sensing method is used to compress and reconstruct, a pure ECG signal can be furthest separated from noise-signal. This paper selects Gaussian random matrix $\boldsymbol{\Phi}$ as measurement matrix, then measures the high-frequency coefficients $\tilde{g}_1$, $\tilde{g}_2$, $\tilde{g}_3$ one by one, namely $y_i = \boldsymbol{\Phi}\tilde{g}_i$. Furthermore, use orthogonal matching pursuit algorithm(OMP) to recover the measurement data $y_i$ respectively, layers of reconstructed high-wavelets coefficients are attained as: $G_1, G_2, G_3$.

Step 4: Finally, it obtains the de-noising ECG signal through the wavelet inverse transform performing on three-level high-frequency coefficients $G_1, G_2, G_3$ (which is reconstructed by compressed sensing) and low-frequency coefficients $d$ (which is decomposed by wavelet) Fig. 4.

## 4 Analysis of Simulation Result

*Experiment 1.* Classic compressed sensing

This set of data selected from MIT-BIH, is internationally acknowledged and could be used as standard ECG database. The classic compressed sensing theory could reconstruct the ECG, using much less data, compared to the classic sampling method. However, because of suppression of noise in the ECG signal, the

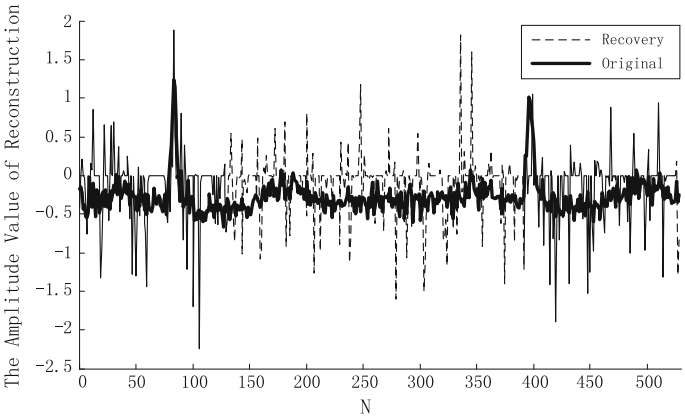



**Fig. 5** Experiment 1. The reconstructed signal by classic compressed sensing

reconstructed ECG signal is blurred, as shown in Fig. 5. This may lead to a false or a missing judgment.

*Experiment 2.* Wavelet transform

In this experiment, the same dataset is used to compare with experiment 1. Signal is reconstructed after wavelet transform is applied. As shown in Fig. 6, we may see the reconstruction effect after wavelet filtering of ECG. Although noise is apparently suppressed in the filtering, some useful information is also filtered. The integrity of signal is affected and hence the result is not ideal.

*Experiment 3.* Compressed sensing and reconstruction based on wavelet transform

In this experiment, the same dataset is used again to compare with experiment 1. We use the compressed sensing algorithm based wavelet transform to reduce noise.

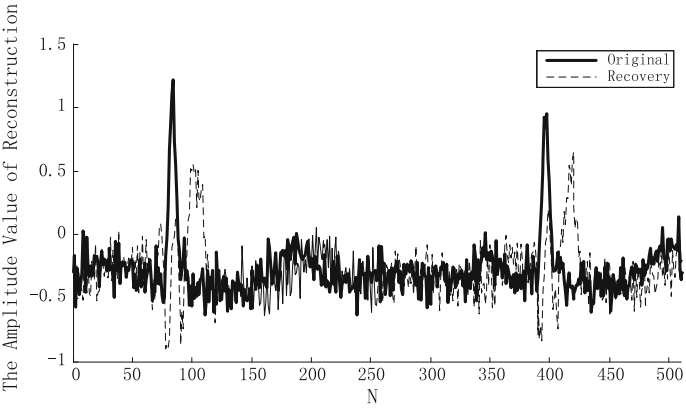



**Fig. 6** Experiment 2. The reconstructed signal by wavelet transform

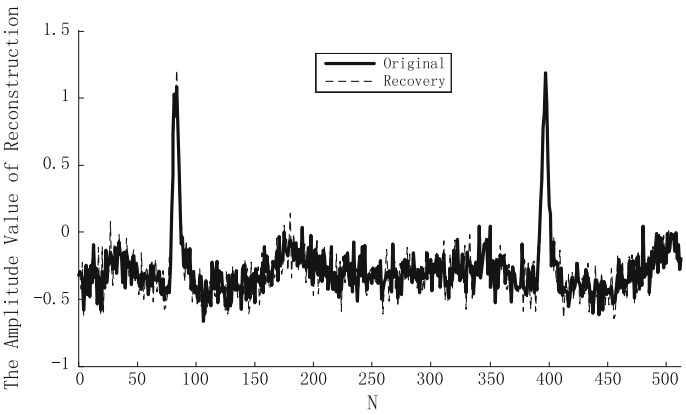



**Fig. 7** Experiment 3. The reconstructed signal by compressed sensing based on wavelet transform

**Table 2** The SNR of the three reconstructed signals

| | Classic compressed sensing | Wavelet transform | Compressed sensing based on wavelet transform |
|---|---|---|---|
| SNR/dB | 15.0848 | 27.0209 | 37.9831 |

The result is shown in Fig. 7. It could be clearly seen that the noise is effectively suppressed, and integrity of the ECG signal is kept as well, hence this method could be used more widely and will have a stronger practical value.

In order to compare the effects of the three methods, this paper also examines the similarities and the differences on the effects of compressed reconstruction on SNR. The definition formula is

$$R_{SNR} = 10lg \frac{\sum_i \hat{s}^2(i)}{\sum_i [\hat{s}(i) - f(i)]^2} \tag{7}$$

where $f(i)$ is the collected original signal, $\hat{s}(i)$ is the reconstructed signal. The higher $R_{\mathrm{SNR}}$ is, the better reconstructed effect can be obtained, as shown in Table 2.

## 5 Conclusion

In this paper, we focus on the shortcoming existed in the classical compressed sensing which has no filtering, it proposes a compressed sensing based on wavelet transform which offers an algorithm for ECG signal de-noising and reconstruction. After that we compare the results with the classic method of compressed sensing

through experiments as well. Above all, the compressed sensing based on wavelet transform algorithm for ECG signal de-noising and reconstruction method can mostly suppress noise similar to ECG amplitude and it can well reserve the feature at the peak point of the original signal. And it obviously improves the effect of the compression and reconstruction, with the ECG signal waveform which is obtained relatively clearly and distinctly. At last, it is very important and meaningful to improve the accuracy of medical diagnosis for its exact judgment. In order to identify more accurate characteristic of ECG signals to serve people, we should strengthen the researching work in real complex environment so that we can use CS earlier in the clinical medical diagnosis.

## References

1. D. Donoho, Compressed sensing. IEEE Trans. Inf. Theory, **52**(4), 1289–1306, 5406–5425 (2006)
2. E. Candes, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inf. Theory **52**(2), 489–509 (2006)
3. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016)
4. Z. Li, Y. Chen, H. Shi et al., NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commu. Netw. **2016**(1), 1–5 (2016)
5. M. Hossein, K. Nadia, A. David, V. Pierre, Compressed sensing for real-time energy-efficient ECG compression on wireless body sensor nodes. IEEE Trans. Biomed. Eng. **58**(9), 2456–2466 (2011)
6. E. Pinheiro, O. Postolache, P. Girao, Compressed sensing implementation in cardiac signals, in *IEEE, 2009 International Workshop on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications*, vol. 9 (Rende, Italy, 2009), pp. 96–101
7. F. Peng, P. Xiong, X. Cai, J. Liu, W. Fu, Noise and interference in ECG signal detection and eliminating methods. Qual. Control Saf. **28**(9), 72–74 (2007)
8. Y. Wang, H. Wang, L. Liu, Noise reduction for heart sound based on wavelet transform. Inf. Electron. Eng. **8**(3), 303–307 (2010)
9. Z. Wang, K. Yin, H. Zhao, Q. Zhang, Pulse signal noise reduction based on wavelet transform. Commu. Technol. **44**(5), 151–153 (2011)
10. M. Lustig, D.L. Donoho, J.M. Pauly, Rapid MR imaging with compressed sensing and randomly under-sampled 3DFT trajectories, in *Proceedings of the 14th. Annual Meeting of ISMRM* (Seattle, WA, 2006)

# A Method of Weak Signal Detection Based on Large Parameter Stochastic Resonance

**Zhixia Wang, Li Guo and Ke Li**

**Abstract** Aiming at weak signal detection based on large parameter stochastic resonance (LPSR), frequency-shifted and rescaling stochastic resonance (FRSR) is used to obtain the weak feature of signals submerged in noise. An improved variable step algorithm is combined to the FRSR in this paper, we use the variable step to take place of a traditional fixed value after shifting and rescaling the frequency. It has been shown marked detection efficiency of the method by the results both in the simulation and engineering application.

**Keywords** Large parameter stochastic resonance · Frequency-shifted and rescaling stochastic resonance · Variable step · Bearing fault detection

## 1 Introduction

Stochastic resonance (SR) indicates a sort of coordination among the signals, noise, and nonlinear systems: signals in certain frequency ranges can be strongly increased by noise when put into a nonlinear system [1]. It is well known that classical SR is feasible to signals with small parameters according to the adiabatic approximation theory, i.e., values of their frequency, amplitude and incidental noise intensity are all smaller than one.

However, actual engineering signals, also named large parameter signals, always exceed the bound of small parameters, so that there would not be noise-induced increases when signals pass through nonlinear systems. Implementations proposed to realize the LPSR include twice sampling stochastic resonance (TSSR) [2–5], frequency-shifted and rescaling stochastic resonance (FRSR) [6] and step-changed stochastic resonance (SCSR) [7, 8]. The previous two methods are based on the idea of turning the high frequency signal to a low one while the last one changing the length of iterative step. In fact, traditional SCSR is magnifying the step directly

Z. Wang (✉) · L. Guo · K. Li
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: w_zhixia@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_67

with a fixed value. Recently, an improved variable step algorithm is proposed by literature [8], in which the author realized SCSR by adjusting the calculating step with the change of input sample points. However, in SCSR, the variable involved is usually very large and have no determination basis, so it is uneasy to estimate. On the other hand, in output spectrogram, amplitudes corresponding to low-frequency components are higher than that of the target frequency, causing interference at a certain degree. As the variable step algorithm can adapt to signal in small and large parameters, we proposed a method of FRSR based on the variable step algorithm. FRSR has a more obvious identification result because the highest peak in the spectrum can be assumed as the target signal, meanwhile alleviating the contradiction between sampling frequency and the number of sample points [6]. Simulation experiments are conducted to verify the method and we also apply it to the practical use of bearing fault detection successfully.

## 2 SR of a Bistable System

A conventional bistable SR system derives from the damped motion of Brownian particle in potential wells. The motion equation of damped harmonic oscillator is expressed in Eq. (1):

$$\ddot{x} = -\gamma\dot{x} - w^2 x \tag{1}$$

$\gamma$ is the damping coefficient of the system, in the overdamped case, we can ignore the section of $\ddot{x}$, then the equation is simplified as first order ordinary differential equation, then SR can be described by Langevin equation as follows:

$$\gamma dx/dt = -dU(\mathrm{x})/\mathrm{dx} + \mathrm{s(t)} \tag{2}$$

$x(\mathrm{t})$ represents the output signal of the system, U(x) is a bistable system with double potential wells, $U(\mathrm{x}) = -\frac{1}{2}ax^2 + \frac{1}{4}bx^4$, parameters a and b are real numbers greater than zero. S(t) used as the system input signal includes a weak periodic signal and its carrying noise, $s(\mathrm{t}) = \mathrm{Asin}(2\pi ft) + \eta(\mathrm{t})$. Suppose $\eta(\mathrm{t})$ as white Gaussion noise, $\eta(\mathrm{t}) = \sqrt{2D}g(\mathrm{t})$, g(t) denotes white noise with zero-mean and one-variance [5], D is the noise intensity. The bistable SR model can be solved by fourth-order Runge-Kutta equation [9]:

$$\begin{cases} \mathrm{k}_1 = h(\mathrm{a}x_n - \mathrm{b}x_n^3 + s_n) \\ \mathrm{k}_2 = h[\mathrm{a}(x_n + \frac{1}{2}\mathrm{k}_1) - \mathrm{b}(x_n + \frac{1}{2}\mathrm{k}_1)^3 + s_n] \\ \mathrm{k}_3 = h[\mathrm{a}(x_n + \frac{1}{2}\mathrm{k}_2) - \mathrm{b}(x_n + \frac{1}{2}\mathrm{k}_1)^3 + s_{n+1}] \\ \mathrm{k}_4 = h[\mathrm{a}(x_n + \mathrm{k}_3) - \mathrm{b}(x_n + \mathrm{k}_3)^3 + s_{n+1}] \\ x_{n+1} = x_n + \frac{1}{6}[\mathrm{k}_1 + 2\mathrm{k}_2 + 2\mathrm{k}_3 + \mathrm{k}_4] \end{cases} \tag{3}$$

where $s_n$ and $x_n$ represent the nth sampling points of the input and output signals respectively [10–13]. h is the iterative step length.

## 3 Large Parameter Stochastic Resonance

For a detected signal with large parameters, its large amplitude A can be converted small by a linear transformation. If it contains noise stronger than the maximum value beyond which RS can not occur, then it is viable to adjust system parameters a, b and $\gamma$ to enlarge the feasible noise range, so that the noise satisfy small parameter requirements relatively. Last but not least, the occurrence of SR depends on the crucial factor of high frequency, which is essentially considered in large parameter SR. FRSR provides a way to raise the usability of LPSR by shifting and rescaling the high frequency [6]. First, use a high-pass filter to eliminate the low-frequency interference, the cut-off frequency $f_c$ is in accordance with estimated driving frequency. Then the filtered signal is modulated by a carrier signal to shift the frequency. Finally, the signal is compressed linearly with a ratio R to meet the requirements of adiabatic approximation. Original frequency $f_o$ is translated to a small one $f'$, and we can recover the $f_o$ by the Eq $f_o = Rf' + f_c$. Step-changed stochastic resonance was first proposed in essay [7], in which LPSR can adapt to weak signal detection by changing the calculation step h. Essay [8] proposed a variable step algorithm for large parameter SCSR:

$$h(\mathrm{i}) = \sqrt{j|s(\mathrm{i})|}/f_s \tag{4}$$

where $s(\mathrm{i})$ stands for ith sampling points while $h(\mathrm{i})$ for ith step length. Based on the fact that the variable step goes for large parameters as well as small ones, the paper associates FRSR with the variable step to ameliorate the large parameter SR.

## 4 Simulation Experiments

To verify the effectiveness of FRSR with variable step, simulation experiments were designed to test the large parameter signals and the results were compared with classical FRSR. Under general conditions, we assign value 1 to the damping

coefficient $\gamma$. And other two system parameters are a = 0.1 and b = 1. The amplitude and frequency of the input sinusoidal signal are A = 0.3, $f_o$ = 40 Hz, and the intensity of the input noise is D = 0.39, thus noise $\eta(t) = \sqrt{2D}g(t)$, input signal is taken as $s(t) = 0.3\sin(2\pi \times 40t) + 4\eta(t)$. Here we note that $f_o$ is set up as a large parameter. The sampling frequency is $f_S = 2000$Hz and 4000 sample points are used for analysis. The cut-off frequency of the high filter, identical to the carrier frequency, is set as $f_c = 38$Hz. Compression ratio is chosen to be R = 100. Classical FRSR usually take the inverse of sampling frequency as the step $h = 1/f_{s1} = 0.05$, here we take place of it with variable step in terms of $h(i) = \sqrt{j|s(i)|}/f_{S1}$, j = 0.1. It is necessary to note that $f_{S1}$ is sampling frequency after scale transformed, $f_{S1} = f_S/R$.

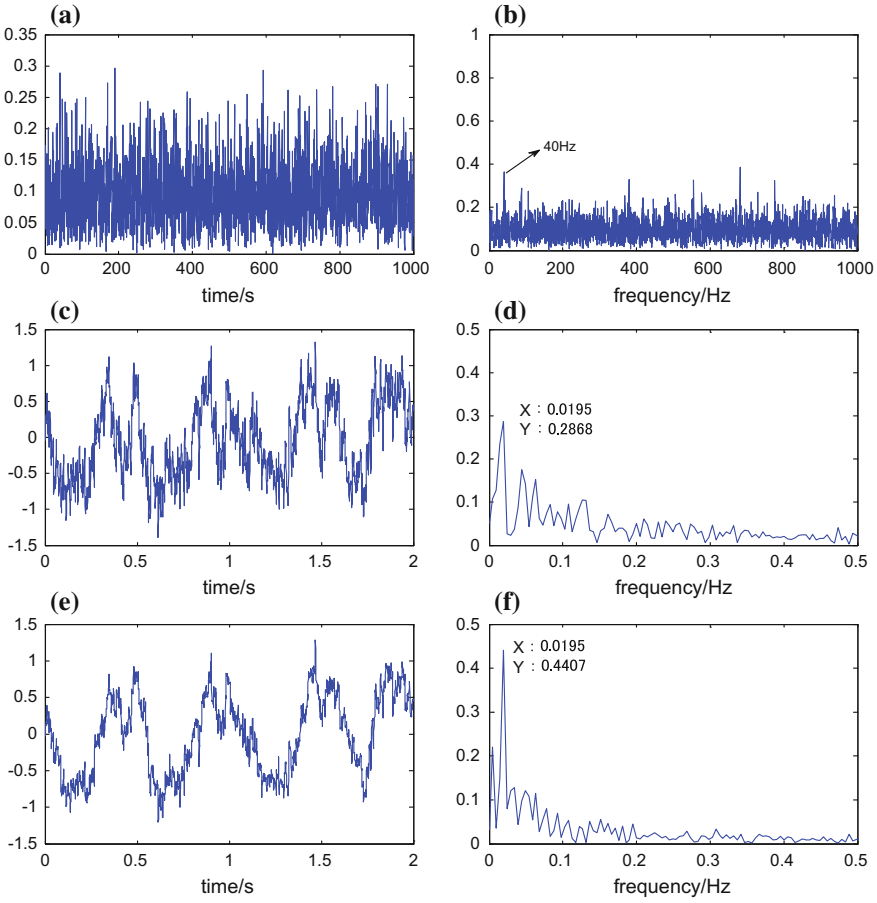



**Fig. 1** **a** and **b** Input signals. **c** and **d** The output waveform and spectrum of FRSR with a fixed step h = $1/f_S$. **e** and **f** The output waveform and spectrum of FRSR based on variable step $h(i) = \sqrt{j|s(i)|}/f_{S1}$, j = 0.1

The input and output signals in time/frequency domain are shown in Fig. 1a–f. Figure 1b is the spectrogram of input signal, in which sinusoidal signal of 40 Hz exists a little obviously but we can't distinguish it from noise spectrum or recognize it as the feature of target frequency. FRSR response is shown in Fig. 1c and d for a fixed step and e and f for variable step. In spectrogram Fig. 1d and f, spectral peak appears at the frequency 0.01953 Hz both, recovering high frequency $f_o = Rf' + f_c = 39.953\text{Hz}$, which can be considered as the target frequency 40 Hz. We can see that the spectral peak is not prominent in Fig. 1d, but significantly amplified in Fig. 1f and amplitudes of around interference frequency are cut down. On the other hand, in Fig. 1e, the noise interference on the waveform periodic signals is reduced clearly than Fig. 1c, signal amplitude has increased to a certain extent, this might be interpreted as a result that more energy is transferred from noise to the periodic signal. Comparisons with FRSR using a fixed step show that, for the same system parameters chosen, FRSR based on variable step have a superior performance.

## 5 Engineering Signal Processing on Bearing Fault Detection

Experiment data for a deep groove ball bearing are provided by Case Western Reserve University. The Test stand fulfilled the detection is shown in Fig. 2. Specifications of the deep groove ball bearing are listed in Table 1. In the experiment, single point fault was introduced to the outer race bearing using electro-discharge machining with fault diameter of 21 mils, depth of 11 mils, and the position of which is at 6 am relative to the load zone. We use a group of fault data from fan end accelerometer with no motor load, approx motor speed $f_r = 1797$ rpm.

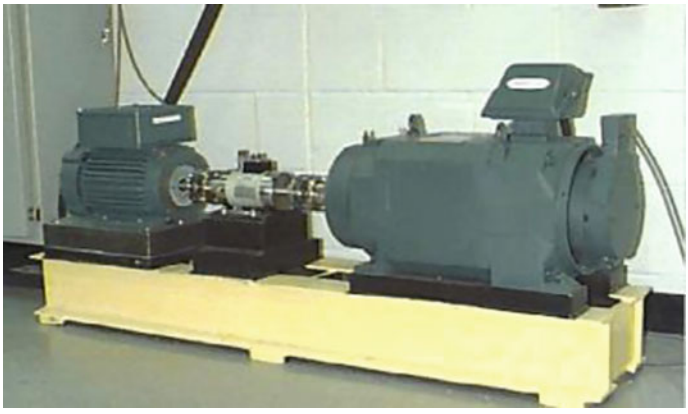


**Fig. 2** Ball bearing fault experiment rig

**Table 1** Specifications of the deep groove ball bearing

| | |
|---|---|
| Model number | 6205-2RS JEM SKF |
| Number of ball (Z) | 9 |
| Inside diameter/mm | 25 |
| Outside diameter/mm | 52 |
| Thickness/mm | 15 |
| Ball diameter (d)/mm | 7.94 |
| Pitch diameter (D)/mm | 39.04 |

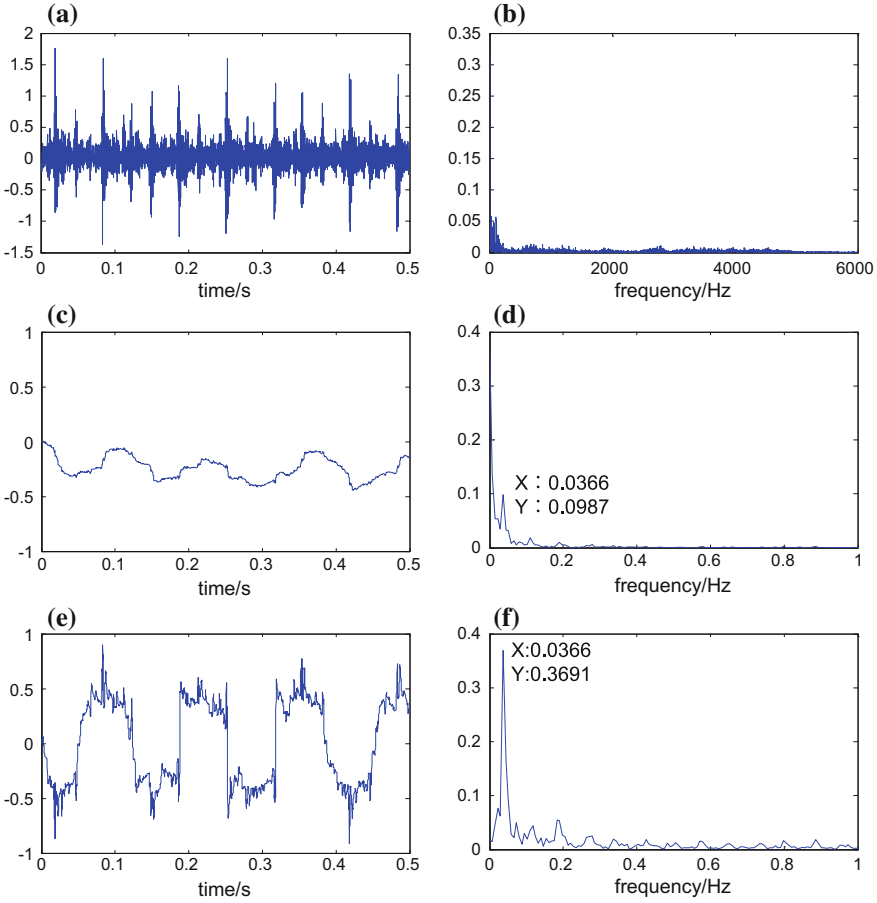



**Fig. 3** **a** and **b** Original experiment data. **c** and **d** The output waveform and spectrum of FRSR with a step $h = 1/f_{S1} = 0.017$. **e** and **f** The output waveform and spectrum of FRSR based on variable step $h(i) = \sqrt{j|s(i)|}/f_{S1}$, j = 1000

The sampling frequency is 12 kHz, and the data length for calculation is 6000 points. The original test data are shown in Fig. 3a, from which we can see the experiment data are a set of weak periodic impact signal. In Fig. 3b, the spectrum of the test data, fault frequency is covered by noise and can not be identified. Formula for calculating the fault frequency of the outer ring of the bearing is:

$$f = \frac{z}{2}(1 - \frac{d}{D}\cos\alpha)f_r \tag{5}$$

where contact angle $\alpha$ is zero, other parameters involved can be found in Table 1. By calculation, fault frequency is obtained $f_o = 107.3$ Hz.

In order to illustrate the differences between the proposed method and classical FRSR, comparisons on engineering signal processing are also presented in this section. In consideration of modulation phenomenon existing in the fault of bearing outer race, Hilbert transform is used for demodulation process firstly. System parameters are still a = 0.1, b = 1. The cut-off frequency of the high filter, i.e., the carrier frequency, is set as $f_c = 38$ Hz. Compression ratio is chosen to be R = 200. The step of classical FRSR is $h = 1/f_{s1} = 0.017$, $f_{S1} = f_S/R = 60$Hz, and the variable step $\mathrm{h(i)} = \sqrt{j|s(\mathrm{i})|}/f_{\mathrm{S1}}$, j = 1000. Output waveforms and spectrograms are depicted in Fig. 3c and d for a fixed step and e and f for variable step, and it is evident that spectral peak at the frequency $f' = 0.0366$ Hz can be found, recovering frequency $f_0 = Rf' + f_c = 107.32$Hz, thus fault frequency of the ball bearing $f_o = 107.3$Hz is detected accurately. Higher spectral peak in Fig. 3d than f demonstrates the advantages of the FRSR based on variable step. And compared with Fig. 3d, time domain waveform is improved obviously in Fig. 3d at the same time. In classical FRSR, the resonance results are sensitive to the rescaling ratio [6], and the output may be not ideal when shifted frequency and the compression ratio have no good match. By adjusting the iterative step length, it is simple to achieve well-resonance output, reducing demands for the match of FRSR parameters.

## 6 Conclusion

A method of large parameter stochastic resonance based on FRSR with variable step is proposed in this paper and the effect is impressive. As shown in the simulation experiments and engineering signal processing, for a given signal, output results of SR has been greatly improved by using variable step instead of a fixed step. The variable step algorithm is suitable for FRSR and lowers the requirements of FRSR parameters. To provide a further promotion, we will carry out the work of optimum parameter estimate by using intelligent algorithms in the next step.

## References


1. M.I. Dykman, P.V.E. McClintock, What can stochastic resonance do? Nature **391**, 344 (1998)
2. L. Yonggang, W. Taiyong, Numerical research of twice sampling stochastic resonance for the detection of a weak signal submerged in a heavy noise. Acta Physica Sinica **52**(10), 2432–2437 (2003)
3. L. Yonggang, W. Taiyong, et al., Scale transformation stochastic resonance for a weak signal detection, in *Proceeding of 11th World Congress in Mechanism and Machine*, Apr 2004, p. 1480
4. W. Taiyong, L. Yonggang, Study of large parameter stochastic resonance in signal processing. *WCCM VI in conjunction with APCOM' 04*, Sep 2004, p. 461
5. Y. Yianhai, L. Yonggang, Z. Qiu, Y. Guo, Large parameter scale transformation stochastic resonance of a bistable system, in 3*rd International Symposium on Instrumentation Science and Technology*, Aug 2004, pp. 1328–1333
6. J. Tan, X. Chen, J. Wang et al., Study of frequency shifted and re-scaling stochastic resonance and its application to fault diagnosis. Mech. Syst. Signal. Process. **23**(3), 811–822 (2009)
7. L. Qiang, W. Taiyong, L. Yonggang et al., Weak signal detection based on step-changed stochastic resonance. J. Tianjin Univ. **39**(4), 432–437 (2006)
8. C. Li, A improved variable step stochastic resonance numerical algorithm. Inf. Commun. **146**, 1–3 (2015)
9. Y. Dingxin, H. Niaoqing, Numerical simulation of stochastic resonance in bistable system for detecting weak signal. J. Natl. Univ. Def. Technol. **6**, 91–94 (2003)
10. Z. Li, K. Liu, Y. Zhao, et al., MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)
11. Z. Li, Y. Chen, H. Shi, et al., NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. **48**, 1–5 (2016)
12. X. Liu, Z. Li, P. Yang, et al., Information -centric mobile ad hoc networks and content routing: a survey. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.04.005
13. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.02.009

# Research on the Algorithm of Wireless Routing Based on Inter-flow Network Coding

**Qiang Liu**

**Abstract** With continuous progress of social economy and science and technology, the continuous developing internet technology has become an indispensable and important constituent part in people's school life and social works. With the increase of network port, the network capacity is gradually decreasing, for which the data information and data transmission speed are also influenced to some different degree. Through the usage of network coding, the above problems can be solved. Furthermore, the handling capacity and reliability of network processing can be improved effectively. When compared to the traditional algorithm of coding-aware routing, the algorithm of wireless based on inter-flow network coding is proposed in this article as it has better handling capacity and reliability.

**Keywords** Algorithm of coding-aware routing · Network coding technology · Algorithm of wireless network coding

## 1 Introduction

With the coming of digital informatization era, the internet technology has been rapidly popularized and developed, especially the popularized applications of mobile internet terminal [1, 2]. With its simple and unique structure, cheap cost, and wide coverage area, the wireless network has become an indispensable and important constituent part in people's work and life [3, 4]. However, the main disadvantages of wireless network is centralized in the low-quality wireless link service, leading to the phenomena of frequent packet loss and reducing the handling capacity and reliability of network [5, 6]. Therefore, the implementation of reasonable measures to improve the handling capacity and reliability of wireless routing is the problem that needs to be solved urgently.

Q. Liu (✉)
College of Music and Film, Tianjin Normal University, Tianjin 300387, China
e-mail: qiangliu_tjnu@yahoo.com; changchuanwu_tjnu@yahoo.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_68

## 2 Summary of Network Coding

With each network node, the network coding technology is to conduct linear and nonlinear processing of information received from each information channel [7, 8]. The middle node plays the function of encoder or signal processor, then it can greatly improve the handling capacity and reliability of network.

## 3 Influences to Wireless Network from Network Coding

Using the method of opportunistic network coding, inter-flow network coding conducts the transmission of data packet for the routing with more coding opportunities to reduce the transmission times of data packet and realize the network coding of two rang for the improvement of network handling capacity [9, 10]. The current improvement of network coding algorithm is mainly based on opportunistic network coding delay-sensitive broadcast transmission algorithm and the analysis of overhearing management policies on network coding [11, 12]. However, there is common defect for the improvement of these algorithms, which is not considering the situation of packet loss with different network flow speed. So, the network handling capacity can be enhanced by improving inter-flow network coding using the characters of links.

## 4 An Algorithm of Wireless Routing Based on Inter-flow Network Coding

Under normal conditions, wireless routing sends request packet to adjacent nodes routing via source nodes, which is transmitted via middle routing nodes. When it reaches to destination nodes [12, 13], the data packet is returned to source node from destination node. After receiving the data packet, the source code will select new path again.

When selecting the new path, the source node will consider the gain effect and determine whether the coding-aware of wireless routing shall be conducted or not by the analysis of nodes with path information in the data packet, then the final path is selected. The existing problem of this operation is unnecessary transmission of data packets when they are in the process of retransmission [14, 15].

Meanwhile, during the transmission process of wireless network data, there is the phenomenon for the loss of wireless links when sending data packet by routing. In order to guarantee the data integrity, the data needs to be retransmitted, leading to the reducing of efficiency [16]. The times of retransmission can be greatly reduced using fixed auxiliary nodes outside the routing to assist in retransmitting the data.

Therefore, the main idea of proposing an Algorithm of Wireless Routing Based on Inter-Flow Network Coding.

Based on coding-aware routing, it utilize the broadcast characters of wireless channel and combines inter-flow network coding technology to improve the handling capacity of wireless routing by transmitting the data using the method of localized opportunistic routing.

Its process is like this: firstly, it establishes the fixed routing from source nodes to destination nodes; then it subdivides the transmission data of source nodes into several same data segments before subdividing several data packets from data segments. Conduct linear network coding to form coding packets and send continuous broadcast to next hop nodes by fixing routing path. Once the reception of several linear independence coding packets by the next hop nodes is confirmed, the next coding packet is sent afterwards.

The next hop nodes summarize the received coding and send it to the next hop node. Continue to send the next segment after conforming the reception of coding. The rest is done in the same way until the destination nodes receive the coding packet of this segment to obtain the original data packet after decoding and send confirmation message to the previous hop node. Then the previous hop node sends out the coding packet of next segment.

Assuming that obey Rayleigh fading wireless channel, signal-to-noise power and decline rate for a, and the distance of two nodes m, n is dmn, Node m transmitted power for wm, Path loss index of k, the wireless channel between two nodes of $p_{mn}$ to transfer success probability.

$$p_{mn} = e^{\frac{d_{mn}^k}{aw_m}} \tag{1}$$

When the node m only sends to a next hop node n, the best transmission power for node m.

$$w_m = \frac{d_{mn}^k}{a} \tag{2}$$

For a sending node m, to make the next hop node receives its sends a packet, the node n number needs to be sent is $N_m$.

$$\phi(w_m) = N_m w_m = \frac{w_m}{p_{mn}} = w_m e^{\frac{d_{mn}^k}{aw_m}} \tag{3}$$

If make $\Phi(w_m)$ minimum, to do partial derivatives on type, make

$\partial(\phi(w_m))/\partial(w_m) = 0$:

$$\frac{\partial(\phi(w_m))}{\partial(w_m)} = \frac{e^{-\frac{d_{mn}^k}{aw_m}} - -\frac{d_{mn}^k}{aw_m}e^{-\frac{d_{mn}^k}{aw_m}}}{\left[e^{-\frac{d_{mn}^k}{aw_m}}\right]^2} = 0$$
$$e^{-\frac{d_{mn}^k}{aw_m}} - -\frac{d_{mn}^k}{aw_m}e^{-\frac{d_{mn}^k}{aw_m}} = 0 \tag{4}$$

The solution to $w_m = \frac{d_{mn}^k}{a}$.

Source node sending to the flow of data into the same size of k packet data segment, and k the packet to the original, after linear combination randomly generated the same coding of the the k original packets of data packet. If the $dp_i(i = 1, 2, 3, \ldots, k)$ is the original data packets, the code package $cp_i = \sum cv_i dp_i$, where $cv_i(i = 1, 2, 3, \ldots, k)$ is the package coding vector. The middle node receives code package, whether to retrieve the first node in the packet forwarding node list. If the forwarding node list storage, conceal and linear combination with coding previously received packets into new code package. Finally, according to a certain algorithm to judge whether or not to broadcast out if not forwarding node list in the package, it discards the coding. Destination node after receiving the code package, retrieval code packages and received before the code package is linearly independent. When the destination node receives the same k linearly independent encoding data package, will automatically be decoded the raw packets:

$$\begin{pmatrix} dp_1 \\ \vdots \\ dp_k \end{pmatrix} = \begin{pmatrix} cv_{11} & \cdots & cv_{1k} \\ \vdots & \ddots & \vdots \\ cv_{k1} & \cdots & cv_{kk} \end{pmatrix}^{-1} \begin{pmatrix} cp_1 \\ \vdots \\ cp_k \end{pmatrix} \tag{5}$$

$dp_i$ is raw packets, $cp_i$ is code package. If the destination node receives the same data segments of k linearly independent coding, decoding the raw packets, along the best route with the highest priority to send an ACK confirm information to the source node.

Respectively in view of the network has 1, 2,…, 10 data flow, (randomly selected a pair of nodes form the source of a stream and obtaining-information), in the case of the flow number, the optimization algorithm of throughput and delay the simulation results are shown in Figs. 1 and 2 (simulation results for running the averaged 10,000). From Figs. 1 and 2. We can see that: first of all, for a given number of flow, the optimized algorithm in throughput and the time delay better than before, it is because after fully considering the network coding expand channel capacity, choosing a path for data transmission with minimum delay. Secondly, in the Internet for several more hours, flow as the increasing of the number of, the time delay is on the rise, but after flow number increases to a certain degree, the time delay showed a trend of gradual decline. The main reason is that, with the increase of flow number coding opportunity increases, encode node packets from multiple streams can be generated a code group, so as to save time to improve efficiency.

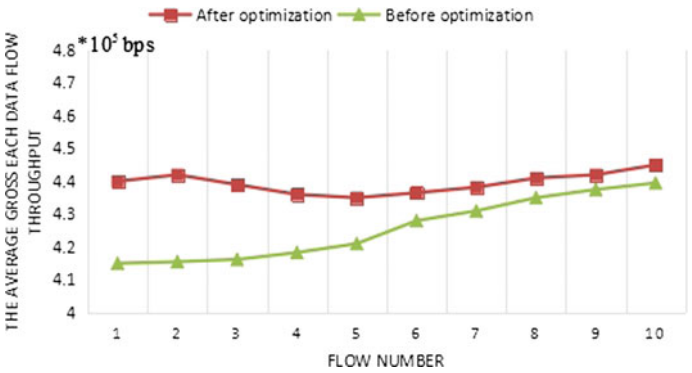



**Fig. 1** Average throughput of each flow

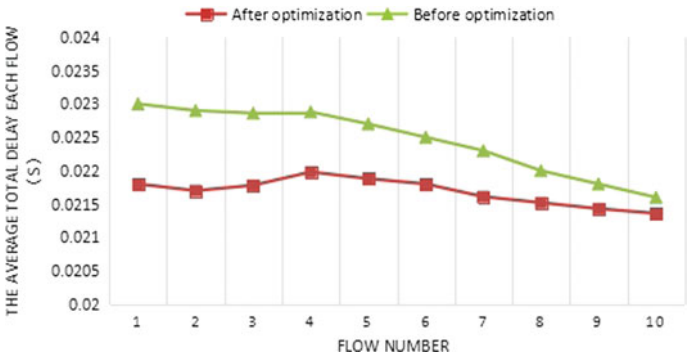



**Fig. 2** The each flow average delay

## 5 Conclusion

Based on network coding is analyzed and the research status of wireless router, this paper introduces the basic principle of network coding and network coding method, by introducing three kinds of wireless router based on network coding is a classic algorithm, this paper expounds the wireless routing mechanism based on network coding. In conclusion of the above, an algorithm of wireless routing based on inter-flow network coding can reduce times of data transmission and improve handling capacity of network under the premise of ensuring the reliability of data transmission.

## References


1. K. Sun, D. Wu, MPC-based delay-aware fountain codes for live video streaming, in *IEEE International Conference on Communications, ICC 2016*
2. H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach. IEEE Signal Process. Mag. **13**(4), 67–94 (1996)
3. X. Liu, Z. Li, Information-centric mobile ad hoc networks and content routing: a survey. Article In Press (2016)
4. R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Trans. Acoust. Speech Signal Process. **37**(7), 984–995 (1989)
5. M. Gruteser, D. Grunwald, Anonymous usage of location based services through spatial and temporal cloaking, in *ACM/USENIX MobiSys* (2003)
6. X. Nie, L.P. Li, A computationally efficient subspace algorithm for 2-D DOA estimation with L-shaped array. IEEE Signal Process. Lett. **21**(8), 971–974 (2014)
7. K. Sun, H. Zhang, D. Wu, H. Zhuang, MPC-based delay-aware fountain codes for real-time video communication. IEEE Internet of Things J. (IoT) (2016)
8. Z. Li, Y. Chen, H. Shi, K. Liu, NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. (2016)
9. G. Ghinita, P. Kalnis, S. Skiadopoulos, PRIVE: anonymous location based queries in distributed mobile systems, in *Proceedings of International Conference on World Wide Web (WWW' 07)*, Banff, Alberta, Canada, 2007, pp. 1–10
10. K. Sun, D. Wu, Video rate control strategies for cloud gaming. J. Vis. Commun. Image Represent. (2013)
11. Z. Li, K. Liu, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11) (2014)
12. T.Q. Xia, Y. Zheng, Q. Wan, X.G. Wang, 2-D angle of arrival estimation with two parallel uniform linear arrays for coherent signals. IEEE Radar Conf. **55**(9), 244–247 (2007)
13. K. Sun, B. Yan, H. Gharavi, Low complexity content-aware image retar-geting, in *IEEE International Conference on Image Processing 2012 (ICIP'2012)*
14. M. Tang, Q. Wu, G. Zhang, L. He, A new scheme of LBS privacy protection, in *IEEE* (2009)
15. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Article in Press (2016)
16. K. Sun, B. Yan, Efficient P-frame complexity estimation for frame layer rate control of H.264/AVC, in *IEEE International Conference on Image Processing 2011 (ICIP'2011)*

# Novel Cumulants-Based Decoherent Method for 2-D DOA Estimation

**Heping Shi, Jihua Cao, Dun Liu and Hua Chen**

**Abstract** A novel decoherence algorithm, called fourth-order cumulants-based improved Toeplitz matrices reconstruction (FOC-ITMR) method is presented to estimate two-dimensional (2-D) direction-of-arrival (DOA) of coherent signals. The FOC-ITMR method fully utilizes the information of received data between the whole two parallel uniform linear arrays (ULAs) and the changing reference element based on FOC. Compared with the previous works, the proposed algorithm can achieve excellent decoherence performance. The theoretical analysis and simulation results demonstrate the effectiveness and efficiency of the proposed algorithm.

**Keywords** Direction-of-arrival (DOA) · Two-dimensional (2-D) · Fourth-order cumulants (FOC) · Coherent signals · Two parallel uniform linear array (ULAs)

## 1 Introduction

Direction-of-arrival (DOA) estimation is a major research issue in array signal processing including radar, communication, sonar, etc. [1–4]. Many DOA estimation techniques, such as multiple signal classification (MUSIC) [5] and estimation of signal parameter via rotation invariance techniques (ESPRIT) [6], have achieved good DOA estimation performance. However, these high-resolution algorithms are not only very sensitive to the noise, but also require the prior

H. Shi (✉)
School of Automotion and Transportation, Tianjin University of Technology and Education (TUTE), Tianjin 300222, China
e-mail: shiheping@tju.edu.cn

J. Cao · D. Liu
School of Electronic Engineering, Tianjin University of Technology and Education (TUTE), Tianjin 300222, China

H. Chen
School of Electronic Information Engineering, Tianjin University, Tianjin 300072, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_69

knowledge of the noise characteristics of the sensors. Fortunately, the fourth-order cumulants (FOC) do not require to know or to estimate the noise covariance as long as the noise is normally distributed.

In practical, highly correlated or coherent signals are common in multipath propagation environments. To decorrelate coherent signals, the spatial smoothing [7] is especially noteworthy. X. Nie et al. [8] have introduced a computationally efficient subspace algorithm for 2-D DOA estimation with L-shaped array. In [9], a fourth-order cumulants-based forward spatial smoothing (FOC-FSS) method has been presented to remedy rank deficiency problem for 2-D DOA estimation. Recently, an alternative decorrelation technique called fourth-order cumulants-based Toeplitz matrices reconstruction (FOC-TMR) algorithm is presented in [10] to obtain good DOA estimation.

Throughout this paper, the following notations are used. $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, and $(\cdot)^\dagger$ denote the conjugation, transpose, conjugate transpose and pseudo-inverse of a matrix, respectively. The notation $cum(x)$ stands for the cumulants of variate $x$, while *arg* is the phase angle operation.

## 2 Signal Model

As illustrated in Fig. 1, the antenna array consists of two parallel ULAs ($X_a$ and $Y_a$) in the $x-y$ plane. Each ULA has $N$ sensors with spacing $d_x$, and the interelement spacing between the two ULAs is $d_y$.

Suppose that $P$ far-field signals $s_i(t)$ ($i = 1, \ldots, P$) impinge on the two parallel ULAs from directions ($\theta_i$, $\beta_i$), where $\theta_i$ and $\beta_i$ are measured relatively to the $x$ axis and to the $y$ axis corresponding to the $i$th signal, respectively. Thus, $\psi_i$, the DOA of the $i$th signal relative to the $z$ axis, can be written by

$$\cos^2\theta_i + \cos^2\beta_i + \cos^2\psi_i = 1 \tag{1}$$

Let the $m$th element of the subarray $X_a$ be the phase reference, and then the observed signals $x_m^k(t)$ at the $k$th element in time $t$ can be expressed as

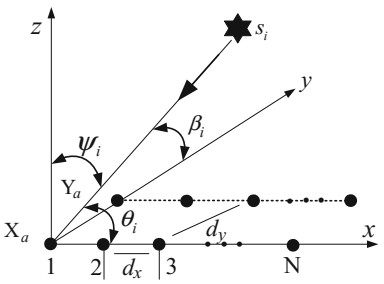



**Fig. 1** Parallel array configuration for 2-D DOA estimation

$$x_k^m(t) = \sum_{i=1}^{Q} \gamma_i s_1(t) e^{-j\frac{2\pi}{\lambda}d_x(k-m)\cos\theta_i} + \sum_{i=Q+1}^{P} s_i(t) e^{-j\frac{2\pi}{\lambda}d_x(k-m)\cos\theta_i} + n_{x,k}(t) \tag{2}$$

where the superscript $(m = 1, 2, \ldots, N)$ of the $x_m^k(t)$ stands for the number of the reference element in subarray $X_a$, and the subscript $k (k = 1, 2, \ldots, N)$ of the $x_m^k(t)$ denotes the number of the element along the $x$ positive axis in subarray $X_a$. $n_{x,k}(t)$ is the additive Gaussian noise of the $k$th element in subarray $X_a$ and $\gamma_i$ indicates the amplitude fading factor of the first $Q$ correlative signals (without loss of generality, we assume $\gamma_1 = 1$).

With a similar processing, employing the $m$th element of the subarray $Y_a$ as the phase reference, and then the observed signals $y_m^k(t)$ at the $k$th element in time $t$ can be expressed as

$$y_k^m(t) = \sum_{i=1}^{Q} \gamma_i s_1(t) e^{-j\frac{2\pi}{\lambda}d_x(k-m)\cos\theta_i} e^{j\frac{2\pi}{\lambda}d_y\cos\beta_i} + \sum_{i=Q+1}^{P} s_i(t) e^{-j\frac{2\pi}{\lambda}d_x(k-m)\cos\theta_i} e^{j\frac{2\pi}{\lambda}d_y\cos\beta_i} + n_{y,k}(t) \tag{3}$$

The observed vectors $\mathrm{X}_a$ and $\mathrm{Y}_a$ can be written as

$$\mathbf{X}^m(t) = [x_1^m(t), x_2^m(t), \ldots, x_N^m(t)]^T \tag{4}$$

$$\mathbf{Y}^m(t) = [y_1^m(t), y_2^m(t), \ldots, y_N^m(t)]^T \tag{5}$$

## 3 Angle Estimation Algorithm

### A. The FOC-ITMR algorithm

The new proposed algorithm named as FOC-ITMR, which is based on reconstructing two Toeplitz matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ using two parallel ULAs, is described in detail in this subsection. Firstly, we define $\mathbf{C}_1$ and $\mathbf{C}_2$ with the cumulants elements $c_k^m$ and $\tilde{c}_k^m$ arranging as follows:

$$\begin{aligned} c_k^m &= cum[x_m^m(t), (x_m^m(t))^*, (x_m^m(t))^*, x_k^m(t)] \\ &= \sum_{i=1}^{P} d_i e^{-j\frac{2\pi}{\lambda}d_x[(k-1)-(m-1)]\cos\theta_i} \end{aligned} \tag{6}$$

$$\begin{aligned}\tilde{c}_k^m &= cum[x_m^m(t), (x_m^m(t))^*, (x_m^m(t))^*, y_k^m(t)] \\ &= \sum_{i=1}^{P} d_i e^{-j\frac{2\pi}{\lambda}d_x[(k-1)-(m-1)]\cos\theta_i} e^{j\frac{2\pi}{\lambda}d_y\cos\beta_i}\end{aligned} \tag{7}$$

The Toeplitz matrices $\mathbf{C}_1$ and $\mathbf{C}_2$ can be constructed, respectively, as follows:

$$\begin{aligned}\mathbf{C}_1 &= \begin{bmatrix} c_1^1 & c_1^2 & \cdots & c_1^N \\ c_2^1 & c_2^2 & \cdots & c_2^N \\ \vdots & \vdots & \cdots & \vdots \\ c_N^1 & c_N^2 & \cdots & c_N^N \end{bmatrix} \\ &= \mathbf{ADA}^H\end{aligned} \tag{8}$$

$$\begin{aligned}\mathbf{C}_2 &= \begin{bmatrix} \tilde{c}_1^1 & \tilde{c}_1^2 & \cdots & \tilde{c}_1^N \\ \tilde{c}_2^1 & \tilde{c}_2^2 & \cdots & \tilde{c}_2^N \\ \vdots & \vdots & \cdots & \vdots \\ \tilde{c}_N^1 & \tilde{c}_N^2 & \cdots & \tilde{c}_N^N \end{bmatrix} \\ &= \mathbf{ADVA}^H\end{aligned} \tag{9}$$

where $\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_P)]$ with $\mathbf{a}(\theta_i) = [1, e^{j(2\pi/\lambda)d_x\cos\theta_i}, \ldots, e^{j(2\pi/\lambda)d_x(N)\cos\theta_i}]^T$ and $\mathbf{D} = diag(d_1, d_2, \ldots, d_P)$. By assuming that $\rho_{4,s_i} = cum[s_i(t), s_i^*(t), s_i^*(t), s_i(t)]$ and $\tilde{\rho}_{4,s_1} = (\sum_{i=1}^{Q}\gamma_i^*)^2(\sum_{i=1}^{Q}\gamma_i)\rho_{4,s_1}$, we get

$$d_i = \begin{cases} \gamma_i\tilde{\rho}_{4,s_1} & i = 1, \ldots, Q \\ \rho_{4,s_i} & i = Q+1, \ldots, P \end{cases} \tag{10}$$

From (8), it is shown that $\mathbf{A}$ is a Vandermonde matrix as long as $\theta_i$ comes from different angles, so $\mathbf{A}$ is a column full-rank matrix, whose columns are linearly independent. From the expression of $d_i$, we can see that $d_i$ is a nonzero constant. That is, $\mathbf{D}$ has a rank of $P$. With $\mathbf{V} = diag[v(\beta_1), v(\beta_2), \ldots, v(\beta_P)]$, we confirm that $\mathbf{V}$ satisfies the condition of full rank for different angles $\beta_i$. Accordingly, the EVD of $\mathbf{C}_1$ yields

$$\mathbf{C}_1 = \sum_{i=1}^{P} \eta_i \mathbf{v}_i \mathbf{v}_i^H \tag{11}$$

where $\{\eta_1, \ldots, \eta_P\}$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_P\}$ are the nonzero eigenvalues and corresponding eigenvectors of the matrix $\mathbf{C}_1$ respectively. The pseudo-inverse of the matrix $\mathbf{C}_1$ is

$$\mathbf{C}_1^{\dagger} = \sum_{i=1}^{P} \eta_i^{-1} \mathbf{v}_i \mathbf{v}_i^H \tag{12}$$

Due to the fact that $\mathbf{A}$ is a column full-rank matrix, from (8), we can attain

$$\mathbf{D}\mathbf{A}_T^H = (\mathbf{A}_T^H \mathbf{A}_T)^{-1} \mathbf{A}_T^H \mathbf{C}_1 \tag{13}$$

Combining (9) with (13), the alternative expression of $\mathbf{C}_2$ can be achieved

$$\begin{aligned} \mathbf{C}_2 &= \mathbf{A}_T \mathbf{D} \mathbf{V} \mathbf{A}_{\mathrm{T}}^H \\ &= \mathbf{A}_T \mathbf{V} \mathbf{D} \mathbf{A}_{\mathrm{T}}^H \\ &= \mathbf{A}_T \mathbf{V} (\mathbf{A}_T^H \mathbf{A}_T)^{-1} \mathbf{A}_T^H \mathbf{C}_1 \end{aligned} \tag{14}$$

Right-multiplying both sides of (14) by $\mathbf{C}_1^{\dagger}\mathbf{A}_{\mathrm{T}}$, we have

$$\mathbf{C}_2 \mathbf{C}_1^{\dagger} \mathbf{A}_T = \mathbf{A}_T \mathbf{V} (\mathbf{A}_T^H \mathbf{A}_T)^{-1} \mathbf{A}_T^H \mathbf{C}_1 \mathbf{C}_1^{\dagger} \mathbf{A}_T \tag{15}$$

Substituted (11) and (12) into (15)

$$\begin{aligned} \mathbf{C}_2 \mathbf{C}_1^{\dagger} \mathbf{A}_T &= \mathbf{A}_T \mathbf{V} (\mathbf{A}_T^H \mathbf{A}_T)^{-1} \mathbf{A}_T^H \Big( \sum_{i=1}^{P} \eta_i \mathbf{v}_i \mathbf{v}_i^H \Big) \Big( \sum_{i=1}^{P} \eta_i^{-1} \mathbf{v}_i \mathbf{v}_i^H \Big) \mathbf{A}_T \\ &= \mathbf{A}_T \mathbf{V} (\mathbf{A}_T^H \mathbf{A}_T)^{-1} \mathbf{A}_T^H \sum_{i=1}^{P} \mathbf{v}_i \mathbf{v}_i^H \mathbf{A}_T \end{aligned} \tag{16}$$

Notice that $\sum_{i=1}^{P} \mathbf{v}_i \mathbf{v}_i^H = \mathbf{I}$, the Eq. (16) can be further rewritten as

$$\begin{aligned} \mathbf{C}_2 \mathbf{C}_1^{\dagger} \mathbf{A}_T &= \mathbf{A}_T \mathbf{V} (\mathbf{A}_T^H \mathbf{A}_T)^{-1} (\mathbf{A}_T^H \mathbf{A}_T) \\ &= \mathbf{A}_T \mathbf{V} \end{aligned} \tag{17}$$

From (17), 2-D angle parameters, which are obtained by performing EVD on $\mathbf{C}_2\mathbf{C}_1^{\dagger}$ denoted as Toeplitz-based generalized DOA matrix, lie in $\mathbf{A}_{\mathrm{T}}$ and $\mathbf{V}$, respectively.

By performing EVD on $\mathbf{C}_2\mathbf{C}_1^{\dagger}$

$$\mathbf{C}_2 \mathbf{C}_1^{\dagger} = \sum_{i=1}^{P} \xi_i \mathbf{u}_i \mathbf{u}_i^H \tag{18}$$

where $\xi_i$ and $\mathbf{u}_i$ are the nonzero eigenvalues and the corresponding eigenvectors of the matrix $\mathbf{C}_2\mathbf{C}_1^{\dagger}$, respectively. Thus, the $P$ bigger signal subspaces can be obtained from (18). It can be verified that the spanned subspaces from the steering matrix $\mathbf{A}_\mathbf{T}$ and the signal subspaces $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_P]$ are same.

Define $\mathbf{h}_i = \mathbf{u}_i/\mathbf{u}_i$ (1), we can get

$$\kappa_i = \frac{1}{N-1}\sum_{m=1}^{N-1} arg\left[\frac{\mathbf{h}_i(m+1)}{\mathbf{h}_i(m)}\right]. \tag{19}$$

Combining $a(\theta_i) = [1, e^{j(2\pi/\lambda)d_x \cos\theta_i}, \ldots, e^{j(2\pi/\lambda)d_x(N-1)\cos\theta_i}]^T$ and $v(\beta_i) = e^{j(2\pi/\lambda)d_y \cos\beta_i}$, the estimate of 2-D DOAs can be obtained

$$\theta_i = \arccos\left(\frac{\lambda}{2\pi d_x}\kappa_i\right) \tag{20}$$

$$\beta_i = \arccos\left[\frac{\lambda}{2\pi d_y} arg(\xi_i)\right]. \tag{21}$$

Till now, 2-D DOAs of incoming signals, namely, $\theta_i$ and $\beta_i$ can be achieved automatically paired according to (20) and (21) without additional computations for parameter pair-matching.

B. Location Analysis

In this subsection, the advantage of the proposed algorithm is discussed. As for two $N \times N$ dimension Toeplitz matrices, the maximum number of signals that can be distinguished is $N - 1$ by the proposed FOC-ITMR method. Assume that the number of each subarray in [10] is 2 $M$ + 1, the FOC-TMR method can distinguish $M$ signals. According to parameters set in [9], the FOC-FSS method can tell the same number of signals as [10]. In other words, for the same number of incoming signals, the FOC-ITMR method makes full of array aperture, and gains the best estimation performance.

## 4 Simulation Results

In this section, simulation results are presented to illustrate the validity of the proposed method. We evaluate the performance of the proposed FOC-ITMR algorithm via comparison with the FOC-FSS [9] and the FOC-TMR [10] algorithms with several experiments. The maximum root-mean-square error (MRMSE) is defined to evaluate the performance of the proposed FOC-ITMR algorithm and the compared algorithms with respect to variables SNR.

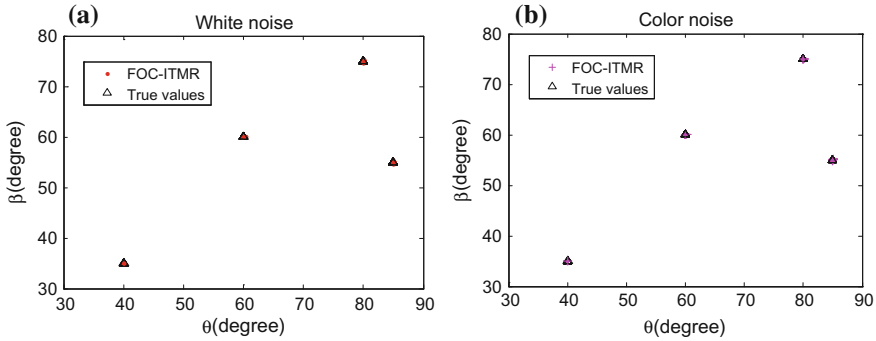



**Fig. 2** 2-D DOA estimation scattergram. **a** White noise. **b** Color noise

$$\text{MRMSE} = \sqrt{\frac{1}{\text{MC}} \sum_{i=1}^{\text{MC}} [\max(\hat{\sigma}_{i,r} - \sigma_i)]^2} \tag{22}$$

where MC denotes the times of Monte Carlo simulation. $\sigma_i$ denotes $\theta_i$ or $\beta_i$, and $\hat{\sigma}_{i,r}$ is the parameter to be estimated of $\theta_i$ or $\beta_i$.

**Experiment 1**: 2-D DOA Estimation Performance
Assume that four non-Gaussian signals with $\theta_i$ and $\beta_i$ impinge from {40°, 80°, 60°, 85°} and {35°, 75°, 60°, 55°}, separately, and the number of sensors in each subarray is $N = 5$ with sensor displacement $d_x = d_y = \lambda/2$. Figure 2 plots the paired results of four targets from 50 Monte Carlo trials with SNR = 15 dB and snapshots $L = 1500$, which shows that the 2-D DOA are paired correctly in our scheme in both white and color Gaussian noise situation. On the contrary, the compared FOC-TMR method and FOC-FSS method cannot distinguish them due to array aperture reduced.

**Experiment 2**: MRMSE and NPS versus SNR
In this simulation, three non-Gaussian signals with $\theta_i$ and $\beta_i$ are incoming from {40°, 85°, 90°} and {45°, 65°, 60°}, separately, and the number of sensors in each subarray is $N = 7$. The number of snapshots is set to 1500, and the SNR is varied from −10 to 25 dB. The $\theta_i$ and $\beta_i$ MRMSE curves of the proposed method and the FOC-TMR and FOC-FSS methods versus SNR are shown in Fig. 3, where 2000 Monte Carlo trials are used. Figure 3 illustrates that the proposed method has the much lower MRMSE than that of the FOC-TMR and FOC-FSS method both in spatially white noise and spatially color noise environments, especially at low SNR. The reason is that the proposed algorithm utilizes the whole ULAs to estimate the source signals, which avoids the loss of array physical aperture and achieves excellent decoherence performance.

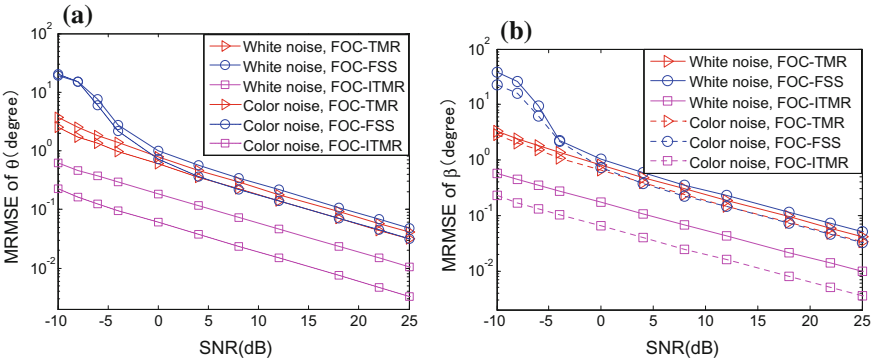



**Fig. 3** The MRMSE versus SNR in both white and color noises. **a** $\theta_i$; **b** $\beta_i$

## 5 Conclusion

In this paper, a novel FOC-ITMR method for coherent signal estimation is proposed. The whole information of the two parallel ULAs is fully utilized by changing the reference element to reconstruct the two new cumulants-based matrices, which avoids the loss of array physical aperture. Therefore, the proposed method gains excellent decoherence performance. Simulation results show the validity of the presented FOC-ITMR algorithm.

**Acknowledgements** This work was supported by the Key Talents Project for Tianjin University of Technology and Education (TUTE) (KYQD16001).

## References

1. H. Krim, M. Viberg, Two decades of array signal processing research: the parametric approach. IEEE Signal Process. Mag. **13**(4), 67–94 (1996)
2. Z. Li, K. Liu, Y. Zhao, Y. Ma, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1423–1426 (2014)
3. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. 1–6 (2016)
4. Z. Li, Y. Chen, H. Shi, K. Liu, NDN-GSM-R: a novel high-speed rail- way communication system via named data networking. EURASIP J. Wirel. Commun. Networking **2016**(48), 1–5 (2016)
5. R.O. Schmidt, Multiple emitter location and signal parameter estimation. IEEE Trans. Antennas Propag. **34**(3), 276–280 (1986)
6. R. Roy, T. Kailath, ESPRIT-estimation of signal parameters via rotational invariance techniques. IEEE Trans. Acoust. Speech Signal Process. **37**(7), 984–995 (1989)
7. S.U. Pillai, B.H. Kwon, Forward/Backward spatial smoothing techniques for coherent signal identification. IEEE Trans. Acoust. Speech Signal Process. **37**(1), 8–15 (1989)
8. X. Nie, L.P. Li, A Computationally efficient subspace algorithm for 2-D DOA estimation with L-shaped array. IEEE Signal Process. Lett. **21**(8), 971–974 (2014)

9. G. Du, N. Tong, Y. Wang, X. Jiang, Joint estimation of frequency and 2-D arrival angles of coherent signals in the presence of colored noise. Congr. Image Signal Process. **5**, 281–285 (2008)
10. H. Chen, C. Hou, Q. Wang, L. Huang, W. Yan, Cumulants-based Toeplitz matrices reconstruction method for 2-D Coherent DOA estimation. IEEE Sens. J. **14**(8), 2824–2832 (2014)

# Prior Structure-Based Sparsity Representation for Compressive Signal Feature Recovery

**Song Kong, Zhuo Sun and Xuantong Chen**

**Abstract** Compressive sampling is a promising solution to reduce required sampling rates for signal reconstruction. In many scenarios, such as cognitive radio and modulation recognition, there are only expecting to acquire useful features rather than original signals. To reconstruct these features from compressive measurements, Compressive Sensing (CS) requires features to be sparse and have a one-dimensional relationship with those measurements. Since most of features are nonlinearly transformed from signals, selecting one with high sparsity and then building a linear mapping between it and measurements become the main challenges. This work proposed a new method to find sparsity representations for signals based on their intra-structure. With this method, two common features, autocorrelation function and fourth order time-varying moment are respectively expressed as another two sparse representations called structure-based sparsity representations. Simulation shows that these representations can work effectively in reducing reconstruction iterations, computing consumption, and memory cost for sensing matrices.

## 1 Introduction

The Shannon/Nyquist sampling theorem specifies that to avoid losing information when capturing a signal, one must sample at least two times faster than the signal bandwidth. This process of massive data acquisition followed by compression is extremely wasteful [1]. In the last 10 years, Compressive Sensing, a new theory, shows that if a signal is sparse or can be compressed under a basis, it can be reconstructed from much fewer numbers of linear measurements than suggested by the Shannon sampling theory [2].

S. Kong (✉) · Z. Sun (✉) · X. Chen (✉)
Beijing University of Posts and Telecommunications, Beijing 100876, China
e-mail: kongsong@bupt.edu.cn

Z. Sun
e-mail: zhuosun@bupt.edu.cn

X. Chen
e-mail: chxt@bupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_70

In many CS scenarios, we only expect to acquire some features of original signals rather than themselves, since reconstructing signals allows for a lot of extra operations, which leads to higher complexity both in time and space domain. Based on CS, there is much valuable work related to signal features reconstruction has been carried out already [3, 4]. However, most features are nonlinear transformations of signals and selecting a suitable transformation that makes the signal sparse in this transformation basis is not straightforward to achieve. Even if the suitable transformation is found, since the CS theory is developed from one-dimensional signal processing, high-dimensional signals need dimensionality reduction and linear mapping with compressed measurements, which are always complicated to calculate [5, 6].

To solve these problems, we proposed a visualized method to find sparsity representation based on intra-structure of the signal. This method performs well in making reconstruction converge rapidly and reducing memory space and computing complexity for sensing matrices.

The rest of the paper is organized as follows. Section 2 builds the general model of structure-based sparsity representation. Section 3 discusses two practical examples of signal feature recovery using proposed method. Experimental results are provided in Sect. 4. And the conclusion are provided in Sect. 5.

## 2 Structure-Based Sparsity Representation

Many natural signals have concise representations when abstracted regularities of distribution. Consider, for example, the planform in Fig. 1a shows self-correlation of OFDM signals. The amplitudes of blocks in the same diagonal line are identical. Based on this distribution, (a) can be expressed as (b) and then common amplitude of each submatrix can be abstracted as (c). Furthermore, a concise representation in Fig. 2: several position matrices and coefficients(many of them may be zero).

Mathematically speaking, we have a vector $x \in R^N$ with elements $x[n], n = 1, 2, ..., N$. (An image or higher dimensional data can be vectorized into a long one-dimensional vector, such as Fig. 3a). Due to its intra-structure, $x$ can be divided into some subvectors:

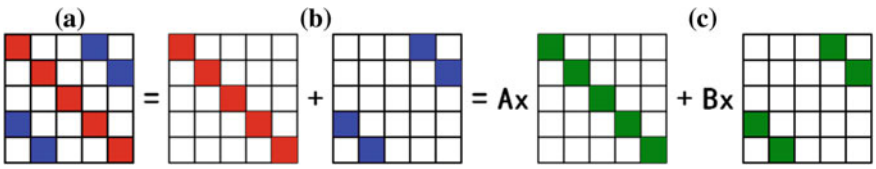



**Fig. 1** A practical example of structure-based sparsity representation. **a** Simplified 2D plane picture of OFDM signal autocorrelation function, blocks in *red* and *blue* represent different amplitudes. **b** Division of autocorrelation function based on intra-structure. **c** Extracting amplitudes, *green* blocks are normalized

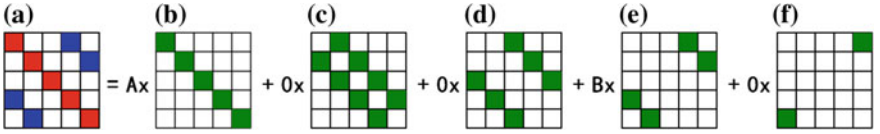



**Fig. 2** A practical example of structure-based sparsity representation. **a** Simplified 2D plane picture of OFDM signal autocorrelation function, blocks in *red* and *blue* represent different amplitudes. **b**–**f** All divisions

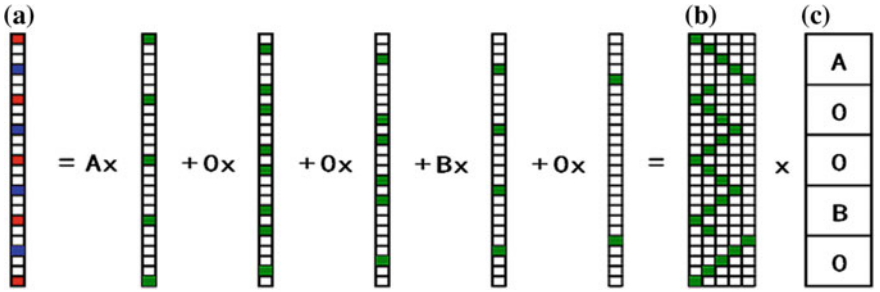



**Fig. 3** Extracting structured basis. **a** Dimensionality reduction. **b** Structured basis. **c** Coefficient vector (sparsity representation)

$$x = \sum_{i=1}^{k} x_i \tag{1}$$

where nonzero elements have the same weighting coefficient $c_i$ in each subvector (if $x_i$ is a zero vector, $c_i$ will be zero). Define a $N \times 1$ vector $\Psi_i$ called identity structured vector to make $x_i = c_i \times \Psi_i$. Then using the $N \times k$ structured basis $\Psi = [\Psi_1 \mid \Psi_2 \mid ...\Psi_k]$ with the vectors $\{\Psi_i\}_{i=1}^{k}$ as columns, the vector $x$ can be expressed as:

$$x = \Psi c \tag{2}$$

where $c$ is the $k \times 1$ column vector with elements $c_i, i = 1, 2, ...$ (such as Fig. 3c shows). Clearly, basis relates $x$ to $c$ which are two equivalent representations. In traditional compressive sensing, basis needs to be orthonormal [7]. For most synthetic signals, DCT, wavelet transform or other common transforms will be appropriate. However, since CS theory is developed from one-dimensional signal processing, the linear mapping process between the vector in transform domain and subsampling measurements introduces complicated operations such as Kronecker product and matrix vectoring which lead to a lower reconstructing accuracy and much time to calculating [8, 9]. In our work, the structured basis will simplify that process and the verification will be given in next section. For real signals, it's difficult to find sufficient orthonormal basis to make sparsity obtained work properly in CS. In fact, these signals have intra-signal correlations (they are structured enough and hence

sparse) [5]. That means our structured basis is sufficient to reconstruct real signals of interest.

Next, we aim to derive the relationship between subsampling vector $y$ and the structure-based sparsity representation $c$. Give the compressive sampling model:

$$y = Ax \tag{3}$$

where $A \in M \times N$ is a linear random sampling matrix. All random samplers used in traditional CS literature would work with high probability [10]. In our work, we chose a random Gaussian matrix as $A$.

Putting together (2) and (3), we reach:

$$y = A\Psi c = \Theta c \tag{4}$$

Now, given the compressived vector $y$, the problem of recovering $x$ boils down to solving $c$ (and hence $x = \Psi c$) from (4). In this work, the orthogonal matching pursuit (OMP) is used to reconstruct $c$.

## 3 Compressive Signal Feature Recovery

### *3.1 Autocorrelation Function Recovery Based on Structured Basis*

CS is widely used in cyclostationary-based wideband spectrum sensing. In previous researches, the sparse cyclic covariance is reconstructed to extract useful features [6, 8, 9]. Although it allows a faster recovery at low sampling costs, the linear relationship between cyclic covariance and compressive measurements is so complicated that results in a lot of computing work. Consider the cyclic covariance is given by the Fourier series of self-correlation which means autocorrelation sparse itself is the key factor of spectrum sensing. Now, there are two projects to solve complexity in reconstructing cyclic covariance: (1) Reconstruct autocorrelation directly from the simple linear relationship between itself and compressive measurements. (2) Represent autocorrelation in structured basis and then recovery structure-based sparsity representation. The first is based on traditional CS which can be solved by OMP algorithm easily so not explained here. Next, we discuss the second method in details.

The autocorrelation matrix $R_x = r_x(n, \tau) = E\{xx^T\}$ of a communication signal $x$ can be shown that:

$$R_x = \begin{bmatrix} r_x(0,0) & r_x(0,1) & \cdots & r_x(0,N-1) \\ r_x(0,1) & r_x(1,0) & \cdots & r_x(1,N-1) \\ \vdots & \vdots & \ddots & \vdots \\ r_x(0,N-1) & \cdots & \cdots & r_x(N-1,0) \end{bmatrix} \tag{5}$$

where the structure of $R_x$ is related to time interval $\tau$ based on prior knowledge. Define $N \times N$ matrices $\{M_\tau\}_{\tau=1,...,N}$ with respect to $\tau$, as follows:

$$M_1 = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & 1 \end{bmatrix}, ..., M_N = \begin{bmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{6}$$

Further, we can express the structured matrix $\Psi$ as $[\Psi_1 \mid \Psi_2 \mid ...\Psi_k]$ where $\Psi_i = vec\{M_i\}$ is the vector stacked by all columns from $M_i$. Following (1) and (2), autocorrelation matrix can be expanded as

$$vec\{R_x\} = \sum_{i=1}^{N} c_i \Psi_i = \Psi c \tag{7}$$

It holds evidently from (3) that $R_y = AR_xA^H$. Meanwhile, $R_y$ is symmetric positive semi-definite which means there are only $M(M+1)/2$ unique terms of $R_y$. We organize them into a vector $r_y$ of length $M(M+1)/2$, as follows:

$$\begin{aligned} r_y = [&r_y(0,0), r_y(1,0), ..., r_y(M-1,0), r_y(0,1), r_y(1,1), \\ &..., r_y(M-2,1), ....., r_y(0,M-1)]^T \end{aligned} \tag{8}$$

Hence, $R_y$ can linearly related to its vectorized counterpart $r_y$ as

$$r_y = Q_M vec\{R_y\} \tag{9}$$

where $Q_M \in \{0, 1/2, 1\}^{(M(M+1)/2)\times M^2}$ is the mapping matrix that maps the entries in $r_y$ with corresponding ones in $vec\{R_y\}$. Accordingly, it can be shown that:

$$r_y = Q_M vec\{AR_xA^H\} \tag{10}$$

Finally, using (7) and (10), we can express the measurement vector $r_y$ as a linear function of the structure-based coefficient vector $c$ as:

$$\begin{aligned} r_y &= Q_M vec\{R_y\} = Q_M vec\{AR_xA^H\} \\ &= Q_M(A \otimes A) vec\{R_x\} = Q_M(A \otimes A)\Psi c \\ &= \Theta c \end{aligned} \tag{11}$$

where $\otimes$ denotes Kronecker product operation and $\Theta = Q_M(A \otimes A)\Psi$ is the sensing matrix. Once $c$ is reconstructed, the autocorrelation $vec\{R_x\}$ can be easily calculated by $vec\{R_x\} = \Psi c$.

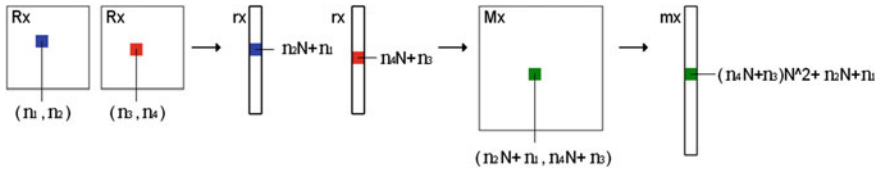



**Fig. 4** The mapping relationship of elements

## *3.2 Fourth-Order Time-Varying Moment Recovery Based on Structured Basis*

This subsection we will show you how to use proposed method to reconstruct fourth-order time-varying moment as an example of high-order features recovery.

Since CS is used for one-dimensional data reconstruction, we adopt the following method to reduce the dimension of fourth-order time-varying moment, which can be expressed as

$$M_x = r_x r_x^T \tag{12}$$

where $r_x = vec\{R_x\} = vec\{E\{xx^T\}\}$ and $M_x$ is a $N^2 \times N^2$ matrix. Further, the one-dimensional expression of $M_x$ is obtained by stacking all columns into a vector:

$$m_x = vec\{M_x\} \tag{13}$$

where $m_x$ is a $N^4 \times 1$ vector and the relation of elements is shown in Fig. 4.

Similar to autocorrelation matrix $R_x$, the structure of $M_x$ is related to time intervals $n_1 - n_2$, $n_3 - n_4$, $n_1 - n_4$, $n_2 - n_3$ and $n_1 - n_2 + n_3 - n_4$ where only three of them are independent. We choose $\tau_1 = n_1 - n_2$, $\tau_2 = n_3 - n_4$ and $\tau_3 = n_1 - n_4$ as factors constructing the structured basis of $m_x$.

Define $K$ $N^4 \times 1$ identity structured vectors $\{\Psi_i\}_{i=1,\ldots,K}$, where $K = (2 \times N - 1)^3$ is with respect to the ranges of $\tau_1$, $\tau_2$ and $\tau_3$. The *j-th* element in $\Psi_i$ can be expressed as:

$$\Psi_i(j) = \begin{cases} 1 & (\tau_3 + N - 1)K^2 + (\tau_2 + N - 1)K + N = i, \\ 0 & \text{else.} \end{cases} \tag{14}$$

Finally, the linear relationship between compressed measurements and fourth-order time-varying moment can be shown that

$$\begin{aligned} m_y &= vec\{M_y\} = vec\{r_y r_y^T\} = (A \otimes A \otimes A \otimes A) m_x \\ &= (A \otimes A \otimes A \otimes A)\Psi c = \Theta c \end{aligned} \tag{15}$$

where $\Theta$ is the $M^4 \times K$ sensing matrix.

### 3.3 Performance Discussion

The core of structure-based sparsity representation is decomposing the signal into several subsignals based on intra-structure and reconstructing one subsignal every iteration, where each subsignal contains at least one element of original signal. That means recovering vector $c$ in (4) needs less iterations contrasting with recovering original signal $x$. For autocorrelation, there are only two nonzero subvectors as Fig. 3 shows which means feature can be recovered roughly after twice iterating on ideal condition based on structured basis and small disturbances at other zero elements will be ignored. However, it needs nearly $2 \times N$ iterations to reconstruct autocorrelation of a N-length signal directly, so proposed method can reduce iterations apparently where $N$ will be large in reality. The same applies to fourth-order time-varying moment which has twenty nine nonzero subvectors need to be reconstructed less than $0.3 \times N^4$ directly reconstructing iterations. Note that there is more than one way to define structured basis and above discussions are based on the way raised in previous two subsections.

Proposed method represents signal in a concise expression which not only reduces reconstruction iterations but also contributes to low computing consumption and memory cost for sensing matrices. Cyclic covariance and autocorrelation are common features in signal detection. The linear relationship between cyclic covariance and sub-Nyquist samples is derived in [8]:

$$r_y = \Phi H^{\dagger} vec\{\widetilde{R}_x^c\} \tag{16}$$

where $\Phi = Q_M(A \otimes A)P_N$ is of size $(M(M+1)/2) \times (N(N+1)/2)$ and $H^{\dagger}$ is the (pseudo) inverse of $H = \sum_{v=0}^{N-1} D_v^T \otimes G_v$

Meanwhile, $r_y$ is linearly related to $vec\{R_x\}$ as follows:

$$r_y = \Phi r_x = Q_M(A \otimes A)vec\{R_x\} \tag{17}$$

From (16) and (17), sensing matrices have the same size $(M(M+1)/2) \times N^2$ resulting in same storage space. But pseudo-operation in (16) increases computing consumption of sensing matrix, so getting cyclic covariance by calculating Fourier transform of autocorrelation needs less consumption. Further, reconstructing autocorrelation using (11) contributes to less memory cost compared with (17).

## 4 Simulations

This section presents simulation results that testify the effectiveness of the proposed algorithm in reducing reconstruction iterations and storage space of the sensing matrix compared with traditional CS. 5000 symbols of the signal are sent under

rectangle filter through AWGN and sampled with 0.375 subsampling rate ($M/N$) at the receiving terminal.

## 4.1 Iterations and Reconstruction Error

In this section, we use two methods to recover autocorrelation, one is reconstructing it directly and another is based on structure-based sparsity representation. We adopt relative error $\epsilon$ defined in (18), where $x^*$ is the solution found for the compressive equation $y = \Theta x$ and $y$ is the real input signal.

$$\epsilon = \frac{\| y - \Theta x^* \|_2}{\| y \|_2} \tag{18}$$

Figure 5 shows the reconstruction error $\epsilon$ versus the iterations for autocorrelation reconstruction based on traditional CS and our proposed algorithm. Compared with the traditional orthonormal basis method, the result shows that the reconstruction under structured basis converges more rapidly. Generally, it will converge after iterating 2–4 times which is much less than the traditional method. This is because every element in structure-based sparsity vector $c$ is equivalent to at least one element in original vector $x$ and that means we can recover several elements of $x$ in every iteration using structured basis method which makes reconstruction more rapid.

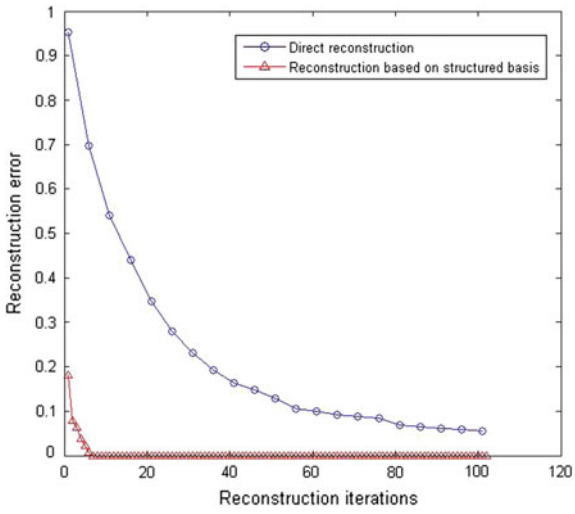



**Fig. 5** Reconstruction error of autocorrelation function based on traditional method and structured basis method

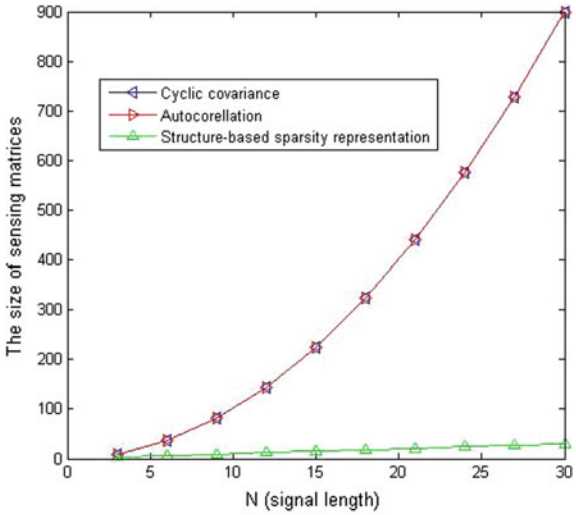



**Fig. 6** The size of sensing matrices (To contrast clearly, we calculate the size ignoring the common part $M \times (M+1)$)

## *4.2 Memory Cost of Sensing Matrix*

Figure 6 shows the memory cost of sensing matrices in (16) (for cyclic covariance reconstruction), (17) (for autocorrelation reconstruction) and (11) (for structure-based representation of autocorrelation reconstruction) versus $N$. Since sensing matrices in three equations have the common part $Q_M(A \otimes A)$, the calculating of matrix size ignores this part. As we can seen, the sensing matrices in (16) and (17) need the same space coinciding with discussions in last section. And the space is growing exponentially along with signal length $N$. However, the space for sensing matrix in (11) grows very slowly. So, our method can effectively reduce memory cost for sensing matrices.

# 5 Conclusion

To solve the problem of finding suitable transformation for signals, we have proposed a structure-based method to get sparsity representation. Two features are reconstructed as practical examples based on our method, one of which is autocorrelation function, and the other is fourth-order time-varying moment. We have built linear relationships referring to these two features. Simulations show that structure-based sparsity representation can work effectively in reducing computing consumption, memory cost for sensing matrices and reconstruction iterations.

## References

1. R.G. Baraniuk, Compressive sensing. Signal Process. Mag. (2007). IEEE
2. B. Li, Z. Zhu, F. Liu, Wavelet-based compressed sensing using low frequency coefficients, in *Fourth International Symposium on Information Science and Engineering* (2012), pp. 119–123
3. B. Huang, L. Gang, Alternating optimization of sensing matrix and sparsifying dictionary for compressed sensing. Trans. Signal Process. (2015). IEEE
4. B. Dongjie, X. Yongle, A sparsity basis selection method for compressed sensing. Trans. Signal Process. (2015). IEEE
5. G. Quer, R. Masiero, D. Munaretto, On the interplay between routing and signal representation for compressive sensing in wireless sensor networks
6. E. Rebeiz, V. Jain, D. Cabric, Cyclostationary-based low complexity wideband spectrum sensing using compressive sampling. ICC (2012). IEEE
7. E.J. Candes, M.B. Wakin, An introduction to compressive sampling. IEEE Signal Process. Mag. **25**(2), 21–30 (2008)
8. Z. Tian, Y. Tafesse, B.M. Sadler, Cyclic feature detection with sub-Nyquist sampling for wideband spectrum sensing. IEEE J. Sel. Top. Signal Process. **6**(1), 58–69 (2012)
9. Z. Tian, Cyclic feature based wideband spectrum sensing using compressive sampling. ICC (2011). IEEE
10. M. Davenporttand, P. Boufounosand, M. Wakin, Signal processing with compressive measurements. IEEE J. Sel. Top. Signal Process. (2007). ISCAS

# A Time Error Model for Correlated Double Sampling PWM Pixel

**Lu Yu, Yun Hao, Zhonghe Chen, Yali Wang and Xihong Ye**

**Abstract** A time error mathematical model of nonlinear response and noise is established in order to achieve low time error of correlated double sampling pulse-width-modulation (CDS PWM) pixel. By analysis and computer, the time error under fixed pattern noise (FPN), reset noise, shot noise, reference noise of pixel-level comparator, an integrated model is established. By means of the model, appropriate selection of high/low reference voltage of pixel-level comparator can decrease time error of CDS PWM. These results serve as a guideline for the design of CDS PWM pixel.

**Keywords** CMOS image sensor · PWM · Nonlinearity · Noise · Time error

## 1 Introduction

In the field of CMOS Image Sensor (CIS), the traditional active pixel sensors (APS) are just in a situation of bottleneck with the decrease of the size and power supply of COMS process. However, digital pixel sensors (DPS) demonstrate excellent performance under the small size process, due to their exposure measurement and quantization in pixels. The researches on pulse-width modulation (PWM) mode can be used in DPS, which reduces the effect of readout noise [1]. It combines with an asynchronous readout method based on the address-event representation (AER), which only outputs the useful part of the visual information, with the advantages of low redundancy, high-resolution, and real-time visual sensory characteristics [2–4].

The pixel of PWM exposure measurement circuit quantitates light intensity by time, so the readout noise and row noise can be eliminated fundamentally [1]. Compared with traditional PWM pixel, correlated double sampling (CDS) PWM pixel inhibits reset noise, and fixed pattern noise (FPN) effectively, and also gets a

L. Yu (✉) · Y. Hao · Z. Chen · Y. Wang · X. Ye
Zhonghuan Information College Tianjin University of Technology, Tianjin, China
e-mail: yuluyouxiang@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_71

lower time error, by comparing the voltage of pixel with two reference voltages [6, 7]. But time error will generates in that nonlinear response and nonlinear noise distort quantization time in CDS PWM. This kind of time error is determined, and image quality is improved by appropriate selection of two reference voltages in pixel-in comparator.

This paper searches on the time error of FPN, reset noise, shot noise, and reference noise of pixel-level comparator, in nonlinear response. A time error model in nonlinear response is proposed for CDS PWM pixel design, which can be used in high-resolution, asynchronous, and AER time-based CIS [4, 5].

## 2 Structure and Principle of CDS PWM Pixel

The structure of CDS PWM pixel is shown in Fig. 1. It is composed of photodiode PD, CMOS reset transistor Mrst, pixel-level comparator with alterable reference voltage, logical control, counter, and memory.

The principle of CDS PWM pixel structure is shown in Fig. 2. Reset transistor Mrst turns on, then reset voltage makes photodiode PD reset. An exposure measurement cycle starts. With exposure process, $V_{pix}$, output voltage of PD, decreases linearly with time t. At the same time, switch $S_1$ is on, $S_2$ is off. $V_{ref}$, the reference voltage of comparator equals to $V_{refh}$. The pixel-level comparator begins to work. When $V_{pix}$ reduces to $V_{refh}$, output of comparator inverts. This time denotes $t_1$. Then Logic and Control controls $S_1$ off, $S_2$ on. So the reference voltage of comparator equals to $V_{refl}$. $V_{pix}$ keeps falling linearly. Output of comparator inverts till $V_{pix}$ equals to $V_{refl}$. This time denotes $t_2$. Finally, the counter quantifies time difference $t_{int}$, which equals to $t_1$ subtracting $t_2$, and puts it in pixel-level memory, waiting for read out.

Time difference $t_{int}$ is defined as

$$t_{\text{int}} = t_2 - t_1 = \int_{V_{refl}}^{V_{refh}} \frac{C_{PD}}{I_{ph}} dV_{pix} = \frac{\left(V_{refh} - V_{refl}\right) \cdot C_{PD}}{I_{ph}} \tag{1}$$

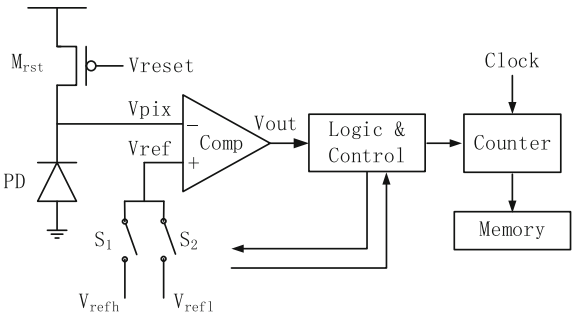



**Fig. 1** Structure of CDS PWM pixel

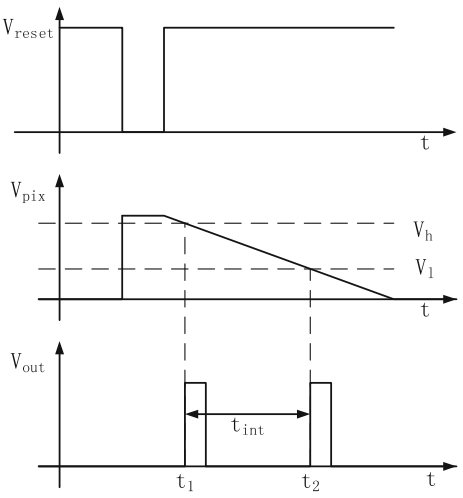



**Fig. 2** Principle of CDS PWM pixel structure

where $C_{PD}$ is photodiode capacitance, Iph is photocurrent. Because Iph is proportional to light intensity L, $t_{int}$ is also proportional to $L^{-1}$. That makes $t_{int}$ reflect light intensity linearly.

## 3 Effect of Nonlinear Response on CDS PWM

Ideally, $V_{pix}$ is linear response to time. But in nonlinear situation, CPD, photodiode capacitance, is not a constant, changing with $V_{pix}$ as

$$C_{PD} = C_{PD0}\left(1 + \frac{V_{pix}}{\phi_0}\right)^{-m} \tag{2}$$

where $C_{PD0}$ is static voltage of PD. m is slope coefficient, which general value is 1/3. $\phi_0$ is gate potential.

So tint can be corrected as

$$t_{\text{int}} = \frac{1}{I_{ph}} \int_{V_1}^{V_2} C_{PD}(V_{pix})\, dV_{pix} = \frac{C_{PD0} \cdot \phi_0}{(1-m) I_{ph}} \left[ \left(1 + \frac{V_2}{\phi_0}\right)^{1-m} - \left(1 + \frac{V_1}{\phi_0}\right)^{1-m} \right] \tag{3}$$

where $V_1$ and $V_2$ are initial point and ending point of $V_{pix}$.

The time error made by nonlinear response with same $V_{refh}$ and $V_{refl}$ is showed by Fig. 3.

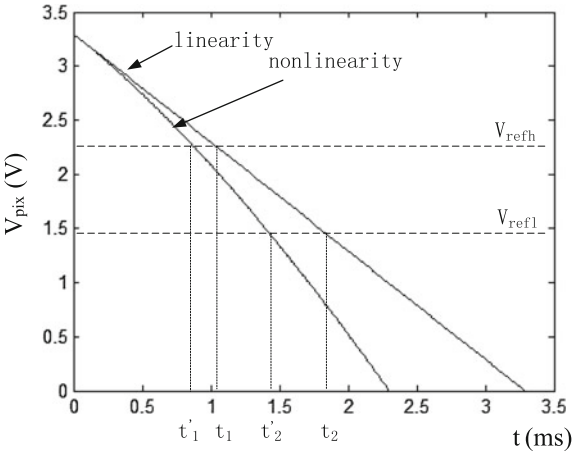



**Fig. 3** Nonlinear response and linear response

Time error parameter $\varepsilon_t$ is defined to measure this error as

$$\varepsilon_t = \frac{\left(t_2^{'} - t_1^{'}\right) - (t_2 - t_1)}{t_2 - t_1} \tag{4}$$

where $t_1$ is the time of comparator inverting when $V_{refh}$ equals to $V_{refh}$ in ideal conditions. Similarly, $t_1$ is the time of comparator inverting when $V_{refh}$ equals to $V_{refl}$ also in ideal conditions, while $t'_1$, $t'_2$ is in nonlinear response.

According to the formula (3) and (4), time error in nonlinear response $\varepsilon_{t,nonlinearity}$ is calculated as

$$\varepsilon_{t,nonlinearity} = \frac{\phi_0}{1-m} \frac{\left(1 + \frac{V_{refh}}{\phi_0}\right)^{1-m} - \left(1 + \frac{V_{refl}}{\phi_0}\right)^{1-m}}{V_{refh} - V_{refl}} - 1 \tag{5}$$

Figure 4 shows $\varepsilon_{t,nonlinearity}$ with different $V_{refh}$ and $V_{refl}$, while m = 1/3, $\phi_0 = 0.7$. $\varepsilon_{t,nonlinear}$ is minished by lower $V_{refh}$ and lower $V_{refh}$.

## 4 Time Errors of Noise on the Nonlinear Response

The noises of CIS come from time-dependent random noise and time-independent pattern noise. Random noise contains thermal noise, shot noise, 1/f noise, reset noise, source follower amplifier noise, column amplifier noise, and readout noise. Pattern noise contains FPN and photo response non uniformity (PRNU).

CDS PWM measures light intensity in pixel, so source follower amplifier noise and readout noise are not generated. As mentioned earlier, FPN and reset noise are eliminated under ideal conditions. But they cannot be ignored in nonlinear response.

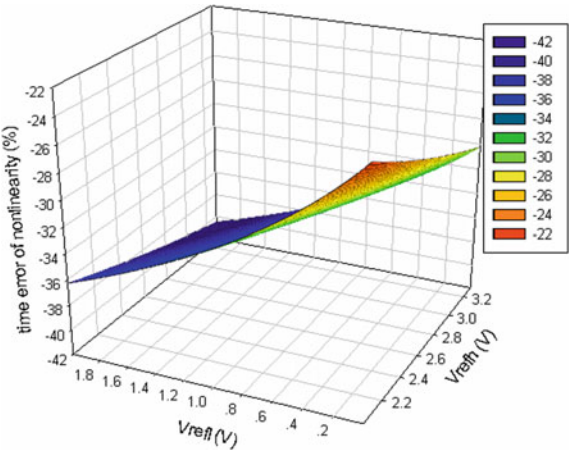



**Fig. 4** Relationship between $V_{ref}$ and $\varepsilon_{t,nonlinear}$

## 4.1 Time Error of FPN

FPN contains pixel FPN and column FPN that is caused by column readout circuit. CDS PWM measures light intensity in pixel, so pixel FPN is the main. The pixel FPN comes from offset voltage in comparator.

The structure of pixel-level PWM comparator is shown as Fig. 5.

Offset voltage of pixel-level comparator. $V_{os}$, is defined as

$$V_{os} = (V_{GS1,2} - V_{TH1,2}) \frac{\Delta(W/L)_2 - \Delta(W/L)_1}{2(W/L)_{1,2}} + (\Delta V_{TH1} - \Delta V_{TH2}) + \frac{g_{m3,4}}{g_{m1,2}} [(V_{GS3,4} - V_{TH3,4}) \times \frac{\Delta(W/L)_4 - \Delta(W/L)_3}{2(W/L)_{3,4}} + (\Delta V_{TH3} - \Delta V_{TH4})] \tag{6}$$

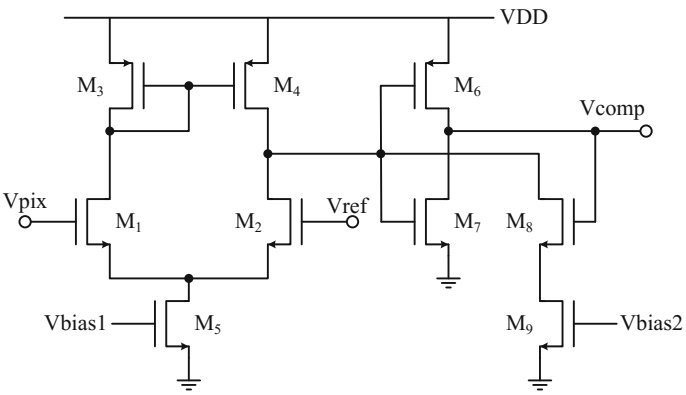



**Fig. 5** Pixel-level comparator schematic

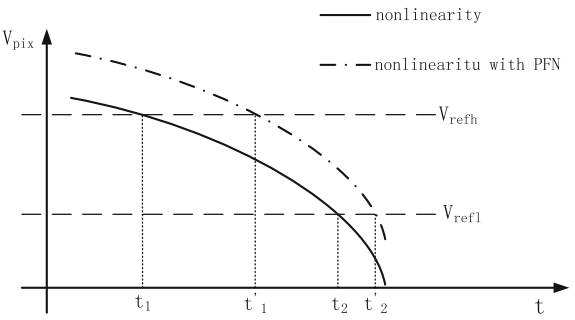



**Fig. 6** Response of nonlinearity with FPN

where $\Delta(W/L)_i = (W/L)_i - W/L$, $\Delta V_{TH} = V_{THi} - V_{TH} = \frac{A_{VTH}}{\sqrt{WL}}$, $A_{VTH}$ is a scaling factor associated with the process.

The response of nonlinearity with FPN can be shown as Fig. 6.

According to the formula (4), time error of FPN can be presented as

$$\varepsilon_{n,FPN} = \frac{\left(1+\frac{V_{refh}-V_{os}}{\phi_0}\right)^{1-m} - \left(1+\frac{V_{refl}-V_{os}}{\phi_0}\right)^{1-m}}{\left(1+\frac{V_{refh}}{\phi_0}\right)^{1-m} - \left(1+\frac{V_{refl}}{\phi_0}\right)^{1-m}} - 1 \tag{7}$$

## 4.2 Time Error of Random Noise

Random noise of CDS PWM includes thermal noise, shot noise, 1/f noise, and reset noise.

Time error parameter $\overline{\varepsilon_n^2}$ is defined to measure random noise in nonlinear conditions.

$$\overline{\varepsilon_n^2} = \frac{\left|-(t_1' - t_1)^2 + (t_2' - t_2)^2\right|}{(t_2 - t_1)^2} \tag{8}$$

### 4.2.1 Time Error of Reset Noise

Reset noise is also known as kTC noise. Reset process of a pixel can be seen as a channel resistance of MOS transistor charge-discharge to PD. The thermal noise in channel is reset noise, which is expressed as

$$\overline{v_{rst}} = \sqrt{\frac{kT}{C_{PD0}}} \tag{9}$$

K = 1.38 × $10^{-23}$ J/K, T is kelvin degree.

According to the formula (8) and (9), time error of reset noise can be presented as

$$\overline{\varepsilon_{n,rsr}^2} = \frac{-\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refh} - \overline{V_{rst}}\right)^{1-m}\right]^2}{\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refl}\right)^{1-m}\right]^2} + \frac{\left[\left(\phi_0 + V_{refl}\right)^{1-m} - \left(\phi_0 + V_{refl} - \overline{V_{rst}}\right)^{1-m}\right]^2}{\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refl}\right)^{1-m}\right]^2} \tag{10}$$

### 4.2.2 Time Error of Shot Noise

Shot noise is a kind of white noise, which is caused by random variation of charge carriers in P–N junction. It is composed of photoelectron shot noise and dark current shot noise.

Photoelectron shot noise can be described as

$$\overline{v_{ph}} = \frac{\sqrt{qI_{ph}t}}{C_{PD}} \tag{11}$$

Dark current shot noise can be described as

$$\overline{v_{dk}} = \frac{\sqrt{qI_{dk}t}}{C_{PD}} \tag{12}$$

where t is expose time, $I_{ph}$ is photocurrent, and $I_{dk}$ is dark current.

Shot noise is:

$$\overline{v_{shot}} = \sqrt{v_{ph}^2 + v_{dk}^2} = \frac{\sqrt{q\left(I_{ph} + I_{dk}\right)t}}{C_{PD}} \tag{13}$$

According to the formula (2), (3), and (13), shot noise can be presented as

$$\overline{V_{shot}^2\left(V_{pix}\right)} = \frac{q\left(I_{ph} + I_{dk}\right)\phi_0}{(1-m)C_{PD0}I_{ph}} \cdot \frac{\left(1 + \frac{V_{rst}}{\phi_0}\right)^{1-m} - \left(1 + \frac{V_{pix}}{\phi_0}\right)^{1-m}}{\left(1 + \frac{V_{pix}}{\phi_0}\right)^{-2m}} \tag{14}$$

Response of nonlinearity with shot noise is as Fig. 7.

According to the formula (8), time error of shot noise can be presented as

$$\overline{\varepsilon_{n,shot}^2} = \frac{-\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refh} - \overline{v_{shot}\left(V_{refh}\right)}\right)^{1-m}\right]^2}{\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refl}\right)^{1-m}\right]^2} + \frac{\left[\left(\phi_0 + V_{refl}\right)^{1-m} - \left(\phi_0 + V_{refl} - \overline{v_{shot}\left(V_{refl}\right)}\right)^{1-m}\right]^2}{\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refl}\right)^{1-m}\right]^2} \tag{15}$$

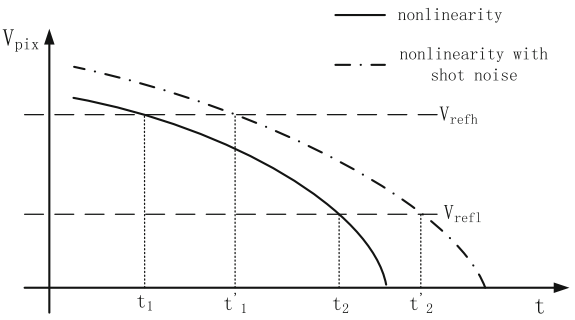



**Fig. 7** Response of nonlinearity in shot noise

### 4.2.3 Time Error of Reference Voltage

Pixel-level comparator plays an important role in measuring light intensity. There are 1/f noise and thermal noise in the comparator. The reference noise voltage, $\overline{v_{noise,ref}^2}$ can describe these noises.

$$\overline{v_{noise,ref}^2} = \int_0^\infty \left[8kT\left(\frac{2}{3g_{m1,2}} + \frac{2g_{m3,4}}{3g_{m1,2}^2}\right) + \frac{2K_N}{C_{ox}(WL)_{1,2}f} + \frac{2K_P}{C_{ox}(WL)_{3,4}f}\frac{g_{m3,4}^2}{g_{m1,2}^2}\right] \cdot |H(f)|^2 df \tag{16}$$

where H(f) is transfer function of the comparator. $K_N$ and $K_P$ are constants related to production process, which magnitude is $10^{-25}V^2F$. $C_{ox}$ is the gate oxide capacitance per unit area.

Variations of $V_{refh}$ and $V_{refl}$ make comparator's invert distort that brings time error shown as Fig. 8.

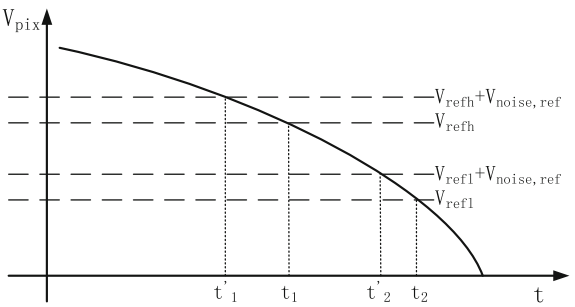



**Fig. 8** Response of nonlinearity in reference noise

According to the formula (8), time error of reference voltage can be presented as

$$\overline{\varepsilon^2_{n,ref}} = \frac{-\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refh} + \overline{v_{noise,ref}}\right)^{1-m}\right]^2}{\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refl}\right)^{1-m}\right]^2} + \frac{\left[\left(\phi_0 + V_{refl}\right)^{1-m} - \left(\phi_0 + V_{refl} + \overline{v_{noise,ref}}\right)^{1-m}\right]^2}{\left[\left(\phi_0 + V_{refh}\right)^{1-m} - \left(\phi_0 + V_{refl}\right)^{1-m}\right]^2} \tag{17}$$

## 5 Model and Analysis of CDS PWM

CDS PWM can effectively inhibit reset noise and FPN in linear response that modifies traditional pixel. But time error will born from nonlinear response and the noise of nonlinear response. The noise of CDS PWM mainly constitutes FPN, reset noise, shot noise, reference voltage noise of comparators.

According to the model in the paper, time error in nonlinear condition can be calculated.

where GSMC 0.18 μm CMOS process technology is used. Area of photodiode is $9 \times 9\ \mu m^2$. $C_{PD0}$ is 36.98 fF. Dark current is 1 nA/cm$^2$. When m = 1/3, $\phi_0 = 0.7$, photoelectron current is $37.09 \times 10^{-12}$ A.

In the range of $V_{refh}$: 2–3.3 V, $V_{refl}$: 0–2 V, $V_{os}$ = 37.9543 mV caused by FPN. Time error is from −0.8730 to −0.3812%.

Higher $V_{refh}$ and higher $V_{refh}$ brings lower $\varepsilon_{n,FPN}$ as Fig. 9 shown.

Equivalent noise voltage and time error of noises are presented as Table 1.

The total time error is

$$\overline{\varepsilon_n} = \sqrt{\overline{\varepsilon^2_{n,rst}} + \overline{\varepsilon^2_{n,shot}\left(V_{refh}, V_{refl}\right)} + \overline{\varepsilon^2_{n,ref}\left(V_{refh}, V_{refl}\right)}} \tag{18}$$

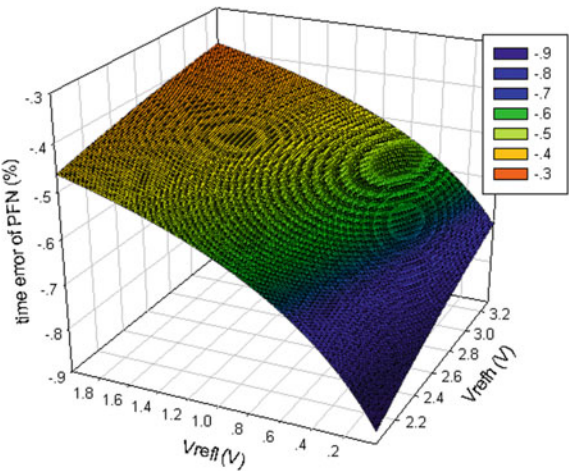



**Fig. 9** Relationship between $V_{ref}$ and $\varepsilon_{n,FPN}$

**Table 1** Random noise with nonlinearity

| Noise | Equivalent noise voltage (mV) | $\overline{\varepsilon_n}$ (%) |
|---|---|---|
| $V_{rst}$ | 0.3346 | 0.0112–0.1661 |
| $V_{shot}(V_{refh})$ | 0–2.8609 | 0.1374–2.6065 |
| $V_{shot}(V_{refl})$ | 2.8609–3.4063 | |
| $V_{noise,ref}$ | 32.5980 | 1.0798–16.1025 |

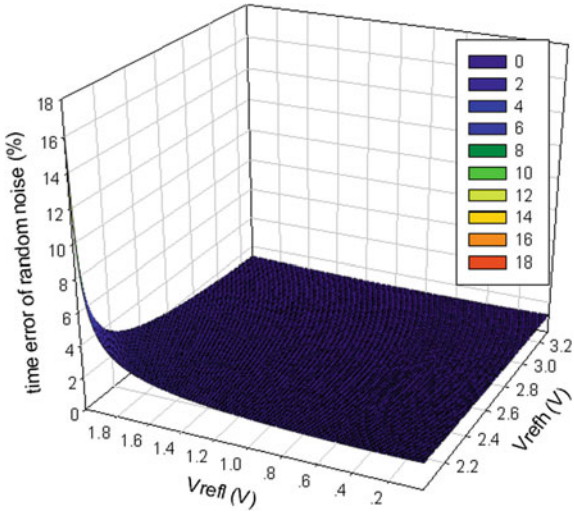



**Fig. 10** Relationship between $V_{ref}$ and $\overline{\varepsilon_n}$

$\overline{\varepsilon_n}$ is related to the value of $V_{refh}$ and $V_{ref}$. It is in the range of 1.0916–16.3129%, and 99.7885% of it is below 5%. Higher $V_{refh}$ and lower $V_{refh}$ brings lower $\overline{\varepsilon_n}$ as Fig. 10 shown.

## 6 Conclusions

To achieve minimum time error, appropriate selection of high/low reference voltage of pixel-level comparator is necessary:

1. Appropriate smaller $V_{refh}$

   Smaller time error is achieved with lower $V_{refh}$. But it also brings lager time error of FPN and random noise. So the value must be compromised.

2. Appropriate smaller $V_{refl}$

Smaller time error is also achieved by lower $V_{refl}$. But it also brings lager time error of FPN and time of exposure that lets to lower frame frequency. So the value must take response of nonlinearity, time error of random noise, and frame frequency into account.

In this paper, a time error mathematical model of nonlinear response and noise is established in order to achieve low time error of CDS PWM pixel. Theoretical analysis and simulation show that appropriate smaller reference voltage of pixel-level comparator can bring minimum time error. These results serve as a guideline for the design of CDS PWM pixel.

## References

1. M. Zhang, A. Bermark, Compressive acquisition CMOS image sensor: from the algorithm to hardware implementation. IEEE Trans. Very Large Scale Integr. VLSI Syst. **18**(3), 490 (2010)
2. C. Posch, D. Matolin, R. Wohlgenannt, A QVGA 143 dB dynamic range frame-free PWM image sensor with lossless pixel-level video compression and time-domain CDS. IEEE J. Solid-State Circuits **46**(1), 259–275 (2011)
3. C. Posch, D. Matolin, R. Wohlgenannt, High-DR frame-free PWM imaging with asynchronous AER intensity encoding and focal-plane temporal redundancy suppression, in *IEEE International Symposium on Circuits and Systems*, (Paris, 2010), pp. 2430–2433
4. J. Xu, D. Li, S. Yao, A time error correction method applied to high-precision AER asynchronous CMOS image sensor. J. Signal Process. Syst. **75**(1), 1–13 (2014)
5. Y. Lu, Y. Suying, X. Jiangtao, An implementation method of real-time vision sensor based on address event representation. Acta Optica Sinica **2013**(1), 251–257
6. C. Posch, D. Matolin, R. Wohlgenannt, High-DR frame-free PWM imaging with asynchronous AER intensity encoding and focal-plane temporal redundancy suppression, in *IEEE International Symposium on Circuits and Systems*, (Paris, 2010), pp. 2430–2433
7. D. Matolin, R. Wohlgenannt, M. Litzenberger et al., A load-balancing readout method for large event-based PWM imaging arraysm in *Proceedings of 2010 IEEE International Symposium on Circuits and Systems (ISCAS)*, (Paris, 2010), pp. 361–364

# A Wide Adjusting Range Frequency-Locking Scheme for Homodyne Coherent Receiver

**Yupeng Li**

**Abstract** A frequency-locking scheme with wide adjusting range in homodyne coherent receiver is presented and experimentally demonstrated. FPGA is used as the core of the control unit to detect and adjust the frequency difference of local light and signal light. With the beat frequency detection, the frequency difference can be well controlled within 2 MHz, and the data signal can be recovered well.

**Keywords** OPLL · Frequency-locking · FPGA · OVCO

## 1 Introduction

Homodyne coherent detection is a good solution for high-speed fiber communication circumstance. With the coherent detection, we can get high sensitivity and eliminate the noise well. OpticalPhase-Locked Loops (OPLL) are crucial system for homodyne coherent detection, and received great attention at the beginning of last 90s [1–3]. However, they never found practical applications because of their complexity and the invention of EDFAs, which greatly solved the receiver sensitivity problems. In recent years, however, coherent detection has come again in several scenarios, such as multilevel optical phase modulation, dispersion compensation, ultra-dense WDM, and microwave photonics.

The previously proposed OPLLs were based on direct laser frequency tuning, which required complex optoelectronic devices and were hard to control.

In this paper, we proposed and experimentally demonstrated a simpler frequency-locking architecture based on Optical Voltage-Controlled Oscillator (OVCO), frequency difference capture (FDC), and loop filter (LF). The frequency tuning is obtained through optical Sub-Carrier (SC) and the control algorithm is executed by FPGA.

Y. Li (✉)
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
e-mail: fx_lyp@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_72

## 2 Theoretical and Simulation Analyses

The structure of the proposed scheme is shown in Fig. 1. The scheme consists of three main parts, the OVCO, the FDC, and the LF. OVCO is used to generate the local light, which is based on a Continuous Wave (CW) tunable laser at frequency $f_{LO}$. The light is externally modulated by the signal coming from an electrical VCO at frequency $f_{VCO}$. The Mach–Zehnder modulator (MZM) is biased at Null point of its transfer function. A sinusoidal carrier-suppressed modulation is obtained, and the spectrum the OVCO output is shown in Fig. 2. Two main SCs at frequency $f_{LO} \pm f_{VCO}$ are generated. In addition, the superfluous optical tones at $f_{LO}$ and $f_{LO} \pm 2 \cdot f_{VCO}$ are generated as well. Using one of the two main SCs, take the one at $f_{LO} - f_{VCO}$ as an example, we can tune an optical frequency of the local light by changing the signal frequency of the electrical VCO output. In other words, by simply changing the voltage applied to the electrical VCO, we can control the frequency of local light conveniently. In addition to OVCO, FDC, and LF are also important to the OPLL, the FDC is used for initial quickly adjust when the frequency difference is about several GHz, and the LF is used for locking when the frequency difference is less than several MHz.

By combining of signal light and local light, optical spectrum of signal light is translated to base-band at the photodiode output. In principle, Because of the superfluous component of the local light, copies of this signal appears around frequencies $f_{VCO}$ and $2 \cdot f_{VCO}$ as well. However, they can be filtered out by the receiver filter, provided that $f_{VCO}$ is larger than the signal spectral width.

Actually, with the proposed architecture, we can obtain optical fine frequency tuning with the speed and stability of an electrical VCO, and reuse the typical RF PLL setups.

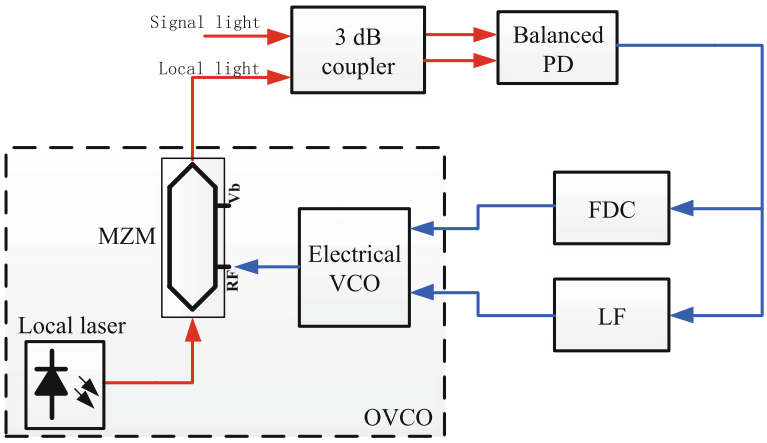



**Fig. 1** The structure of frequency-locking scheme

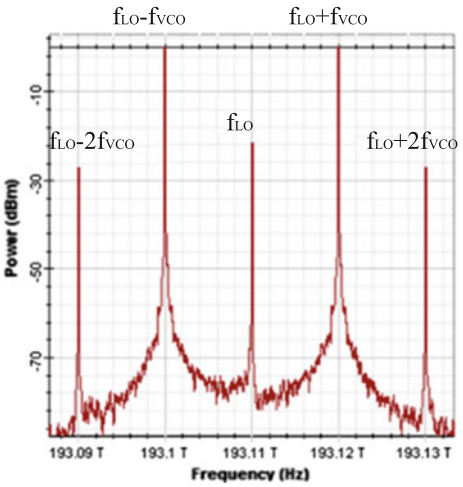



**Fig. 2** The spectrum of the OVCO output

Other details can be found in Fig. 1. A 3 dB coupler is used for combining the signal light and local light, and the balanced photodiode (BPD) is used to change the optical signal to electrical signal. The resulting electrical signal is processed by FDC and LF.

We took a simulation to verify the scheme, Fig. 3 is the simulation diagram. The lindwidth of the laser is 10 kHz, the BPSK signal is modulated by a phase modulator, the data bit is 5 Gb/s. The MZM is biased at Null point, and the lower frequency SC is chosen for beating with the signal light. The central frequency of the signal light is 193.1 THz, and the central frequency of local laser is 193.11 THz. After modulated by the 10 GHz electrical sine signal, the frequencies of the SCs are 193.1 THz and 193.12 THz respectively, as shown in Fig. 4a. With the low pass filter (LPF), the high-frequency component is eliminated and the original signal is well recovered (Fig. 4b). The simulation results verify the function of the scheme.

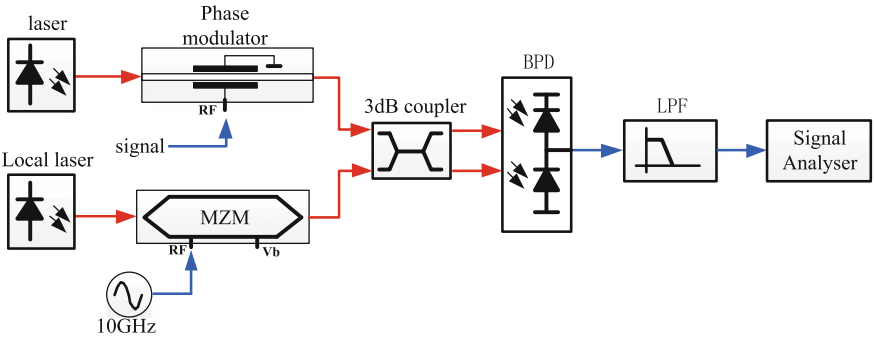



**Fig. 3** Simulation diagram of the coherent receiver

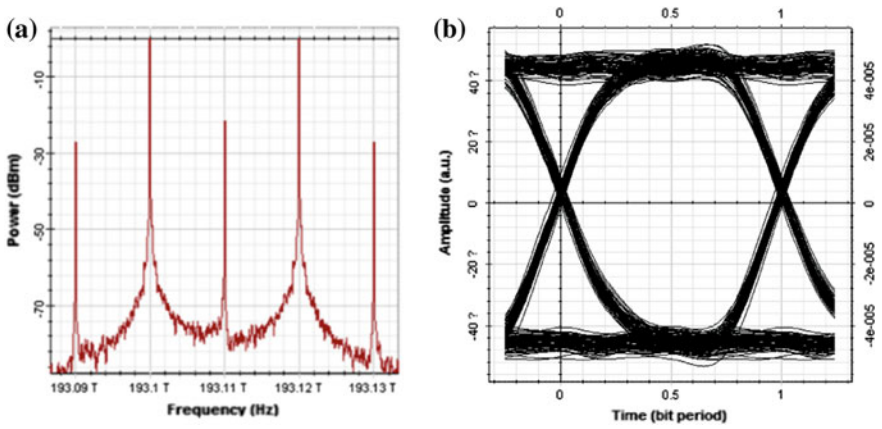



**Fig. 4** The spectrum of OVCO output and the signal recovered after LPF

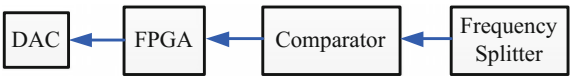
## 3 Experimental Results

For practical implementation, the frequency difference of the signal light and local light is usually too large. In order to make the scheme more efficiency, we add a frequency difference capture (FDC) component for a wider range adjustment. The control unit of the FDC is a commercial FPGA. The beat signal is sent to FPGA after processed by frequency splitter and comparator. The FPGA control the DAC to adjust the frequency of the local light according to the frequency of the beat signal.

In the experiment, we use two 1550 nm tunable CW lasers with linewidth 10 kHz, an electrical VCO with 1000 MHz/V tuning coefficient and 10 GHz central frequency, and the tuning range is from 5 to 14.5 GHz. So the tuning range of local is about 10 GHz, a relatively large range. At the beginning of the experiment, we adjust the local laser to make the frequency difference of the signal light and local light less than 10 GHz, then the FDC component can response the frequency difference signal and adjust local light frequency automatically. With the feedback control, the frequency difference can decrease to 2 MHz within 5 s, then the signal can be processed by the loop filter and a perfect signal can be obtained at the output (Fig. 5).

Figure 6 is the spectrum of the beat signal at different stage. Figure 6a is the frequency difference at about 400 MHz, and (c) is less than 2 MHz. The results show that the FDC worked well and the OPLL is effective. When the frequency

**Fig. 5** The structure of FDC



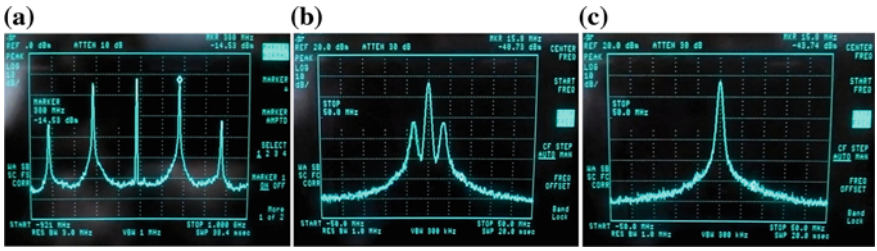



**Fig. 6** The electrical spectrum of the beat signal

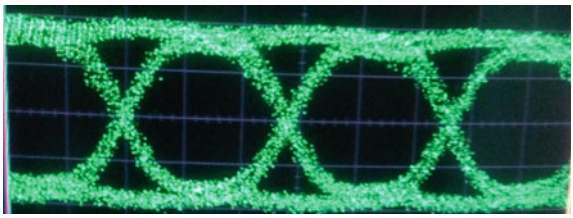


**Fig. 7** The recovered signal eye diagram

difference is less than 2 MHz, the loop filter will continue to process the beat signal. When the signal light and local light is well locked, we can get a clear eye diagram at the output (Fig. 7).

## 4 Conclusion

We proposed a wide adjusting range frequency-locking scheme and demonstrate it with simulation and experiment. The results show that the scheme can effectively lock the frequency of the signal light and the local light with rapid speed and accuracy. We believe that the proposed architecture can pave the way for the future transmission systems.

**Acknowledgements** This work is supported by the Doctor Fund of Tianjin Normal University (52XB1505).

## References

1. M. Fice et al., Homodyne coherent optical receiver using an optical injection phase-lock loop. J. Lightwave Technol. **29**(8), 1152–1164 (2011)
2. H. Sun, K.T. Wu, K. Roberts, Real-time measurements of a 40 Gb/s coherent system. Opt. Express **16**(2), 873–879 (2008)
3. N.-K. Kjøller, K. Røge, P. Guan, H.C. Mulvad, M. Galili, L.K. Oxenlowe, A novel phase-locking-free phase sensitive amplifier based regenerator. OFC, W4C.2 (2015)

4. S. Camatel, V. Ferrero, Design, analysis and experimental testing of BPSK homodyne receivers based on subcarrier optical phase-locked loop. J. Lightwave Technol. **26**(5), 552–559 (2008)
5. D. Ly-Gagnon, K. Katoh, K. Kikuchi, Unrepeatered optical transmission of 20 Gbit/s quadrature phase-shift keying signals over 210 km using homodyne phase-diversity receiver and digital signal processing. Electron. Lett. **41**(4), 206–207 (2005)
6. L.H. Enloe, J.L. Rodda, Laser phase locked loop. Proc. IEEE **25**(2), 165–166 (1965)
7. M. Lu, H. Park, E. Bloch, A. Sivananthan, A. Bhardwaj, Z. Griffith, L.A. Johansson, Highly integrated optical heterodyne phase-locked loop with phase/frequency detection. OE **20**(9), 9736–9741 (2012)
8. R.E. Best, *Phase-Locked Loops: Design, Simulation, and Applications* (R.R. Donnelley &Sons, Inc., Tsinghua University Press, Beijing, 2003), pp. 18–23

# Adaptive Down Sampling by Improved Methods of FRI

**Yao Shi, Bo Yu, Min Jia, Zhizhong Zheng and Qing Guo**

**Abstract** In modern signal processing, sampling is a key step. The number of sampling points directly affects the computation of subsequent signal processing. Nyquist sampling theorem uses twice the highest frequency of signal to sample the signal. In fact, for sparse signal, not all of the points are necessary. FRI is a highly efficient sampling method, but FRI can merely deal with discrete signals. By the improved methods of FRI, FRI theory can be extended to process continuous ECG signals. What's more is, sampling scheme put forward by this paper can change the number of points according to the application. If the requisite degree of accuracy is low, less points are needed. Finally, simulation experiment shows that this method can not only reduce sampling rate greatly, but also can ensure the accuracy of recovery.

**Keywords** Adaptive · Down sampling · FRI · ECG signals

Y. Shi · M. Jia · Q. Guo (✉)
Communication Research Center, Harbin Institute of Technology, Harbin, China
e-mail: qguo@hit.edu.cn

Y. Shi
e-mail: 1928368042@qq.com

M. Jia
e-mail: jiamin@hit.edu.cn

B. Yu
Shenzhen Academy of Aerospace Technology, Shenzhen, China
e-mail: yb@htrfid.com

Z. Zheng
School of Arts and Science, Zhengzhou Institute of Information Science and Technology, Zhengzhou, China
e-mail: zhengzoom@126.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_73

## 1 Introduction

According to classical Nyquist sampling theorem, when the signal's highest frequency is $f_{\max}$ and sampling frequency is $f_s$, if $f_s \geq 2f_{\max}$, the signal can be recovered perfectly after sampling. In order to reduce sampling rate, some theorems are put forward, such as FRI [1–3]. FRI is a highly efficient sampling method. FRI is the abbreviation of "finite rate of innovation," mainly to handle discrete signals, such as Diracs, B spline, and E spline [4, 5]. The essence of FRI theory is the information acquisition, rather than aimlessly sampling as Nyquist sampling.

However, the types of signal which FRI theory [3, 6] can handle are limited, as FRI can merely deal with discrete signals. This paper puts forward a new method to extend the types of signal which FRI theory can handle. By the improved methods of FRI, FRI theory can not only deal with discrete Dirac, but also process high-frequency continuous signal. What's more is, sampling scheme put forward by this paper can change the number of sampling points according to the application. If the requisite degree of accuracy is low, less points are needed, at the same time, sampling scheme can ensure that the chosen points are the most important, so that the original signal can be recovered by as few points as possible.

First, the spectrum of the signal is divided by frequency into two parts. In the frequency domain, the information of high-frequency signal is recorded by the spectral line. In the time domain, the low-frequency part is recorded by the line fitting algorithm. Then the improved FRI theory can be used for subsequent processing and signal recovery.

The outline of the paper is as followed. In Sect. 2, the main steps of comprehensive signal analysis in time and frequency domain are introduced. When analyzing the signal in time domain, a new method, line fitting algorithm is put forward. In frequency domain, spectrum lines are used to record the signal. In Sect. 3, the classical finite rate of innovation theory is reviewed. The adaptive sampling method is introduced to collect different numbers of sampling points to satisfy different accuracy requirement. How to use FRI theory to process the signal is introduced in Sect. 4. In order to verify the efficiency of sampling scheme and the accuracy of the recovery, the simulation results are given in Sect. 5.

## 2 Comprehensive Signal Analysis in Time and Frequency Domain

The signal spectrum can be got by FFT. Through the signal spectrum, the signal is divided into high-frequency and low-frequency components. The high frequency is analyzed in the frequency domain, while the low-frequency is studied in the time domain.

### 2.1 In the Frequency Domain

The high-frequency parts of the signal change rapidly in the time domain. So a great number of points are needed to record the information if the high-frequency parts are analyzed in time domain. Instead, if the spectrum lines are used to record the information of the high-frequency parts in the frequency domain, a number of sampling points are greatly reduced. In the frequency domain, the frequency value is expressed by $f_k$, and the magnitude is represented by $a_k$.

### 2.2 In the Time Domain

The low-frequency parts of the signal sometimes contain many frequency components. A large number of spectrum lines are needed to record the information if the low-frequency parts are analyzed in frequency domain. As the low-frequency parts of the signal change slowly in the time domain, if the low-frequency parts are analyzed in the time domain, the number of sampling points will be greatly reduced.

This paper puts forward a new method, line-fitting algorithm, to collect the information in time domain. The main steps of the algorithm are as followed. First, the waveform of the signal is divided into many tiny segments. Then the starting moment and the slope of the segment are used to record the waveform. The starting moment is $t_k$ and the slope of the segment is $a_k$. So $t_k$ and $a_k$ can be used to describe the low-frequency parts of signal very well in the time domain.

## 3 The Classical Finite Rate of Innovation Theory

$t_k$ and $a_k$ are obtained through the above steps, then the signal can be processed by FRI theory. The steps of processing signal with FRI are as follows.

First, let the signal $x(t)$ go through a specific filter $h(t)$, then the output signal $y(t)$ can be expressed by the following formula:

$$y(t) = h(t) * x(t) = \int h(m)y(t-m)dm \tag{1}$$

Here Gauss function, Singer function, B spline, and E spline function can be selected as $h(t)$. $x(t)$ is a signal composed by pulse stream, which is determined by the position $\{t_k\}_{k=0}^{K-1}$ and amplitude $\{a_k\}_{k=0}^{K-1}$. $t_k$ and $a_k$ can be gotten from time domain and frequency domain respectively.

Then taking samples from $h(t)$, the discrete samples are obtained, which are given by

$$\begin{aligned} y_n &= y(nT) = \langle x(t), \varphi(\frac{t}{T} - n)\rangle \\ &= \int_{-\infty}^{\infty} x(t)\varphi\left(\frac{t}{T} - n\right)dt \end{aligned} \tag{2}$$

Deform the formula (2),

$$\begin{aligned} y_n &= \langle x(t), \varphi_B(nT - t)\rangle \\ &= \frac{1}{B}\sum_{m \le M = \left[\frac{B\tau}{2}\right]} \hat{x}_m e^{j2\pi\frac{mn}{N}} \end{aligned} \tag{3}$$

That is,

$$\hat{x}_m = B\hat{y}_m \tag{4}$$

$\hat{x}(\omega)$ is the continuous time Fourier transform of the function $x(t)$,

$$\hat{x}(\omega) = \int_{-\infty}^{\infty} x(t)e^{-jt\omega}dt \tag{5}$$

Calculate Fourier series coefficients,

$$\hat{x}_m = \frac{1}{\tau}\hat{p}\left(\frac{2\pi m}{\tau}\right)\sum_{k=0}^{K-1} a_k e^{-j2\pi m\frac{t_k}{\tau}} \tag{6}$$

Deform the formula (5),

$$\hat{x}_m\hat{p}^{-1}\left(\frac{2\pi m}{\tau}\right) = \frac{1}{\tau}\sum_{k=0}^{K-1} a_k u_k^m \tag{7}$$

$u_k = e^{-j2\pi\frac{t_k}{\tau}}$ and $\hat{p}^{-1}$ is the inverse multiplicative of $p$. For simplicity, assume that $\hat{p}(\frac{2\pi m}{\tau}) = 1$.

In order to obtain the value of $u_k$, the Z—transformation of the filter must be calculated first,

$$\hat{h}(z) = \sum_{m=0}^{K} h_m z^{-m} = \prod_{m=0}^{k-1}(1 - u_k z^{-1}) \tag{8}$$

That is, the root of $\hat{h}(z)$ is equal to the value of $u_k$. $h_m$ satisfies:

$$\begin{aligned} h_m * \hat{x}_m &= \sum_{i=0}^{K} h_i \hat{x}_{m-i} = \sum_{i=0}^{K} \sum_{k=0}^{K-1} a_k h_i u_k^{m-i} \\ &= \sum_{k=0}^{K-1} a_k u_k^m \sum_{i=0}^{K} h_i u_k^{-i} = 0 \end{aligned} \tag{9}$$

As $\hat{h}(u_k) = 0$, The filter is called a devastating filter, because it returns signal to zero.

Here $h_0 = 1$, the formula can be written in matrix form:

$$\begin{pmatrix} \hat{x}_{-1} & \hat{x}_{-2} & \cdots & \hat{x}_{-K} \\ \hat{x}_0 & \hat{x}_{-1} & \cdots & \hat{x}_{-K+1} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}_{K-2} & \hat{x}_{K-3} & \cdots & \hat{x}_{-1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_K \end{pmatrix} = - \begin{pmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \vdots \\ \hat{x}_{K-1} \end{pmatrix} \tag{10}$$

This shows that at least $2K$ of continuous values are needed to solve the above system. From the above equation, $\{h_m\}$ can be obtained. The roots of Z transform of $\{h_m\}$ is $u_k$, $t_k = -\frac{\mathrm{In}(u_k)}{j2\pi} \times \tau$ then we can get $t_k$. Amplitude $a_k$ can be got by the following equation:

$$\frac{1}{\tau} \begin{pmatrix} 1 & 1 & \cdots & 1 \\ u_0 & u_1 & \cdots & u_{K-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_0^{K-1} & u_1^{K-1} & \cdots & u_{K-1}^{K-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_{K-1} \end{pmatrix} = \begin{pmatrix} \hat{x}_0 \\ \hat{x}_1 \\ \vdots \\ \hat{x}_{K-1} \end{pmatrix} \tag{11}$$

In the time domain, $t_k$ is the initial time of each line, $a_k$ is the slope of each line. In the frequency domain, $t_k$ is the position of the spectrum line, $a_k$ is the amplitude of spectrum line. Then the high-frequency part of FRI signal is returned into time domain, and the part from frequency domain and the part from time domain are put together. Then the original signal can be recovered.

## 4 Processing Signals with Adaptive Sampling and Finite Rate of Innovation

Through above steps, a group of $(f_k, a_k)$ and $(t_k, a_k)$ are gotten. In some situations, the degree of accuracy is not so important; the number of sample points can be reduced.

In fact, the largest $a_k$ can be chosen to describe the signal approximately. For example, if the requisite degree of accuracy is low, merely 30% points are needed; whose $a_k$ is largest in all points. If the requisite degree of accuracy is high, 80% or more points are needed. In this way, different numbers of points are sampled according to the application.

Then FRI theory can be used to process the signals. The flowchart of adaptive down sampling by improved methods of FRI is shown as follows (Fig. 1).

Specifically, through the filter, original signals are divided into two parts. High-frequency parts of the signals are recorded by spectrum lines, while low-frequency parts of the signals are recorded by line fitting algorithm. Then adaptive sampling and FRI theory are used to process the signal. FRI theory can deal with discrete signals, whose time delay is $t_k$ and amplitude values is $a_k$. Through the above steps, a continuous signal can expressed by $f_k$ and $a_k$ in the frequency domain and $t_k$ and $a_k$ in the time domain. So FRI theory also can be used to deal with the continuous signals to reduce sampling rate. $f_k$ in the frequency domain can be considered as $t_k$ in FRI theory, and $a_k$ in the frequency domain can be considered as $a_k$ in FRI theory. $t_k$ and $a_k$ in the time domain can be considered as $t_k$ and $a_k$ in FRI theory. So the FRI theory can be used to restore the signal.

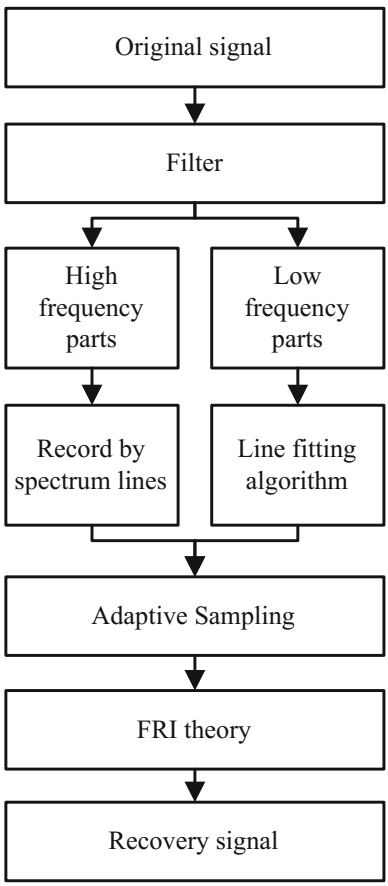



**Fig. 1** Adaptive down sampling by improved methods of FRI

## 5 Experiments and Results

A record of ECG signal is selected as original signal. The waveform of the original ECG signals is shown in Fig. 2.

In order to get the Nyquist sampling frequency, the spectrum of the ECG signals are drawn in Fig. 3.

The signal's highest frequency is about 500 Hz. If using the Nyquist sampling theorem, at least 1000 points per-second have to be collected.

If using the improved methods of FRI, the high frequency is recorded by spectrum lines, and through spectrum analysis, nine points are needed to record the high frequency in frequency domain. The low frequency is analyzed by line fitting algorithm and 30 points are needed in time domain. So 39 points can describe the

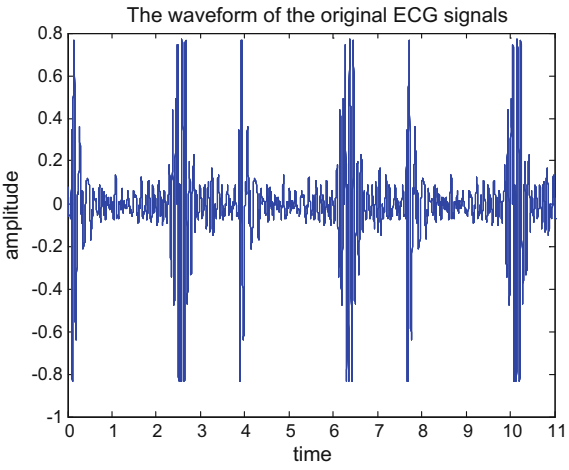



**Fig. 2** The waveform of the original ECG signals

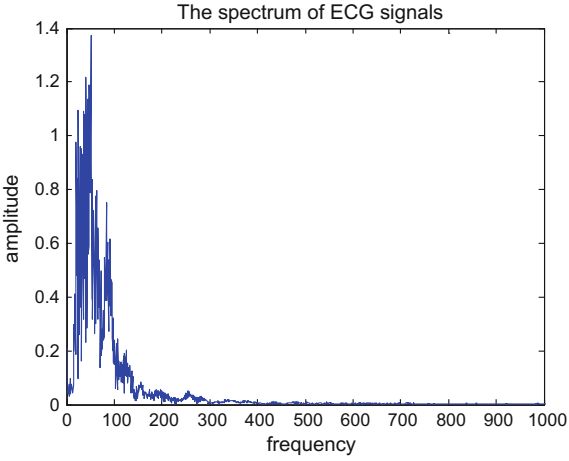



**Fig. 3** The spectrum of the ECG signals

original signal very well; however, in some cases, sampling points can be further reduced. If the requisite degree of accuracy is low, only 50% or less of all the points are needed. However, the accuracy of recovery decreases accordingly. In fact, the percent of points can be flexibly changed according to the requirement of accuracy. The process of sampling is shown in Fig. 4.

The number of points needed by different sampling scheme is sorted out in Table 1.

After the processing of FRI theory, the recovery waveform of different sampling scheme is drawn in Fig. 5.

In order to compare recovery performance of the improved methods of FRI with Nyquist sampling theorem, the minimum mean square error of three sampling scheme are drawn in one picture. The result shows that, though the improved methods of FRI use much fewer points, its recovery performance is close to Nyquist sampling theorem, especially when the SNR becomes larger. If the requirement of accuracy is lower, less points are needed, so that the efficiency of sampling can be further improved (Fig. 6).

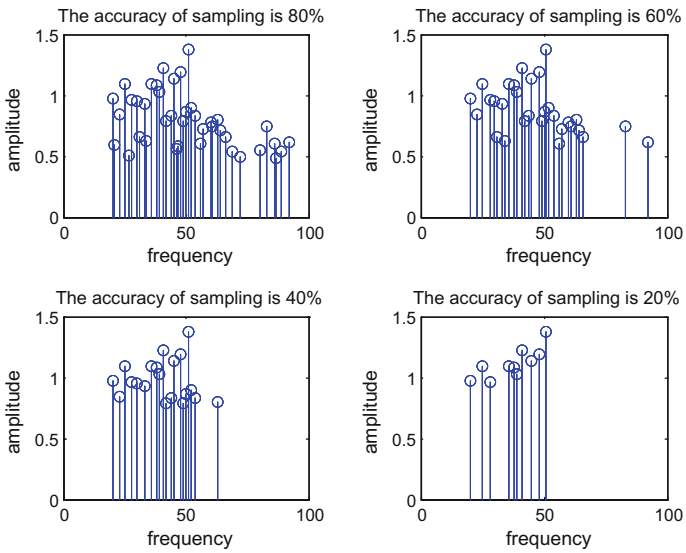



**Fig. 4** The process of sampling the ECG signals

**Table 1** The number of points needed by different sampling scheme

| Sampling scheme | Nyquist sampling theorem | The improved methods of FRI | |
|---|---|---|---|
| | | High accuracy | Low accuracy |
| The number of points needed | 11,000 | 39 | 20 |

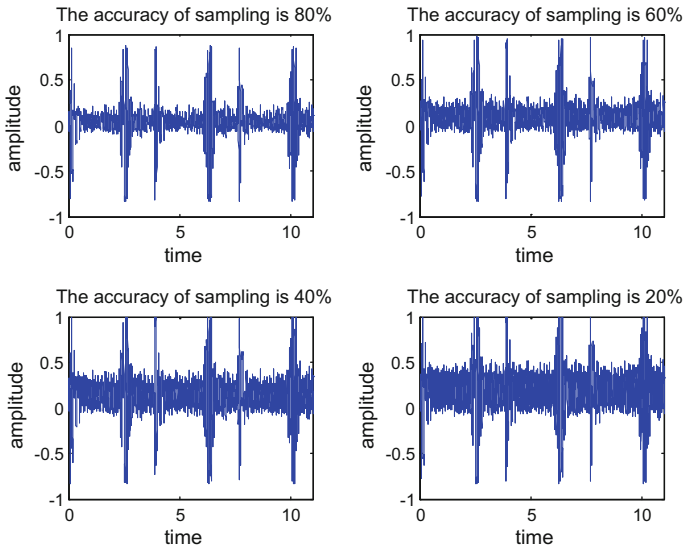
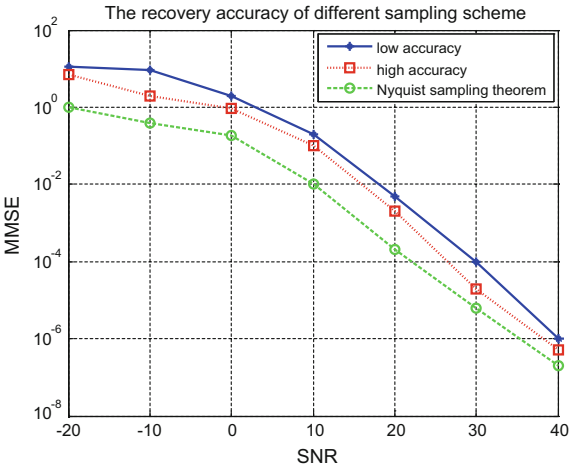



**Fig. 5** The recovery waveform of different sampling scheme

**Fig. 6** The recovery performance of the algorithm




## 6 Conclusion

Adaptive down sampling by improved methods of FRI is proposed in this paper. The proposed methods mainly have three advantages. First, classical FRI theory can merely deal with discrete signals, while the proposed method can deal with continuous ECG signals. Second, the proposed method can not only reduce sampling rate greatly, but also can ensure the accuracy of recovery. The experimental results

show that the proposed method can use much less points than Nyquist Sampling Theorem to get approximate recovery performance. Finally, the proposed method can collect different numbers of sampling points according to the different requirement of accuracy in actual situations. On the premise of the signal recovery precision, less points are collected to improve the sampling efficiency.

**Acknowledgements** This work was supported by the National Natural Science Foundation of China under Grant No. 61671183 and 91438205.

## References

1. M. Vetterli, P. Marziliano, T. Blu, Sampling signals with finite rate of innovation. IEEE Trans. Signal Process. **50**(6), 1417–1427 (2002)
2. P.L. Dragotti, M. Vetterli, T. Blu, Sampling moments and reconstructing signals of finite rate of innovation: Shannon meets strang-fix. IEEE Trans. Signal Process. **55**(5), 1741–1757 (2007)
3. T. Blu, P.L. Dragotti, M. Vetterli, P. Marziliano, L. Coulot, Sparse sampling of signal innovations. IEEE Signal Process. Mag. **25**(2), 31–40 (2008)
4. J.A. Uriguen, P.L. Dragotti, T. Blu, On the exponential reproducing kernels for sampling signals with finite rate of innovation, in *Proceedings of the Ninth International Workshop on Sampling Theory and Applications*, Singapore, May 2011
5. J.A. Uriguen, T. Blu, P.L. Dragotti, FRI sampling with arbitrary kernels'. IEEE Trans. Signal Process. **61**(21), 5310–5323 (2013)
6. H. Pan, T. Blu, P.L. Dragotti, Sampling curves with finite rate of innovation. IEEE Trans. Signal Process. **62**(2), 458–471 (2014)

# Part VIII
# Patten Recognition, Deep Learning and Learning Automata

# Face Recognition Based on Local Gabor Binary Patterns and Convolutional Neural Network

**Xudie Ren, Haonan Guo, Chong Di, Zhuoran Han and Shenghong Li**

**Abstract** Enhancing the robustness to changes caused by facial aging in automatic face recognition system is still an important problem worth researching. Compared with the external factors, such as illumination, posture and expression, facial aging which can produce variations in both shape and texture of the face has more complex effects. In this paper, we propose a method based on Local Gabor Binary Patterns and Convolutional Neural Network (LGBP-CNN) to improve the performance of age invariant face recognition problem. For each face image, this method first extracts shape, texture and local neighbor relationship features with multi-orientation and multi-scale Gabor filters as well as local binary patterns (LBP) operators. Then, we utilize one kind of Deep Learning model-convolutional neural network which has shown brilliant performance on face recognition area to avoid the dimension curse problem brought by Gabor filtering and further extract features. Such kind of method has robustness to changes of illumination, posture, expression, shape and texture by combining Gabor transform, LBP and convolutional neural network. Experiments are implemented on the FG-NET database and the results can outperform the state of the art ones, which verify the validity of the proposed method in age invariant face recognition problem.

**Keywords** Face recognition · Age invariant · Gabor filter · Local binary pattern · Convolutional neural network

X. Ren (✉) · H. Guo · C. Di · Z. Han · S. Li (✉)
Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: renxudie@sjtu.edu.cn

S. Li
e-mail: shli@sjtu.edu.cn

H. Guo
e-mail: haonan2012@sjtu.edu.cn

C. Di
e-mail: dichong95@sjtu.edu.cn

Z. Han
e-mail: hzrtom@sjtu.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_74

## 1 Introduction

Face recognition is attractive in the fields of pattern recognition and image processing, and it is widely applied in security monitoring, digital communication and so on. Most researches on face recognition focus on compensating the influence caused by external factors, such as illumination, posture and expression [7]. While, internal factors include facial aging have a stronger effect on face recognition accuracy. The shape and texture of faces changed dramatically with age, which can degrade face recognition performance. Therefore, facial aging problem is still a great challenge in automatic face recognition and worth working on.

Compared with the tasks on external factors, facial aging has received less attention. Park et al. [8] present a 3D aging modeling technique for age-invariant face recognition which adapts view-invariant 3D face models to the given 2D faces. They model shape and texture separately at different ages and apply second level Principal Component Analysis (PCA) to remove the correlation between them. Du et al. [5] demonstrate a sparse-constrained non-negative matrix factorization (NMFsc) algorithm which decomposes face images into basis matrices and coefficient matrices to heighten the recognition accuracy. By adjusting the sparseness for matrices, they establish facial feature model with a strong ability of local representation and employ an improved prototype method to carry on the facial aging simulation.

The above approaches are based on simulating by adding virtual aging samples to improve the performance. Another direction is based on feature extraction, which can provide robustness shape and texture features for classifier to implement facial aging recognition, e.g., using SIFT and multi-scale local binary patterns (MLBP) as local descriptors to build feature spaces [6]. In [9], a texture embedded discriminative graph matching (TED-GM) model is introduced to address the problem of age invariant face recognition by applying Gabor binary pattern histogram sequence to encode the discriminative and compact feature.

Nevertheless, these methods do not sufficiently establish robust high-level feature space for recognizing and cannot globally understand the age invariant information of images, thus leaving the space for enhanced performance. Recently, Convolutional Neural Network (CNN) and Deep Learning techniques have obtained continuous development and breakthrough in image processing tasks. Convolutional architectures have been applied to large-scale face recognition system and achieved the state of the art performance. Motivated by above facts, this paper formulate a method based on Local Gabor Binary Patterns and Convolutional Neural Network (LGBP-CNN) to improve the performance of age invariant face recognition problem. Compared with the other existing approaches, our method presents the following features:

1. Gabor filtering: Gabor wavelets have been successfully utilized in face recognition area [10] and Gabor features can be robust to illumination, posture and expression variations. Therefore we extract Gabor maps of face images with multi-orientation and multi-scale Gabor filters to obtain local details.

2. Local Binary Patterns: LBP operator does well in local features extraction which makes it popular in recognition tasks. It is straightforward to calculate and realize, so we apply it to further strength the local information on Gabor maps.
3. Convolutional Neural Network: Multi-orientation and multi-scale Gabor filters can not only contain more sufficient information of face images but also cause the dimension curse problem. CNN can dramatically reduce the number of parameters needed to train by weight sharing which represents this structure possesses more tolerance to dimension. We use CNN to solve the dimension curse problem and further extract high-level features for age invariant face recognition.

Experimental results on the FG-NET database [1] demonstrate the effectiveness and robustness of the proposed method to age variation in face recognition.

The rest of the paper is organized as follows: Sect. 2 provides an introduction to the proposed LGBP-CNN method for age invariant face recognition. Experimental results on the FG-NET database are presented and discussed in Sect. 3. Section 4 concludes the paper.

## 2 Local Gabor Binary Patterns and Convolutional Neural Network

In this paper, we propose an age invariant face recognition method based on Local Gabor Binary Patterns and Convolutional Neural Network (LGBP-CNN) to improve the performance. For each face image, we first extracts robust local face features with multi-orientation and multi-scale Gabor filters as well as local binary patterns (LBP) operators. Then, the method employ convolutional neural network which has been widely applied for image processing and pattern analysis to remit the dimension curse problem caused by Gabor filtering and further extract high-level features for recognition. In the following section, we describe every step in age invariant recognition detailedly.

### *2.1 Gabor Magnitude Maps*

The characters of Gabor wavelets resembles to the ones of receptive field in vision cells, which signifies they can be an outstanding simulation of human recognition function. Gabor wavelet transformations from different orientations and scales can reflect the pixel gray value change information in local range of face image, in other words, they possess excellent locality in both frequency domain and spatial domain. Gabor wavelets also are capable of tolerating a certain degree of image rotation and deformation as well as not sensitive to light which make them conducive to feature extraction in different environments. Taking account of the above brilliant features of Gabor wavelets, we employ Gabor filters to decompose the input face images.

Let $E(x, y)$ represents the gray level distribution of a face image and Gabor maps of the input face image can be obtained by applying the following function:

$$G(z) = E(x, y) * \psi_{\mu,\nu}(z) \tag{1}$$

where $z = (x, y)$, $*$ denotes the convolution operator and $\mu$, $\nu$ define the orientation and scale of the Gabor filter kernels.

The Gabor filters in this paper are defined as follows:

$$\psi_{\mu,\nu}(z) = \frac{\|k_{\mu,\nu}\|^2}{\sigma^2} e^{-(\|k_{\mu,\nu}\|^2 \|z\|^2 / 2\sigma^2)} [e^{ik_{\mu,\nu} z} - e^{-(\frac{\sigma^2}{2})}] \tag{2}$$

where $\|\cdot\|$ denotes the norm operator and the wave vector $k_{\mu,\nu}$ is indicated as:

$$k_{\mu,\nu} = k_\nu e^{i\phi_\mu} \tag{3}$$

where $k_\nu = k_{max}/f^\nu$ and $\phi_\mu = \pi\mu/8$, $f$ is the spacing factor between filters in the frequency domain.

Pay attention to that the phase information of the above Gabor transformation is time-varying, we only employ the magnitude of each Gabor map to compose a set of Gabor Magnitude Maps (GMMs).

## 2.2 Local Binary Patterns

The values of the Gabor transformation vary tardily with displacement, thus the details in the GMMs can be further improved. LBP is a non-parametric operator and can encode the local information of images into a series of patterns, which makes it conducive to our local feature extraction task. The primary LBP operator $LBP(P, R)$ labels each pixel of an image by comparing value with its P neighbors on a circular neighborhood with the radius R. As depicted in Fig. 1, the comparative result between each pixel and its neighbor can be calculate as a binary number for negative

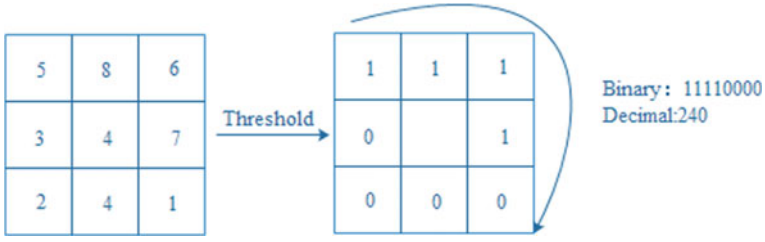



**Fig. 1** Primary LBP operator

values assigned as 0 and the others assigned to 1. Then the entire results of the center pixels can be encoded to a decimal number.

$LBP(P, R)$ provides $2^P$ output values, but only some certain patterns contain more information. Uniform LBP is an effective subset of the abovementioned $2^P$ patterns which denotes there are at most 2 bitwise transitions from 0 to 1 or vice versa in the decimal number encoding process. We apply it in our paper to take place of the primary LBP operator.

## *2.3 Convolutional Neural Network*

The features extracted by Local Gabor Binary Patterns (LGBP) are produced in multi-orientation and multi-scale, which can generate a high dimension and cause the dimension curse problem. To directly take advantage of the original LGBP features, we apply Convolutional Neural Network (CNN), an efficient deep model, to decrease the subsequent task complexity and further extract high-level features for age invariant face recognition. CNN utilizes the concepts of receptive fields, weight sharing and sub-sampling (pooling) to reduce the complexity of the network structure and the number of parameters, in other words, this deep structure possesses more tolerance to input dimension. Receptive field is equivalent to construct a number of spatially localized filters which can obtain some salient features of the input. Weight sharing can reduce the number of parameters which needs to be trained. Pooling provides a guarantee of translation invariance.

A typical CNN is consisted of alternating convolution and sub-sampling layers, then turns into fully connected layers when approaching to the last output layer. It usually adjusts all the filter kernels (convolution kernels) by back-propagation algorithm [11], which is based on stochastic gradient descent algorithm, to decrease the gap between the network output and the true value. Overall, the convolution layer (C layer) obtains the local features by connected with local receptive fields. The sub-sampling layer (S layer) is a mapping feature layer which is used for pooling operation and completing the secondary extraction calculations. Each C layer is followed by an S layer, and the special twice feature extraction structure makes convolutional neural network have strong distortion tolerance on the input images.

The specific structure of CNN in this paper includes an input layer (Input), two C layers (C1, C2), two S layers (S1, S2) and a fully connected output layer (Output) for recognition. Since a convolution kernel of the convolution layer can only extract one characteristic of input feature maps, it requires multiple convolution kernels to extract different features.

The entire process of LGBP-CNN based face recognition is illustrated in Fig. 2.

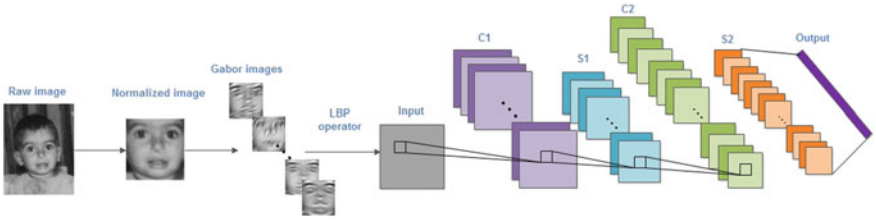



**Fig. 2** Entire process of LGBP-CNN

## 3 Experimental Results

Experiments are implemented on the FG-NET database to verify the validity of the proposed method in this paper. We also compare the results with the other state of the art methods on the same database to evaluate the effectiveness of LGBP-CNN. The database, settings and results are shown in the following.

### *3.1 Database*

We employ the FG-NET Aging database [1] to evaluate the proposed method for it is a well-known public domain database in facial aging analysis area. This database is universal which means it is convenient for us to compare the results with other methods. The FG-NET database contains 1,002 face images of 82 individuals at different ages ranging from 0 to 69. The image size is about $400 \times 500$ in pixel and images of an individual in the FG-NET database are displayed in Fig. 3.


**Fig. 3** Images of an individual at different ages in the FG-NET database

**Table 1** Parameters in LGBP-CNN

| Parameter | | | Value |
|---|---|---|---|
| LGBP | Gabor | Orientation | 8 |
| | | Scale | 5 |
| | LBP | P | 8 |
| | | R | 1 |
| CNN | Learning Rate | | 1 |
| | Kernel Size | | $5 \times 5$ |
| | Polling Size | | $2 \times 2$ |

### *3.2 Settings*

As shown in Fig. 3, the raw face images should be preprocessed prior to applying LGBP-CNN. Each image is transformed to gray-scale and normalized according to the location of eyes. After preprocessing, we make the centers of two eyes horizontal and align accurately. As for feature extraction stage, some parameter settings need to be done and the details in this stage are described in Table 1.

### *3.3 Results and Analysis*

To test the effectiveness and robustness of the proposed method in this paper for facial aging, we evaluate it on all 1,002 images in the database. In addition, we compare the rank-1 recognition accuracy with the state of the art ones including simulation-based ones [4, 8] and feature-based ones [2, 6, 9], and the exact accuracies are listed in Table 2. Evidently, the performance of LGBP-CNN model has higher accuracy than others on the FG-NET database. The reason probably lies in that we use CNN to dispose the original robustness LGBP features without losing information as well as to further extract high-level features for recognition. These observations impressively demonstrate the effectiveness and robustness of the proposed LGBP-CNN age invariant face recognition method.

## 4 Conclusion

In order to enhance the performance of age invariant face recognition problem, this paper proposes a method based on Local Gabor Binary Patterns and Convolutional Neural Network, which takes full advantage of the shape, texture and local neighbor relationship information of images. By combining Gabor transform, LBP and convolutional neural network, this method is insensitive to variation caused by facial aging.

**Table 2** Recognition accuracy comparison among several models

| Paper | Model | Database (subjects, image) | Recognition accuracy (%) |
|---|---|---|---|
| Geng et al. [4] | Learn aging patterns on concatenated PCA coefficients of shape and texture from full face across a series of ages | FG-NET(10,10) | 38.1 |
| Park et al. [8] | Learn aging patterns based on PCA coefficients in separated 3D shape and texture | FG-NET(82,1002) | 37.4 |
| Li et al. [6] | Multi-feature discriminative analysis combined with several distance based classifiers | FG-NET(82,1002) | 47.5 |
| Bereta et al. [2] | Diverse local texture descriptors combined with several distance based classifiers | FG-NET(82,1002) | <45.0 |
| Yang et al. [9] | Texture embedded discriminative graph matching | FG-NET(82,1002) | 64.47 |
| This paper | LGBP-CNN | FG-NET(82,1002) | 72.4 |

Experiments are conducted on the FG-NET Aging database to examine the performance of the LGBP-CNN and the results achieved demonstrate that this method can outperform the state of the art ones and extract robustness features.

Further work will be focused on increasing the dimension capacity of CNN to multi-orientation and multi-scale Gabor filters by investigating better global optimization techniques and speeding up the convergence of the cost function.

**Acknowledgements** This research work is funded by the National Science Foundation of China (61271316), Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, and Chinese National Engineering Laboratory for Information Content Analysis Technology.

## References

1. The fg-net aging database, http://www.fgnet.rsunit.com/
2. M. Bereta, P. Karczmarek, W. Pedrycz, M. Reformat, Local descriptors in application to the aging problem in face recognition. Pattern Recogn. **46**(10), 2634–2646 (2013)
3. J. Bouvrie, Notes on convolutional neural networks. Neural Nets (2006)
4. X. Geng, Z.H. Zhou, K. Smithmiles, Correction to "automatic age estimation based on facial aging patterns". IEEE Trans. Pattern Anal. Mach. Intell. **29**(12), 2234–2240 (2007)
5. D.U. Jixiang, C. Zhai, Y.E. Yongqing, An age-span face recognition method based on an nmf algorithm with sparseness constraints. Caai Trans. Intell. Syst. (2012)
6. Z. Li, U. Park, A.K. Jain, A discriminative model for age invariant face recognition. IEEE Trans. Inf. Forensics Secur. **6**(3), 1028–1037 (2011)

7. S. Liao, A.K. Jain, S.Z. Li, Partial face recognition: alignment-free approach. IEEE Trans. Softw. Eng. **35**(5), 1193–1205 (2013)
8. U. Park, Y. Tong, A.K. Jain, Age-invariant face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**(5), 947–954 (2010)
9. H. Yang, D. Huang, Y. Wang, Age invariant face recognition based on texture embedded discriminative graph model, in *IEEE International Joint Conference on Biometrics* (2014), pp. 1–8
10. W. Zhang, S. Shan, W. Gao, X. Chen, Local gabor binary pattern histogram sequence (lgbphs): a novel non-statistical model for face representation and recognition, in *Tenth IEEE International Conference on Computer Vision* (2005), pp. 786–791
11. D. Zipser, R.A. Andersen, A back-propagation programmed network that simulates response properties of a subset of posterior pa. Nature **331**(6158), 679–684 (2010)

# Research on Recognition Technology of Human Lower Limbs Feature Based on the Random Forest Algorithm

**Yankai Liu and Meijuan Yu**

**Abstract** This paper according to the background of a game project of family service entity robots quickly follow, introduces the problems existing in the most human recognition methods first. Then proposes a laser scanner as a hardware device, takes the method for recognition by random forest algorithm through the extraction of multiple different characteristics of the human lower limbs. And the experiment results show that the method is feasible and the recognition rate is high.

**Keywords** Random forest · Laser scanner · Human lower limbs recognition

## 1 Introduction

Home service robot is an important part of the national robot competition over the years of which technical essence concludes three aspects: first is self-localization and navigation ability, second is capability for human–computer interaction, third is human body recognition ability [1]. These three aspects ask for human recognition as the important fundamental. Only if robot can identify the human body better, they can complete the other services based on the function better.

## 2 Existing Methods and Problems

The most of current identification methods are mainly based on the face skin and human shape [2]. The recognition based on the human face skin color coverts the image collected by the vision sensor into a specific color space. Judge whether it

Y. Liu · M. Yu (✉)
College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China
e-mail: amy_1121@163.com

Y. Liu
e-mail: ooaround@foxmail.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_75

belonging to the human body according to the color of the image to be detected. This method requires strict environment conditions. The recognition rate is low when the background is complex or the ambient light is dim [3]. Human body recognition based on the shape feature of human body is a kind of information that represents the body contour according to the existing human characteristics, such as symmetry and edge density, star vector representation, Harr-Like wavelet feature, Histograms of Oriented Gradient, and so on [4–7]. The performance of orientation gradient histogram is better in the existing method. However, the number of feather vectors extracted by this method is large which makes the computation of feature extraction and classifier training and classification are relatively high [8, 9]. In addition, the identification of the entire body contour demands to maintain sufficient relative distance between the robot and the human. When it is within 2 m, it cannot identify because the vision sensor cannot collect the whole body contour [10–13].

## 3 The Identification of Human Lower Limbs Characteristics Based on Random Forest Algorithm

In recent years, the emergence of laser scanner and the reduction of the cost have provided a new assistant tool for the research of the robot. Two-dimensional laser scanners can return the information of obstacle in the form of lattice. The human body has a relatively fixed pattern on the retuned dot matrix by a laser scanner. The recognition of the individuals based on the feature is not strict because of the characteristic of the lower limbs of the human body (legs). The human body can be detected by the gait characteristics in the detection range of laser scanner. It provides the position of human body processed by the laser scanner for the vision sensor and then it identifies the characteristics of a specific person through the machine learning algorithm in the direction of the human body.

### *3.1 Human Lower Limbs Detection Based on Random Forest Algorithm*

Random forest algorithm carries on the modeling of decision tree for each bootstrap sample using methods of sampling of bootstrap to extract multiple samples from the original samples. And then combine the expectation of multiple decision trees and obtain the final results by voting [14]. Using this algorithm to identify the characteristics of the lower limbs of the human body, we must use the laser scanner to scan the lower limbs of the human body first. And then deal with the information of lower extremity dot matrix. Break up the dot matrix to get cloud information of 2D point. And deal with the information of dot matrix through the learning method of machine, and label and store the information to the database. This passage has

picked up 1,600 multi-frame laser scanning pictures and more than 600 without the human lower limbs from ten participants in the indoor environment in different postures and dressing to train the random forest classifier.

To get further specific information about the human lower limbs, we need to make the noise reduction and segmentation of the point cloud obtained by scanning. Based on the statistical properties of the probability density algorithm, remove the outliers in point cloud dates, classify the processed cloud as the sample to be classified through the machine learning methods such as forest classification. After that, we can get the relative specific information about human lower limbs by the identified class as the point cloud collection of human lower limbs matching with human gait. And then find the position of the relative position of the scanner legs according to calculation of the detection of the human lower limbs and match with the corresponding position in the visual system and do the human body detection through the visual system in this position. Finally, the human body features (mainly the upper parts of the human body and face features) are extracted and realize the classification and recognition of the human body with the feature information about the lower limbs. In addition, in the process of information collection, the gait information of the two feet is collected and processed at the same time and mix up with the information of lower limbs collected by the laser sensor realizing the comprehensive analysis of lower limbs to further realize the recognition and tracking of the lower limbs of the human body.

## *3.2 Matching the Characteristics of the Human Lower Limbs*

After the training data being classified by the random forest classifier, only the part of the suspected human lower limbs is left. Then we obtain the recognized the lower limbs of the human body by matching the lower limbs of the human body purely. The matching methods of this paper for the lower limbs of the human body characteristics are based on the stride, lower extremity diameter, and lower limbs surface orientation of human lower limbs.

(a) Stride,the characteristics of stride refers to reflect he characteristics of the mutual relationship between the collocation around the foot while walking, including length and width and angle of left and right feet. We only use the characteristic information of length and width of left and right feet, because laser radar feedback data for two-dimensional data. In order to deal with concisely,this paper combine the length of left and right as stride and width of left and right as step width.

According to the prior knowledge and a lot of data proving that the walking flat of healthy people generally follow a length of about 50–80 cm. The main different between individuals are about the length of legs, the longer of the legs the longer of the length. Stride is usually two times long as length 100–160 cm. The distance

between the left and right foot called step width in walking which measure the mid point of heel as the reference. Healthy people are about 8 ± 3.5 cm.

(b) The lower limbs diameter, in this paper, the diameter of the lower limb is mainly the diameter of the leg below the knee joint. Due to the reasons of personal wearing and the error of laser radar data, this paper uses the information of leg diameter (height × 0.2 ± 3) cm.
(c) Lower limbs surface orientation, due to the characteristics of laser radar imaging, laser radar-scanning data should be set toward the scanning direction of the convex arc. The normal angle of arc jointed by the dot of double lower limbs of human body should be within a range. This feature is mainly used to test whether the two legs are correct, and the characteristics of the laser radar can be very good to exclude the interference like human legs.

## 4 The Results and Analysis of the Recognition of the Lower Limbs of the Human Body

This passage has picked up 1,600 multi-frame laser scanning pictures and more than 600 without the human lower limbs from ten participants in the indoor environment in different postures and dressing to train the random forest classifier. To avoid the influence of other subjects on the training data, the scanning range of the laser radar is set to 50 cm–4 m no feedback as beyond while collecting training data. No other obstacles in the sampling range.

After a long period of testing, the use of random forest classifier can make the correct identification of almost 99% of the lower limbs, but there will be the error identification about others similar to the human body. Figures 1 and 2 are for the human lower limbs before and after the recognition. In Fig. 1, the blue dot is circled

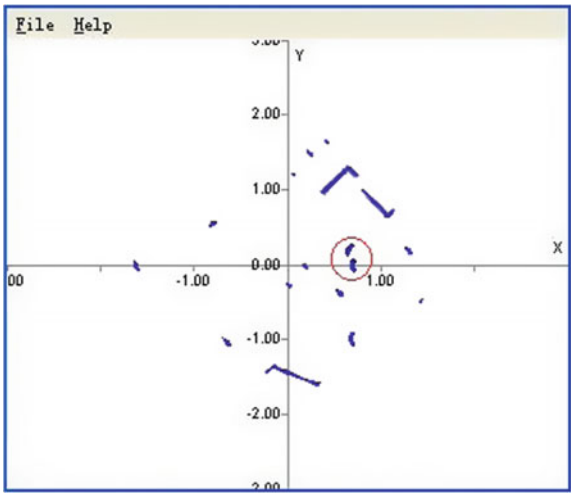



**Fig. 1** Before identification

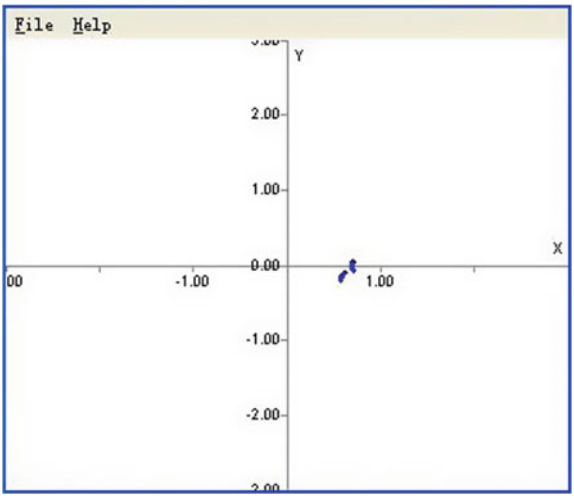
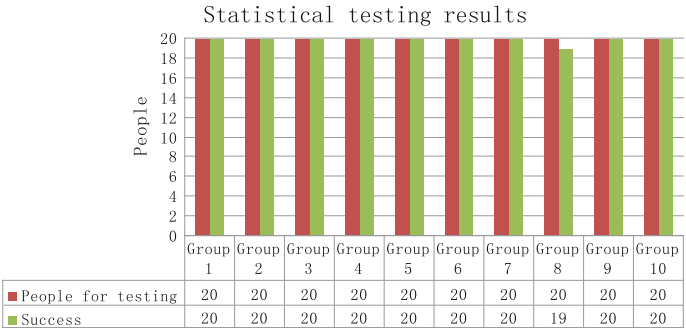



**Fig. 2** After identification




| | Group 1 | Group 2 | Group 3 | Group 4 | Group 5 | Group 6 | Group 7 | Group 8 | Group 9 | Group 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| People for testing | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 20 |
| Success | 20 | 20 | 20 | 20 | 20 | 20 | 20 | 19 | 20 | 20 |

**Fig. 3** Results of testing

by the red is the characteristic of the human lower limbs. In Fig. 2, it is the laser radar-scanning frame after identifying and removing the redundant points, which can be seen clearly.

In addition, this experiment was carried out in 10 groups of 20 people. In Fig. 3, it can be seen that only the eight group is failed, others are all succeed.

## 5 Summary and Outlook

This paper makes the family serving robot that quickly following the game project as the background. At first, the paper introduces the method of human body recognition and the existing problems in most of the participating colleges and after

that, the methods of recognizing human lower limbs based on laser scanner and the fast method for its target is presented. In the practical level, it is indicated as the recognition of the lower limbs of the human body by the random forest algorithm. Prove the feasibility sufficiently by experiments. In the indoor environment of this paper, it is able to effectively identify and match the lower limbs of the test subjects. Otherwise, in the case of other obstacles, it can identify the information about lower limbs of a single body.

## References

1. H. Zongbo, *The research on some problems of domestic mobile service robots*, Harbin Institute of Technologym, vol 12 (2006), pp. 24–34
2. Z. Hou, C. Han, Survey of visual tracking technology. J. Autom. (04), 603–617 (2006)
3. X. Hu, *Research and Implementation of Target Recognition and Face Recognition for Home Service Robot in Complex Scenes*, Hefei University of Technology, vol. 3 (2011), pp. 16–18
4. Xie Wei, Multi spectral palm print coding technique for human identification. Comput. Eng. **42**(3), 226–231 (2016)
5. D.P. Tao, L.W. Jin, Y.F. Wang, X.L. Li, Rank preserving discriminant analysis for human behavior recognition on wireless sensor networks. IEEE Trans. Circuits Syst. Video Technol. **10**(1), 813 (2014)
6. Z. Li, L. Song, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. 2 Mar 2016. In Press, Corrected Proof
7. Xuan Liu, Zhuo Li, Information-centric mobile ad hoc networks and content routing: A survey. Ad Hoc Networks, In Press, Corrected Proof, Available online 19 April 2016
8. X. Wang, Q. Wu, X. Lin, Z. Zhuo, *Finding Specific Person Using Spike Neural Network Based on Texture features. Intelligent Computing Theory,* LNCS 8588 (Springer, 2014), pp. 488–494
9. Z. Li, K. Liu, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2014)
10. L. Xiaojin, *Moving Human Body Recognition Algorithm Based on Visual Color Processing Mechanism*, vol. 5, Fujian Normal University (2015), pp. 17–19
11. X.G. Wang, M. Wang, W. Li, Scene-specific pedestrian detection for static video surveillance. IEEE Trans. Pattern Anal. Mach. Intell. **36**(2), 361 (2014)
12. Z. Liu, G. Feng, W. Chen, Gait Recognition Based on Local Binary Patterns and Identify Common Vector. Comput. Sci. **40**(9), 262–265 (2013)
13. Z. Li, Y. Chen, NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel Commun. Netw. **2016**, 48 (2016)
14. F. Kuangnan, J. Wu, A Review of Research on Random Forest Method. Stat. Inf. Forum **26** (03), 32–38 (2011)

# A Learning Automaton-Based Algorithm for Influence Maximization in Social Networks

**Jinchao Huang, Hao Ge, Ying Guo, Yan Zhang and Shenghong Li**

**Abstract** Influence maximization problem is to find a small set of influential nodes in a social network, whose activation of information propagation can be maximized using one of the propagation models such as independent cascade model. This paper proposes a new method to solve this problem based on discretized linear automaton algorithm and simple greedy algorithm. Every allowable node in the social network is regarded as an action and the goal of the learning automaton is to select an optimal subset of actions. To speed up the convergence, firstly, the problem space is reduced by excluding the nodes lacking influence; then, the original learning process is divided into two parts: the sampling process and the learning process. Besides, a new scheme for determining the best parameters is presented, which is more applicable to practical problems. The obtained results show that the new proposed algorithm is efficient in real-life social networks. The results are close to the ones obtained by the greedy algorithm in terms of accuracy, and the new method greatly improves the speed of selecting the optimal subset.

**Keywords** Learning automata · Influence maximization · Social network · Independent cascade model

J. Huang · H. Ge (✉) · Y. Guo · Y. Zhang · S. Li
Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
e-mail: sjtu_gehao@sjtu.edu.cn

J. Huang
e-mail: hjc2015@sjtu.edu.cn

Y. Guo
e-mail: guoying2014vip@sjtu.edu.cn

Y. Zhang
e-mail: jessiezhang1993@sjtu.edu.cn

S. Li
e-mail: shli@sjtu.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_76

## 1 Introduction

Investment in social influence has been found to be useful for marketing purposes recently. Social influence means change of thoughts, beliefs, and views of a person as a result his/her interaction with another person or a group [9]. The purpose is to identify a number of people, in order to attain the best marketing effect. In [5], Kempe, Kleinberg, and Tardos provided the first systematic study of influence maximization as a discrete optimization problem and proposed two basic stochastic influence cascade models, the linear threshold (LT) model and the independent cascade (IC) model. They also presented the greedy algorithm for the IC model. However, the greedy algorithm used the Monte-Carlo simulations on influence cascade to estimate the influence spread, which made the greedy algorithm rather slow and dissatisfied with required efficiency in social networks with considerable nodes.

Some recent work [1, 4, 6] proposed heuristic algorithms designed for the IC model. Although these heuristics improved the speed, their accuracy was not high in the real social networks. To overcome the inefficiency of the influence maximization solutions, several other methods were proposed to improve the greedy algorithm, such as CELF [8] and CELF++ [2].

Learning automata is a powerful mathematical model for many discrete optimization problems. In [10], Afshin and Keyhan firstly attempted to use learning automata to solve the influence maximization problem. They only put forward the concept, but they did not present the specific details of the experiment. In this paper, we propose a new algorithm based on the learning automata called LA-IM for influence maximization in social networks. In our algorithm, the allowable nodes of the social network are regarded as the actions of the learning automata, so that the problem is transformed into the problem that the learning automaton selects an optimal subset. Our main contributions are summarized as follows:

- We use a new learning automata to solve the influence maximization problem, where learning automaton take use of the discrete linear reward-inaction scheme.
- To increase convergence speed, we divide the original learning process into two parts: sampling process and learning process, where different algorithms are used in different process.
- Different from the traditional method of finding the best parameter, we propose a new criterion to determine whether the current parameter is the best, and the new method is more applicable to the real-life cases.

The rest of this paper is organized as followed. In Sect. 2, we introduce the related background of influence maximization and learning automata. Section 3 presents the detailed description of our proposed algorithm. The obtained results and discussions are reported in Sect. 4. Finally, conclusions and future works are mentioned in Sect. 5.

## 2 Independent Cascade Model

### 2.1 Introduction of Social Networks and Greedy Algorithm

As in [7], let $G = (V, E, p)$ be an influence graph, where $v \in V$ are users or nodes, and $e \in E$ are the links or edges between them. Each edge $e = (i, j)$ between them. Each edge $e = (i, j)$ between nodes $i$ and $j$ is associated with an influence probability $p_{ij} \in [0, 1]$. This means that the node $i$ is active in time $t$ and node $j$ is successfully activated by node $i$ in time $t + 1$ with probability $p_{ij}$.

In the independent cascade model, every node only has two states, active state and inactive state. Once a node is active, it cannot return to the inactive state. When the node $i$ is active in time $t$, the influence is independently propagated at $t + 1$ from node $i$ to its current neighbors, which are in inactive state in time $t$, with probability $p_{ij}$. And the node $i$ only has one chance to active its inactive neighbors. A node can be independently activated by any of its newly activated neighbors. The propagation process ends when there are no inactive nodes or no nodes with the ability of activating other nodes in the social networks. Now, we define the influence of a set of nodes $A$, denoted $\sigma(A)$, to be the expected number of active nodes at the end of the process. The influence maximization problem is to find a $k$-node set that can achieve maximum influence, where $k$ is a parameter.

Table 1 shows the generic greedy algorithm proposed by Kempe et al. in [9].

### 2.2 Learning Automata

As shown in [12], the learning automaton learns the optimal action by interacting with a random environment. At instant $n$, the LA selects one action $\alpha(n)$ and then sends this action to an environment which will feedback a scalar reinforcement signal, $\beta(n)$. After that, the LA receives this response and takes this response into consideration to update the LA's state. This cycle repeats until the LA selects one optimal action. A discretized linear-based LA can be described by a quadruple $< A, B, \phi, L >$. They are explained as follows:

**Table 1** Algorithm greedy

| Algorithm1 Greedy$(k, \sigma(\cdot))$ |
|---|
| 1:Initialize $S = \emptyset$ |
| 2:for $i = 1 : k$ do |
| 3:select $u = argmax_{w \in V \setminus S}(\sigma(S \cup \{w\}) - f(S))$ |
| 4:$S = S \cup \{u\}$ |
| 5: end for |
| 6:output S |

- $A = \{\alpha_1, \alpha_2, \ldots, \alpha_r\}$ is the set of actions of stochastic automaton;
- $B = \{\beta_1, \beta_2, \ldots, \beta_r\}$ is the set of responses from stochastic environment. In this paper, we consider $B = \{0, 1\}$, where "1" means reward and "0" means penalty;
- $\phi = \{\varphi_1, \varphi_2, \ldots, \varphi_s\}$ is the set of internal states of stochastic automaton. Furthermore, the state consists of $P$, where $P$ is a probability distribution over the set of actions. We have $P(t) = (p_1(t), p_2(t), \ldots, p_r(t))$, where $p_i(t)$ is the probability of selecting action $\alpha_i$ at $t$ and it satisfies $\Sigma_{i=1}^{r} p_i(t) = 1$;
- $L$ is a learning scheme to update $\phi$, namely $P$.

The Discretized Linear Reward-Inaction ($DL_{ri}$) automaton was the first discretized automaton presented in the literature [11]. The probability space $[0, 1]$ is divided into $N$ intervals, where $N$ is resolution parameter and is recommended to be an even integer. Assume the index of the chosen action is $m$, the specific description is as follows. If $\beta = 1$, the automaton updates the probability using the equations: $p_i(t+1) = max\{p_i(t) - \Delta, 0\}$ *for* $i \neq m$, $p_m(t+1) = 1 - \Sigma_{i \neq m} p_i(t+1)$. The $DL_{ri}$ automata is $\varepsilon$-optimal in all random environment.

## 3 LA-IM Algorithm

In this section, we propose a learning automata-based algorithm called LA-IM for the influence maximization problem in social networks. The details of our algorithm are presented as follows:

### *3.1 Improvements of the Original Method*

In a social network, if all nodes are considered as effective nodes, algorithm will slow down. In [3], Guisheng et al. showed most of the nodes are connected to a few other nodes and only a small group of nodes are connected to many other nodes. To reduce the problem space, we use the same method as in [10]. First, mean degree of the total nodes of the network is calculated, and the nodes with lower than mean degree of the graph are ignored in the problem space. The nodes in the problem space are called allowable nodes, and only these nodes are regarded as effective actions of the learning automaton.

One of the most important parts of learning automata algorithms is the updating equations. Once the updating equations are determined, the algorithm is determined. Now, we describe two kind of updating equations, $DL_{ri-k}$ and $DL_{rp-k}$, which are derived from the discretized linear algorithm.

Different from the discretized linear reward-inaction algorithm, $DL_{ri-k}$ proposed in this paper rewards probabilities of all the selected actions according to the updating equations shown in Table 2, where assume the current seedset is $S_m$.

**Table 2** Algorithm $DL_{ri-k}$

| |
|---|
| 1:If $\beta = 1$ do |
| 2:$p_i(t+1) = max\{p_i(t) - \Delta, 0\} \quad for \quad i = 1,2,\ldots,r$ |
| $p_j(t+1) = p_j(t) + (1 - \Sigma_{i=1}^{r} p_i(t+1))/k \quad for \quad j \in S_m \qquad (1)$ |
| 3:else do |
| 4:$p_i(t+1) = p_i(t)$ for i=1,…,r |

Different from $DL_{ri-k}$, $DL_{rp-k}$ not only updates the current chosen action set when the response is reward. It also rewards the optimal action set when the response is penalty. When the response $\beta = 1$, update the current seed set using Eq. 1. When the response $\beta = 0$, update the optimal seed set using Eq. 1.

## *3.2 Application of LA-IM*

Assume the action number is $r$, then the probability of selecting each node at time $t = 0$ is initialized to $1/r$. In each iteration, the automaton chooses $k$ different actions randomly based on the action probability vector. Then the chosen actions are sent as a seed set S to the independent cascade model (environment) to calculate the influence value $\sigma(S)$. If the value $\sigma(S)$ is the best, so far, the feedback of the environment is $\beta = 1$ and vice versa. Then the automaton updates its action probability vector according to the updating equations. The iteration ends until the automaton has selected the best $k$ actions. Namely the sum of the top $k$ probability of the actions is over 0.999.

To increase convergence speed, the original learning process is divided into two parts, sampling process and learning process. In the process of sampling, a heuristic method [10] is used to fully sample the actions. After receiving the feedback from the environment, the automaton updates the probability vector by the $DL_{ri-k}$ updating equation. In the process of the learning, the $k$-seed nodes are selected randomly based on their probabilities. And the learning automaton updates its action probability vector according to the $DL_{rp-k}$ updating equations. The specific description is given in Table 3.

## *3.3 Analysis of LA-IM*

Because of using the heuristic method, we can believe that the approximate optimal seedset has been found after finishing the sampling process. That is to say LA-IM is effective in the social network. Both processes of the algorithm use Monte-Carlo to estimate the value of influence spread. Instead of executing 10,000 Monte-Carlo experiments for every seedset, LA-IM only needs 100 times. It is because LA-IM does not need to make the final decision after one traversal. So it does not require high accuracy of the $\sigma(S)$. The reduction of the execution number of Monte-Carlo

**Table 3** Algorithm LA-IM

Input:Graph $G = (V, E, p)$, Size $k, \alpha, \beta$, $GenerationsCount$;
Output:set of top-k influence nodes $S$;
Initialize: $p(0) = 1/r, for 1 \leq i \leq r, |S| = k, bestNodes = \emptyset, N = \emptyset, gens = 0; Inf = 0, bestInf = 0;$

1:Sampling process:
While ( $t < 2500$ ) do
  if $\sigma(N_n) \geq \theta$ $N_n \rightarrow N$ (put $bestNodes$ from previous generation to current seedset)
  While($N.Count < k$)
    The learning automaton selects an action randomly based on the probability vector
  $Inf$=calculate $\sigma(N)$
  If ($Inf > bestInf$)  ($DL_{ri-k}$)
    Reward $N$
    $bestInf = Inf$
    $N \rightarrow S$
  End while
2:Learning process:
While(learning automaton has not converged)
  While($N.Count < k$)
    learning automaton pick up an action randomly based on the probability vector
Inf=calculate $\sigma(N)$
  If ($Inf > bsetInf$)  ($DL_{rp-k}$)
    Reward $N$
    $bestInf = Inf$
    $N \rightarrow S$
  Else
    Reward $S$
  End while

experiments greatly decrease the time LA-IM takes to select the optimal seedset in a social network. Therefore, our proposed algorithm improves the speed of the convergence.

## 4 Experiments and Results

In real-life cases, there is no criterion for deciding the final result is correct or not. So the traditional method to find the best parameter is not applicable any more. In this paper, we increase the parameter $N$ to find the optimal influence spread in the social network, where the objective functions is the influence spread. When the influence spread does not change, we think we have found the best parameter. We compare the LA-IM algorithm with greedy algorithm on a real-world social network, NetworkScience. It contains 1589 nodes as well as 2742 edges, and has significant community structure. The experiment executes on the IC model with $p = 0.1$. The best parameters corresponding to different k are shown in Table 4. Figure 1 shows the performance of the LA-IM algorithm in comparison with greedy algorithm. The execution time of the two algorithms are shown in Fig. 2.

**Table 4** Best parameters corresponding to different k in the network science network

| k | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| $n$ | 23 | 20 | 23 | 25 | 23 | 20 | 23 | 23 | 23 | 31 |

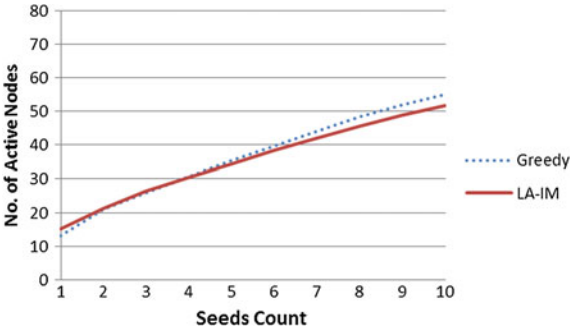



**Fig. 1** Comparing LA-IM and the greedy algorithm (in terms of accuracy)

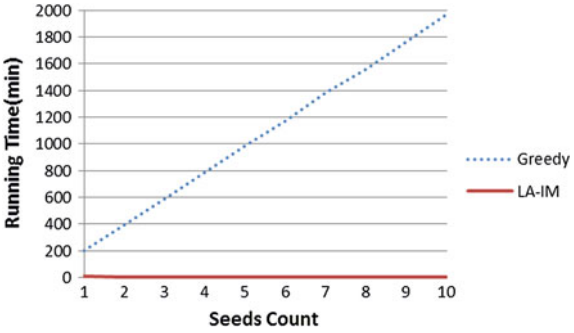



**Fig. 2** Comparing LA-IM and the greedy algorithm (in terms of time)

From Fig. 1, we know when the number of initial nodes $k < 4$, the LA-IM algorithm performs very closely to the greedy algorithm; as the number of initial active nodes k exceeds 4, the performance of the LA-IM falls behind that of the greedy algorithm, but it still keep a high accuracy.

Seen from Fig. 2, the LA-IM greatly improves the speed and the execution time of the LA-IM algorithm does not increase with the seeds count. For example, to produce a 5-element seed set, the greedy needs 989 min, while the new method only needs 5.9 min (approximately 170 times reduction). This difference will increase with increasing the number of elements of the seed set.

## 5 Conclusion

In this paper, we propose a learning automaton-based method to solve the influence maximization problem. To improve the speed without losing accuracy, the algorithm is divided into two process and it uses a heuristic method. First, we analyze the

effectiveness of the new algorithm. Then, in Sect. 4, the obtained results demonstrate the proposed method greatly improves the speed and keeps a high accuracy. In the future work, we consider taking use of the learning automaton with estimator, which may show a better performance.

**Acknowledgements** This research work is funded by the National Science Foundation of China (61271316), Key Laboratory for Shanghai Integrated Information Security Management Technology Research, and Chinese National Engineering Laboratory for Information Content Analysis Technology.

## References

1. W. Chen, C. Wang, Y. Wang, Scalable influence maximization for prevalent viral marketing in large-scale social networks, in *KDD 2010* (2010), pp. 1029–1038
2. A. Goyal, W. Lu, L.V.S. Lakshmanan, Celf++: optimizing the greedy algorithm for influence maximization in social networks, in *International Conference on World Wide Web, WWW 2011* Hyderabad, India, March 28–April 2011, pp. 47–48
3. Y. Gui-Sheng, W. Ji-Jie, D. Hong-Bin, L. Jia, Intelligent viral marketing algorithm over online social network, in *2011 Second International Conference on Networking and Distributed Computing (ICNDC)* (2011), pp. 319–323
4. K. Jung, W. Heo, W. Chen, Irie: scalable and robust influence maximization in social networks **5**(1), 918–923 (2011)
5. D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2003), pp. 137–146
6. J. Kim, S.K. Kim, H. Yu, Scalable and parallelizable processing of influence maximization for large-scale social networks? in *IEEE International Conference on Data Engineering* (2013), pp. 266–277
7. S. Lei, S. Maniu, L. Mo, R. Cheng, P. Senellart, Online influence maximization, in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2015), pp. 645–654
8. J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. Vanbriesen, N. Glance, Cost-effective outbreak detection in networks, in *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2007), pp. 420–429
9. A. Manstead, M.E. Hewstone, S.T. Fiske, *The Blackwell Encyclopedia of Social Psychology* (Blackwell Reference, 1995)
10. A. Mohammadi, K. Khamforoosh, Influence maximization in social networks using learning automata (2015)
11. B.J. Oomen, Absorbing and ergodic discretized two-action learning automata. IEEE Trans. Syst. Man Cybern. **16**(2), 282–293 (1986)
12. B.J. Oommen, M. Agache, Continuous and discretized pursuit learning schemes: various algorithms and their comparison. IEEE Trans. Syst. Man Cybern. Part B Cyber. A Publ. IEEE Syst. Man Cybern. Soc. **31**(3), 277–287 (2001)

# Last-Position Elimination-Based Parallel Learning Automata

**Yuyang Huang, Hao Ge, Jinchao Huang, Fanming Wang and Shenghong Li**

**Abstract** The updating scheme for Learning Automata (LA) is important. Thathachar and Arvind first proposed parallel operation of leaning automata (LA), which was a promising mechanism that could maintain the accuracy while reducing convergence time. In this paper, we implement this mechanism which helps to break the limit of the convergence speed of single LA. In contrast to existing scheme, the proposed scheme eliminates the worst performed LA in the sequence of interactions, until there is only one LA left. We compare Last-Position Elimination-Based Parallel Learning Automata (LEPLA) scheme with the classic one incorporating two pursuit schemes, discretized generalized pursuit algorithm (DGPA) and discretized pursuit algorithm with reward-inaction ($DP_{RI}$), respectively. Simulations prove that the proposed scheme gets an evidently higher accuracy and faster convergence than the classic ones.

**Keywords** Artificial intelligence · Learning automaton · Elimination · Parallel

Y. Huang (✉)
Shanghai Starriver Bilingual School, Shanghai, China
e-mail: markhuanghyy@outlook.com

H. Ge · J. Huang · S. Li
Department of Electronic Engineering, Shanghai Jiao Tong University, Shanghai, China
e-mail: sjtu_gehao@sjtu.edu.cn

J. Huang
e-mail: hjc2015@sjtu.edu.cn

S. Li
e-mail: shli@sjtu.edu.cn

F. Wang
Riffle Institute, Singapore, Singapore
e-mail: fanmingwang@gmail.com.sg



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_77

## 1 Introduction

Learning Automaton (LA) is a self-adaptive machine that can find the optimal action from a stochastic environment, and the goal of LA is to maximize the expected reward under the random environment. Usually, it has a finite number of actions which can have interactions with the environment, as is shown in Fig. 1 [1]. In each cycle, it selects an action $\alpha(t)$ and then gets a random response $\beta(t)$ that is either a reward or a penalty. LA uses environment's response to update the state probability vector of choosing the next action. Therefore, LA has a broad range of applications, such as game playing [2], congestion avoidance in wired networks [3], and tutorial–like systems [4].

Besides single LA, there is another system called multi-LA. Parallel operation of LA [5] is one of multi-LA systems and it targets to improve the speed by using the parallel nature of the environment. Different from the learning process of single LA that at an instant only one action is selected and one feedback is received from the environment. Parallel operation of LA sends several actions collectively as an input and uses all feedback signals to update the internal state of LA. An example of parallel operation of LA is illustrated in Fig. 2.

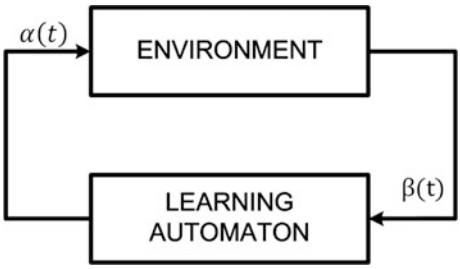



**Fig. 1** A single LA that interacts with a random environment

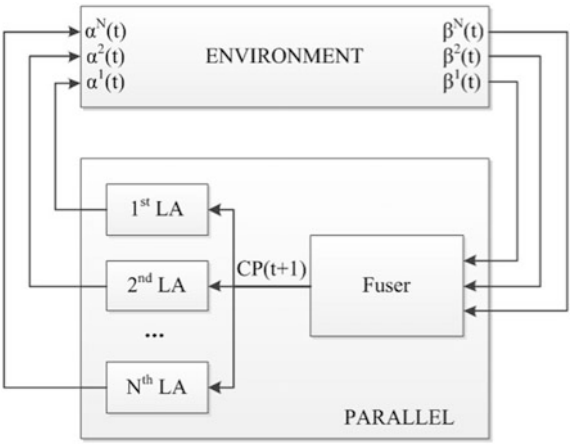



**Fig. 2** The framework of parallel operation of LA

In parallel operation of LA, at any time instant t, N actions are selected and N responses are received simultaneously.

In this paper, we propose another parallel framework which improves the classical parallel algorithm introduced in [5]. Also, numerical simulations are performed to verify the effectiveness of the proposed mechanism.

The remainder of this paper is organized as follows. The related parallel mechanism is proposed in Sect. 2. New parallel framework is proposed in Sect. 3. Various analyses are presented in Sect. 4. Eventually, Sect. 5 concludes the paper.

## 2 Related Works

Thathachar and Arvind presented a parallel version of estimator-based LA [1, 6], which intended to reduce computation time. In this paper, mathematical symbols are consistent with "A Novel Parallel Framework for Pursuit Learning Schemes" [7]. Assuming the parallel system consists of $N$ identical LA, they have a common action set $A = \{\alpha_1, \alpha_2, \alpha_3, \ldots, \alpha_r\}$, a common probability vector $CP(t) = [cp_1(t), cp_2(t), \ldots, cp_r(t)]$, and a shared estimator vector $D(t)$.

**Step 1** Each LA, say *mth* LA, selects an action $\alpha^m(t)$ based on the $CP(t)$ independently, and all other LAs each obtains its own reinforcement signal $\beta^m(t)$;

**Step 2** A fuser combines all selected actions and feedbacks to update the common probability; $cp_i(t+1) = cp_i(t) + \lambda \times [R_i(t) - cp_i(t) \sum_{1}^{N} R_i(t)]$

**Step 3** Update the common estimator

$$W_i(t+1) = W_i(t) + R_i(t) \tag{1}$$

$$Z_i(t+1) = Z_i(t) + C_i(t) \tag{2}$$

where $R_i(t) = \sum_{j=1}^{N} \beta^j(t) \times I\{\alpha^j(t) = \alpha_i\}$

and $C_i(t) = \sum_{j=1}^{N} I\{\alpha^j(t) = \alpha_i\}$;

**Step 4** If $\max\{CP(t+1)\} = 1$ then converge. Else go to **Step 1**.

However, this parallel algorithm is not widely used in practical areas because in this parallel framework, LA is acting as action selector instead of learning module. The task of each LA is to choose the action based on the $CP(t)$, while updates are done by the fuser.

## 3 Proposed Parallel Learning Automata

As is pointed out previously, the sluggish learning speed is one of the most important problems of classical parallel learning automata. In the classical parallel algorithm, the parallel learning automata can reduce the convergence time approximately for a factor of $N$, which is the number of modules. The reason is that probability vector updates with multiple samples received by the parallel scheme. However, this algorithm does not target to decrease the number of interactions with the environment, but to save time of computation. In this paper, the focus point is to decrease the number of interactions through parallel operation.

In Last-Position Elimination-Based Learning Automata (LELA) [8], the team has proved that the last-position elimination-based algorithm is effective on single LA, which can lead to a faster convergence than usual optimal pursuit algorithm, such as Discretized Generalized Pursuit Algorithm (DGPA) [9] and Discretized Pursuit scheme ($DP_{RI}$) [10, 11]. The reason behind it is that avoiding high penalized actions can lead to a higher reward probability. In the same way, if a parallel operation can abandon the worst performance LA in the framework with certain condition, the probability of getting reward will increase while the number of interactions will decrease.

Therefore, we propose a parallel scheme that has a common estimator and every module has its own probability vector and estimator. After a fixed number of loops, the parallel framework checks the number of rewards of each LA to find the one with the fewest rewards. Then, the worst one will be eliminated from the framework. At the same time, common estimator will reset each LA's estimator. Last-Position Elimination-Based Parallel Learning Automata can not only save computing space, but also decrease the total number of iterations. The parallel framework is explained here.

**Last-Position Elimination-Based Parallel Learning Automata Parameters**

| | |
|---|---|
| n | Resolution parameter |
| N | Size of the module |
| $W_i(t)$ | The number of events that the $i$th action has been rewarded up to time instant $t$, for $i = 1, 2, \ldots, r$ in the whole parallel framework |
| $Z_i(t)$ | The number of events that the $i$th action has been selected up to tome instant $t$, for $i = 1, 2, \ldots, r$ in the whole parallel framework |
| $\Delta = 1/r/n$ | Smallest step size |
| $\rho^l(t)$ | The performance of $l$th LA, at time $t$ |
| $t_m$ | The number of loop when it is time to eliminate the worst performance LA |

**Method**

**Initialize** each LA, say $l$th LA to be: $p_i^l = 1/r$, for $i = 1, 2, \ldots, r$ and $l = 1, 2, \ldots, N$

**Initialize** $\beta^l = 0$, for $l = 1, 2, \ldots, N$

**Initialize** $W_i(t) = 0$ and $Z_i(t) = 0$
**Initialize** $\rho^l = 0$, for $l = 1, 2, \ldots, N$
Repeat
**Step 1** at time t, each LA, say *l*th LA,, selects an action $\alpha^l = (t)$ based on its own probability vector $P(t)$ independently and gets a signal $\beta^l = (t)$. *l*th LA updates its own estimator by

$$W_i^l(t+1) = W_i^l(t) + \beta(t)$$
$$Z_i^l(t+1) = Z_i^l(t) + 1$$
$$d_i^l(t+1) = \frac{W_i^l(t+1)}{Z_i^l(t+1)}$$

**Step 2** The fuser receives all feedbacks

$$W_i(t+1) = W_i(t) + \sum_{j=1}^{N} \beta^j(t)$$
$$Z_i(t+1) = Z_i(t) + \sum_{j=1}^{N} 1 = Z_i(t) + N$$

**Step 3** The fuser gets each LA's performance

$$\rho^l(t+1) = W_i^l(t+1)$$

**Step 3**: Calculate the common estimator's value, reset each LA's estimator to the common estimator's value

**If** $t = \mathrm{t_m}$ **Then**
$$d_i(t+1) = W_i(t+1)/Z_i(t+1)$$
$$d_i^l(t+1) = d_i(t+1),\ \mathrm{l} \in \mathrm{N}$$
Find $\mathrm{l} \in \mathrm{N}$ , such that
$$\rho^l(t_m) = min\ \{W_i^l(t_m)\}$$
**If** $\mathrm{N} \geq 2$ **Then** Remove $l^{th}$ LA
$$N = N - 1$$
**Endif**
**Endif**

**Step 4**: **If** $\max\{P_i^l(t) = 1\}$, **Then** CONVERGE to the action whose $p = max\{P(t)\}$. **Else** go to step 1. **Endif**
**End Last-Position Elimination-Based Parallel Learning Automata**

Different from classical parallel framework, the Last-Position Elimination-Based Parallel Learning Automata (LEPLA) gives each learning automaton its own probability vector, other than a common probability vector.

# 4 Experimental Results

## 4.1 Simulation Results

In this part, the proposed Last-Position Elimination-Based Parallel Learning Automata (LEPLA) is compared with the classical parallel scheme; two pursuit LA algorithms, Discretized Generalized Pursuit Algorithm (DGPA) and Discretized Pursuit scheme ($DP_{RI}$), are embedded in these two frameworks, respectively.

Two ten-action (r = 10) problems, $E_A, E_B$, are famous benchmark environments used right now [1]. The reward probabilities for the actions of the five benchmark environments are:

$$E_A = \{0.70, 0.50, 0.30, 0.20, 0.40, 0.50, 0.40, 0.30, 0.50, 0.20\};$$
$$E_B = \{0.10, 0.45, 0.84, 0.76, 0.20, 0.40, 0.60, 0.70, 0.50, 0.30\};$$

Among all simulations performed, the algorithm is considered to converge if the probability of choosing an action is greater or equal to a threshold $T(0 < T < 1)$. If the converged action is the one with the highest probability of being reward, the system is recognized as converging correctly.

"Best" parameters of a pursuit algorithm are defined as the values that not only yield the quickest convergence speed but also make sure the system converged to the correct action with a continuous *NE* experiments. In our experiments, simulations are performed with the same threshold $T$ and the number of experiments *NE* [12]. That value is $T = 0.999$ and $NE = 750$. After finding the best parameters, we use 250,000 experiments to evaluate the average accuracy and convergence rate.

Indication of comparison is defined here. Accuracy is the result of number of correctly convergence/number of experiments. We compare the proposed one with single operated LA and classical parallel. The result of improvement is calculated as following:

$$improvement = \frac{\text{Interaction}_{LELPA} - \text{Interaction}_{original}}{\text{Interaction}_{original}} * 100\% \tag{3}$$

$$interaction = N * Iteration \tag{4}$$

## 4.2 Performance Analysis

From the above simulation results shown in Tables 1 and 2, we may draw the following conclusions.

First of all, the proposed parallel operation is superior to the classical operation with the number of interactions. Since we have already multiplied N to the number of iterations, the highest improvement is up to 65.9%. For example, in $E_A$ when

**Table 1** Average convergence speed and accuracy over 250,000 experiments of the generalized pursuit algorithm (DGPA) working in different parallel frameworks

| EA | Classic parallel | | | | Proposed parallel | | | | Proposed one compares with | |
|---|---|---|---|---|---|---|---|---|---|---|
| N | Parameter | Iteration | Interaction | Accuracy | Parameter | Interaction | Elimination period | Accuracy | classic (%) | single LA (%) |
| 1 | 216 | 783 | 783 | 0.995 | 201 | 788 | \ | 0.995 | −0.6 | 0.0 |
| 2 | 108 | 442 | 884 | 0.995 | 51 | 506 | 197 | 0.995 | 42.8 | 35.7 |
| 3 | 72 | 328 | 984 | 0.995 | 22 | 457 | 87 | 0.995 | 53.6 | 42.0 |
| 4 | 54 | 273 | 1092 | 0.995 | 21 | 372 | 36 | 0.995 | 65.9 | 52.8 |
| EB | Classic parallel | | | | Proposed parallel | | | | Proposed one compares with | |
| N | Parameter | Iteration | Interaction | Accuracy | Parameter | Interaction | Elimination period | Accuracy | classic (%) | single LA (%) |
| 1 | 881 | 2364 | 2364 | 0.995 | 932 | 2514 | \ | 0.995 | −6.3 | 0.0 |
| 2 | 440 | 1161 | 2322 | 0.995 | 298 | 1608 | 536 | 0.995 | 30.8 | 36.9 |
| 3 | 294 | 789 | 2367 | 0.995 | 125 | 1442 | 272 | 0.995 | 39.1 | 43.5 |
| 4 | 220 | 578 | 2312 | 0.995 | 81 | 1371 | 154 | 0.995 | 40.7 | 46.2 |

**Table 2** Average convergence speed and accuracy over 250,000 experiments of the discretized pursuit reward-inaction algorithm (DPri) working in different parallel frameworks

| EA | Classic parallel | | | | Proposed parallel | | | | Proposed one compares with | |
|---|---|---|---|---|---|---|---|---|---|---|
| N | Parameter | Iteration | Interaction | Accuracy | Parameter | Interaction | Elimination period | Accuracy | classic (%) | single LA (%) |
| 1 | 28 | 753 | 753 | 0.997 | 42 | 1007 | \ | 0.997 | −33.7 | 0.0 |
| 2 | 10 | 331 | 662 | 0.997 | 16 | 642 | 229 | 0.997 | 3.0 | 36.2 |
| 3 | 5 | 251 | 753 | 0.997 | 12 | 599 | 96 | 0.997 | 20.5 | 40.6 |
| 4 | 3 | 171 | 684 | 0.997 | 9 | 585 | 58 | 0.997 | 14.5 | 41.9 |
| EB | Classic parallel | | | | Proposed parallel | | | | Proposed one Compares with | |
| N | Parameter | Iteration | Interaction | Accuracy | Parameter | Interaction | Elimination period | Accuracy | classic (%) | single LA (%) |
| 1 | 55 | 1445 | 1445 | 0.997 | 59 | 1481 | \ | 0.997 | −2.5 | 0.0 |
| 2 | 26 | 727 | 1454 | 0.997 | 45 | 1369 | 237 | 0.997 | 5.8 | 7.5 |
| 3 | 18 | 497 | 1491 | 0.997 | 21 | 1249 | 231 | 0.997 | 16.2 | 15.6 |
| 4 | 12 | 385 | 1540 | 0.997 | 21 | 1286 | 124 | 0.997 | 16.5 | 13.2 |

$DP_{RI}$ is embedded in parallel framework and the parallel scale is 4, classical scheme needs 1092 interactions to converge, while the proposed one only needs 372 interactions. What we want to clarify is that when the parallel scale is 1, the proposed one is surpassed by the original one, but technically speaking, using one LA in one iteration cannot be deemed to be "parallel operation".

Moreover, with the increase of modules, the proposed parallel continuously decreases its interactions. Different from conventional parallel scheme, which nearly maintains the total number of interactions, the proposed one requires fewer interactions, and finally outperforms classical one. In Table 1, we can notice that the advantage increases from 40 to 60%. In Table 2, the advantage changes from 3 to 16%. The improvement is nearly monotone increasing with parallel scale.

## 5 Conclusion

In this paper, we introduce a novel parallel framework for pursuit learning automata. By implementing last-position elimination philosophy [8], this framework outperforms the classic parallel structure in the aspect of required number of interactions. The main reason for improvement is that the proposed scheme discards LA which gets relatively fewer number of rewards [7]. Thus, the framework obtains a higher accuracy. Simulations experiments have confirmed the efficiency of the proposed parallel system. A parallel mechanism that can fully use the potential of large parallel scale deserves further investigation.

**Acknowledgements** This research work is funded by the National Science Foundation of China (61271316), Shanghai Key Laboratory of Integrated Administration Technologies for Information Security and Chinese National Engineering Laboratory for Information Content Analysis Technology.

## References

1. M.A. Thathachar, P.S. Sastry, *Networks of Learning Automata: Techniques for Online Stochastic Optimization* (Springer Science and Business Media, 2011)
2. B.J. Oommen, M. Agache, Continuous and discretized pursuit learning schemes: various algorithms and their comparison. Syst. Man Cybern. Part B Cybern. **31**, 277–287 (2001)
3. O. Tilak, R. Martin, S. Mukhopadhyay, Decentralized indirect methods for learning automata games. **41**(10) (2011)
4. S. Misra, B.J. Oommen, S. Yanamandra, M.S. Obaidat, Random early detection for congestion avoidance in wired networks: a discretized pursuit learning-automata-like solution. Syst. Man Cybern. Part B Cybern. 66–76 (2010)
5. B.J. Oommen, M.K. Hashem, Modeling a student classroom interaction in a tutorial-like system using learning automata. Man Cybern. Part B Cybern. 29–42 (2010)
6. M.A.L. Thathachar, P.S. Sastry, Estimator algorithms for learning automata, in *Platinum Jubilee Conference*, vol. 12 (1986), pp. 29–32

7. H. Ge, Y. Wang, J. Li, W. Jiang, S. Li, A novel parallel framework for pursuit learning. Nucl. Phys. B 1–20 (2016)
8. J. Zhang, C. Wang, M. Zhou, Last-position elimination-based learning automata. Cybernetics 2484–2492 (2014)
9. M. Agache, B.J. Oommen, Generalized pursuit learning schemes: new families of continuous and discretized learning automata. Syst. Man Cybern. Part B Cybern. 738–749 (2002)
10. M.A.L. Thathachar, B.J. Oommen, Discretized reward-inaction earning automata. J. Cybern. Inform. Sci. **2**, 24–29 (1979)
11. B.J. Oommen, J.K. Lanctot, Discretized pursuit learning automata. Syst. Man Cybern. 931–938 (1990)
12. R.S. Sutton, A.G. Barto, *Reinforcement Learning: An Introduction*

# Research on Dynamic Gesture Recognition Based on Multi Feature Fusion

**Meijuan Yu and Yankai Liu**

**Abstract** Gesture recognition is an indispensable part of the human–computer interaction technology. In this paper, the research on dynamic gesture recognition technology based on multi feature fusion is studied. While identifying the posture using SVM, the dynamic gesture track feature is extracted and recognized using Gaussian pyramid optical flow algorithm. Then we'll get the final recognition results by information fusion on decision level. Finally through the contrast experiment to prove the dynamic gesture recognition algorithm in the paper has higher gesture recognition rate.

**Keywords** Dynamic gesture · Multi feature fusion · SVM · Gauss pyramid optical flow algorithm

## 1 Introduction

With the continuous development of human–computer interaction technology, gesture recognition [1] has become a popular research topic. The gesture is a natural and intuitive man–machine communication mode. In interaction between human and computer, gesture recognition is an essential and indispensable technology to realize the human–computer interaction of new generation, the so-called gesture recognition is simply refers that the computer according to the trajectory recognition of arm and hand to figure out its meaning, including static gesture recognition and dynamic gesture recognition. In contrast, the static gesture recognition is relatively simple, at present, the common used gesture recognition technology include the template matching method, neural network method and statistical analysis

M. Yu · Y. Liu (✉)
College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China
e-mail: ooaround@foxmail.com

M. Yu
e-mail: amy_1121@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_78

method, the accuracy of these techniques in recognizing static gestures are still good, but for the recognition of dynamic gesture, the recognition accuracy is supposed to be improved.

## 2 Dynamic Gesture Recognition Based on Multi Feature Fusion

The focus of dynamic gesture recognition is the accurate extraction of gesture features, mainly consists of two parts, hand shape feature, and gesture trajectory. This paper firstly uses gesture profile of the first frame of the gesture image to obtain the profile matrix, and then uses the Fourier transform [2] to convert the hand shape characteristic matrix into the feature vector with one dimension, using SVM algorithm to recognize the gesture features. Then obtain the trajectory characteristics of hands by Gaussian pyramid optical flow algorithm, and fusing the recognition results of trajectory feature and hand type feature in decisional degree [3, 4], to get the final recognition results.

### *2.1 Hand Feature Extraction and Recognition Based on SVM*

Support vector machine [5] is the VC dimension theory and structural risk minimization principle established on the statistical learning theory, according to the limited sample information between the complexity and the study ability of the model to seek the best compromise, so as to obtain the best generalization performance, SVM is able to automatically search for the samples (the support vectors) that have better distinguish ability in classification [6], the optimal segmentation surface constructed by these samples to furthest the interval between the classes.

The project make uses SMO (Sequential Minimal Optimization) [7] to conduct training when realize classifier by SVM, selecting the sigmoid function as the kernel function. Using heuristic search algorithm to improve the SMO, so as to realize the intelligent parameter selection, effectively reduced the workload of manual selection of parameters. Specific steps are as follows:

(a) Ergodic non boundary data samples, when ergodic data samples violated the KKT conditions, it is selected as one of the optimal multipliers a1.
(b) Continue to traverse the other samples, take the advantage of heuristic search algorithm, make the step size of traverse between the sample and a1 as the standard, to find the sample data with maximum optimal step size a2, and to optimize and update.
(c) Repeat step (a) and (b), until all non boundary samples are satisfied with KKT conditions.

(d) Checking all samples of data sets, to judge whether they are satisfied with the KKT condition, if there are samples that not satisfied with it, then repeat steps (a)–(c), until all the sample data meet the KKT condition, then the algorithm finishes.

## 2.2 *Extraction and Modeling of Optical Flow Features of Gesture*

This paper adopts the Gauss Pyramid optical flow algorithm [8, 9], to extract and modeling of optical flow characteristics of dynamic gesture, optical flow vector obtained by calculation can be expressed by a four-dimensional space, in which the $(x, y)$ represents the position information of the optical flow vector in the image, $(\mathrm{u}, \mathrm{v})$ representing the component size of the optical flow vector in the horizontal direction. There are M * N optical flow vector in an optical flow field (in which the M * N is the resolution of the image). The point of extraction and modeling of gesture flow feature lies in the reduction processing of dimension of high original feature vector, that is to remove the redundant feature vector or the feature vector without obvious motion feature, this paper extracts the optical flow feature gesture video sequence. Here we will reduce the dimension of the original feature.

Transfer the component $(\mathrm{u}, \mathrm{v})$ of optical flow vectors into polar coordinates $(r, \theta)$. Through the test, $\theta$ has the better recognition ability, so this paper may ignore the feature of r, choose $\theta$ to represent the velocity of the optical flow. Thus we transfer the optical flow field into the obtained three-dimensional vector $(x, y, \theta)$ which is similar to the common gray image. Then it represents the optical flow field with a one-dimensional array, that is also expresses the movement behavior. Using the histogram intersection [10] to calculate the similarity between the optical flow field [11], so as to achieve recognition of movement.

# 3 Experiment Results Analysis

The experiment first use a camera to get the dynamic gesture sequence, then extract and recognize the hand type features and optical flow features, take the example of dynamic hand gesture recognition "palm forward," detailing the process gesture recognition based on multi feature fusion.

(a) Collect the video sequence of the gesture by the camera, as shown in Fig. 1, dynamic monitoring image of the forward gesture.
(b) Extract the first frame of the image, to segment gesture based on skin color, and to conduct the diffusion filter and contour extraction treatment of the image, the

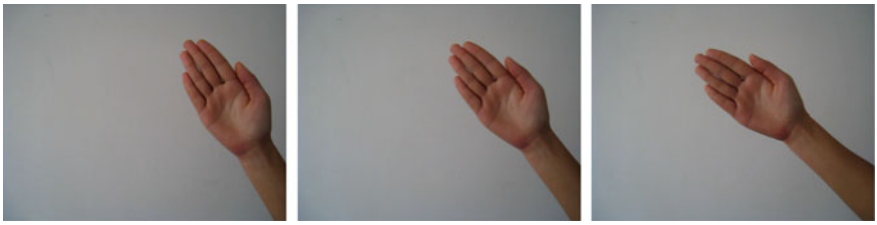


**Fig. 1** Dynamic monitoring image of the forward gesture

**Fig. 2** Gesture contour image

first obtained frame gesture image after treatment as shown in Fig. 2, using SVM algorithm to recognize the hand gesture contour features.

(c) To extract and recognize the trajectory feature of dynamic hand gesture, computation of optical flow is sensitive to noise, so it is necessary to pretreat the read video images, so as to reduce the optical flow calculation and improve the accuracy of optical flow calculation. This paper used the frame difference method to determine the motion region of the object. As shown in Fig. 3, they are binary differential images of adjacent frames in gesture sequences.

Binary differential images obtained by frame difference method still have obvious salt and pepper noise, it is needed the process of removing the noise. Fig. 4, Fig. 5, respectively, represent the optical flow field of Fig. 1 before and after

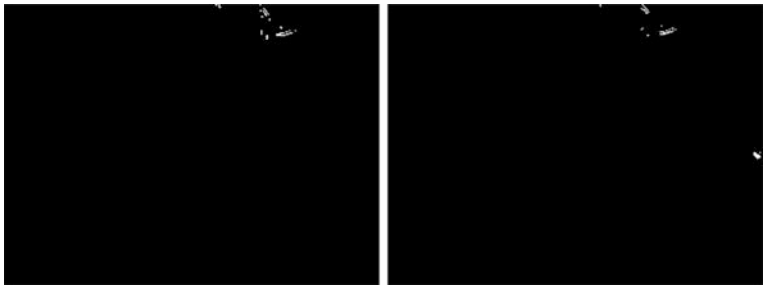


**Fig. 3** Binary differential images obtained by frame different method

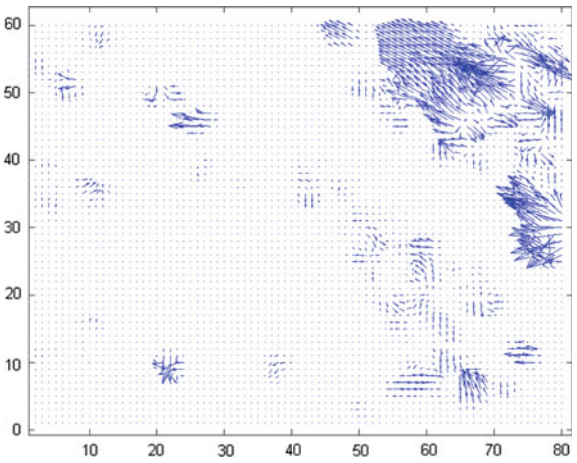



**Fig. 4** Optical flow field of two images before preprocessing

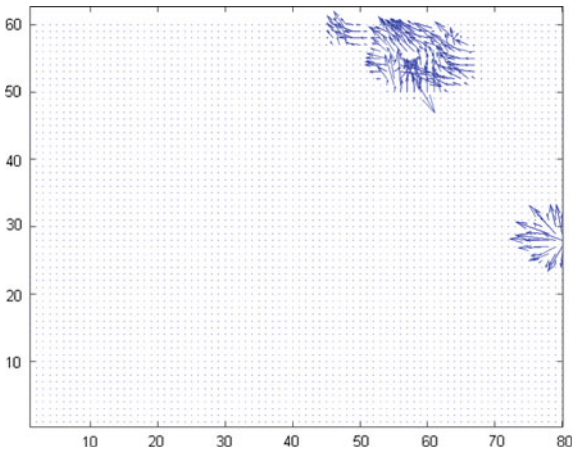



**Fig. 5** Optical flow field of two images after preprocessing

the image preprocessing. Obviously, the fine accurate optical flow field without noise was obtained after pretreatment.

This test collected four kinds samples of dynamic hand gestures, respectively, including, palms forward, palms down, fist backward, eight upward, because people have differences in doing the different hand postures, each kind of dynamic gesture samples provided from 10 individuals and each of them do 10 times, that is to say, each gesture recognition has a total of 100 test samples, then, to contrast the dynamic hand gesture recognition based on the traditional support vector machine (SVM) to the multi feature fusion of gesture recognition, it is concluded that the recognition results as shown in Fig. 6. The experiment shows that, in the four kinds of recognition, the algorithm designed by this paper only occurs once error in "eight upward," the recognition rate has a certain of improvement.

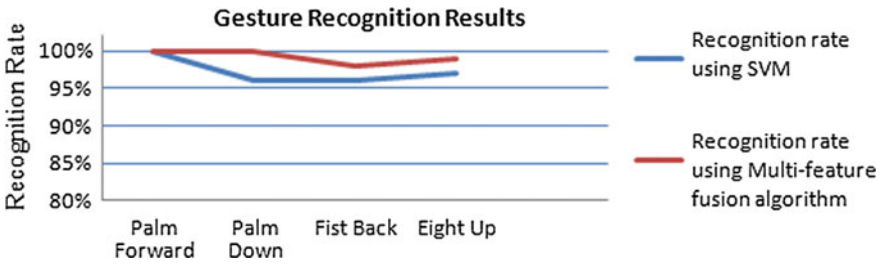



**Fig. 6** Comparative experiments identification results

## 4 Conclusion

In this paper, we recognize the hand posture and motion features by adopting the SVM and Gaussian pyramid optical flow algorithm, respectively, and realize the dynamic hand gesture recognition by decision level fusion technology, using the contrast experiment to verify the algorithm that the dynamic hand gesture recognition technique based on decision level fusion has higher recognition rate, and it will has sound application in the field of human–computer interaction in the future.

## References

1. T.B. Moeslund, L. Nørgaard, A brief overview of hand gestures used in HCI [R]. Technical report No. CVMT 03–02
2. Y. Wang, Face recognition based on fractional fourier transform [D] (Henan: Zhengzhou University, 2015)
3. L.I. Kuncheva, J.C. Bezdek, R.P.W. Duin, Decision templates for multiple classifier fusion: an experimental comparison. Pattern Recognit. **34**(2001), 299–314 (2013)
4. Z. Li, L. Song, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme [J]. Ad Hoc Networks, In Press, Corrected Proof (2016), Accessed 2 Mar 2016
5. L. Huan, W. Shitong, Binary-class classification algorithm of multiple-access acquisited objects based on SVM [J]. CAAI Trans. Intel. Syst. **9**(4), 1–9 (2014)
6. Z. Li, Y. Chen, NDN-GSM-R: a novel high-speed railway communication system via named data networking [J]. EURASIP J. Wireless Commun. Netw. **2016**, 48 (2016)
7. J.C. Platt, Sequential minimal optimization: a fast algorithm for training support vector machines [R]. Technical Report No. 4 (2015)
8. L. TU, S. Zhong, Q. Peng, Moving object detection based on Gaussian pyramid [J]. J. Cent. South Univ. (Science and Technology) **44**(7), 2778–2786 (2013)
9. X. Liu, Z. Li, Information-centric mobile ad hoc networks and content routing: a survey [J]. Ad Hoc Networks, In Press, Corrected Proof (2016), Accessed 19 Apr 2016
10. Y. Zhu, Y. Jia, Y. Wang, Noisy image compressive sensing based on nonlinear diffusion filter [C], in *2013 3rd International Conference on Materials Engineering for Advanced Technologies (ICMEAT 2013)* (2013)
11. Z. Li, K. Liu, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter [J]. IEEE Commun. Lett. **18**(11), 1915–1918 (2014)

# A Method of Target Identification with UWB Based on S-Transform and Improved Artificial Bee Colony Algorithm

**Gang Lei and Ting Jiang**

**Abstract** Ultra-wideband signal (UWB) has been used in communication, location, and identification. In this paper, a novel target identification method is proposed. The UWB received signals is processed by time-frequency analysis method: S-transform. S-transform is an extension of short time Fourier transform and continuous wavelet transform, and it has good time-frequency characteristics. Then we use improved artificial bee colony (ABC) algorithm to optimize the penalty factor and kernel parameter of support vector machine (SVM), and finish the target identification. In view of the basic artificial bee colony algorithm has the problem of slow convergence speed. We propose a probability selection method based on quadratic function to optimize the algorithm.

**Keywords** UWB · Time-frequency analysis · S-transform · Improved artificial bee colony algorithm

## 1 Introduction

UWB signal has the advantages of being not sensitive to the channel fading, low transmitting power spectral density and high location accuracy. In recent years, the UWB technology is widely used in communication and radar detection. UWB radar uses the echo signals to analyze the target. In this paper, we use the received signals to identify the target during the communication between UWB devices.

The traditional way of target recognition in UWB is first to extract the features in the received signals, then input the features into classifier. Wang Y uses fuzzy pattern recognition method to classify the targets [1]. He J introduces the neural

G. Lei · T. Jiang (✉)
Key Laboratory of Universal Wireless Communication, Beijing University of Posts and Telecommunications, Beijing, China
e-mail: tjiang@bupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_79

network classifier [2]. Ying R puts the features of the signal into support vector machine (SVM) to complete recognition [3]. These classification methods have achieved good results. But the signal is processed in time domain. This led to long time process and low efficiency. For stationary signal, feature extraction is able to meet the requirement. But in reality, only a small number of signals are stationary. UWB signal is non-stationary, the statistic characteristics of non-stationary signals are changed by time. So we need to know the change tendency and the value of the frequency spectrum of the signals. Based on this requirement, time-frequency analysis method is proposed. Short time Fourier transform (STFT) method is simple, and it is easy to implement, but the frequency resolution is low. The wavelet transform (WT) can accomplish the multi-resolution analysis of the signal, but the adaptive ability is poor. Tang Q introduces the smooth pseudo Wigner Ville Distribution (WVD) to process the signal in time-frequency domain [4], by changing window to suppress the interference. But at the same time, this method widens the signal, reduces the resolution of the signal. In this paper, we introduce S-transform to process the received signal in time-frequency domain. S-transform is an extension of short time Fourier transform and continuous wavelet transform, and it has good time-frequency characteristics.

The default hyper-parameters of the classifier are not usually optimal. In paper [3], the author picks the parameters of the SVM by the method of permutation and combination, so the theoretical basis of the method is not strong. In this paper, we introduce the improved artificial bee colony algorithm (IABC) to optimize the penalty factor and kernel parameter of SVM, then we compare the recognition accuracy and convergence speed with the basic ABC algorithm.

This paper is organized as follows. In Sect. 2 the real UWB communication scenario is established. In Sect. 3 S-transform is introduced to analyze the signals and IABC algorithm is introduced to optimize the SVM parameters. In Sect. 4 the method is applied to various scenarios, and we compare the identification accuracy and convergence speed with the basic ABC algorithm. Finally, the conclusions are given in Sect. 5.

## 2 Establish the UWB Scenario

The signal collecting scene is established in a small forest near our school. Our work is based on the received signals of UWB communication system.

Two UWB communication nodes are placed at fixed positions in the forest. One is transmitter (TX), the other is receiver (RX). Four different scenarios are measured in the UWB communication scenario. They are no targets, wooden board, iron cabinet and human. The measurement topology is shown in Fig. 1.

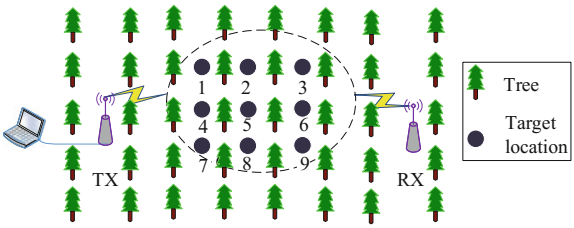



**Fig. 1** The experimental environment model

## 3 Methodology

### *3.1 Theory of S-Transform*

S-transform was put forward by Stockwell in the study of geophysical data [6, 7]. It is a reversible local time-frequency analysis method. The S-transform of the signal is defined as follows.

$$S(\tau,f)=\int_{-\infty}^{\infty}x(t)w(\tau-t,f)\exp(-2\pi ift)dt \tag{1}$$

$$w(\tau-t,f)=\frac{|f|}{\sqrt{2\pi}}\exp[\frac{-f^2(\tau-t)^2}{2}] \tag{2}$$

where $w(\tau-t,f)$ is Gaussian Window, $\tau$ is to control the parameters of Gaussian Window. From the formula it can be seen S-transform is different from short time Fourier transform at the length of Gaussian Window. The height and the width of Gaussian Window in S-transform are changed with frequency. Thus it can overcome the defects of the fixed height and width of Gaussian Window in short time Fourier transform.

S-transform can be realized by FFT quickly. The discrete expression of S-transform is shown as follows.

$$S[m,n]=\sum_{k=0}^{N-1}X[n+k]\cdot \mathrm{e}^{-2\pi^2k^2/n^2}\cdot \mathrm{e}^{\mathrm{j}2\pi km/N}\ (n\neq 0) \tag{3}$$

$$S[m,n]=\frac{1}{N}\sum_{k=0}^{N-1}x[k]\quad (n=0) \tag{4}$$

where

$$X[n]=\frac{1}{N}\sum_{k=0}^{N-1}x[k]\cdot \mathrm{e}^{-\mathrm{j}2\pi kn/N} \tag{5}$$

Processing the sample points by S-transform. The transformation results are a two-dimensional matrix. The columns are corresponding sampling time, the rows are corresponding frequency values, and the matrix elements are corresponding amplitude. So an S-transform can be discretely identified in a different frequency band.

## 3.2 Theory of IABC Algorithm

ABC algorithm is an algorithm inspired by bee colony behaviors. It is put forward by Karaboga group to optimize the algebra problem in 2005 [8–10]. In ABC algorithm, the bees are divided into three distinct groups: employed bees, onlookers, and scouts. We need to optimize the parameters of SVM before using it. The parameters including kernel function parameter and penalty factor. The influence of different kernel functions for SVM is not obvious [11]. In this paper, we choose RBF radial basis function. Then we use IABC algorithm to optimize the kernel parameter and penalty factor. The following is the process of IABC algorithm.

(1) *Initialize the bee populations*

ABC algorithm generates $S_N$ employed bees and $S_N$ initial solutions randomly. Then calculate the fitness of each solution according to the fitness function

$$V_{fitness} = \begin{cases} 1/(1+V_{obj}) & V_{obj} \geq 0 \\ 1+\text{abs}(V_{obj}) & V_{obj} \leq 0 \end{cases}, \tag{6}$$

$$V_{obj} = 1 - V_{acc} \tag{7}$$

where $V_{acc}$ is the classification accuracy of the training samples tested by SVM.

(2) *Employed phase*

Each employed bee searches the neighboring field of the original solution. The searching method is as follows:

$$v_{ij} = x_{ij} + \phi_{ij}(x_{ij} - x_{kj}), \tag{8}$$

where $k \in \{1, 2, \ldots, S_N\}$, $j \in \{1, 2, \ldots, D\}$ is the subscript which is selected randomly. $k \neq i$ and $\phi$ is a random number in $[-1, 1]$. Each loop the employed bee will find a new solution, and replace the old solution if the fitness is better.

(3) *Onlooker phase*

Each onlooker bee selects a solution to search the neighboring field. In Basic ABC algorithm, the probability of selecting according to the fitness of the solution

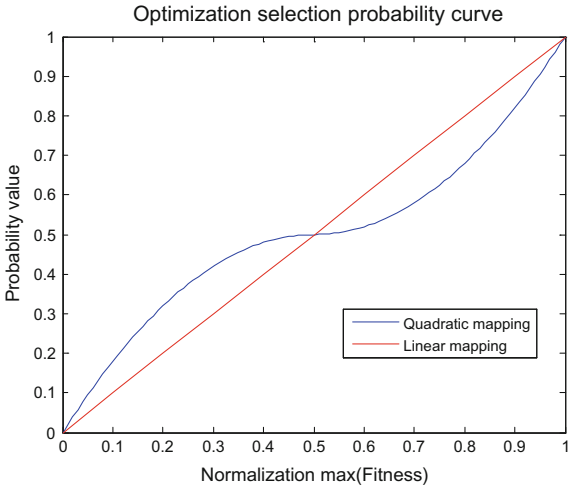



**Fig. 2** Quadratic function mapping curve

$$p_i = \frac{V_i}{\sum_{n=1}^{S_N} V_n}, \tag{9}$$

where $V_i$ is the fitness of each solution and the commonly used selection method is roulette algorithm. In this case, the probability of the food source is selected according to fitness. This lead to the bees quickly group to the local high fitness food source, and destroy the diversity of the bee colony and the algorithm may full into local optimum and gets slow convergence speed.

In order to ensure the diversity of bee colony, we improve the roulette algorithm. We improve the low fitness food source selection probability and reduce the high fitness food source selection probability by quadratic function mapping. The mapping curve is shown in Fig. 2. We first normalize the fitness vector by $\max(Fitness)$. Therefore the mapping curve can be defined as

$$\begin{cases} p_i = -2V_i^2 + 2V_i \; 0 \le V_i \le 1/2 \\ p_i = 2V_i^2 - 2V_i + 1 \; 1/2 \le V_i \le 1 \end{cases} \tag{10}$$

(4) *Scout phase*

If an employed bee cannot improve the fitness of the solution after a given number of loops, it may fall into the local optimum. Then it becomes a scout bee and abandon its original source. The scout bee will generate a new solution randomly.

# 4 Analysis

## *4.1 S-Transform of the Sample Point*

The received UWB signal normalized waveforms in different scenarios are shown in Fig. 3.

As can be seen from the received signals, the differences between the waveforms are not obvious. So it is not suitable for inputting into the classifier directly. Then we use S-transform to process the normalization waveforms. The result of the S-transform can be discretely identified in a different frequency band. In an S-transform, we set the time interval between samples as 1. The distribution is shown in Fig. 4. The x coordinate is sample points, when the scale between samples is 1, it means the time scale is 1. The y coordinate is the multiple of the fundamental frequency, and the z coordinate is the amplitude of S-transform.

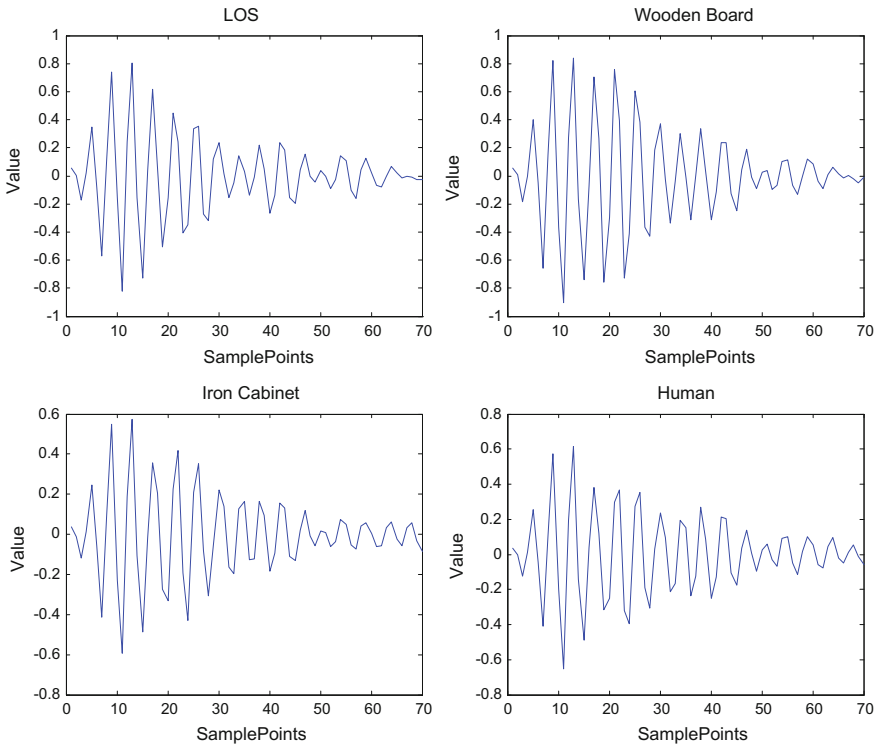



**Fig. 3** UWB received signals

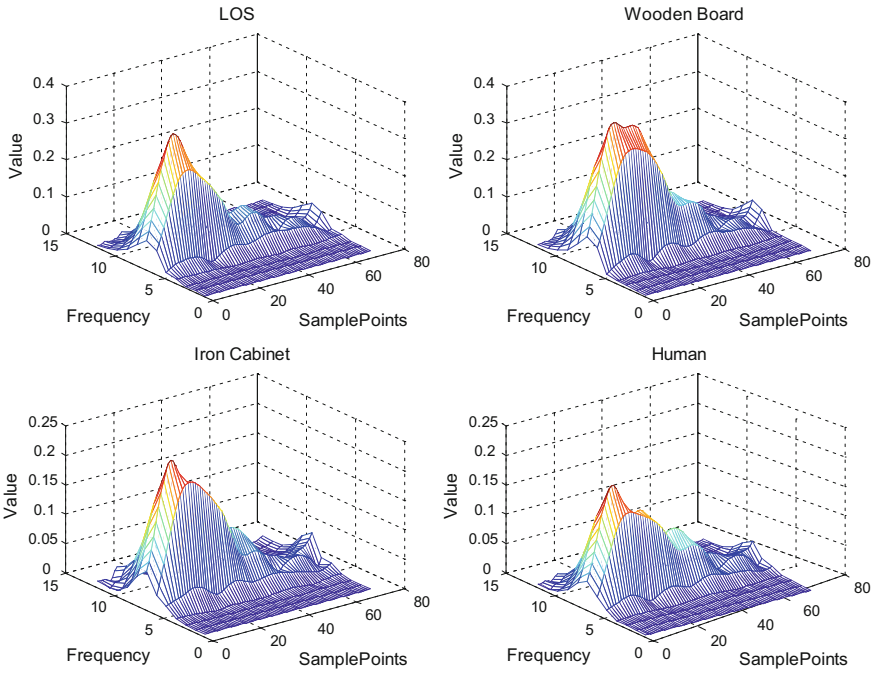



**Fig. 4** S-transform time-frequency distribution

## *4.2 Optimization and Comparison*

After S-transform, we can directly use the shape of the waveforms to represent the different obstacles. First, we change each waveform matrix into a vector. The rows of each S-transform matrix are 12, and the columns are 70. So we can get a vector length of 840. Database of the targets consists of 3200 groups of sample data. Each obstacle database consists of 800 groups of sample data. For each obstacle database, we randomly choose 600 groups as training data, and the rest is test data. Firstly, we use the training data to optimize the SVM.

(1) *Initialize the control parameters of IABC algorithm*

Before optimization, we performed simulation of the SVM when $\gamma$ varies in [0, 40] with 0.02 of each step, $C$ varies in [1, 10] with 1 of each step. The recognition accuracy results are illustrated in Fig. 5.

According to Fig. 5, in order to obtain a fine performance when using SVM, we use IABC algorithm to find the best $C$ and best $\gamma$. The population size $S_N$ is 20. The dimensions of the problem $D$ is 2 ($C$ and $\gamma$). The maximum number of iterations is 20 ($MCN$). And the maximum iterations of food source is 5 ($limit$). The search

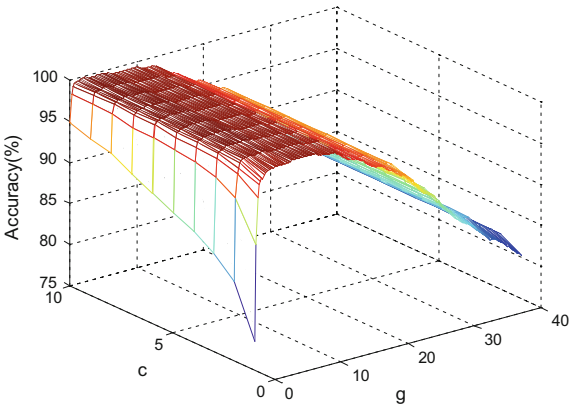



**Fig. 5** Accuracy distribution when g and C varies

**Table 1** Recognition accuracy of different Algorithms (%)

| Target type | Default | Basic ABC | IABC |
|---|---|---|---|
| LOS | 98.0 | 99.5 | 99.5 |
| Wood board | 91.0 | 97.0 | 98.5 |
| Iron cabinet | 87.5 | 96.0 | 98.0 |
| Human | 91.0 | 92.0 | 96.5 |
| Total | 92.125 | 96.125 | 98.125 |

scope of penalty factor $C$ is in [1, 10] and kernel parameter $\gamma$ is in [0.1, 40]. Through the IABC algorithm, we can get the best $C$ at 7.9645 and best $\gamma$ at 5.9874.

(2) *Comparison between algorithms*

Then we use the same experimental data to test the basic ABC algorithm and IABC algorithm. The results are shown in Table 1. LOS means no targets. And the default means no algorithm to optimize the classifier.

As we can see from these results, optimization algorithms can effectively improve the recognition accuracy of SVM. When using IABC algorithm to optimize the parameters of SVM, we can get higher accuracy of the classification. And we compare the convergence speed of the two algorithms. The convergence curves are shown in Fig. 6.

From Fig. 6 it can be concluded that in the late period of optimization, the IABC algorithm has the ability of rapid convergence and jumping out of local optimal solution and global searching. Compared with the basic ABC algorithm, the IABC algorithm has better ability of optimization.

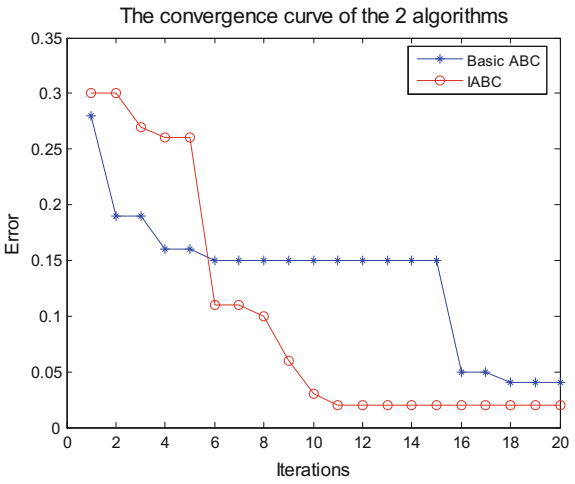



**Fig. 6** The convergence curves of ABC and IABC algorithms

## 5 Conclusions

In this paper, a new method of target identification is proposed. We use S-transform to analyze the UWB received signals. Then the IABC algorithm is introduced to optimize the SVM. The identification results demonstrate that the target detection and identification based on S-transform is effective. The IABC algorithm can give consideration to both convergence speed and global optimum, thus we can obtain higher identification accuracy compared basic ABC algorithm.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (61671075), National Natural Science Foundation of China (61631003), and National Natural Science Foundation of China (61171176).

## References

1. Y. Wang, T. Jiang, A method of target identification with UWB based on genetic algorithm and fuzzy pattern recognition [C], in *Communications Workshops (ICC), 2013 IEEE International Conference on. IEEE*, (2013), pp. 936–940
2. J. He, T. Jiang, Z. Xing, A method of target detection and identification based on RPROP and UWB channel characteristic parameters [C], in *Globecom Workshops (GC Wkshps), 2012 IEEE*, (IEEE, 2012), pp. 1460–1463
3. R. Ying, T. Jiang, Z. Xing, Classification of transmission environment in UWB communication using a support vector machine [C], in *Globecom Workshops (GC Wkshps), 2012 IEEE*, (IEEE, 2012), pp. 1389–1393
4. Q. Tang, T. Jiang, A new method of target identification in UWB communication system based on smooth pseudo Wigner Ville distribution and semi-supervised clustering [C], in *The Proceedings of the Second International Conference on Communications, Signal Processing, and Systems*, (Springer International Publishing, 2014), pp. 469–476

5. B. Boashash, Time frequency analysis [M]. Gulf Professional Publishing, (2003)
6. R.G. Stockwell, L. Mansinha, R.P. Lowe, Localization of the complex spectrum: the S transform [J]. IEEE Trans. Signal Process **44**, 998–1001 (1996)
7. R.G. Stockwell, Why use the S-transform? [J]. Field Inst. Commu. **52**, 279–309 (2007)
8. D. Karaboga, B. Basturk, A powerful and efficient algorithm for numerical function optimization: artificial bee colony (ABC) algorithm [J]. J. Global Optim. **39**(3), 459–471 (2007)
9. D. Karaboga, B. Basturk, A comparative study of artificial bee colony algorithm [J]. Appl. Math. Comput. **214**(1), 108–132 (2009)
10. D. Karaboga, C. Ozturk, A novel clustering approach: artificial bee colony (ABC) algorithm [J]. Appl. Soft Comput. **11**(1), 652–657 (2011)
11. O. Chappele, V. Vapnik, O. Bousquet et al., Choosing multiple parameters for support vector machine [J]. Mach. Learn. **46**(1), 131–160 (2002)

# Dynamic Hand Gesture Recognition Based on Parallel HMM Using Wireless Signals

**Jiabin Xu and Ting Jiang**

**Abstract** Dynamic hand gesture recognition plays an important role in human–computer Interaction. This paper proposes a novel method for dynamic hand gesture recognition using wireless signals. Through the analysis of wireless frame structure, the preamble's signal of 802.11a is collected through Software Defined Radio platform and reserved as the data source. In addition, more than one time-domain feature sequences perform unique shape for different dynamic hand gesture. These sequences are split into single cycle (time-series) and the unavoidable electronic interference is reduced through discrete wavelet transform. At the same time, due to fuzziness of dynamic hand gesture, the amplitude and duration for the same dynamic hand gesture are not exactly same. Therefore, the parallel HMM models which represent for different hand gestures and features are built for recognition. The result shows that the average recognition rate is about 90.5% for dynamic hand gesture recognition.

**Keywords** Dynamic gesture recognition · Software defined radio · Time-series · Parallel HMM models · Wireless signals

## 1 Introduction

According to the survey on about 600 people including students, researchers, layman, professionals, etc., most of them interact for more than 80% time with HCI technologies [1]. Nowadays, the hand gesture, voice, and face recognition are the most common techniques in HCI. But compared with the other two ways, hand gesture recognition is much more convenient and effective. The most common way to recognize hand gesture is through the camera. However, due to the requirement for precise position of hand including the distance and the angle relative to the

J. Xu · T. Jiang (✉)
Key Laboratory of Universal Wireless Communication, Beijing University of Posts and Telecommunications, Beijing, China
e-mail: tjiang@bupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_80

camera, it is not feasible in our real life. Moreover, another common way requires the user to wear some sensor devices [2], which is uncomfortable for the user and needs additional overhead. Recently, great progress has been made in gesture recognition based on wireless signals. Fadel Adib and Dina Katabi [3] used Wi-Fi signal to identify simple gestures made behind the wall. Another gesture recognition system that leverages wireless signals to enable whole-home sensing and recognition of human gestures called Wisee was proposed by University of Washington [4]. In December 2013, MIT developed the WiTrack which can achieve 3D motion tracking using wireless signals despite whether the body is behind the wall [5]. Furthermore, Vital-Radio which can track users' breathing and heart rates had been developed with the help of wireless signals [6].

However, taking into consideration that Wi-Fi is mainly designed for communication, the TOAs from different directions for Wi-Vi [3] are difficult to collect. Moreover, the Wisee used the feature of OFDM symbol to extract the Doppler shift with a big amount of calculation. On the other hand, this paper utilized the characteristic structure of communication signal to extract the feature which represents for different hand gestures. Otherwise, the consecutive sequences represented for different features are separated with discrete Fourier transform and the unavoidable signal interference for Wi-Fi is reduced with preprocessing, then a series of parallel HMM models which represent for different dynamic hand gestures are used for recognition. The result shows that this method can achieve high efficiency and accuracy on dynamic hand gesture recognition.

This paper is organized as follows: in Sect. 2, we give a brief introduction of the experimental platform and scene. Section 3 discusses how we process the signal and build the model. Section 4 presents the result of the model. In Sect. 5, we deliver a conclusion and talk about the future work.

## 2 The Experimental Platform and Scene

### *2.1 Characteristic of Communication Signal*

In 802.11a, the PLCP preamble is described in Fig. 1.

These training symbols always remain same in each frame for the aim of timing estimation, channel estimation, etc. So when dynamic hand gestures are performed between transmitter and receiver, the impact of propagation channel is reflected on

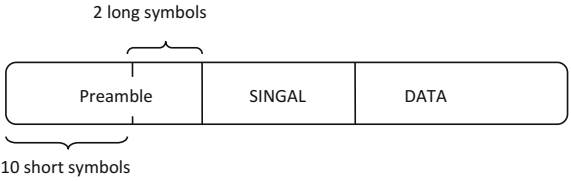



**Fig. 1** The structure of OFDM frame

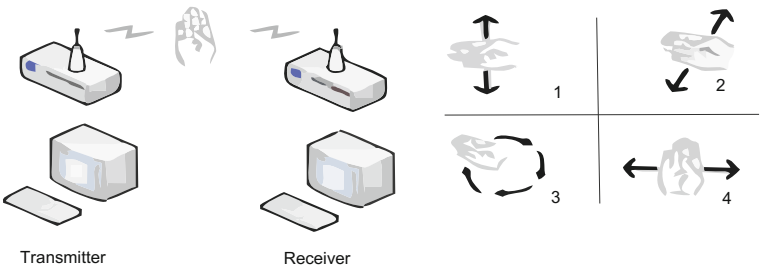



**Fig. 2** Scene overview and four dynamic hand gestures

the change of the training symbols. Therefore, the properties of the long training symbols can be used as a basis for the recognition.

However, due to the consumption on energy detection, only complete sampling points of the two long training symbols are captured as our data. All the work is implemented on SORA, a commercial software reconfigurable radio platform on PC. Complete wireless communication demos including 802.11a/b/g have been provided in the form of software for the further work. Furthermore, UMXDot11 is one of the demos which can achieve the function of sending and receiving frames based on 802.11a protocol. Finally, the two long training samples are extracted from the processing of frequency estimation.

### *2.2 Experimental Scene*

Two platforms of SORA are required in this experiment, one acts as transmitter while the other one acts as receiver. The transmitting and receiving antennas of SORA are separated with a distance of one meter within the same horizontal plane, which is limited by the antenna transmit power and other wireless signal. Moreover, the hand should be placed between the antennas as shown in Fig. 2.

Four different dynamic hand gestures are selected in our work including gesture-1 (moving up and down), gesture-2 (moving back and forth), gesture-3 (moving clockwise), and gesture-4 (moving left and right). What's more, in order to find correct information of gestures, all of the dynamic gestures should perform periodically.

## 3 Feature Extraction and Model Building

In previous work, the energy of the two long training symbols are extracted as feature and the SVM machine improved by dynamic time warping algorithm was built to classify different gestures [7]. In our work, more useful information is found

through the time-domain feature of two long training symbols. On the other hand, the work on extracting single-cycle (time-series) is simplified with discrete Fourier transform. Taking into account that recognition through dynamic time warping takes relatively more time, the hidden Markov model is applied on the recognition of dynamic hand gestures.

## 3.1 The Time-Domain Feature

In 802.11a protocol, the long preamble consists of two long training symbols. At the same time, each of the long training symbols occupies 3.2 us. Taking into account the sampling rate 20 M/s, there are 128 complex sampling points which are captured with UMXDot11 demo for each frame. Thus the sampling points for each frame can be described as follows:

$$r_i = I_i + Q_i, i = \{1, 2, 3, \ldots, 128\} \tag{1}$$

Once the sampling points are written into files, different features in time-domain can be calculated such as the received energy, the additional delay, the RMS delay spread, the max amplitude, the average amplitude, the standard deviation, the index that exceeds 85% of the received energy, the number of points that decline 10 dB of the max amplitude.

Subsequently the time-domain features for 4096 successive frames are drawn into sequences. Moreover, there are three kinds of time-domain features varied with

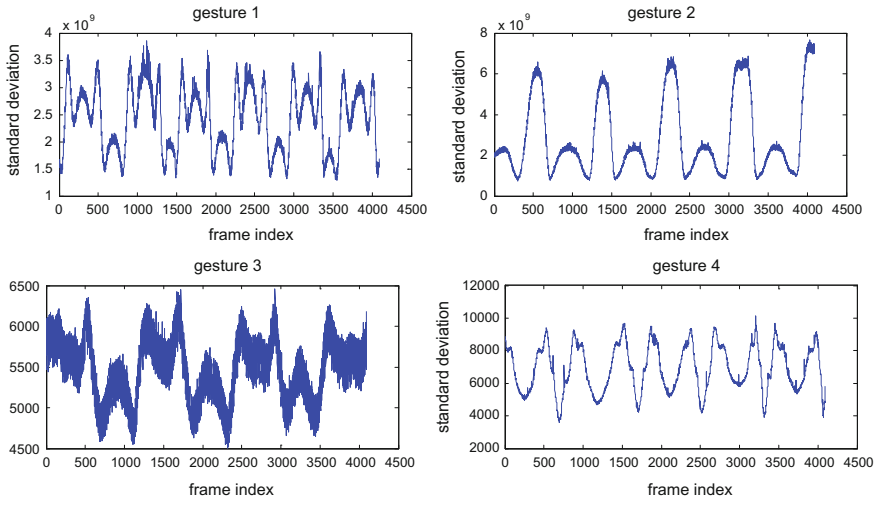



**Fig. 3** The sequence of the standard deviation for each gesture

the dynamic hand gestures. As we can see in Fig. 3, the sequences of the standard deviation have unique shape for different dynamic hand gestures.

The features behind these three kinds of sequences can be calculated by the following formulas:

(1) The received energy

$$E_{\mathrm{n}} = \sum_{i=1}^{128} |r_i|^2 \tag{2}$$

(2) The standard deviation

$$S_n = \frac{\sum_{i=1}^{128} \left( |r_i| - \frac{\sum_{i=1}^{128} |r_i|}{128} \right)^2}{128} \tag{3}$$

(3) The index that exceeds 85% of the received energy

$$N_n = \min\left( \sum_{i=1}^{128} |r_i|^2 > 0.85 E_n \right) \tag{4}$$

Actually apart from the time-domain feature, frequency offset and channel estimation can also be calculated through the long preamble [8] for recognition of dynamic hand gestures. Moreover, the frequency-domain feature may supply enough information for dynamic hand gestures. However, these are not in the scope of this paper.

## 3.2 *Extract Single-Cycle*

As Fig. 2 shows, different gestures have different cycles in the sequence of the feature. For example, gesture-1 always has six cycles while gesture-2 always has four cycles. For the purpose of recognition, single-cycle (time-series) should be extracted from these periodic sequences. Here, discrete Fourier transform (DFT) is applied to analyze the main period of these sequences.

Suppose that the feature sequence for 4096 successive frames is given by:

$$x(n), n = 0, 1, \ldots 4095 \tag{5}$$

Then 4096 points discrete Fourier transform is adopted on the feature sequence as the formula shows:

$$X(k) = \sum_{n=0}^{4095} x(k) W^{nk}, k = 0, 1, \ldots 4095 \tag{6}$$

where $W = exp^{-j\frac{2\pi}{4096}}$.

As we all know, the output $|X(k)|, k = 0, 1, \ldots 4095$ represents a series of frequencies whose corresponding periods are an integer multiple of the sampling interval. Moreover, the main period can be inferred by finding the maximum amount in these outputs.

In experiment, one dynamic hand gesture usually takes about 1.5–4.5 s, while the UMXDot11 demo usually receives about 300 frames in 1 s. As a result, the period should be around 450–1350. Actually, only seven $|X(k)|$ corresponding to the following periods needed to be calculated:

$$periods = \{1365, 1024, 819, 682, 585, 512, 455\} \tag{7}$$

Note that the period resolution is not enough in practice. So when finding the biggest period power among these seven kinds of periods, the periods around the biggest period power should be taken into consideration. For example, if period 585 has the biggest power, we have to take the periods 512–682 into consideration.

### *3.3 Preprocess of the Time-Series*

The length of the time-series for each feature is about 450–1350 points. However, the membership degree of the time-series in HMM decreases with increasing length. In other word, the difference between the two time-series is difficult to distinguish when the time-series are too long. So it is not suitable to build HMM models on these raw time-series, because both the length and the noise of time-series reduce the identification accuracy.

In this paper, finally the discrete wavelet transform (DWT) is adopted to reduce the noise and shorten the length with the shape keeping same. According to experimental results, the approximate coefficients of order 3 basically keep the shape when the decomposition level is set to 3 and db1 is chosen as the wavelet.

After the work above, the time-series have a length of about 150 points. Moreover the noise of these time-series has greatly reduced. Finally, the time-series

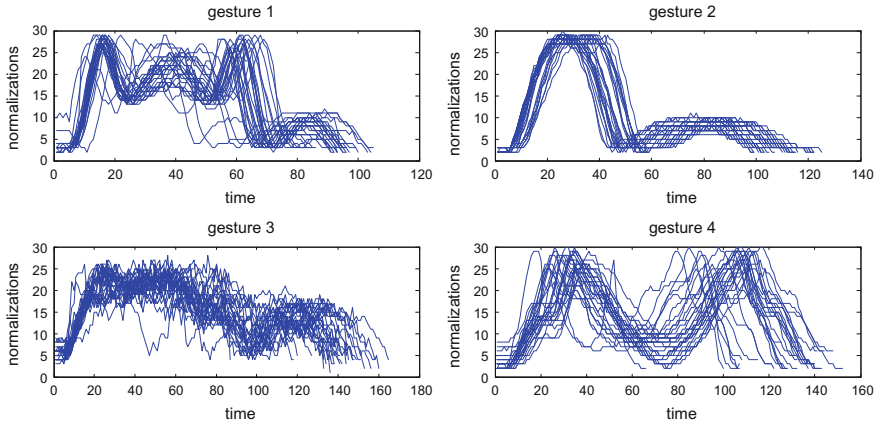



**Fig. 4** The time-series for each gesture

are normalized and rounded to the integer as the data source for hidden Markov model. In Fig. 4, 30 time-series for each gesture are extracted from the sequences with the help of discrete Fourier transform (DFT) and discrete wavelet transform (DWT).

### *3.4 Model Performance Analysis*

Note that each gesture takes different time to finish, which means the length of time-series are different. So the similarity of time-series with different length needs other model instead of classifier. As we all know, dynamic time warping (DTW) and hidden Markov model (HMM) are the two most common algorithms for pattern matching. The research trend transited from DTW to HMM in approximately 1988–1990 [9]. So far the hidden Markov model has been widely used in speech and handwriting recognition [10]. In fact, each dynamic hand gesture can be divided into three time-series which represent for the three kinds of time-domain features, so there are 12 HMM models are trained for recognition.

## 4 Experiments Results

At first, only one time-domain feature of the long preamble for each dynamic gesture is extracted for training and recognition. As shown in Fig. 5, the received energy has a higher accuracy for gesture-2 and gesture-4 while has a lower accuracy for gesture-1 and gesture-3. On the other hand, the index that exceeds 85% of the received energy has a higher accuracy for gesture-1 and gesture-3 while has a lower

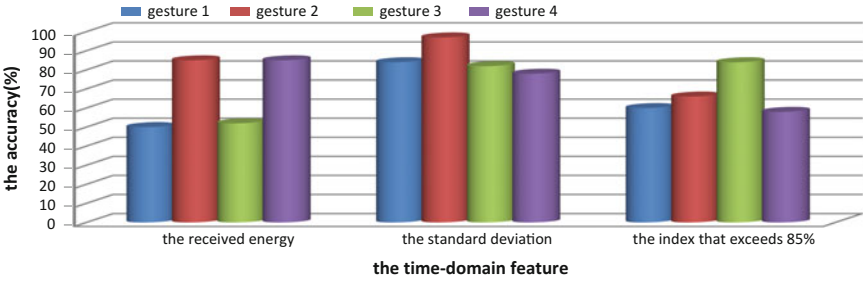



**Fig. 5** Training and recognition with only one feature

**Table 1** The distribution of recognition rates

| Gesture index | I | II | III | IV |
|---|---|---|---|---|
| I | 87 | 1 | 2 | 10 |
| II | | 99 | | 1 |
| III | 3 | | 90 | 7 |
| IV | 1 | 13 | | 86 |

accuracy for gesture-2 and gesture-4. So we can conclude that different features perform various accuracy for the same dynamic hand gesture, these three features can provide more information than one feature.

In this paper, three features are extracted for training and recognition. As seen in Fig. 5, the standard deviation always keeps high accuracy for these four kinds of gestures, so if these three kinds of features get three unique results with the help of parallel HMM, the result of the standard deviation is chosen as the result. Finally the result is shown in Table 1.

From the experiment results, we can conclude that the average recognition rate is about 90.5%. Moreover, the recognition with parallel HMM costs less than one second, which is much better than the model with dynamic time warping.

**Acknowledgements** This work was supported by National Natural Science Foundation of China (61671075), National Natural Science Foundation of China (61631003), and National Natural Science Foundation of China (61171176).

## References

1. A.K. Dubey, K. Gulabani, R. Rathi, Empirical study to appraise consciousness of HCI technologies, in *2014 International Conference on Signal Propagation and Computer Technology (ICSPCT)* (IEEE, 2014), pp. 447–450
2. R. Xu, S. Zhou, W.J. Li, MEMS accelerometer based nonspecific-user hand gesture recognition. Sens. J. IEEE **12**(5), 1166–1173 (2012)
3. F. Adib, D. Katabi, *See Through Walls with WiFi*! (ACM, 2013)

4. Q. Pu, S. Gupta, S. Gollakota et al., Whole-home gesture recognition using wireless signals, in *Proceedings of the 19th annual international conference on Mobile computing & networking* (ACM, 2013), pp. 27–38
5. F. Adib, Z. Kabelac, D. Katabi et al., 3D tracking via body radio reflections, in *Usenix NSDI*, vol. 14 (2014)
6. F. Adib, H. Mao, Z. Kabelac et al., Smart homes that monitor breathing and heart rate, in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (ACM, 2015), pp. 837–846
7. T.J. Zhouge, A new method of dynamic gesture recognition using Wi-Fi signals based on DWT and SVM improved by DTW, in *2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, vol. 12 (2015)
8. T.J. Huangwen, Applications of software radio for hand gesture recognition by using long training symbols, in *2015 9th International Conference on Signal Processing and Communication Systems (ICSPCS)*, vol. 12, Cairns (Barrier Reef), Australia (2015)
9. C. Fang, From dynamic time warping (DTW) to hidden markov model (HMM). University of Cincinnati (2009)
10. T. Takiguchi, S. Nakamura, K. Shikano, HMM-separation-based speech recognition for a distant moving speaker. IEEE Trans. Speech Audio Process. **9**(2), 127–140 (2001)

# Automatic Target Recognition for SAR Images Based on Fuzzy Logic Systems

**Xuhong Feng and Qilian Liang**

**Abstract** Automatic target recognition (ATR) generally refers to the autonomous or aided target detection and recognition by computer processing of data from a variety of sensors such as forward looking infrared (FLIR), synthetic aperture radar (SAR), inverse synthetic aperture radar (ISAR), laser radar (LADAR), millimeter wave (MMW) radar, multispectral/hyperspectral sensors, low-light television (LLTV), video, etc. SAR is a coherent active imaging method that utilizes the motion of a radar mounted on a vehicle such as an aircraft (airborne SAR) or a satellite (spaceborne SAR) to synthesize the effect of a large aperture radar. Motivated by the unique character of fuzzy logic system, simultaneously handling numerical data and linguistic knowledge, and the promising knowledge-based approach, we propose an FLS-based approach to SAR ATR.

## 1 Introduction to ATR

Automatic target recognition (ATR) generally refers to the autonomous or aided target detection and recognition by computer processing of data from a variety of sensors such as forward looking infrared (FLIR), synthetic aperture radar (SAR), inverse synthetic aperture radar (ISAR), laser radar (LADAR), millimeter wave (MMW) radar, multispectral/hyperspectral sensors, low-light television (LLTV), video, etc. [1]. It is an extremely important capability for targeting and surveillance missions of defense weapon systems operating from a variety of platforms. ATR is aimed at reducing the workload for human operators (e.g., image analysts, pilots,

X. Feng (✉)
College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin 300387, China
e-mail: fengxuhong75@sina.com

Q. Liang
Department of Electrical Engineering, University of Texas at Arlington, Arlington TX 76019-0016, USA
e-mail: liang@uta.edu



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_81

tankers) who are tasked with a large scope of activities ranging from assessing the battlefield/battlespace situation over large areas and volumes to targeting individual targets on land, sea, or air [1]. It is generally agreed that significant progress has been made in the development of sensors and processing hardware systems. Great progress is being made in the development of algorithms using a variety of techniques.

ATR data can be in the form of nonimaging one-dimensional sensor returns, such as ultra-high range-resolution radar returns (UHRR) for air-to-air automatic target recognition and vibration signatures from a laser radar for recognition of ground targets. The ATR data can be two-dimensional images. The most common ATR images are infrared, but current systems must also deal with synthetic aperture radar (SAR) images. Finally, the data can be three dimensional, such as sequences of multiple exposures taken over time from a nonstationary world. Targets move, as do sensors, and that movement can be exploited by the ATR. Hyperspectral data, which are views of the same piece of world looking at different spectral bands, is another example of multiple image data: the third dimension is now wavelength and not time [2].

ATR system design usually consists of four stages. The first stage is to select the sensor or sensors to produce the target measurements. The next stage is the preprocessing of the data and the location of the data (segmentation). The human retina is a ruthless preprocessor. Physiologically motivated preprocessing and segmentation is demonstrated along with supervised and unsupervised artificial neural segmentation techniques. The third design step is feature extraction and selection: the extraction of a set of numbers which characterize regions of the data. The last step is the processing of the features for decision-making (classification).

According to [3], the details of ATR discrimination features used by Linclon Laboratory at MIT include textural, size, contrast, and polarimetric features. The Lincoln Laboratory ATR system has three main stages: detection (or prescreening), discrimination, and classification. In the detection stage, a two-parameter CFAR (constant false alarm rate) detector is used as a prescreener to select candidate targets in a SAR image on the basis of local brightness. In the discrimination stage, a target-sized 2-D matched filter accurately locates the position of candidate targets and determines their orientation. Then, discrimination features (including textural, size, contrast, and polarimetric features) are calculated and used to reject cultural and natural-clutter false alarm. In the classification stage, a 2-D pattern-matching algorithm provides additional false alarm rejection and then classifies the remaining detections by target type (i.e., tank, howitzer).

In recent advances, a knowledge-based ubiquitous and persistent sensor network (KUPS) was proposed for threat assessment in [4], in which “sensor” is a broad characterization. It refers to diverse data or information from ubiquitous and persistent sensor sources such as organic sensors and human intelligence sensors. In [5], ATR was studied using waveform design and diversity. In [6], ATR was studied in a networked scenario. In [7], opportunistic sensing, to select a small subset of sensors from sensor networks, was studied from information theory point of view. Radar sensor networks were evaluated as an application example in

opportunistic sensing, and radar channel modeling was studied in [8]. Further, situation understanding was investigated based on ATR results.

The rest of this paper is organized as follows. In Sect. 2, we give an overview of synthetic aperture radar. In Sect. 3, we present the motivation of fuzzy logic system for SAR ATR. The SAR ATR scheme was proposed in Sect. 4 with fuzzy rule design and computing. Section 5 concludes this paper.

## 2 Synthetic Aperture Radar

In recent years, there has been a growing interest in synthetic aperture radar (SAR) imaging for applications ranging from remote sensing to surface surveillance and ATR [9].

The first proposal for rudimentary radar was made soon after Maxwell's celebrated theory of electromagnetism having been confirmed by Herz in 1887. The first demonstration of a radar was made by Hulsmeyer in Germany in 1904 [3].

For true land-mapping purposes, however, the limited size of the antenna presented a formidable constraint. The advent of the microwave power amplifier suggested to a number of workers the possibility of coherently recording and analyzing the returns as the airborne radar was carried through a relatively long distance. An extremely analogous principle was established by the radio astronomers at the same time and it was they who coined the term "synthetic aperture". The first "synthetic aperture radar" (SAR) using this principle date from about 1954 and were due to a number of workers in the USA, in particular Wiley and Cutrona [10].

Synthetic Aperture Radar (SAR) is a coherent active imaging method that utilizes the motion of a radar mounted on a vehicle such as an aircraft (airborne SAR) or a satellite (spaceborne SAR) to synthesize the effect of a large aperture radar. The motion of the vehicle results in the collection of echoed signals from the target when illuminated with the radar's radiation pattern from various views. The SAR imaging (inverse) problem is to integrate these echoed signals for high-resolution imaging [11]. In a similar fashion, Inverse SAR which is a relatively new variant of SAR, rather than moving the radar to image the target, the inherent angular or translational motion of the target is exploited.

Today, the SAR has already become a viable technology in remote sensing problem of radar, producing high-resolution maps and images in real time. Recognizing targets in SAR images is a difficult problem for conventional computer vision systems. First, all SAR features are nonattached; they are not tightly related with surface makers nor explicit object geometry such as edges. With slight movements of the observer, features suddenly appear, disappear, and abruptly change their shapes. Second, in SAR image recognition, objects are often intentionally hidden from an observer. For example, enemy tanks are often hidden under trees. Thus, for SAR recognition scenarios, we must handle unstable nonattached features, occlusion, and camouflage [2].

An ATR system in the presence of high-resolution polarimetric SAR has three basic stages: the detection of region of interest, the matching of the detection with the stored target templates, and the classification. One of known techniques for the SAR target detection and estimation is the Constant False Alarm Rate (CFAR) detector. It identifies potential targets in the image on the basis of signal amplitude. An improved target detection performance is to use fully polarimetric (HH, HV, and VV) imagery processed by the polarimetric whitening filter (PWF) [12, 3]. The detected objects are then matched against stored templates for classification. To construct target templates, one approach is to decompose each target class into a number of angel-indexed subclasses in order to handle the unknown orientation. With this approach, in order to achieve a reasonable performance, a large number of performance reference images are required for the classifier. Another approach is to use a neural network based technique to automatically "learn" the representative templates. The goal is to reduce the number of reference images while maintaining a reasonably high correct classification rate [13]. SAR imagery has been successfully used for ATR [2, 12–14, 15].

## 3 Motivation of Proposing FLS-Based SAR ATR

Historically, target recognition in SAR images has been using three different approaches: statistical pattern recognition (a lot of works have been done on this), artificial neural network (e.g., [13]), and model-based vision (e.g., [2]). According to [1], researchers now are developing knowledge-based system.

Perlovsky et al. [14] describes a model-based neural network approach and applies it to several target detection/segmentation problems. The basic neural network architecture consists of an association subsystem that computes weights that associate data with models. The general idea is to estimate the clutter model parameters for each image independently, and to classify those pixels with smallest likelihood of belonging to the clutter model (i.e., the outliers) as targets. The approach combines *a prior* knowledge (physical laws of electromagnetic scattering) with adaptations to the actual environment. The results are presented using SAR images.

In essence, a mode-based system analyzes each image in detail and identifies each part of a signature's contribution toward recognition, while pattern recognition and artificial neural network based recognition system handle a target signature as a whole. The capability of part analysis in the model-based vision approach provides the potential for the robustness with respect to partial occlusion of the target and cluttered backgrounds. For these reasons, Katsushi et al. regard the model-based approach is the most promising [16]. But we can devise a knowledge-based system combing the advantages of model-based approach, and using the linguistic knowledge as well as numerical data. This system is fuzzy logic system (FLS).

A fuzzy logic system (FLS) is unique in that it is able to simultaneously handle numerical data and linguistic knowledge. It is a nonlinear mapping of an input data

(feature) vector into a scale output, i.e., it maps numbers into numbers [17]. FLS has been successfully used into pattern recognition, image segmentation [16, 18].

To the authors' knowledge, no SAR ATR approach has been proposed based on Fuzzy Logic Systems (FLS).

## 4 FLS-Based Approach to SAR ATR

Sometimes it is really difficult to judge if a target is something based on the whole signature of a target, so the model-based approach recognizes the target by analyzing each part of a signature, but sometimes it is also hard to say this part match that of the stored model. In order to achieve robust recognition under severe occlusion, camouflage, and unstable SAR features, we introduce FLS into the knowledge-based approach.

We can use FLS to evaluate the fitness of each part in a signature corresponding to that of the stored models, then based on the contribution of each part, we can make decision if the object is a tank or truck, or something else by defuzzifier.

We use FLS to measure the similarity between the parts of SAR image and model. In FLS design, we use the FBF model designing method which was described in [17].

The relations can be represented by linguistic knowledge using IF-THEN sentences:

IF part 1 is A1, part 2 is B1, part 3 is C1, …, THEN the target is weapon 1.
IF part 1 is A2, part 2 is B2, part 3 is C2, …, THEN the target is weapon 2.
…
IF part 1 is A$i$, part 2 is B$i$, part 3 is C$i$, …, THEN the target is weapon $i$.

If represented by mathematical formula and using Gaussian membership function, product inference, height defuzzifier, the defuzzifier function can be written as

$$f(x) = \frac{\sum_{l=1}^{M} \bar{y}^l \prod_{k=1}^{n} \exp\left(-\frac{(x_k - m_{F_k^l})^2}{2\sigma_{F_k^l}^2}\right)}{\sum_{l=1}^{M} \prod_{k=1}^{n} \exp\left(-\frac{(x_k - m_{F_k^l})^2}{2\sigma_{F_k^l}^2}\right)} = \sum_{l=1}^{M} \bar{y}^l p^l(x),$$

where $p^l$ denotes the $l$ th basis function, and here $x_k$ denotes the numerical data of SAR image of the target, and $m_{F_k^l}$ denotes the numerical data of stored model value. We can adjust the difference by changing $\sigma_{F_k^l}$.

Detecting a group of features is rarely possible in the SAR domain, where most features are quite unstable. We can extract two kinds of primitive features from SAR image: points and line segments. All feature points are detected by applying an interest operator. All line segments are detected by a line detector based on edge

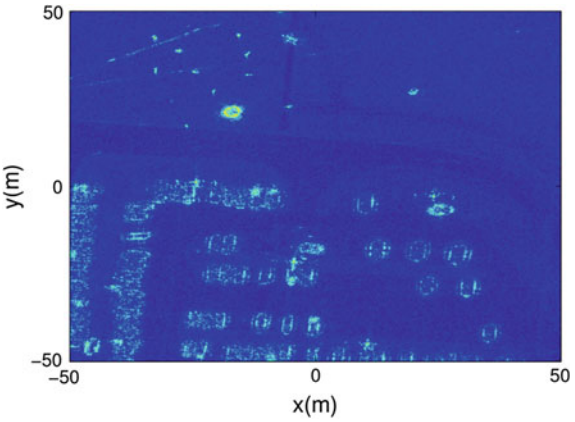



**Fig. 1** A SAR image in car parking slot

detecting technique. The most easily detectable features from SAR images are isolated peaks [2].

In Fig. 1, as an example, we show a SAR image with a number of cars in parking slots. The "Gotcha Volumetric SAR Data Set, Version 1.0," was from US Air Force, which consists of SAR phase history data collected at X-band with a 640 MHz bandwidth with full azimuth coverage at 8 different elevation angles with full polarization. The imaging scene consists of numerous civilian vehicles and calibration targets. The optical image is not possible to be obtained at night or a raining day. However, the SAR image is able to be constructed at any time and any weather condition. A FLS-based classifier could be designed to identify all the targets based on each target feature, such as boundaries via edge detection, the sides of each detected Region of Interests (RoI), and also the raw data size information. An FLS is rule-based and it can combine the linguistic (e.g., shapes) and numerical (e.g., sizes, raw data values, etc.) knowledge.

## 5 Conclusions

Automatic target recognition generally refers to the autonomous or aided target detection and recognition by computer processing of data from a variety of sensors. SAR is a coherent active imaging method that utilizes the motion of a radar mounted on a vehicle such as an aircraft (airborne SAR) or a satellite (spaceborne SAR) to synthesize the effect of a large aperture radar. Motivated by the unique character of fuzzy logic system, simultaneously handling numerical data and linguistic knowledge, and the promising knowledge-based approach, we propose an FLS-based approach to SAR ATR. Our fuzzy rules combine the linguistic languages which could classify SAR images successfully.

## References


1. B. Bhanu, D. Dudgeon, E. Zelnio, A. Rosenfeld, D. Casasent, I. Reed, Guest editorial: introduction to the special issue on automatic target detection and recognition. IEEE Trans. Image Process. **6**(1), 1–4 (1997)
2. K. Ikeuchi et al., Model-based SAR ATR system. Proc. SPIE **2757**, 376–387 (1996)
3. R. Voles, Resolving resolutions: imaging and mapping by modern radar, in *IEE Proceedings-F*, vol. 140, no. 1 (1993), pp. 1–11, Feb 1993
4. Q. Liang, X. Cheng, KUPS: knowledge-based ubiquitous and persistent sensor networks for threat assessment. IEEE Trans. Aerosp. Electron. Syst. **44**(3) (2008)
5. Q. Liang, Automatic target recognition using waveform diversity in radar sensor networks. Pattern Recogn. Lett. (Elsevier) **29**(2), 377–381 (2008)
6. Q. Liang, X. Cheng, S. Samn, NEW: network-enabled electronic warfare for target recognition. IEEE Trans. Aerosp. Electron. Syst. **46**(2), 558–568 (2010)
7. Q. Liang, X. Cheng, S. Huang, D. Chen, Opportunistic sensing in wireless sensor networks: theory and applications. IEEE Trans. Comput. **63**(8), 2002–2010 (2014)
8. Q. Liang, Radar sensor wireless channel modeling in foliage environment: UWB versus narrowband. IEEE Sens. J. **11**(6), 1448–1457 (2011)
9. C. Fosgate, Multiscale segmentation and anomaly enhancement of SAR imagery. IEEE Trans. Image Process. **6**(1), 7–21 (1997)
10. C. Wiley, Synthetic aperture radars. IEEE Trans. AES **21**, 440–443 (1985)
11. M. Soumekh, *Fourier Array Imaging* (Prentice Hall, Englewood Cliffs, NJ, 1994)
12. L. Novak, S. Halversen, G. Owirka, M. Hiett, Effects of poarization and resolution on SAR ATR. IEEE Trans. Aerosp. Electron. Syst. **33**(1), 102–115 (1997)
13. K.-C. Chang, Y.-C. Lu, High resolution polarimetric SAR target classification with neural network, in *Proceedings of IEEE Conference on Fuzzy System* (1995), pp. 1681–1688
14. L. Perlovsky et al., Model-based neural network for target detection in SAR images. IEEE Trans. Image Process. **6**(1), 203–216 (1997)
15. A. Mahalanobis et al., Multi-class SAR ATR using shift-invariant correlation filters. Pattern Recogn. **27**(4), 619–626 (1994)
16. A. Moghaddamzadeh et al., A fuzzy region growing approach for segmentation of color images. Pattern Recogn. **30**(6), 867–881 (1997)
17. J.M. Mendel, Fuzzy logic systems for engineering: a tutorial. Proc. IEEE **83**(3), 345–377 (1995)
18. S. Chen et al., Neural-fuzzy classification for segmentation of remotely sensed images. IEEE Trans. Signal Process. **11**, 2639–2654 (1997)
19. S. Rogers et al., Automatic target recognition using neural networks. Proc. SPIE **2492**, 346–360 (1995)
20. L. Novak et al., Radar target identification using spatial matched filters. Pattern Recogn. **27**(4), 607–617 (1994)
21. M. Stevens, J. Beveridge, Precise matching of 3-D target models to multisensor data. IEEE Trans. Image Process. **6**(1), 126–142 (1997)
22. C. Olson, D. Huttenlocher, Automatic target recognition by matching oriented edge pixels. IEEE Trans. Image Process. **6**(1), 103–113 (1997)
23. D. Ballard, Generalizing the Hough transform to detect arbitrary patterns. Pattern Recogn. **13**(2), 111–122 (1981)
24. W. Grimson, *Object Recognition by Computer* (MIT Press 1990)
25. Q. Liang, Situation understanding based on heterogeneous sensor networks and human-inspired favor weak fuzzy logic system. IEEE Syst. J. **5**(2), 156–163 (2011)

# A New Pruning Method to Train Deep Neural Networks

**Haonan Guo, Xudie Ren and Shenghong Li**

**Abstract** Deep neural networks are very powerful models for machine learning tasks. However, suffering from overfitting and gradient vanishing problems, they are difficult to train. We proposed a method of gradually pruning the weakly connected weights to train deep neural networks and an effective strategy to identify the weak connections. Our method can improve the conventional stochastic gradient descent and can get even better performance than the widely used dropout method for deeper models.

**Keywords** Deep neural networks · Overfitting · Stochastic gradient descent

## 1 Introduction

Deep learning is a subfield of machine learning, which has dramatically improved the state of the art in image classification [1], object detection [2], speech recognition [3], and many other domains in the recent decades. Deep learning models are composed of simple but nonlinear modules (neural networks in most cases) layer by layer, each of which can learn a transformation that maps the representation in the last layer to a slightly more abstract level. So these deep models have great power to discover the representations needed for detection or classification tasks automatically from raw data [4].

The architectures of deep neural networks differ widely across applications, among which convolutional neural networks and LSTM recurrent networks are the most widely used. Basically, learning relies on the minimization or maximization of a scalar objective function, and the optimization is typically performed by stochastic

H. Guo (✉) · X. Ren · S. Li (✉)
Shanghai Jiao Tong University, Shanghai 200240, China
e-mail: haonan2012@sjtu.edu.cn

X. Ren
e-mail: renxudie@sjtu.edu.cn

S. Li
e-mail: shli@sjtu.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_82

gradient descent, in which the gradient is evaluated using the backpropagation procedure [5]. Whereas, with a non-convex loss function and millions of parameters, it was believed to be too difficult to train deep neural networks for a long time [6]. Untile 2006, Hinton [7] introduced the idea of greedy layer-wise pretraining to train deep belief networks, and since then, ReLUs (Rectified Linear Units) [8], dropout [9], new initialization strategies [10] and many other tricks have been proposed and improved the efficiency and reliability of the training of deep models.

However, there are still some challenges on the training of deep models. One is the problem of overfitting: since there are so many layers and millions of parameters in the deep model, once the complexity of the model exceed that needed for fitting the training data distribution, it will do much worse on the test data than on the training data. And there are still no efficient strategies of choosing a proper depth and width of the neural network on account of particular data distribution.

The other is the problem of vanishing gradients in backward propagation. As the gradient information is backpropagated, repeated multiplication with small weights renders the gradient information ineffectively small in earlier layers. Several approaches exist to reduce this effect in practice, for example, through careful initialization [11], hidden layer supervision [12], but there are still no ideal performance.

In this work, we proposed a new method to train sparsely connected deep models from initially fully connected architectures through pruning the weakly connected weights. The advantages of our method is that through pruning the small weights we can solve the vanishing gradients problem to some extend and decrease the amount of parameters so as to reduce the threat of overfitting. Besides, our method can spend little training time than traditional methods benefit from the gradually decreasing of the amount of parameters. Our method is motivated by the work of [13], which introduced the weight prone method to compress the size consumes of deep neural networks in order to make it suitable for embedded mobile applications. We improved this method and use it to optimize the training process and defeat overfitting. We will introduce some related works in Sect. 2 and propose our method in Sect. 3. In Sect. 4, we show some experiments on MNIST dataset to prove the effectiveness of our method.

## 2 Background

Deep learning models are composed of building blocks layer by layer, which are generally neural networks. One layer of neural network is formed by lots of computation modules, which we called neural. Each neuron consists in taking a weighted sum of input features, followed by an element-wise nonlinearity:

$$y = f(\mathbf{W} \cdot \mathbf{x} + b)\,, \tag{1}$$

where $\mathbf{W}$ is a k-dimensional weight vector, b is a scalar bias term, $\mathbf{x}$ is a k-dimensional vector of input features, $f(\cdot)$ denotes an element-wise nonlinearity

such as the rectifier linear $max(\cdot, 0)$, or hyperbolic tangent (tanh), and y denotes the scalar output of the neuron. So the entire model can be formalized as a function of input features parameterized by weights and biases, i.e., $P(Y|\mathbf{x}, \theta)$, where $\theta$ denotes the assemble of the parameters **W** and b, and Y denotes the output of the model.

The output should be associate with a loss function, such as a cross-entropy loss for the softmax classification layer. Learning relies on finding the proper parameters $\theta$ in the loss surface to minimize the loss function. And the optimization is typically performed by stochastic gradient descent, in which the gradient is evaluated using the backpropagation procedure.

## *2.1 Dropout*

Many methods have been developed for reducing overfitting. These include stopping the training as soon as performance on a validation set starts to get worse, introducing weight penalties of various kinds such as L1 and L2 regularization and soft weight sharing [14]. Dropout gets the most dramatic performance so far [9].

The key idea is to randomly drop units (along with their connections) from the neural network during training. On each presentation of each training case, each hidden unit is randomly omitted from the network with a probability of 0.5, so a thinned networks is trained each time. At test time, we use the mean network that contains all of the hidden units but with their outgoing weights halved to compensate for the fact that twice as many of them are active. This significantly reduces overfitting and gives much improvements over other regularization methods.

Our method has some similarities with dropout on the way of selecting thinned networks during training. Whereas we select a deterministic and in some sense optimal network while the dropout method selects a random network for training each time. And we prone the weak connections of the units while the dropout method omits a unit with all its connections. The experiment shows that our method can even get better performance than dropout for deeper models.

## *2.2 Advantages of Sparsity*

Sparsity has been a key element of deep learning models such as autoencoders [15] and Deep Belief Networks [16]. Sparse representations are generally more robust to small input changes and more likely to be linearly separable, or more easily separable with less nonlinear machinery, simply because the information is represented in a high-dimensional space.

To realize the property of sparsity, one can adopt the method of penalizing the L1 or L2 norm of the whole weight vector. Our method can prune weight through forcing the weak connections to zero, so it is a more strong sparsity constraint than L1 and L2 normalization. We can not only obtain a sparse connection but also sparse representations.

## 3 Methods

We proposed the method of gradually pruning the weakly connected weights to train deep neural networks. The pruning process is iterative and along with the training process as shown in Table 1. Unlike conventional stochastic gradient descent algorithm, after the parameter updates with each mini-batch of training data, we find out the low-weight connections with weights below a threshold. Once a connection is continuously marked as low-weight for a critical times, we identify the connection as weakly connected and prune it in the following training process, i.e., set the value of the weight to zero in the feedforward propagation and set the gradient to zero in the backward propagation. The prune operation should be proceed all the way with the training process until the sparseness of the network conforms to our target. Then we train the network without pruning until we get the best performance.

### *3.1 Weak Connections*

The key point of our method is to find out the weakly connected weights and our strategy is to preset a threshold value for each layer of the deep neural network. After each parameter updates with a stochastic selected mini-batch, we compare the weights with the threshold value for each layer and mark out the low-weight connections with weights below the threshold. Simultaneously, we should also preset a critical time. Once a connection is continuously marked as low-weight beyond the critical time, we then identify the connection as weakly connected and prune it. So we set two parameters: threshold value (different for each layer) and critical time for the identifying of the weak connections.

The threshold values can seriously impact the amount of the pruned connections when the prune operation was first processed. If they are set to be too large, a huge amount of weights will be pruned after the first prune process, while much of them are still around the initial values and have not been fully trained. Since our prune operation is irreversible, it can get a bad connected architecture although it is sparse.

**Table 1** Training processes of our method

| |
|---|
| Step 1: the follow steps are processed iteratively to prune the weak connections |
| a. select a stochastic mini-batch from the training data set |
| b. calculate the gradients of the weights and biases by backpropagation |
| c. update the weights and biases with a proper learning rate |
| d. identify the weak connections and prune them until the desired sparseness being achieved |
| Step 2: train the thinned networks with the conventional SGD method |

And if the threshold values are set to be too small, it will take too much time to get the architecture with the same sparseness. So we should tune the threshold values properly on account of the mean value of the weights for each layer and the ultimate sparseness we want to achieve.

Conversely, the critical time have much less impact on the prune process since we found that once a weight were below the threshold values for a continuous times, it will be stubborn to be weak and hard to beyond the threshold values. So a moderate critical time has the same effect with a longer one. We set the critical time to be 600 iterations for all our experiments.

### 3.2 Model Description

Consider a neural network with L hidden layers. Let $l \in \{1, 2, \ldots, L\}$ index the hidden layers of the network. Let $y^{(l)}$ denote the vector of outputs from layer $l$, $W^{(l)}$ and $b^{(l)}$ are the weights and biases at layer $l$. We introduce a matrix $M^{(l)}$ for each layer to mark the weakly connected weights. If a weight is indentified to be weak connection, we set the corresponding element of $M^{(l)}$ to be zero. Apparently $M^{(l)}$ has the same size as $W^{(l)}$.

Then the update of the weights can be formalized with the following equations:

$$W^{(l)} = W^{(l)} + \eta \nabla W^{(l)};, \tag{2}$$

$$W^{(l)} = W^{(l)} * M^{(l)} \tag{3}$$

$\nabla W^{(l)}$ denotes the gradient of the loss function (cross-entropy for the softmax classifier) with respect to the weight vectors at layer $l$. $\eta$ is the learning rate and $*$ denotes an element-wise product.

$M^{(l)}$ should be updated with the strategy described in Sect. 3.1 at each iteration along with the update of weights and biases. The feedforward process in the layer $l + 1$ can be described as:

$$y^{(l+1)} = f(W^{(l)} y^{(l)} + b^{(l)}) \tag{4}$$

$f(.)$ denotes the activation function and we adopt tanh in our experiments.

### 3.3 Regularization

To accelerate the pruning process, we adopt the L2 regularization in the pruning phase. Since L2 regularization can penalize large parameters, and push the weights to be weak connections. We adopt a relatively large L2 parameter during the pruning phase to accelerate the prune process and a small L2 parameter in the training phase without pruning to improve the training performance.

# 4 Experiments

We test our method on deep neural networks for classification problem on the MNIST data set. The MNIST data set consists of $28 \times 28$ pixel handwritten digit images. The task is to classify the images into 10 digit classes. There are 60000 examples in the training set and 10000 in the testing set. We select 10000 examples from the training set for validating, which can help us find the optimal hyper-parameters (learning rate, threshold values for pruning and L2 parameter in our experiments). Some examples of the data set are shown in Fig. 1. Our deep models have the input layer of 784 dimensions and we employ the z-score operation before they are fed into the input layer.

## *4.1 Results on Models with Different Hidden Layers*

We test our method on neural networks with different hidden layers (2–5) and different hidden units. The results are shown in Table 2. For all the experiments, we choose an ultimate sparseness rate of 0.5 for our pruning method. We found that if trained without any tricks, the deeper model can even get a worse performance in consequence of overfitting. Our method can defeat overfitting to some extent and improve the performance of SGD for all the different models. For the deeper model with 4 and 5 hidden layers, our method even get better results than dropout.

**Fig. 1** Some examples selected from the training set

**Table 2** Comparison of different models on MNIST dataset

| Training method | Architecture | Error % |
|---|---|---|
| SGD | 2 hidden layers, 1000 units | 2.12 |
| SGD + dropout | 2 hidden layers, 1000 units | 1.53 |
| SGD + pruning | 2 hidden layers, 1000 units | 1.67 |
| SGD | 3 hidden layers, 500 units | 2.41 |
| SGD + dropout | 3 hidden layers, 500 units | 1.77 |
| SGD + pruning | 3 hidden layers, 500 units | 1.86 |
| SGD | 4 hidden layers, 500 units | 2.39 |
| SGD + dropout | 4 hidden layers, 500 units | 2.14 |
| SGD + pruning | 4 hidden layers, 500 units | 2.03 |
| SGD | 5 hidden layers, 200 units | 2.12 |
| SGD + dropout | 5 hidden layers, 200 units | 2.12 |
| SGD + pruning | 5 hidden layers, 200 units | 2.12 |

**Table 3** Comparison of different models on MNIST dataset

| Sparseness rate | 0.8 | 0.7 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 |
|---|---|---|---|---|---|---|---|
| Error % | 2.09 | 2.02 | 1.94 | 1.86 | 2.03 | 2.11 | 2.20 |

## *4.2 Effect of Sparseness Rate*

The ultimate sparseness rate is a tunable hyper-parameter in our method. We trained a model with different sparseness rate to explore the effect of varying this hyper-parameter. We use a 784-500-500-500-10 architecture. The results are showed in Table 3. We found that the best performance is achieved with the sparseness rate near 0.5, and the lower sparseness rate will get a worse performance.

## *4.3 Sparse Connection Versus Dense Connection*

We also made some experiments to demonstrate that the sparse architecture can get better performance than a dense architecture with the same amount of parameters. We train a 784-500-500-500-10 architecture with the ultimate sparseness rate of 0.5 and a 784-250-250-250-10 architecture without pruning. The latter get a test error of 2.07, while the former sparse architecture can get a better test error of 1.86. Many other similar experiments were conducted and all get the similar results, demonstrating the superiority of our pruning method.

## 5 Conclusion

We proposed a method of gradually pruning the weakly connected weights to train deep neural networks. And we proposed an effective strategy to identify the weak connections. The advantage of our method is that we can get a spare architecture and reduce the redundant parameters, so we can reduce overfitting caused by large and deep models. We can also solve the gradient vanishing problem to some extend through pruning the small weights. Our method has some similarities with the widely used method of dropout. The experiments demonstrate that our method can improve the conventional SGD and can get even better performance than dropout for deeper models.

**Acknowledgements** This research work is funded by the National Science Foundation of China (61271316), Shanghai Key Laboratory of Integrated Administration Technologies for Information Security, and Chinese National Engineering Laboratory for Information Content Analysis Technology.

## References

1. K. He, X. Zhang, S. Ren, J. Sun, Delving deep into rectifiers: surpassing humanlevel performance on imagenet classification (2015), arXiv:1502.01852
2. P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, Y. LeCun, Overfeat: integrated recognition, localization and detection using convolutional networks (2013), arXiv:1312.6229
3. G. Hinton et al., Deep neural networks for acoustic modeling in speech recognition. IEEE Signal Process. Mag. **29**, 82–97 (2012)
4. Y. Bengio, Learning deep architectures for AI. Found. Trends Mach. Learn. **2**(1), 1–127 (2009)
5. Y. LeCun, L. Bottou, G. Orr, K. Muller, Efficient backprop, in *Neural Networks: Tricks of the Trade* (Springer, 1998)
6. G. Tesauro, Practical issues in temporal difference learning. Mach. Learn. **8**, 257–277 (1992)
7. G.E. Hinton, S. Osindero, Y.W. Teh, A fast learning algorithm for deep belief nets. Neural Comput. **18**(7), 1527–1554 (2006)
8. X. Glorot, A. Bordes, Y. Bengio, Deep sparse rectifier neural networks, in Aistats, vol. 15, no. 106 (2011), p. 275
9. N. Srivastava, G.E. Hinton, A. Krizhevsky et al., Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
10. I. Sutskever, J. Martens, G.E. Dahl et al., On the importance of initialization and momentum in deep learning, in *ICML*, vol. 28, no. 3, pp. 1139–1147 (2013)
11. X. Glorot, Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, in *International Conference on Artificial Intelligence and Statistics* (2010), pp. 249–256
12. C.Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu, Deeply-supervised nets (2014), arXiv:1409.5185
13. S. Han, J. Pool, J. Tran et al., Learning both weights and connections for efficient neural network, Advances in *Neural Information Processing Systems* (2015), pp. 1135–1143
14. S.J. Nowlan, G.E. Hinton, Simplifying neural networks by soft weight-sharing. Neural Computation **4**(4) (1992)

15. H. Lee, A. Battle, R. Raina, A. Ng, Efficient sparse coding algorithms, in *NIPS06* (MIT Press, 2007), pp. 801–808
16. H. Lee, C. Ekanadham, A. Ng, Sparse deep belief net model for visual area V2, *NIPS07* (MIT Press, Cambridge, MA, 2008), pp. 873–880

# An Ideal Local Structure Learning for Unsupervised Feature Selection

**Yanbei Liu, Kaihua Liu and Deliang Liu**

**Abstract** In this paper, we propose a novel *Ideal Local Structure Learning (LSL) for unsupervised feature selection* method, which performs local structure learning and feature selection simultaneously. To obtain more accurate information of data structure, an ideal local structure with block diagonal constraint is introduced. Furthermore, a simple yet effective iterative algorithm is presented to optimize the proposed problem. Experiments on various benchmark datasets demonstrate the superiority of LSL compared with the state-of-the-art algorithms.

**Keywords** Unsupervised feature selection · Local structure learning · Similarity matrix · High-dimensional data

## 1 Introduction

High-dimensional data often contains quite a lot of irrelevant and noise features, which may cause damage to the data processing. To overcome these issues, feature selection technique has been considered as an effective tool to reduce the dimensionality. Because the labeling of samples is usually expensive and we cannot always obtain the labels beforehand, unsupervised feature selection holds great potential in real-world applications.

Most existing unsupervised feature selection methods have been proposed. Laplacian Score (LS) [1] selects features that best preserve the local manifold structure.

Y. Liu (✉) · K. Liu
School of Electronic Information Engineering, Tianjin University, Tianjin, People's Republic of China
e-mail: liuyanbei@tju.edu.cn

K. Liu
e-mail: liukaihua@tju.edu.cn

D. Liu
Shijiazhuang Mechanical Engineering College, Shijiazhuang, People's Republic of China
e-mail: liudeliang82@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_83

Cai et al. [2] proposed Multi-Cluster Feature Selection (MCFS), which selects valuable features using spectral regression to best preserve the local manifold structure. Li et al. [3] jointly use nonnegative constraint and $\ell_{2,1}$ norm to perform Nonnegative Discriminative Feature Selection (NDFS). Robust Unsupervised Feature Selection (RUFS) jointly selects features via a nonnegative matrix factorization and local structure learning [4]. However, the similarity matrices obtained by these methods are usually not the idea local neighbor assignment, and the block diagonal structure obtained is fragile and will be destroyed when the signal noise ratio is small.

In this paper, we propose a novel method, called *An Ideal Local Structure Learning (LSL) for unsupervised feature selection* to select the most discriminate features by exploring the local structure learning and feature selection simultaneously. In the new model, different from the most existing technology of fixing the input data structure associated to the similarity matrix, we learn a more reasonable data similarity matrix with block diagonal property which has exactly $c$ connected components, i.e., the $c$ clusters. So the similarity matrix obtained by local structure learning can be more accurate. Furthermore, to solve the proposed problem, we present a simple yet effective iterative algorithm. Experiments on five real-world datasets demonstrate that the proposed LSL has a better performance than other state-of-the-art unsupervised feature selection methods.

## 2 Local Structure Learning for Unsupervised Feature Selection

The most related work for unsupervised feature selection is Nonnegative Discriminative Feature Selection (NDFS) [3]. The objective function of NDFS is formulated as follows:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{F}} Tr(\mathbf{F}^T\mathbf{L}\mathbf{F}) + \beta\left(\|\mathbf{X}\mathbf{W}-\mathbf{F}\|_F^2 + \gamma\|\mathbf{W}\|_{2,1}\right) \\ &s.t.\ \mathbf{F}^T\mathbf{F} = \mathbf{I}, \mathbf{F} \geq 0, \end{aligned} \tag{1}$$

where $\mathbf{L} = \mathbf{B} - \mathbf{A}$ is the Laplacian matrix, in which $\mathbf{B}$ is the diagonal degree matrix with $\mathbf{B} = \sum_{j=1}^{n} A_{ij}$. $\mathbf{A} = \{A_{ij}\}$ is the similarity matrix measuring the spatial closeness of samples. $\gamma$ and $\beta$ are weighting parameters.

Inspired by [5], an ideal local structure $\mathbf{S}$ can be obtained by constraining its corresponding Laplacian matrix to be $rank\,(\mathbf{L}) = n - c$. Under this constraint, the learned similarity matrix is block diagonal with proper permutation. The optimization term is formulated as follows:

$$\begin{aligned} &\min_{\mathbf{F},\mathbf{S}} \|\mathbf{S}-\mathbf{A}\|_F^2 + 2\alpha Tr(\mathbf{F}^T\mathbf{L}_S\mathbf{F}) \\ &s.t.\ \Sigma_j s_{ij} = 1, s_{ij} \geq 0, \mathbf{F}^T\mathbf{F} = \mathbf{I}, \end{aligned} \tag{2}$$

Putting Eqs. (1) and (2) together, the proposed method LSL is to solve the following optimization problem:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{F},\mathbf{S}} \|\mathbf{S}-\mathbf{A}\|_F^2 + 2\alpha Tr(\mathbf{F}^T\mathbf{L}_S\mathbf{F}) + \beta\left(\|\mathbf{X}\mathbf{W}-\mathbf{F}\|_F^2 + \gamma\|\mathbf{W}\|_{2,1}\right) \\ &s.t.\ \Sigma_j s_{ij} = 1, s_{ij} \geq 0, \mathbf{F}^T\mathbf{F} = \mathbf{I}, \mathbf{F} \geq 0, \end{aligned} \tag{3}$$

where the first two terms learn the pseudo cluster labels using spectral analysis via the ideal local learning regularization while the last two terms simultaneously learn the feature selection matrix by joint $\ell_{2,1}$ norms minimization.

## 3 Optimization Algorithm

To optimize the objective function, we first rewrite the objective function of LSL as follows:

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{F},\mathbf{S}} \|\mathbf{S}-\mathbf{A}\|_F^2 + 2\alpha Tr(\mathbf{F}^T\mathbf{L}_S\mathbf{F}) + \beta\left(\|\mathbf{X}\mathbf{W}-\mathbf{F}\|_F^2 + \gamma\|\mathbf{W}\|_{2,1}\right) + \frac{\rho}{2}\left\|\mathbf{F}^T\mathbf{F}-\mathbf{I}_c\right\|_F^2 \\ &s.t.\ \Sigma_j s_{ij} = 1, s_{ij} \geq 0, \mathbf{F} \geq 0, \end{aligned} \tag{4}$$

where $\rho > 0$ is a parameter to control the orthogonality condition. In practice, $\rho$ should also be large enough to insure the orthogonality satisfied.

**Update S**: With fixed **F** and **W**, the problem (4) is transformed into

$$\begin{aligned} &\min_{\mathbf{W},\mathbf{F},\mathbf{S}} \|\mathbf{S}-\mathbf{A}\|_F^2 + 2\alpha Tr(\mathbf{F}^T\mathbf{L}_S\mathbf{F}) \\ &s.t.\ \Sigma_j s_{ij} = 1, s_{ij} \geq 0, \end{aligned} \tag{5}$$

The problem (5) is equivalent to the following problem:

$$\min_{\Sigma_j s_{ij}=1, s_{ij}\geq 0} \sum_{i,j}(s_{ij}-a_{ij})^2 + \alpha\sum_{i,j}\left\|\mathbf{f}_i-\mathbf{f}_j\right\|_2^2 s_{ij}). \tag{6}$$

Note that the problem (6) is independent for different $i$, denote $v_{ij} = \left\|\mathbf{f}_i-\mathbf{f}_j\right\|_2^2$, and denote $\mathbf{v}_i$ as a vector with the $j$-th element equal to $v_{ij}$, so the problem can be written in vector form as

$$\min_{\mathbf{s}_i^T\mathbf{1}=1, s_{ij}\geq 0} \left\|\mathbf{s}_i - (\mathbf{a}_i - \frac{\alpha}{2}\mathbf{v}_i)\right\|_2^2. \tag{7}$$

This problem can be solved by an efficient iterative algorithm [6].

**Update W**: With fixed **F** and **S**, the problem (4) is transformed into

$$\min_{\mathbf{W}} \|\mathbf{XW} - \mathbf{F}\|_F^2 + \gamma \|\mathbf{W}\|_{2,1} \tag{8}$$

Taking the derivative of problem (8) w.r.t **W**, and setting the derivative to zero, we have

$$\mathbf{W} = (\mathbf{X}^T\mathbf{X} + \gamma\mathbf{G})^{-1}\mathbf{X}^T\mathbf{F}. \tag{9}$$

Here **G** is a diagonal matrix with $G_{ii} = \frac{1}{2||\mathbf{w}_i||_2}$.

**Update F**: Substituting **W** by Eq. (9), the problem (4) is transformed into

$$\min_{\mathbf{F}\geq 0} Tr(\mathbf{F}^T\mathbf{QF}) + \frac{\rho}{2}\left\|\mathbf{F}^T\mathbf{F} - \mathbf{I}_c\right\|_F^2 \tag{10}$$

where $\mathbf{Q} = \alpha\mathbf{L}_S + \beta(\mathbf{I}_n - 2\mathbf{X}(\mathbf{X}^T\mathbf{X} + \gamma\mathbf{G})^{-1}\mathbf{X}^T)$ and $\mathbf{I}_n \in \Re^{n\times n}$ is an identity matrix. Following [7], we introduce multiplicative updating rules. Letting $\phi_{ij}$ be the Lagrange multiplier for constraint $F_{ij} \geq 0$ and $\mathbf{\Phi} = [\phi_{ij}]$, the Lagrange function is

$$Tr(\mathbf{F}^T\mathbf{QF}) + \frac{\rho}{2}\left\|\mathbf{F}^T\mathbf{F} - \mathbf{I}_c\right\|_F^2 + Tr(\mathbf{\Phi}\mathbf{F}^T). \tag{11}$$

Setting its derivative with respect to $F_{ij}$ to 0 and using the Karush–Kuhn–Tucker (KKT) condition [8] $\phi_{ij}F_{ij} = 0$, we obtain the updating rules:

$$F_{ij} \leftarrow F_{ij}\frac{(\rho\mathbf{F})_{ij}}{(\mathbf{QF} + \rho\mathbf{F}\mathbf{F}^T\mathbf{F})_{ij}}. \tag{12}$$

Then, we normalize **F** such that $(\mathbf{F}^T\mathbf{F})_{ii} = 1, i = 1, \ldots, c$.

## 4 Experiments

Five datasets are used to compare the performances of different algorithms. These datasets include two face image datasets (ORL[1] and COIL20 [2]), one spoken letter recognition dataset (Isolet1[2]) and two biological datasets (Colon[3] and Lung). Following [2, 3], we evaluate the performances of algorithms in term of clustering Accuracy (ACC) and Normalized Mutual Information (NMI). In our experiment, we tune the parameters by a "grid-search" strategy from $\{10^{-4}, 10^{-3}, \ldots, 10^3, 10^4\}$. The number of selected features is set as $\{50, 100, 150, 200, 250, 300\}$ for all the datasets. We

[1] http://www.face-rec.org/databases/.

[2] http://www.ics.uci.edu/mlearn/MLSummary.html.

[3] http://featureselection.asu.edu/datasets.php.

**Table 1** Clustering results (ACC% ± std) of different feature selection algorithms

| | Colon | Lung | Isolet | ORL | COIL20 |
|---|---|---|---|---|---|
| LS | 58.1 ± 0.00 | 67.7 ± 4.03 | 47.7 ± 2.64 | 46.8 ± 3.07 | 54.6 ± 3.41 |
| MCFS | 53.2 ± 0.00 | 75.6 ± 9.58 | 51.2 ± 3.19 | **54.3 ± 3.81** | 57.7 ± 5.10 |
| NDFS | 63.7 ± 1.37 | 72.2 ± 1.01 | 58.5 ± 3.20 | 51.3 ± 2.49 | 61.6 ± 3.82 |
| LSL | **64.2 ± 2.34** | **80.7 ± 4.72** | **58.7 ± 2.89** | 53.2 ± 2.03 | **63.7 ± 5.30** |

**Table 2** Clustering results (NMI% ± std) of different feature selection algorithms

| | Colon | Lung | Isolet | ORL | COIL20 |
|---|---|---|---|---|---|
| LS | 1.83 ± 0.86 | 49.0 ± 2.72 | 63.7 ± 1.26 | 68.7 ± 1.91 | 69.9 ± 2.42 |
| MCFS | 2.38 ± 1.05 | 54.2 ± 9.89 | 67.3 ± 1.41 | **73.2 ± 2.41** | 71.4 ± 2.19 |
| NDFS | 4.24 ± 8.93 | 50.3 ± 5.07 | 72.9 ± 1.43 | 72.1 ± 1.43 | 75.5 ± 3.03 |
| LSL | **6.22 ± 4.16** | **58.2 ± 4.83** | **73.0 ± 1.77** | 72.1 ± 1.46 | **76.4 ± 1.99** |

report the best clustering results of all the algorithms using different optimal parameters. After selecting the features, we use K-means to cluster samples. Because the results of K-means clustering depend on initialization, we repeat the experiments 20 times. Finally, the average results with standard deviation are reported.

The performances of different feature selection algorithms which are evaluated by ACC and NMI are shown in Tables 1 and 2. We can see from the two tables that the performances of LSL have the best clustering results on most datasets, which indicates LSL selects the most discriminative features. Specially, the embedded approaches MCFS, NDFS, and LSL achieve better performance than LS that ranks features one by one. Second, with the discriminate analysis and considering of noise data, NDFS and LSL outperform MCFS. Finally, The proposed LSL performs feature selection and local structure learning simultaneously, and hence it obtains better results.

## 5 Conclusion

In this paper, we propose a novel method, called An Ideal Local Structure Learning (LSL) for unsupervised feature selection to select the most discriminate features by exploring the local structure learning of data and feature selection simultaneously. Instead of fixing the input data structure associated to the similarity matrix, a new similarity matrix with block diagonal property is introduced into our model. To solve the proposed objective function, a simple and efficient algorithm is presented. Extensive experiments on five benchmark datasets have validated the effectiveness of the proposed method.

**Acknowledgements** This work was partially supported by Major Program of National Natural Science Foundation of China (Grant no. 13&ZD162) and Applied Basic Research Programs of China National Textile and Apparel Council (Grant no. J201509).

## References

1. X. He, D. Cai, P. Niyogi, Laplacian score for feature selection, in *Proceedings of Advances in Neural Information Processing Systems*, vol. 18 (2005)
2. D. Cai, C. Zhang, X. He, Unsupervised feature selection for multi-cluster data, in *Proceedings of the International Conference on Knowledge Discovery and Data Mining* (2010), pp. 333–342
3. Z. Li, Y. Yang, J. Liu, X. Zhou, H. Lu, Unsupervised feature selection using nonnegative spectral analysis, in *Proceedings of Association for the Advancement of Artificial Intelligence* (2012)
4. M. Qian, C. Zhai, Robust unsupervised feature selection, in *Proceedings of International Joint Conference on Artificial Intelligence* (2013), pp. 1621–1627
5. F. Nie, X. Wang, M.I. Jordan, H. Huang, The constrained laplacian rank algorithm for graph-based clustering, in *Proceedings of Association for the Advancement of Artificial Intelligence* (2016)
6. J. Huang, F. Nie, H. Huang, A new simplex sparse learning model to measure data similarity for clustering, in *Proceedings of the International Conference on Artificial Intelligence* (2015), pp. 3569–3575
7. Y. Liu, R. Jin, L. Yang, Semi-supervised multi-label learning by constrained non-negative matrix factorization, in *Proceedings of the International Conference on Artificial Intelligence* (2006), pp. 421–427
8. H.W. Kuhn, A.W. Tucker, Nonlinear programming, in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (1951), pp. 481–492

# Person Re-Identification Based on Kernel Large Margin Nearest Neighbor Classification

**Linlin Yang, Jian Cheng and Haijun Liu**

**Abstract** Person re-identification is a process of matching person images of same identity across nonoverlapping camera views at different locations and times. In this paper, we introduce how to use the kernel trick to improve the performance of large margin nearest neighbor (LMNN) classification for person re-identification. Since the classification ability of LMNN is weak for those person features with nonlinear distribution, KLMNN combining kernel trick and LMNN is introduced to extend linear distance metric to nonlinear cases. Three kernel-based methods and two indicators are applied to evaluate the performance of KLMNN.

**Keywords** Person re-identification · Kernel trick · KLMNN

## 1 Introduction

The task of person re-identification can be formalized as the problem of matching a given query image against a gallery. As illustrated in Fig. 1, given a query image captured by a camera, we attempt to search for the image which has the same identity of the query image in the gallery captured by other cameras. The research on person re-identification has two directions: (1) capturing a reliable person representation, (2) learning a suitable distance metric. In this paper, we focus on the latter.

Person re-identification is a challenging task because a person observed in different camera views often undergoes significant variations in illumination, viewpoint, background changes, occlusions, and low resolution. Many existing person re-identification approaches try to learn a linear projection matrix for distance metric [1]. However, its performance may degrade when the given data points has nonlinear distribution in original space. In this case, kernel trick is introduced to

L. Yang · J. Cheng (✉) · H. Liu
School of Electronic Engineering, University of Electronic Science and Technology of China, 2006 Xiyuan Ave, Chengdu 611731, China
e-mail: chengjian@uestc.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_84


**Fig. 1** The process of person re-identification. Given a query image, we attempt to search for the image which has the same identity of the query image in the gallery

map the original data points to a high-dimensional space which can be called as kernel space.

One of the successful applications in kernel trick is Support Vector Machine (SVM) [2] classifier which seeks for an optimal separating hyper-plane in kernel space. In addition, the kernel-based method Kernel PCA (KPCA) [3] has also been widely applied to extract nonlinear principal components. To alleviate the limitation on linear mapping capability, Mignon used kernel Pairwise Constrained Component Analysis (PCCA) [4] to map the original data points into a high-dimensional space without explicitly computing the mapping for person re-identification.

Inspired by the use of kernel trick for classification, KLMNN combining kernel trick and LMNN [5] is introduced for person re-identification. Three kernels and four challenging person re-identification datasets are used to evaluate the performance of person re-identification.

## 2 Person Re-Identification Based on KLMNN

In this section, KLMNN combining kernel trick and LMNN is introduced to extend linear distance metric to nonlinear structured data without high computational complexity.

### *2.1 The Framework of Kernel Method for Person Re-Identification*

Let $X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{d \times n}$ denote the feature matrix extracted from the person images, where $d$ is the feature dimension and $n$ is the number of the training samples. As illustrated in Fig. 2, it can be seen that the framework of kernel method for person re-identification has three steps,

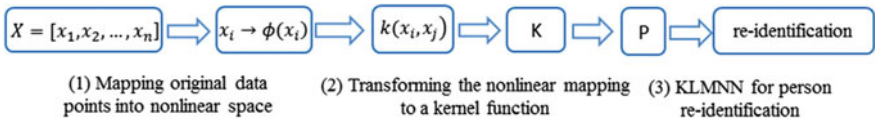



**Fig. 2** The framework of KLMNN for person re-identification

(1) Mapping original data points into nonlinear space

A nonlinear mapping ϕ(x) should be applied to map the original data points to a high-dimensional space, where those data points could be linearly separated.

(2) Transforming the nonlinear mapping to a kernel function

Kernel function is introduced to replace the nonlinear mapping ϕ(x). Choosing the kernel function is equal to choosing the nonlinear mapping ϕ(x). Once a kernel function is chosen, the corresponding kernel matrix K can be acquired.

(3) KLMNN for person re-identification

KLMNN is used to learn a linear projection matrix in kernel space. A Mahalanobis distance metric in kernel space is applied to measure the similarity for person images re-identification.

## 2.2 KLMNN

Based on the framework of KLMNN for person re-identification, we will introduce how to embed kernel trick to LMNN. About LMNN, it attempts to minimize the distance between $x_i$ and its target neighbors, and to maximize the distance between $x_i$ and its imposters by a large margin in original space. The target neighbors of $x_i$ can be simply defined as the k nearest neighbors sharing the same class label of $x_i$. The imposters of $x_i$ are those data points which are the nearest neighbors of $x_i$, but having a different class label of $x_i$. LMNN applied Mahalanobis distance to compute the similarity,

$$D_M(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j) \tag{1}$$

where $M$ is a positive semi-definite matrix, so it can be decomposed as $L^T L$. We can rewrite Eq. (1) as follows,

$$D_M(x_i, x_j) = (x_i - x_j)^T L^T L (x_i - x_j) = (e^i - e^j)^T X^T L^T L X (e^i - e^j) \tag{2}$$

where $e^i = [0, 0, \ldots, 1 \ldots 0]^T$ is a column vector and the $i$th dimension of $e^i$ is 1.

The performance of LMNN may degrade when the given data has nonlinear distribution. To solve this problem, kernel trick is introduced to map the input data

points to a high-dimensional space. Usually, the kernel function can be written as follows:

$$k(x_i, x_j) = \phi(x_i)^T \phi(x_j) \quad i, j = 1, 2, \ldots n \tag{3}$$

By replacing inner product with kernel function, we acquire three kernel functions which are widely used in practice,

$$\text{Linear kernel:} \quad k_L(x_i, x_j) = x_i^T x_j \tag{4}$$

$$\chi^2\text{kernel:} \quad k_{\chi^2}(x_i, x_j) = \sum_{m=1}^{d} \frac{2x_i^m x_j^m}{\left(x_i^m + x_j^m\right)} \tag{5}$$

$$\text{RBF} - \chi^2\text{kernel:} \; k_{\chi^2}^{RBF}(x_i, x_j) = e^{-D_{\chi^2}^2(x_i, x_j)} \tag{6}$$

where $x_i^m$ is the $m$th dimension of $x_i$, $D_{\chi^2}^2(x_i, x_j) = \sum_{m=1}^{d} \frac{\left(x_i^m - x_j^m\right)^2}{\left(x_i^m + x_j^m\right)}$ is the corresponding distance metric of $\chi^2$. Then the kernel matrix K can be acquired

$$\mathrm{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \cdots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \cdots & k(x_2, x_n) \\ \vdots & \vdots & \vdots & \vdots \\ k(x_n, x_1) & k(x_n, x_2) & \cdots & k(x_n, x_n) \end{bmatrix}_{n \times n} \tag{7}$$

$k_i = [k(x_1, x_i), k(x_2, x_i), \ldots, k(x_n, x_i)]^T$ is the $i$th column vector in K. After acquiring kernel matrix, we need to use the kernel matrix K instead of X and the projection matrix P instead of $L$ in Eq. (2) to compute corresponding distance metric,

$$D_P(k_i, k_j) = \left(e^i - e^j\right)^T K^T P^T P K \left(e^i - e^j\right) \tag{8}$$

The idea of KLMNN learns a projection transformation P which minimizes the distance between each input and its target neighbors in Eq. (9) and maximizes the distance between each input and its imposters in Eq. (10) in kernel space

$$\varepsilon_{pull}(P) = \sum_{ij} \eta_{ij} \left(e^i - e^j\right)^T K^T P^T P K \left(e^i - e^j\right) \tag{9}$$

$$\varepsilon_{push}(P) = \sum_{ijl} \eta_{ij} (1 - \mu_{il}) \left[1 + D_P(k_i, k_j) - D_P(k_i, k_l)\right]_+ \tag{10}$$

where $[z]_+ = max(z, 0)$, if $x_j$ is the target neighbor of $x_i$, $\eta_{ij} = 1$, others $\eta_{ij} = 0$, and if $x_l$ is the imposters of $x_i$, $\mu_{il} = 0$, others $\mu_{il} = 1$.

The loss function of KLMNN is the combination of $\varepsilon_{pull}(P)$ and $\varepsilon_{push}(P)$,

$$\varepsilon(P) = (\alpha)\varepsilon_{pull}(P) + (1-\alpha)\varepsilon_{push}(P) \tag{11}$$

The trade-off constant $\alpha \in (0, 1)$, which can be tuned via cross validation.

## 3 Experiments

### 3.1 Datasets and Image Representations

We demonstrate the performance of KLMNN on VIPeR [6], iLIDS [7], 3DPeS [8], and CAVIAR4REID [9] datasets. VIPeR dataset contains 1,264 person images of 632 pedestrians captured from two nonoverlapping viewpoints. It contains a high degree of viewpoint and illumination variations as illustrated in Fig. 3a. iLIDS dataset has 476 images of 119 pedestrians acquired from nonoverlapping cameras. Since this dataset was collected at an airport, the images often have severe occlusions caused by pedestrians, luggage, and trolleys as illustrated in Fig. 3b. 3DPeS dataset which includes 1,011 images of 192 pedestrians is collected by 8 nonoverlapping outdoor cameras with significantly different viewpoints that can be seen in Fig. 3c. CAVIAR4REID dataset includes 1,220 person images of 72 pedestrians captured by two different cameras. Typical challenges on this dataset are viewpoint and pose changes, different light conditions, occlusions, and low resolution as shown in Fig. 3d.

All person images of four datasets are normalized to 128 × 48. A 32 × 32 sub-window with a step of 16 × 16 is used to extract feature on all original images excluding 3DPeS dataset which extracts feature in the mask area. Within each sub-window, we extract 16-bins color histograms from the RGB, YUV, and HS color channels, 8-neighbors of radius 1 and 16-neighbors of radius 2 uniform Local


**Fig. 3** Typical person image pairs from **a** VIPeR, **b** iLIDS, **c** 3DPeS, and **d** CAVIAR4REID dataset (images on each column are the same person)

Binary Patterns histograms. The histograms are normalized with the $l_1$ norm in each color histograms and texture histograms. Finally, all histograms are concatenated together to form a single vector.

### 3.2 Experimental Details and Evaluation

In our experiments, each dataset was randomly divided into two subsets: training set and test set. We randomly selected all images of p individuals to set up the test set, and the rest of the individuals were used for training. This partition was repeated 10 times. Each test set was composed of a gallery set and a query set. Under each partition, one image of each individual in the test set was randomly selected as the gallery set, the rest of the test images were used as query set. This process was also repeated 10 times. All results are reported as the mean of the $10 \times 10$ folds. We evaluate the KLMNN with several existing metric learning algorithms including KISSME [10], LFDA [11] and LMNN [5] on four datasets. PCA was applied to original data points before using KISSME, LFDA, and LMNN. The dimension can keep 95% energy after PCA. As for KISSME, the ratio of negative/positive pairs in the train set is 10. The trade-off constant $\alpha$ in loss function is 0.5 in all experiments. The target neighbor on VIPeR, iLIDS, 3DPeS, and CAVIAR4REID datasets is 1, 1, 1, 3, respectively.

The Cumulative Match Characteristic curve (CMC) is a kind of tool for evaluating the performance of person re-identification, $\mathrm{CMC(r)} = \sum_{i=1}^{r} p(i), p(i)$ is the probability of correct match at rank $i$. In addition, we also report the proportion of uncertainty removed (PUR) scores. $\mathrm{PUR} = \frac{\log(N) + \sum_{i=1}^{N} P(i) \log p(i)}{\log(N)}$, N is the size of gallery set. PUR score compares the uncertainty under a randomized rank and a ranking algorithm.

### 3.3 Performance Analysis

Figure 4 shows the CMC curves of each algorithm on four datasets. The KLMNN based on the linear, $\chi^2$ and $R_{\chi^2}$ kernels are indicated by KLMNN-L, KLMNN $-\chi^2$ and KLMNN $-R_{\chi^2}$. The results are also summarized in Table 1 along with the PUR scores. The CMC curves and PUR score indicate that KLMNN is better than the existing metric learning algorithms including KISSME and LFDA. The comparison between LMNN and KLMNN indicates that KLMNN successfully learns an effective metric in a discriminant space.

On VIPeR dataset, the highest CMC and PUR scores are achieved using the KLMNN $-R_{\chi^2}$ method. Even though the improvement is obvious when compared with other algorithms, the accuracy at $\mathrm{r} = 1$ is the lowest among these four datasets. This can be explained by that VIPeR has only two images per individual in training

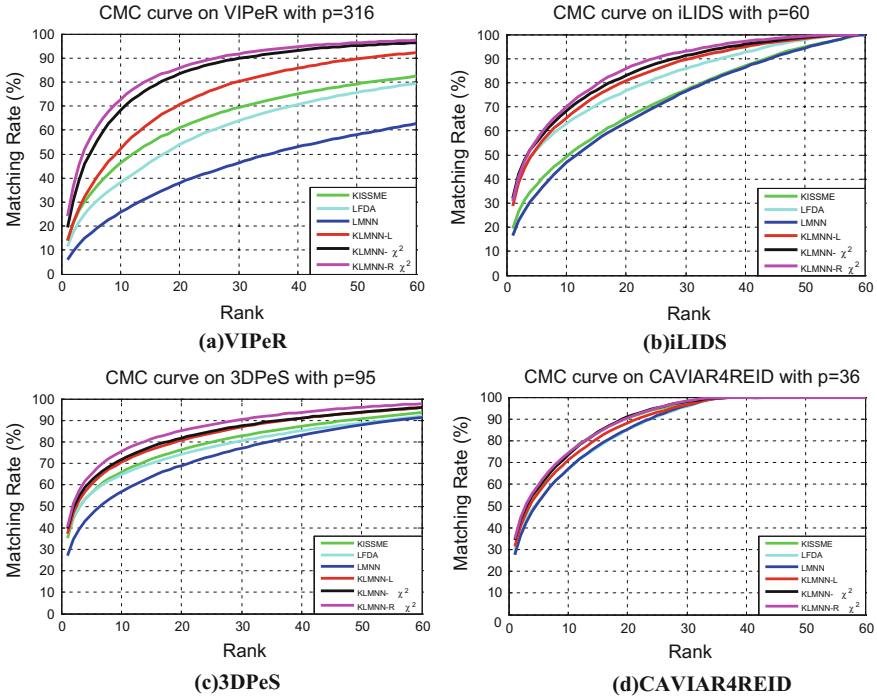



**Fig. 4** CMC curve on four person re-identification datasets

set, while other dataset has at least two images in the training set. At the same time, it shows that VIPeR dataset is more difficult than the other datasets.

On iLIDS dataset, though the highest accuracy at r = 1 is KLMNN $-\chi^2$, KLMNN $-R_{\chi^2}$ achieves good performance according to the iLIDS CMC curve and PUR score when considering the whole performance.

The performance of all algorithms on 3DPeS is the best among the four datasets. This can be explained that the original image use the mask to extract the foreground before extracting feature.

On CAVIAR4REID dataset, the improvement of KLMNN $-\chi^2$ and KLMNN $-R_{\chi^2}$ at r = 1, 5 10, 20 is not so clear when compared with LFDA. This is may be caused by the low resolution of original person images. Compared against the LFDA, KLMNN $-R_{\chi^2}$ improves about 6.6% at PUR score. This demonstrated the effectiveness of our proposed method.

**Table 1** CMC at r = 1, 5, 10, 20 and PUR scores on four datasets

| | r | KISSME | LFDA | LMNN | KLMNN-L | KLMNN $-\chi^2$ | KLMNN-$R_{\chi^2}$ |
|---|---|---|---|---|---|---|---|
| VIPeR | 1 | 14.30 | 11.64 | 5.94 | 14.03 | 19.54 | **24.13** |
| | 5 | 33.64 | 28.10 | 17.20 | 36.67 | 51.03 | **57.37** |
| | 10 | 46.31 | 38.13 | 25.79 | 52.29 | 68.31 | **72.76** |
| | 20 | 60.84 | 53.92 | 37.98 | 70.55 | 83.34 | **85.91** |
| | PUR | 26.14 | 22.28 | 12.32 | 33.16 | 42.40 | **46.14** |
| iLIDS | 1 | 19.46 | 31.32 | 16.41 | 28.70 | **32.12** | 30.71 |
| | 5 | 37.69 | 52.15 | 33.88 | 52.77 | 55.42 | **56.40** |
| | 10 | 49.32 | 63.12 | 46.89 | 65.41 | 68.32 | **69.99** |
| | 20 | 65.22 | 76.75 | 63.44 | 80.80 | 82.99 | **85.93** |
| | PUR | 12.23 | 23.54 | 10.06 | 24.02 | 27.13 | **27.90** |
| 3DPeS | 1 | 35.34 | 36.92 | 27.01 | 37.39 | 39.61 | **40.62** |
| | 5 | 56.40 | 56.22 | 46.25 | 60.32 | 62.08 | **65.09** |
| | 10 | 66.07 | 64.77 | 56.76 | 70.54 | 71.83 | **75.56** |
| | 20 | 76.37 | 74.11 | 68.99 | 80.87 | 81.80 | **85.21** |
| | PUR | 30.85 | 30.58 | 22.37 | 34.77 | 36.52 | **39.72** |
| CAVIAR4REID | 1 | 27.55 | 33.97 | 27.58 | 31.50 | 34.47 | **34.94** |
| | 5 | 57.36 | 58.62 | 51.77 | 56.31 | 58.74 | **59.88** |
| | 10 | 74.15 | 73.29 | 67.04 | 71.11 | 73.51 | **74.30** |
| | 20 | 90.51 | 89.69 | 85.33 | 88.26 | **90.94** | 90.52 |
| | PUR | 20.96 | 18.19 | 17.26 | 21.07 | 24.21 | **24.80** |

## 4 Conclusion

In this paper, we applied Kernel Large Margin Nearest Neighbor for person re-identification. Experimental results suggest that the original data points can be correctly classified by using the kernel trick. Three kernels have different classification performance. In generally, KLMNN $-R_{\chi^2}$ performed better than KLMNN $-\chi^2$ which, in turn, performed better than LMNN-L. Though the performance of LMNN-L is the worst in three kernels, it is better than LMNN, which indicates that KLMNN can successfully learn an effective distance metric in kernel space. Comparison against two existing approaches (KISSME, LFDA) also demonstrated the effectiveness of our proposed KLMNN for person re-identification on VIPeR, iLIDS, 3DPeS, and CAVIAR4REID datasets.

**Acknowledgements** This work was supported by the National Science Foundation of China (NO. 61671125 and NO. 61201271), and the State Key Laboratory of Synthetical Automation for Process Industries (NO. PAL-N201401).

## References


1. P.M. Roth, M. Hirzer, M. Köstinger et al., Mahalanobis distance learning for person re-identification, in *Person Re-Identification* (Springer, London, 2014), pp. 247–267
2. E. Osuna, R. Freund, F. Girosi, Support vector machines: training and applications (1997)
3. B. Schölkopf, A. Smola, K.R. Müller, Nonlinear component analysis as a kernel eigenvalue problem. Neural Comput. **10**(5), 1299–1319 (1998)
4. A. Mignon, F. Jurie, A new approach for distance learning from sparse pairwise constraints, in *2012 IEEE Conference on Computer Vision and Pattern Recognition*
5. K.Q. Weinberger, L.K. Saul, Distance metric learning for large margin nearest neighbor classification. J. Mach. Learn. Res. **10**, 207–244 (2009)
6. D. Gray, H. Tao, Viewpoint invariant pedestrian recognition with an ensemble of localized features, in *Computer Vision–ECCV 2008* (Springer Berlin Heidelberg, 2008), pp. 262–275
7. W.S. Zheng, S. Gong, T. Xiang, Associating groups of people, in *The British Machine Vision Conference*, vol. 2 (2009), p. 6
8. D. Baltieri, R. Vezzani, R. Cucchiara, 3dpes: 3d people dataset for surveillance and forensics, in *Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding* (ACM, 2011), pp. 59–64
9. D.S. Cheng, M. Cristani, M. Stoppa et al., Custom pictorial structures for re-identification, in *The British Machine Vision Conference*, vol. 1, no. 2 (2011), p. 6
10. M. Koestinger, M. Hirzer, P. Wohlhart et al., Large scale metric learning from equivalence constraints, in *2012 IEEE Conference on Computer Vision and Pattern Recognition*
11. S. Pedagadi, J. Orwell, S. Velastin et al., Local fisher discriminant analysis for pedestrian re-identification, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 3318–3325

# People Relation Extraction of Chinese Microblog Based on SVMDT-RFC

**Ge Zhou, Xiao Peng, Chenglin Zhao and Fangmin Xu**

**Abstract** People relation extraction is a significant topic in information extraction field. While in traditional study, the feature of extraction lexical and semantic was attached importance to, and the function of classifier was neglected, furthermore, there is great difference between microblog language materials and that of tradition. When it mentioned traditional classification algorithm, its low correctness and the inaccuracy to identification of fuzzy sample become the reason of being used little. In this paper, the traditional classification algorithm was improved. Using SVMDT-Random Forest and we designed, the fuzzy sample classifying ability increased, which remedied the shortcomings of SVM and Random Forest effectively. By testing the microblog language materials, the result indicated that this method can improve the performance of people relation extraction.

**Keywords** People relation extraction · Chinese microblog · Multi-class classifier · SVMDT-RFC

## 1 Introduction

With the rapid development of the Internet, the form of information dissemination has all series types of differences, such as blogs, microblogs become the dominant mode of transmission. People relation extraction is a special and huge problem, and the existing approaches of solving the problems are mainly in the following three ways:

G. Zhou (✉) · C. Zhao · F. Xu (✉)
Key Lab of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing, China
e-mail: gezhou@bupt.edu.cn

F. Xu
e-mail: xufm@bupt.edu.cn

X. Peng
The State Radio Monitoring_center Testing Center, Beijing, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_85

(1) Relation extraction method based on pattern matching. This method is mainly used in natural language processing methods and matches the entity relation with the already dealt templates. Hamish Cunning and some scholars realized the relation extraction after defining some rules and then matching between the templates [1]. Someone like Pantel proposed a method to realize the semantics extraction by Espresso algorithm and better the common templates [2].
(2) Relation extraction method based on machine learning. Han Bing, fulfilled relation extraction via support vector machines [3]. And Huang Weichun improved the people relation extraction efficiency by reducing dimensions of feature vectors by character selection method [4].
(3) Entity relation extraction method based on ontology. Based on ontology method, syntactic analysis method, lexical frequency matrix analysis, and verb-centered theory were used to extract all types of conceptions and relations existing in this paper [5].

The main method to study on people relation extraction is machine learning. Due to the irregularities of language materials and the indetermination of context, there must be more fuzzy samples. In accordance to the problems above, we can cope with complex classification using the improved Random Forest Classifier, which aims to enhance the classification ability of fuzzy sample and then to improve the overall performance of classifier.

## 2 Algorithm Design of Sentiment Analysis

### *2.1 SVM Decision Tree Point Classification Algorithm Based on Class Separation Distances*

The essence of Random Forest Classifier algorithms is an ensemble classifier. Selecting the appropriate meta classifier is very significant. It is really a worthy question that introducing a appropriate unit classifier.

SVM (Support Vector Machine) is type of classifier about distance, which has very good performance, and it works in binary classification. People relation extraction is a problem of multi-classification. For a multi-classification matter, it can be average between any two categories. So we can transform the multi-classification to a binary classification.

SVM decision tree is a kind of classifier to combine the SVM decision tree and binary decision tree as a whole tree structure.

When designing the SVM classifier, the difference of misclassification should be ensured as little as possible, that is to say, if the node which is split mistakenly has a certain distance with the root of the tree, which will be favorable to classification of fuzzy samples. Thus, it is liable to divide the samples which are easy to divided, then the difficult ones. For the sake of this aim, classification algorithm of every decision node has to make sure that separability reach high. According to this idea,

we measure the separability between categories in term of the distance of sort center [6], every division method has to ensure the separability of the right set and wrong set be highest and the distance of their center be furthest.

Assume the positive set classified by node classification as $S_1$, positive set type number as $N_1$, negative set as $S_2$, negative set type number as $N_2, X_i$ represents the No. $i$ sample set, and $i = 1, 2, \ldots, N_1 + N_2, X_i$ the number of samples is $n_i$, sample vector space is $x$,the splitting method based on within-class distance:

**Algorithm1 Splitting Algorithm of SVM Decision Tree Based on Within-Class Distance**

(1) To calculate every classification set center:

$$C_i = \frac{1}{n_i} \sum_{x_k \in X_i} x_k \tag{1}$$

(2) Assume that J is the number of class nodes splitting method, for every kind of splitting method, calculating it in accordance with the Step (3) and (4):

(3) Firstly, to calculate the center of positive set $S_1$ center and negative set $S_2$ center:

$$e_1^j = \frac{1}{N_1} \sum_{i \in S_1} C_i \tag{2}$$

$$e_2^j = \frac{1}{N_2} \sum_{i \in S_2} C_i \tag{3}$$

To calculate the Euclidean distance from positive set center to negative set center:

$$d_{S_1 S_2} = \left\| e_1^j - e_2^j \right\| \tag{4}$$

(4) To calculate the average distance from the sample center to set center in $S_1$ and $S_2$:

$$d^j_{S_1} = \frac{1}{N} \sum_{i \in S_1} \left\| C_i - e_1^j \right\| \tag{5}$$

$$d^j_{S_2} = \frac{1}{N} \sum_{i \in S_2} \left\| C_i - e_2^j \right\| \tag{6}$$

The optimal separation solution $f$ can be achieved:

$$f = \max_{j=1,2,\ldots,J} \left( d_{S_1 S_2}^j + d^j_{S_1} + d_{S_2}^j \right) \tag{7}$$

## 2.2 Weighted Random Forest Voting Based on Margin

Random Forest Classification (RFC) is classifiers combination, including many decision tree classification models.

For the voting process, the Weighted Random Forest Voting based on Margin was designed in this paper. The classification margin was defined the difference between the right votes on this sample of ensemble classifier and the most votes determined other category [7]. If the margin of the sample is greater than zero, it indicates that the number that the meta classifier made the right decision is more than that of the wrong ones.

Assume sample space $S = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$, the Random Forest Model $H = \{h_1, h_2, \ldots, h_k\}$, $h_k$ represents the No. $k$ meta decision trees, for a given sample $(x_n, y_n)$, on every meta decision tree in random forest, the calculation method of classification margin is as below:

In one process of iteration, the classification interval of a certain meta decision tree can be shown as below:

$$mg(h_k) = \sum_{i=1}^{i<=n} (f(h_k, x_i, y_i))s, \tag{8}$$

$$f(h, x, y) = \begin{cases} 1 & (h(x) = y) \\ 0 & (h(x) != y) \end{cases} \tag{9}$$

The average classification interval of certain iteration is as below:

$$am(S, H) = \frac{1}{H}\left(\sum_{k=1}^{k<=H} mg(h_k)\right) \tag{10}$$

If $mg(h_k) > am(S, H)$ represents the big contribution to classification of the meta classification interval, we improve the weight of meta classification in the next iteration. If $mg(h_k) < am(S, H)$ represents the small contribution to classification of the meta classification interval, we decrease the weight of meta classification in the next iteration.

Therefore, the set $H^k_{weight}$ of meta classification with weight of every iteration, then repeat the operation above till up to a certain end condition

According to what was shown above, to introduce SVM decision tree into the random forest and to utilize the high accuracy of node splitting algorithm of SVM decision tree improved the adjustment ability of single decision tree in random forest algorithm, and made up for the overfitting flaws of SVM decision tree.

The frame of Random Forest Algorithm based on SVM Decision Tree (SVMDT-RFC) can be described as follows:

**Algorithm2 Random Forest Based on SVM Decision Tree**

**Input**: training set: $S = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$, forest tree scale NTree, the end condition F

**Output**: Weighted Random Forest Based on SVM Decision Tree

(1) Initialize a random forest with NTree SVM decision tree in training set S.

(2) Extract the combination feature, and train the SVM decision tree, adjust the node splitting algorithm in accordance with the distance between classes.

(3) According to the class margin, adjust the weighting of every tree dynamically, and then normalize weighting.

(4) Conduct the next iteration, if which meets the stop condition F, the training stops, and get the last random forest simulator, otherwise go back to Step 2 and go on iteration.

## 3 Experiment and Result Analysis

### *3.1 Experimental Dataset*

After manual screening and filtering, 30,000 microblogs were chosen as experimental data. Six types of people relations were selected, they were Friend Relation, Family Relation Teacher–Student Relation, Classmates Relation, User Relation, Fans Relation respectively. 90% of the data was used as training corpus, and 10% as testing corpus. There existed two pairs of entity relation: Friend Relation and Classmate Relation, Teacher–Student Relation and Fans Relation, in which were some certain fuzzy samples.

### *3.2 Experimental Performance Evaluation Index*

Performance indexes in this experiment adopted used precision, recall rate and comprehensive measurement index F value as evaluation criterions for microblog sentiment polarity classification, and computational formula of the three indexes were as following:

$$Precision = \frac{\text{Number of this kind of microblogs which are judged as correct}}{\text{Number of microblogs which are judged as this kind}} \quad (11)$$

$$Recall = \frac{\text{Number of this kind of microblogs which are judged as correct}}{\text{Number of microblogs which should be judged as this kind}} \quad (12)$$

$$\mathrm{F} = \frac{2 \times Precsion \times Recall}{Precsion + Recall} \quad (13)$$

## 3.3 Experimental Design and Result Analysis

In order to prove the validity of people relation extraction based on Random Forest Classification Algorithm (SVMDT-RFC) of SVM decision tree, the traditional Support Vector Machine (SVM) and Radom Forest Classification were experimented as well, and we evaluated and analyzed the results of three types of methods.

At beginning, three groups of experiments were conducted; Group 1 used SVM classifier, Group 2 RFC classifier, Group 3 SVMDT-RFC which was designed by myself. The result was shown above. Table 1 is the classification effect corresponding to different relations. After analyzing the experiment results, it can be concluded that:

(a) Compared with two traditional classifier, F1 was in a descending order with SVMDT-RFC > RFC > SVM, SVM is worst at classification, which coincidences with its character; while SVMDT-RFC performs best in the condition of multi-classification, moreover, it enhanced the overall effect of people relation extraction and attained the expected results.

**Table 1** Extraction effect on different relation type using different classifier

| Classifier | Relation type | Precision | Recall | F1 |
|---|---|---|---|---|
| SVM | Classmate relation | 65.29 | 64.67 | 64.98 |
| | Friend relation | 66.70 | 66.13 | 66.41 |
| | Family relation | 71.68 | 72.11 | 71.89 |
| | Teacher–student relation | 62.47 | 63.86 | 63.16 |
| | Fans relation | 64.79 | 65.09 | 64.94 |
| RFC | Classmate relation | 67.37 | 66.97 | 66.17 |
| | Friend relation | 68.86 | 67.85 | 68.35 |
| | Family relation | 72.31 | 72.96 | 72.63 |
| | Teacher–student relation | 64.13 | 64.83 | 64.48 |
| | Fans relation | 66.05 | 66.57 | 66.31 |
| SVMDT-RFC | Classmate relation | 73.48 | 72.94 | 73.21 |
| | Friend relation | 73.79 | 73.25 | 73.52 |
| | Family relation | 73.42 | 74.10 | 73.76 |
| | Teacher–student relation | 68.29 | 69.58 | 68.92 |
| | Fans relation | 69.79 | 70.12 | 69.95 |

**Table 2** The language material has people relation extraction effect on people relation extraction in fuzzy samples

| Classifier | Language material length | Precision | Recall | F1 |
|---|---|---|---|---|
| RFC | 0 < L < 25 | 60.27 | 60.86 | 60.56 |
| | 25 < L < 60 | 64.31 | 64.27 | 64.29 |
| | 60 < L < 90 | 68.86 | 68.11 | 68.48 |
| | 90 < L < 140 | 73.85 | 74.06 | 73.95 |
| SVMDT-RFC | 0 < L < 25 | 64.39 | 64.92 | 64.65 |
| | 25 < L < 60 | 71.59 | 71.08 | 71.33 |
| | 60 < L < 90 | 74.64 | 73.94 | 74.29 |
| | 90 < L < 140 | 74.07 | 74.53 | 74.30 |

(b) Under the "family relation" language material condition with barely exist fuzzy samples, SVMDT-RFC had a little progress compared with the two traditional classifiers. While the two pairs of relation, in which language materials, the effect was obvious. The analysis was obtained that SVMDT-RFC can improve the classification effect with the matter of fuzzy sample classification.

In order to study the ability of SVMDT-RFC working on fuzzy samples, another experiment was conducted. The proposed length of microblog language material is L, according to the official rules of microblog, L must be less than 140 words. So on the scale of $0 < L < 140$, four kinds of length were set. And the experiment result is below (Table 2):

According to the result above, we analyzed it and attained the conclusions as below:

(a) Both of the classifiers' effect improved with the language material length getting longer. It can be explained that longer the language material length is, richer the meaning of the word, higher the feature vector, the power is better, which leads to the serious effect.
(b) With all types of language material length, SVMDT-RFC has more obvious effect than RFC, propose the improvement effect as S, that is:

$$S_{25<L<60} > S_{60<L<90} > S_{0<L<25} > S_{90<L<140} \tag{14}$$

As is shown above, SVMDT-RFC is not good at people relation extraction with long length, and it is appropriate to accomplish people relation extraction with medium length and short length language materials.

## 4 Conclusion

In the light of the deficiency of classification in entity relation extraction, combining with the advantages of SVM decision tree and random forest, and SVMDT-RFC was proposed to improve the classification performance, especially for the identification of fuzzy samples. Compared with the traditional classification, SVMDT-RFC can improve the overall effect of people relation extraction, and make the fuzzy samples classification more obvious in language materials, moreover, it fits the people relation extraction with medium length and short language materials. The next period of work will be conducted from two aspects, one is how to descend dimension to feature vector using a better feature selection method; the other is how to optimize SVMDT-RFC so as to improve the efficiency of people relation extraction.

**Acknowledgements** This work was supported by the Key Program of National Natural Science of China (Grant No. 61431008), BUPT Youth Innovation Project (BUPT-2015RC01).

## References

1. H. Cunningham, D. Maynard, K. Bontcheva et al., A framework and graphical development environment for robust NLP tools and applications (ACL, 2002), pp. 168–175
2. P. Pantel, M. Pennacchiotti, Espresso: leveraging generic patterns for automatically harvesting semantic relations, in *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics* (Association for Computational Linguistics, 2006), pp. 113–120
3. B. Han, H.-F. Lin, Characters extraction based on support vector machine, in *Study of the Problem of Computing Technology and Language in China, the Seventh International Conference on Chinese Information Processing* (2007)
4. W. Huang, S. Fan, L. Xiong, M. Zhong, People relation extraction method based on feature selection. Sci. Technol. Eng. (3), 254–259 (2015)
5. L. Karoui, M.A. Aufaure, N. Bennacer, Analyses and fundamental ideas for a relation extraction approach, in *2007 IEEE 23rd International Conference on Data Engineering Workshop* (IEEE, 2007), pp. 880–887
6. Y.-P. Zhu, R. Dai, Text classifier based on SVM decision tree. Pattern Recogn. Artif. Intell. **18**(4), 412–416 (2005)
7. R.E. Schapire, Y. Freund, P. Bartlett et al., Boosting the margin: a new exmethodation for the effectiveness of voting methods. Ann. Stat. 1651–1686 (1998)

# SVM-Based Sentiment Analysis Algorithm of Chinese Microblog Under Complex Sentence Pattern

**Jundong Zhang, Chenglin Zhao, Fangmin Xu and Peiying Zhang**

**Abstract** With the development of Web2.0 era, as local information publishing and social networking platform of Twitter, microblog has become an important medium for people to share and propagate information. Sentiment classification for microblog has also become research hotspot in natural language processing field. By analyzing existing sentiment classification features and complex sentence patterns of microblog and directing at defects of current microblog sentiment classification in feature selection and extraction, this paper combined semantic relation between complex sentences and sentence features of complete sentence based on proposing features of sentence-level fine-grained embedding features and semantic features under complex sentence pattern so as to conduct effective analysis of microblog sentiment features under complex sentence context. It used SVM classification model to conduct comparative experiment, and results indicated that feature selection method proposed in this paper could improve performance of microblog sentiment analysis.

**Keywords** Sentiment analysis · Support Vector Machine (SVM) · Complex sentence pattern · Classification

J. Zhang (✉) · C. Zhao (✉) · F. Xu (✉)
Key Lab of Universal Wireless Communications, Beijing University of Posts and Telecommunications, Beijing, China
e-mail: zdbeijing66@bupt.edu.cn

C. Zhao
e-mail: zdbeijing66@126.com

F. Xu
e-mail: xufm@bupt.edu.cn

P. Zhang (✉)
China University of Petroleum (East China), Qingdao, China
e-mail: zhangpeiying@bupt.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_86

# 1 Introduction

With the development and popularization of mobile Internet and social networking services in the twenty-first century, microblog is increasingly penetrating into life of many netizens. These text messages have enormous commercial and social values. Hence, sentiment analysis technology for people to acquire and mine a great number of sentiment messages has emerged at the right moment and has rapidly become a research hotspot in domestic and overseas natural language processing field.

In terms of research methods, sentiment analysis of microblog is mainly classified into two major types: classification method based on sentiment knowledge and classification method based on machine learning. The former, based on dictionary of sentiment knowledge, conducts weighting by combining sentiment polarity units to obtain text sediment values. In 2005, Wilson [1] et al. used linguistic knowledge to classify semantic sentiment polarity of phrases. In 2007, Xu Linhong [2] et al., by analyzing sentence word and structure, studied 9 semantic features influencing statement sentiment and established sentiment dictionary. In 2015, Wang Zhitao [3] et al. constructed network new word dictionary automatically through statistical information, established multi-level rule sets at different semantic levels. Classification method based on machine learning establishes classification model through annotated training corpus to classify sentiment polarities. In 2002, Pang [4] et al. applied machine learning method to sentiment analysis for the first time; in 2012, Xie Lixing [5] et al. used SVM to conduct strategic analysis of hierarchical structure, did verifying analysis of combination and selection of features; in 2015, Liu Longfei [6] et al. used convolutional neural network of character-level word vector and word-level word vector to do analysis of microblog sentiment tendency.

At present, existing microblog sentiment analysis algorithms mainly have the following deficiencies: (1) neglect semantic and structural analysis of fine-granularity at sentence level; (2) lack sufficient consideration of semantic information under complex context. For (1), according to distribution features of microblog sentences, this paper defined embedding features of sentences; for (2), this paper conducted feature extraction of semantic relation between complex sentences and sentence features of complete sentences, and then made effective analysis of complex sentiment change. Through experiment of original microblog datasets, it verified effectiveness of this algorithm.

# 2 Algorithm Design of Sentiment Analysis

## *2.1 Algorithm Design of Sentiment Analysis*

Figure 1 is the overall design flowchart of SVM-based microblog sentiment analysis algorithm proposed in this paper. First, it conducted grouped manual annotation of original corpus to obtain microblog annotated corpus. The algorithm is divided into

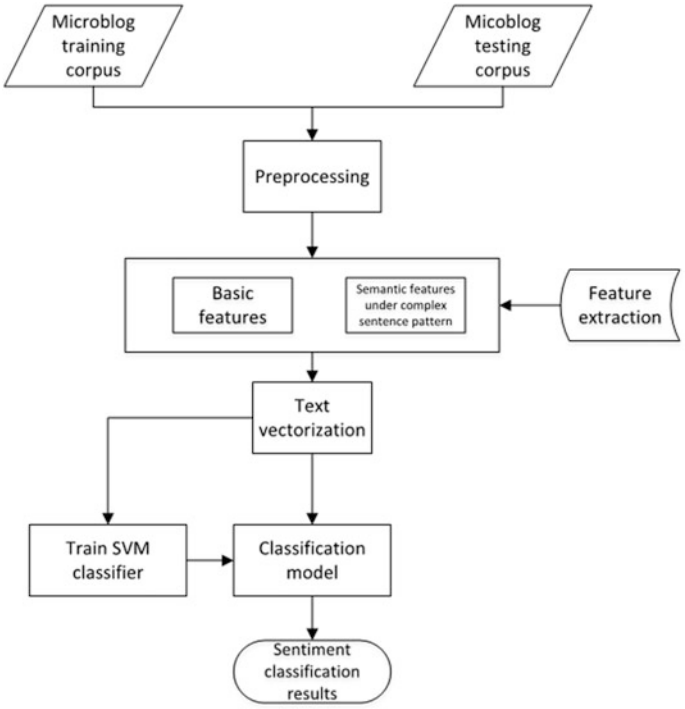



**Fig. 1** Flowchart of SVM-based microblog sentiment analysis algorithm

training phase and testing phase, and it firstly established classification model through microblog training corpus. Preprocessing part of this paper was mainly filtering URL link, # topic label #, @ user, stop word, and others. It did feature extraction of preprocessed microblog text, and then extracted existing basic microblog features and semantic features under complex sentence pattern proposed in this paper. After feature weight calculation of extracted features, it used vector spatial model (VSM) to transform microblog text into feature vectors. It inputs feature vectors into SVM trainer to conduct training of classification model, and then constructed a good classification model. In testing phase, feature vectors of testing corpus were input into SVM classification model which was constructed in training phase, and then final sentiment classification results of microblog were obtained.

## *2.2 Feature of Analysis*

As the design of classifier model has been quite mature at present, this paper laid research emphasis on analysis, selection and extraction of features. Classifier used SVM classification model.

First, according to researches conducted by previous scholars [5, 7–9], this paper concluded features with good effect as basic microblog features and used them in subsequent algorithm basis and experimental comparison. It totally concluded 7 major types of features and subdivided into 13 small class features. According to microblog features and defects of existing features, this paper gradually introduced features of embedded sentences and semantic features under complex sentence pattern as expansion of basic features to improve performance of sentiment classification.

**Features of Embedded Sentence**. Extraction of existing microblog features took microblog text as the entirety to analyze semantic relation features it contained while neglecting semantic feature extraction at sentence level of the text. If analysis was made on the entirety of the text, feature information of positive and negative sentiment words and exclamation mark were used only at discourse level. Hence, directing at problem (1) in introduction section and based on analyzing distribution of numbers of microblog sentences, this paper conducted sentence-level division and defined embedding features at sentence level of more fine-granularity.

According to sentence separator regular expression W = {。|;|!|?|\\.|;|!|\\?| ~}, this paper obtained microblog complete sentence set $S_i$, and this paper made statistics of proportions of microblog messages containing different numbers of sentences, which were concretely shown in Fig. 2.

It is found through statistics of microblog sentence distribution that most microblogs contain multiple sentences, and when number of sentences was 9, occupied proportion was 428413/432701 ≈ 99.009%. Hence, this paper defined upper limit of threshold value as $SC = 9$ for number of sentences. Therefore, this paper defined feature combination model of embedded sentences as shown in Eq. (2).

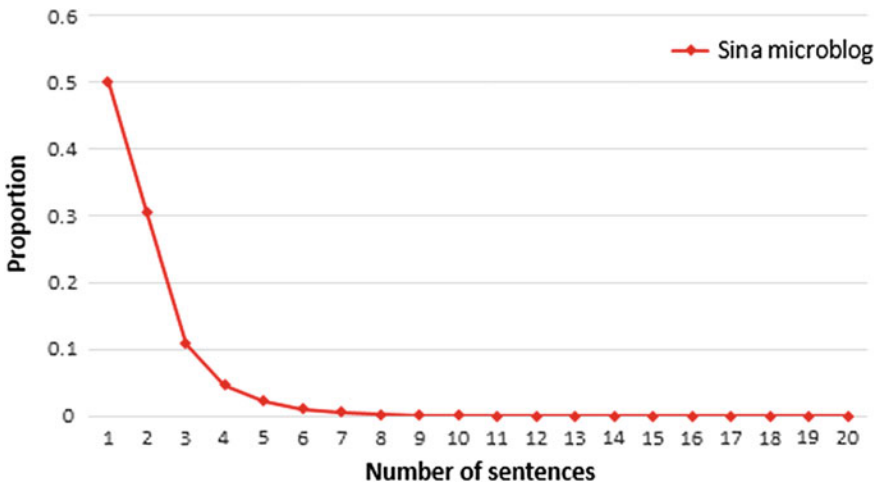



**Fig. 2** Statistics of Sina microblog sentences distribution

$$SF = \{SF_1, SF_2, \ldots, SF_n\}(1 \le n \le SC) \tag{2}$$

Whereby $SF_i$ was feature of embedded sentences defined by this paper, threshold value $SC$ of number of sentences was regarded as combination dimensionality, and combination feature $SF$ at sentence level which described complete microblog was constructed.

**Semantic Features under Complex Sentence Pattern**. By comprehensive analysis of existing basic features, it is found that they did not take full consideration of complex context information as well as multi-sentiment co-occurrence and change in microblog. Directing at problem (2) in introduction section, this paper described sentiment information under complex context from two aspects: (1) Relation between complex sentences. Complex sentence pattern expressed more complex meanings through pause and associated word and so on, usually containing two or more mutually independent subject-predicate structure. (2) Sentence pattern of complete sentence. Final punctuations of one complete sentence usually were exclamation mark "!", question mark "?" whereby "!" was enhancement of tone while "?" which matched with rhetorical question marker in the sentence would similarly enhance intensity of this sentence. This paper defined "semantic features under complex sentence pattern."

*Features of Relation between Sentences*. Sub-clause set of complex sentence was obtained according to regular expression W = {,|, |\\s+}. It is found through research and analysis that relationships between sub-clauses of complex sentences could be determined as adversative, progressive, hypothetical, and concessive relation, all of which would result in change of sentimental intensity of sub-clauses. This paper used semantic dependency analysis technology provided by Language Technology Platform (LTP) researched and developed by Research Center for Social Computing and Information Retrieval of Harbin Institute of Technology to extract xml format contents, semantic dependency analysis described the word through semantic framework borne by the word, it could transcend constraint of syntactic structure at sentence surface layer and directly obtain deep semantic information. Corresponding feature weights were expressed by 4-digit decimal number transformed from 0/1 binary number, specifically shown in Eq. (3):

$$E_{sentence\text{Relation}} = \begin{cases} 0001_2 = 1_{10}, & concessive\ relation \\ 0010_2 = 2_{10}, & hypothetical\ relation \\ 0100_2 = 4_{10}, & adversative\ relation \\ 1000_2 = 8_{10}, & progressive\ relation \end{cases} \tag{3}$$

1 in binary digit in Eq. (3) represented there existed corresponding relation between sub-clauses, 0 expressed nonexistence, and $E_{sentence\text{Relation}}$ expressed the final feature weight.

*Symbol Features of Sentence Patterns*. By matching "!" and "?", we extracted identifications of exclamatory sentence and rhetorical sentence, corresponding

features weights were expressed by 2-digit decimal numbers transformed from 0/1 binary number, specifically shown in Eq. (4):

$$E_{sentence\text{Pattern}} = \begin{cases} 01_2 = 1_{10}, & exclamatory\ sentence \\ 10_2 = 2_{10}, & rhetorical\ sentence \end{cases} \tag{4}$$

*Semantic Features under Complex Sentence Pattern.* This paper integrated sentiment polarity phrases, relation between sentences, symbol features of sentence patters and features of emoticons into features of embedded sentence at sentence level already defined in this paper, and redefined *SF* in Eq. (2) as "semantic features under complex sentence pattern", whereby feature combination element namely features of embedded sentence $SF_i$.

## 3 Experiment and Result Analysis

### 3.1 Experimental Dataset

This paper obtained original corpus used for microblog sentiment analysis through crawler, filtered out about 30,000 original microblogs and selected 5,000 testing corpuses and residual 23,004 ones as training corpus. It conducted manual annotation in way of grouped discussion.

### 3.2 Experimental Performance Evaluation Index

This paper adopted SVM classification model. Performance indexes adopted extensively used precision, recall rate, and comprehensive measurement index F value [10] as evaluation criterions for microblog sentiment polarity classification, and computational formula of the three indexes were as following:

$$Precision = \frac{\text{Number of this kind of microblogs which are judged as correct}}{\text{Number of microblogs which are judged as this kind}} \tag{5}$$

$$Recall = \frac{\text{Number of this kind of microblogs which are judged as correct}}{\text{Number of microblogs which should be judged as this kind}} \tag{6}$$

$$\mathrm{F} = \frac{2 \times Precsion \times Recall}{Precsion + Recall} \tag{7}$$

### 3.3 Experimental Design and Result Analysis

In order to verify effectiveness of SVM judgment method proposed in this paper, we combined three methods in Table 1 to respectively conduct experiments of testing data and then implemented performance index evaluation and analysis of testing results.

The following analyses were made through F values in three experiments in Table 1.

In the last two experiments, after semantic features under complex sentence pattern of this paper were imported, overall F value increased from 0.670 to 0.707. Increasing amplitudes of F values in positive and neutral microblogs were quite obvious, F value of positive microblog increased from 0.706 to 0.762, the effect was the most obvious, while increasing amplitude of F value of negative microblog was small, which, besides being related to corpus that it relied on, mainly resulted from sentiment expression habits of words: in terms of negative sentiment expression, people usually used concise and clear sentence patterns while rarely using complex sentence pattern. Hence, analysis of semantic features in negative microblog under complex sentence patterns didn't have good effect.

In comparison between the method in this paper namely experiment III and experiment I based on sentiment knowledge with superior performance at present, F values of positive, neutral, and negative microblogs in experiment III as well as average F value were elevated to certain degree, but elevating degree didn't reach very obvious range. This was because sentiment analysis method based on semantic rules and expression used in experiment I could make certain analysis of semantics and structures in simple sentence and complex sentence and had corrected polarity judgment to certain degree. However, as the method in experiment I was matching based on sentiment dictionary in essence, its reliance degree on knowledge

**Table 1** Experimental analysis of microblog sentiment analysis

| Experimental method | | Precision | Recall | F |
|---|---|---|---|---|
| Semantic rule + sentiment dictionary + emoticons | Positive | 0.718 | 0.721 | 0.719 |
| | Negative | 0.581 | 0.612 | 0.596 |
| | Neutral | 0.756 | 0.706 | 0.730 |
| | Average | 0.685 | 0.679 | 0.682 |
| Basic microblog features (SVM) | Positive | 0.706 | 0.673 | 0.689 |
| | Negative | 0.584 | 0.561 | 0.572 |
| | Neutral | 0.739 | 0.756 | 0.747 |
| | Average | 0.676 | 0.663 | 0.670 |
| Basic microblog features + semantic features under complex sentence pattern (SVM) | Positive | 0.762 | 0.739 | 0.750 |
| | Negative | 0.606 | 0.598 | 0.602 |
| | Neutral | 0.771 | 0.759 | 0.765 |
| | Average | 0.713 | 0.701 | 0.707 |

dictionary was higher than the method proposed in this paper, which influenced classification performance to certain degree.

SVM classification method proposed in this paper combined current basic microblog features based on SVM classification method, which not only avoided deficiency of traditional SVM classification model in semantic feature extraction at sentiment word level but also processed sentiment analysis phenomena of multi-sentiment co-occurrence and change under complex context in microblog at more fine-granularity level very well.

## 4 Conclusion

Based on combining existing methods and microblog features, this paper mainly made the following improvements: (1) based on statistics of microblog sentence distribution, it defined features of embedded sentences and conducted semantic and structural analysis of microblog at more fine-granularity sentence level; (2) In consideration of complex and rich context of microblog, through extraction of relation between complex sentences and features of sentence pattern and effective analysis of co-occurrence and change of complex sentiment, and based on combining features of embedded sentences, it redefined semantic features under complex sentence pattern and processed sentiment expression in microblog more comprehensively. Through comparative experiments, it verified effectiveness of the method proposed in this paper. The next step of work will continue to optimize feature selection, do feature analysis of elements like repost, comment and time in information of microblog social contact context, expand scope of feature extraction.

**Acknowledgements** This work was supported by the Key Program of National Natural Science of China (Grant No. 61431008), BUPT Youth Innovation Project (BUPT-2015RC01), Shandong Provincial Natural Science Foundation, China (Grant No. ZR2014FQ018).

## References

1. T. Wilson, J. Wiebe, P. Hoffmann, Recognizing contextual polarity in phrase-level sentiment analysis, in *Conference on Human Language Technology and Empirical Methods in Natural Language Processing* (Association for Computational Linguistics, 2005)
2. X. Linhong, L. Hongfei, Calculation of discourse sentiment based on semantic features and subject. Comput. Res. Dev. **44**(z2), 356–360 (2007)
3. W. Zhitao, Y. Zhiwen, G. Bin et al., Chinese microblog sentiment analysis based on dictionary and rule set. Comput. Eng. Appl. **51**(8), 218–225 (2015)
4. B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques. Comput. Sci. 79–86 (2002)
5. X. Lixing, Z. Ming, S. Maosong, Multi-strategy chinese microblog sentiment analysis and feature extraction based on hierarchical structure. J. Chin. Inf. Process. **26**(1), 73–83 (2012)

6. L. Longfei, Y. Liang, Z. Shaowu et al., analysis of microblog sentiment tendency based on convolutional neural network. J. Chin. Inf. Process. **29**(6), 159–165 (2015)
7. Y. Jing, L. Shiping, SVM-Based text words and phrases sentiment analysis. Comput. Appl. Softw. **28**(9), 225–228 (2011)
8. L. Tingting, J. Donghong, Microblog sentiment analysis based on SVM and CFR multi-feature combination. Appl. Res. Comput. **32**(4), 978–981 (2015)
9. Z. Zhilin, Z. Chengqing, A study of chinese microblog sentiment classification method based on diversified features. J. Chin. Inf. Process. **29**(4), 134–143 (2015)
10. C.J.C. Burges, A tutorial on support vector machines for pattern recognition. Data Min. Knowl. Disc. **2**(2), 121–167 (1998)

# A Target Discrimination Method Based on Iterative Manifold SVM

**Chunning Meng, Shengzhi Sun, Heng Xu and Mingkui Feng**

**Abstract** To improve the false reject rate of discriminator for automatic target recognition based on synthetic aperture radar, we propose a new target discrimination method based on a modified manifold support vector machine (SVM). Covariance matrix features which combines texture features and their correlation information are used, and the distinguishability of these features are proved to be good by our experiment. An iterative manifold SVM discriminator is designed to better match the covariance matrix features in the non-euclidean space. The center of the hypersphere in SVM instead of the Karcher mean is selected as the base point by a novel iterative algorithm. Experimental results on RADARST-2 database demonstrate the superiority of the proposed method.

**Keywords** Synthetic aperture radar · Target discrimination · Support vector machine · Manifold

## 1 Introduction

With the advantage of high resolution, high penetration, and full-time and full-weather imaging capability, synthetic aperture radar (SAR) has been widely used in military and civil fields [1]. There are three phases in SAR automatic target recognition (ATR): detection, discrimination, and recognition [2]. In the phase of detection, regions of interest (ROI) including target, natural clutter, and artificial clutter are extracted. Natural clutter is rejected in the phase of discrimination, and

C. Meng (✉) · S. Sun · M. Feng
Department of Electronic Technology, China Maritime Police Academy, Ningbo 315801, China
e-mail: mengchunning123@163.com

S. Sun
Academy of Equipment, Beijing 101416, China

H. Xu
Shanghai Aerospace Electronic Technology Institute, Shanghai 201108, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_87

artificial clutter is removed by recognition. In this paper, we focus on the phase of discrimination, which can be subdivided into feature extraction, feature selection and discriminator design. Lincoln features were widely applied in SAR ATR [3, 4], but single Lincoln feature is not fine enough for rejecting the natural clutter, feature selection methods for the feature combination are required. Therefore, the discrimination performance was limited by the feature selection algorithms [5, 6]. The discriminators used in SAR ATR were one class classification [7], in which case only the label of the target samples were need for the training. Quadratic distance discriminator (QDD) [6], which assumed that the target features vector is subject to Gaussian distribution, was usually used as discriminators, and the target class was discriminated by calculating its probability. Edge-based Support Vector Data Description (SVDD) [8, 9] has good performance when the training data are fewer or biased, in which kernel function is used to map the feature data to higher dimensional space without prior distribution of the data. We designed a region covariance matrix feature for discriminator without feature selection, which combines texture features and the correlation information among these features [9]. In order to avoid the assumption of the prior distribution, One class SVM was used as discriminator, which is equivalent to SVDD when the sigmoid kernel function is used. However, the region covariance matrix feature space was Riemannian manifold while SVM was defined in Euclid space. To cope with this problem, Tuzel et al. [10] transformed the feature data to approximated tangent space at a base point on the manifold, and the base point was localized by the Karcher mean. Recently, similar mapping methods were researched by Yun et al. [11] and Nguyen et al. [12], and how to choose better base point was still not considered. We proposed a novel iterative manifold based SVM discriminator (IMSVM), which works by mapping the data in manifold space to tangent space of more appropriate base point. The experiments show that the proposed discriminator is suitable for the region covariance matrix features of the target and natural clutter.

## 2 Feature Extraction and Discriminator Design

In order to adopt the integration image to improve the calculation speed, the minimum rectangular region Rc was selected from candidate target chip images caught by Adaboost [13]. The region covariance matrix feature was extracted form $R_c$. The region covariance matrix of the target is defined as

$$C_{R_c} = \frac{1}{N-1}\sum_{i=1}^{N}(z_i-\mu)(z_i-\mu)^T \tag{1}$$

where $Z_i$ is the feature vector of the ith pixel in region $R_c$, $N$ is the sum of pixels in $R_c$, $\mu$ is the mean vector of the feature vectors of all pixels in $R_c$. $Z_i$ is defined as

$$z_i = \left[ I(x,y), \frac{\partial I(x,y)}{\partial x}, \frac{\partial I(x,y)}{\partial y}, \frac{\partial^2 I(x,y)}{\partial x^2}, \frac{\partial^2 I(x,y)}{\partial y^2}, \sqrt{\left|\frac{\partial I(x,y)}{\partial x}\right|^2 + \left|\frac{\partial I(x,y)}{\partial y}\right|^2}, \sqrt{\left|\frac{\partial^2 I(x,y)}{\partial x^2}\right|^2 + \left|\frac{\partial^2 I(x,y)}{\partial y^2}\right|^2} \right]^T \tag{2}$$

where $(x, y)$ is the coordinate of the ith pixel, $I(x, y)$ is the intensity of the ith pixel, $\frac{\partial I(x,y)}{\partial x}$ and $\frac{\partial I(x,y)}{\partial y}$ are derivatives of the intensity along the horizontal direction and the vertical direction respectively, $\frac{\partial^2 I(x,y)}{\partial x^2}$ and $\frac{\partial^2 I(x,y)}{\partial y^2}$ are second derivatives of the intensity along the horizontal direction and the vertical direction respectively. As shown in Eq. (2), both the intensity information and the edge information described by the derivatives are contained in the feature vector of each pixel, which shows that $C_{R_c}$ combines various textural features. Moreover, elements of diagonal in $C_{R_c}$ are the variance of each element in $Z_i$, and others describe the correlation information of different elements in $Z_i$. Generally, the clutter chip images have little intensity change while the intensity change of the target region images is big. Therefore, the target can be theoretically discriminated by the region covariance matrix features.

Traditional SVM defined in Euclid space is not suitable for the region covariance matrix features. However, covariance matrix is a positive definite symmetric matrix, which can be regarded as a second-order tensor [14]. The tensor space can be described as a Riemannian manifold [15]. In manifold $M$, select a base point $X \in M$, the tangent space of the base point $X$ is denoted by $T_X M$, then there is a one-to-one correspondence between $M$ and $T_X M$ near the point $X$. Tangent space is isomorphic to Euclid space, so SVM can be used by mapping the covariance matrix features to the tangent space. If the base point determined by the Karcher mean is selected to establish tangent space, the distance relations in the original manifold space can be remained in the tangent space to some extent [10]. However, in this tangent space, only the distances between this base point and any points are constant, not any two points. Therefore, the original manifold can only be described by this tangent space approximately. To solve this problem, we select the center of the hypersphere in SVM as the mapping base point. Because not the distances of any two points, but the distances between the center of the hypersphere and any points is the key to the discriminative rules of SVM. So, in this tangent space, if SVM discriminator is adopt, the structure of the original manifold can be remained. Based on the above analysis, we proposed a novel manifold SVM, whose base point for mapping is sought by a novel iterative algorithm. The iterative process was as follows. First, the initial mapping base point $P_0$ was selected, and the training data were mapped to the tangent space $(T_{P_0} M)$ of $P_0$, then SVM was trained in this space to obtain the center of the suprasphere $a_0$. Second, point $a_0$ was mapped to the original manifold $M$ to get the point $P_1$, and the training data were mapped to the tangent space $(T_{P_1} M)$ of $P_1$, then SVM discriminator was trained in this space to obtain the center of the suprasphere $a_1$. Finally, the iteration was carried out

following the second steps until point $P_n$ coincided with point $a_{n-1}$. The optimal classification criterion was used by IMSVM, so the base point sought by the iteration was optimal. In addition, the experiment showed that the iteration will be converged after 12 times of iterations in most cases.

## 3 Experiments

The experimental data were captured by C band SAR of RADARSAT-2 satellite on March, 2010. The original scene of the images was a port near Tokyo, the area of which was 20 square kilometers. The size of the corresponding single look complex image is $16036 \times 11955$ pixels. Both the azimuth direction and the range direction have a resolution of about 3 meters. 64 ship target chip images and 440 clutter chip images were extracted to verify the proposed method. MATLAB toolbox prtools4.1 and dd_tools1.6.3 were used in our experiment.

Figure 1 shows the histograms of intensity variance and the first order derivative variance of row direction in the covariance matrices. Apparently, the feature values in the clutter chip images are smaller and more concentrated than that in the target chip images, which shows that both the intensity variance and the first-order derivative variance features are distinguished easily. This is due to the fact that there are more sharp edges in the target chip images.

Because a covariance matrix can be regarded as a point in the manifold space. The separability of the target and clutter chip images was analyzed by their locations in the manifold space. Figure 2 show the histograms of the distances from tangent vectors of covariance matrices in the corresponding tangent plane to the different base points. As can be seen from the two figures, the distances corresponding to the target are clearly different from the distances corresponding to the clutter, which indicate that their locations in the manifold space can be distinguished easily. Moreover, the distances between the target curve and the clutter curve are different under different base points. In Fig. 2, the peak distance of the

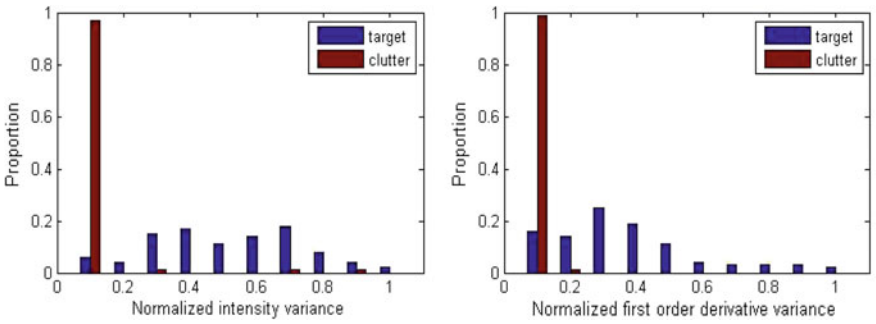



**Fig. 1** Histograms of the intensity variance features and the first order derivative variance of row direction in the covariance matrices

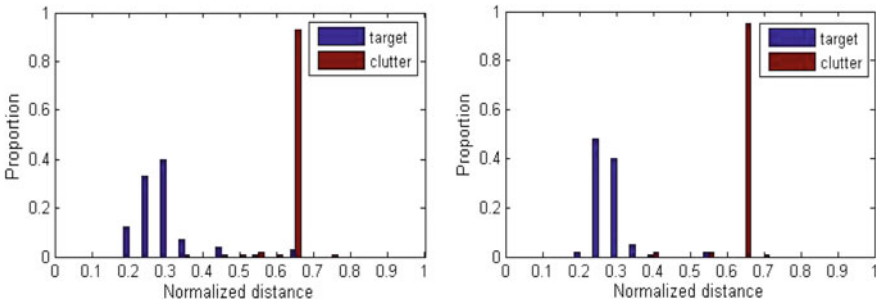



**Fig. 2** Histogram of the distances between the tangent vector of the covariance matrices and the different base point. On the *left*, the base point is defined by the Karcher mean. On the *right*, the base point is localized by IMSVM

target curve are 0.3 and 0.25 in the left and right respectively, and the peak distances of the clutter curves are same (1.15) in the two figures. Therefore, the distance between the target curve and the clutter curve achieved by IMSVM is bigger than the results achieved by the Karcher mean. The results prove that the target can be discriminated easier if selecting the base point by IMSVM.

In order to compare with the existing methods, ROC curve and its area under curve (AUC) were used as the evaluation criteria. Our method was denoted as IMSVM +CovFtr, the method combining QDD and Lincoln features was denoted as QDD +LincolnFtr, the method combining the original SVM and Lincoln features was denoted as SVM+LincolnFtr, the method combining SVM (the Karcher mean is selected to establish tangent space) and the covariance matrix features was denoted as MSVM+CovFtr. The mean ROC curve was obtained by 100 independent experiments. During each experiment, 40 samples were selected randomly as the training set, and the remaining 24 samples were the testing data. All the clutter chip images were used to test during each experiment. Gaussian kernel function was used in SVM, MSVM and IMSVM, and the parameter is set to 0.48. Five Lincoln features, which had good rejection performance, were selected by AUC. As shown in Table 1, the AUC values of these five features were big enough when QDD was used.

Figure 3 shows the ROC curves of the four methods, which describe the change of the detection rates under different false alarm rates. Table 2 shows the AUC values and the false alarm rates of the four methods when the detection rates is 0.95. It is observed that IMSVM+CovFtr outperformed MSVM+CovFtr, which proves that selecting the mapping base point by IMSVM is better than selecting the point by the Karcher mean. The AUC of SVM+LincolnFtr is greater than QDD+LincolnFtr and MSVM+CovFtr have an advantage over SVM+LincolnFtr, which

**Table 1** AUC values of the selected five features when QDD was used

| Lincoln feature | Standard deviation | Peak CFAR | Mean CFAR | Percentage CFAR | Rockwell feature |
|---|---|---|---|---|---|
| AUC | 0.9679 | 0.9348 | 0.9706 | 0.9675 | 0.9503 |

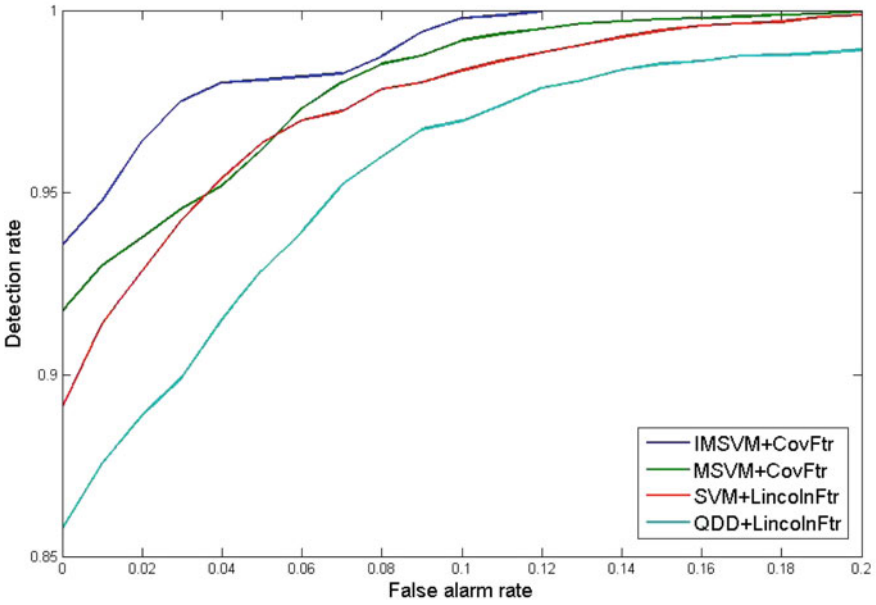



**Fig. 3** Receiver operating characteristic curve for the four methods

**Table 2** Quantitative performance analysis when the detection rate is 0.95

| | AUC | False alarm probability |
|---|---|---|
| IMSVM+CovFtr | 0.9992 | 0.0113 |
| MSVM+CovFtr | 0.9927 | 0.0369 |
| SVM+LincolnFtr | 0.9906 | 0.0365 |
| QDD+LincolnFtr | 0.9863 | 0.0684 |

proof the superiority of SVM and covariance matrix features. The ROC curve of our method is located at the upper left, whose AUC values (0.9992) is the maximum. And the false alarm rate of our method is the minimum when the detection rate is 0.95. The results show that our method is more effective.

## 4 Conclusion

In summary, we propose a new target discrimination method based on IMSVM and region covariance matrix features. Without selecting the features, the error caused by feature selection algorithm was avoided. The optimal mapping base point was sought by iteration in IMSVM algorithm, which is proposed to better match the manifold space of the covariance matrix features. We show improved results over previous works on Radarsat-2 SAR database.

**Acknowledgements** This work is supported by National Natural Science Foundation of China (Grant No. 61401105), Education Sciences Planning Project (DIA150308), Higher Education Reform Project of Zhejiang province (JG2015207) and Technology research project of Ministry of public security of China (2015JSYJC029).

## References

1. C. Malik, R. Malhotra, A survey on various classification methods for SAR images. J. Netw. Commun. Emerg. Technol. **6**(5), 68–72 (2016)
2. L.M. Novak, G.J. Owirka, C.M. Netishen, Performance of a high-resolution polarimetric SAR automatic target recognition system. Lincoln Lab. J. **6**(1), 11–24 (1993)
3. L.M. Novak, S.D. Halversen, G.J. Owirka, M. Hiett, Effects of polarization and resolution on SAR ATR. IEEE Trans. Aerosp. Electron. Syst. **33**(1), 102–115 (1997)
4. S. Chehresa, A. Amirkhani, G.A. Rezairad et al., Optimum features selection for oil spill detection in SAR image. J. Indian Soc. Remote Sens. 1–13 (2016)
5. G. Chandrashekar, F. Sahin, A survey on feature selection methods. Comput. Electr. Eng. **40** (1), 16–28 (2014)
6. G. Gao, An improved scheme for target discrimination in high-resolution SAR image. IEEE Trans. Geosci. Remote Sens. **49**(1), 277–294 (2011)
7. S.S. Khan, M.G. Madden, A survey of recent trends in one class classification. Lect. Notes Comput. Sci. **6206**, 188–197 (2009)
8. D.M.J. Tax, R.P.W. Duin, Support vector data description. Mach. Learn. **54**(1), 45–66 (2004)
9. A. Masjedi, Y. Maghsoudi, M.J. Valadanzoej, A new method for contextual classification of polarimetric SAR data based on combining SVM and MRF. J. Geomatics Sci. Technol. **5**(2), 1–16 (2015)
10. O. Tuzel, Pedestrian detection via classification on Riemannian manifolds. IEEE Trans. Pattern Anal. Mach. Intell. **30**(10), 1713–1727 (2008)
11. Y. Yun, K. Fu, I.Y.H. Gu et al., Visual object tracking with online learning on Riemannian manifolds by one-class support vector machines, in *2014 IEEE International Conference on Image Processing (ICIP)* (IEEE, 2014), pp. 1902–1906
12. H.Q. Nguyen, H.J. Yang, T.N. Thieu, Feature extraction from covariance by using kernel method for classifying polysomnographys data, in *Proceedings of the 9th International Conference on Ubiquitous Information Management and Communication*, vol. 99 (ACM, 2015), pp. 1–7
13. P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in *Proceeding of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1 (2001), pp. 511–518
14. D. Cai, X. He, J. Han, Subspace learning based on tensor analysis. Computer Science Department, UIUC, UIUCDCS-R-2005-2572, Technical Report, May 2005, pp. 1–15
15. X. Pennec, P. Eillard, N. Ayache, A Riemannian framework for tensor computing. Int. J. Comput. Vision **66**(1), 41–66 (2006)

# Scene Character Recognition via Bag-of-Words Model: A Comprehensive Study

**Zhong Zhang, Hong Wang and Shuang Liu**

**Abstract** In this paper, we focus on the feature representation methods under the framework of bag-of-words model (BoW) for scene character recognition. We investigate three kinds of methods: the original BoW methods, the stroke-based methods, and the context-based methods. Specifically, various feature representation methods are introduced, their merits and demerits are explored, and existing problems are discussed. Finally, we empirically evaluate them on several widely used databases (Chars74k, ICDAR2003 and SVHN).

**Keywords** Scene character recognition · Bag of words · Feature representation

## 1 Introduction

Scene text recognition plays an important role in the field of computer vision and pattern recognition with the development of vision-based applications, for instance, translation, human–computer interaction, web content analysis, etc. The common scene-text-extraction system consists of two major components: text detection and text recognition. In the text detection stage, regions containing texts are localized in the whole image, and then based on the cropped text blocks, the scene text is recognized [1]. In recent years, scene text detection has achieved great successes while scene text recognition, especially scene character recognition still demands further consideration because of various interference factors in scene images, such as blur,

Z. Zhang (✉) · H. Wang · S. Liu
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission,
Tianjin Normal University, Tianjin, China
e-mail: zhong.zhang8848@gmail.com

H. Wang
e-mail: hongwang0925@gmail.com

S. Liu
e-mail: shuangliu.tjnu@gmail.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_88

non-uniform illumination and distortion. The fundamental problem for scene character recognition lies in feature representation. The bag-of-words (BoW) model regards an image as a collection of unordered descriptors extracted from local patches, quantizes them into irrelevance codewords, and then calculates a histogram representation for image classification. The BoW model is widely used in many fields, such as human action recognition [2, 3] and image retrieval [4, 5]. The BoW model is one of the key feature representation for scene character images and has achieved promising results on several public databases, including Chars74k [6], ICDAR2003 [7] and SVHN [8]. However, there is few comprehensive study concerning the BoW model for scene character representation. This paper makes such a survey, in which various feature representation methods are introduced, their merits and demerits are explored, and existing problems are discussed. We believe that this work will greatly benefit both beginners and practitioners in the field.

For clarity, we classify the representative feature representation methods under the framework of BoW model into three parts: the original BoW methods, the stroke-based methods and the context-based methods.

- The original BoW methods. These methods learn dictionary from training sets, and encode the local patches in order to obtain the minimum reconstruction error. Numerous coding methods have been proposed, such as hard voting (HV), sparse coding (SC) and locality-constrained linear coding (LLC). Although the BoW model is simple and effective, it ignores the spatial order of local descriptors, which severely limits the descriptive power of the image representation.
- The stroke-based methods. This kind of method is a middle-level representation for scene character recognition. Some methods [1, 9] utilized the intrinsic characteristics of characters and explored the strokes at different scales. Then stroke detectors are trained and the maximal output of detectors are treated as features which contain more spatial information. As a result, this method leads to accurate and robust recognition results.
- The context-based methods. Gao et al. [10] proposed to learn co-occurrence local strokes for robust character recognition by using a spatiality embedded dictionary (SED). The spatial information is incorporated into dictionary directly by reserving a local response region for every codeword. Hence, the features contain more precise spatial information for robust character recognition.

In this paper, we focus on the issue of feature representation under the framework of BoW model. The various feature representation methods are introduced, their advantages and disadvantages are explored, and the main problems and challenges are summarized. Futhermore, we provide a detailed comparison on the Chars74k, ICDAR2003 and SVHN databases. The rest of this paper is organized as follows: in Sect. 2 the three methods are over viewed, and the experimental study is presented in Sect. 3. Finally, we draw the conclusion in Sect. 4.

## 2 Review Methods

### 2.1 Original BoW Methods

We denote $F = [f_1, f_2, \ldots, f_N] \in \mathbb{R}^{M \times N}$ to be the $N$ $M$-dimensional descriptors extracted from a training image, $L = [l_1, l_2, \ldots, l_K] \in \mathbb{R}^{M \times K}$ to be a dictionary with $K$ codewords and $A = [a_1, a_2, \ldots, a_N] \in \mathbb{R}^{K \times N}$ to be the corresponding coding of the $N$ features. The dictionary $L$ is used to code the descriptors with different coding schemes. This section reviews three typical coding methods, i.e., HV, SC, and LLC.

⋄ The HV [11]. In this coding scheme, each code has only one nonzero element. That is to say, one descriptor needs single nearest codeword to reconstruct it with solving the following constrained least-squares fitting problem:

$$\begin{aligned} &\arg\min_A \textstyle\sum_{i=1}^{N} \|f_i - La_i\|_2^2 \\ &s.t. \|a_i\|_{l_0} = 1, \|a_i\|_{l_1} = 1, a_i \geq 0, \forall i \end{aligned} \tag{1}$$

where $\|a_i\|_{l_0}$ is the $L_0$ norm of $a_i$, $\|a_i\|_{l_1}$ is the $L_1$ norm of $a_i$ and $\|s\|_{l_2}$ denotes the $L_2$ norm of $s$. Specifically, $\|a_i\|_{l_0} = 1$ means only one non-zero element in $a_i$. $\|a_i\|_{l_1}$ defined as the sum of the absolute values of its elements. In practice, the only nonzero element is found by optimizing the object function. The principle of HV is uncomplicated, yet HV will lead to a significant quantization error.

⋄ The SC. Different from the HV, the SC [12] presented the encoding process in a soft manner. Specifically, there are a small group of nonzero elements in SC with respect to the following objective function:

$$\arg\min_A \sum_{i=1}^{N} \|f_i - La_i\|_2^2 + \beta \|a_i\|_{l_1} \tag{2}$$

The parameter $\beta$ controls the sparsity of $a_i$. The advantages of the algorithm not only possess a unique solution but also achieve much less quantization loss. However, in SC the regularization term of $l_1$ norm is not smooth. The SC procedure may chooses widely varying codewords for similar descriptors to favor the sparsity, thus losing the correlations among codes.

⋄ The LLC. To ameliorate the SC method, Wang et al. [13] proposed the LLC method. They focus on the locality rather than the sparsity because the locality must result in the sparsity. The criteria of LLC are listed in the following:

$$\begin{aligned} &\arg\min_A \textstyle\sum_{i=1}^{N} \|f_i - La_i\|_2^2 + \gamma \|d_i \odot a_i\| \\ &s.t. \|a_i\|_{l_1} = 1, \forall i \end{aligned} \tag{3}$$

In Eq. (3), $\odot$ is a multiplication of the corresponding elements in the two matrices. $d_i \in \mathbb{R}$ is the Euclidean distance between the descriptor and the codeword. According to the experimental results on the public databases, LLC is the best in the three

methods. On the one hand, LLC is more effective and more accurate for feature representation than HV. On the other hand, in contrast with SC, LLC captures the correlations of the similar descriptors and it has a analytic solution while SC needs complex optimization process.

## 2.2 Stroke-Based Methods

The stroke-based methods are middle-level representation under the framework of BoW model and capture more structure information than those original BoW methods. The whole process of stroke-based method can be divided into two layers.

The first layer is stoke bank building. As Gao et al. stated in [1], they first manually label the key points for 62 classes (0–9, a–z, A–Z ) and select the discriminative stokes by quoting the extension factor. Second, they collect the positive and negative samples with the number of $N_{pos/neg}$ to train the stroke detectors. Then, the HOG and the multi-class linear SVMs are integrated to obtain the $N_m = \sum_{i=1}^{N_c} N_{stoke,ci}$ stoke detectors in the stoke bank. $N_{stoke,c_i}$ denotes the number of stoke detectors in the class $c_i$.

The second layer is the feature representation. Given a training image, the maximal value of the corresponding detector is denoted as $a_{c_{i,j}}$. Then the confidence vector for the given training image is expressed as follows:

$$f = (a_{c_1,1}, \ldots, a_{c_1,N_{stroke,c_1}}, \ldots, a_{c_N,N_{stork,c_{N_c}}}) \tag{4}$$

We can obtain more spatial information from images, but artificially annotating key points is time-consuming and nontrivial. In general, these methods are obviously superior to other methods, such as the spatial pyramid (SP) [14]. In SP, images are separated into a sequence of increasingly finer blocks and the feature vectors in different blocks are catenated together to obtain the final vector. We will list the detailed experimental results in Sect. 3.

## 2.3 Contex-Based Methods

Although Shi et al. [15] and Bai et al. [9] have proposed to choose stokes randomly, such as DMSDR (Discriminative Multi-Scale Stroke Detector Based Representation) and DSEDR (Discriminative Spatiality Embedded Dictionary Learning Based Representation), the stroke-based methods are still complex and the results are easily affected by the choosing strokes. Gao et al. [16] proposed another method to learn co-occurrence local strokes for robust character recognition by using a spatiality embedded dictionary (SED). The SED combines every codeword with a

particular response region. The spatial information is incorporated into SED directly via preserving the homologous local response region of the codeword. Hence, the features contain more precise spatial information for robust character recognition.

## 3 Experimental Study

### *3.1 Databases and Settings*

In this section, three public scene text character databases are employed for empirical evaluation: Chars74k, ICDAR2003 and SVHN. The Chars74k [6] database contains 62 classes (0–9, a–z, A–Z) collected from natural scene images. In the experiment, we split training and testing database as in [6, 15, 17], i.e., in every class, we randomly select 30 images in which 15 images are used for training and the rest of images are used for testing. The ICDAR2003 [7] database contains 6185 training samples and 5430 testing samples in 62 classes (0–9, a–z, A–Z). The samples cover different conditions in natural scene, for instance, nonuniform illumination, distortion, and complex backgrounds. The SVHN [8] database derives from street view images. As stated in [15], we choose the cropped digits which contains 73257 training samples, 26032 testing samples, and 531131 additional less difficult digits for training for experiment.

For the stroke-based methods, the images are scaled to $W * H = 32 \times 64$. The positive and negative samples for stoke detectors are normalized to $16 * 16$ and the ratio number of positive and negative samples $N_{pos/neg}$ is set to 2. We adopt the 36-dimensional HOG features with bin number 9, cell size $8 * 8$ pixels and block size $2 * 2$. For the context-based methods, the images are also normalized to $W * H = 32 \times 64$ and the images are parted into $8 * 16$ blocks. The 36-dimensional HOG features are extracted from every blocks with cell size $2 * 2$ pixels. Besides, although Wang et al. [13] and Yang et al. [12] have testified that LLC is the best method in the the original BoW methods, it is still inappreciable when compare with the advanced stroke-based methods and context-based methods. Considering the limited space, we will not repeat them here.

### *3.2 Experimental Discussion*

The results of our experiments on the ICDAR2003 and the Chars74k databases are shown in Table 1. From the results, several conclusions can be drawn. First, the DSEDR algorithm achieves the highest classification accuracies on the ICDAR2003 and Chars74k databases. Second, the performance of the DMSDR method is better than other methods except the SED and the DSEDR and it shows that the context-based methods are more effective than the stroke-based methods. Third, although

**Table 1** Character recognition results on ICDAR2003 and Chars74k databases

| Algorithm | Date | ICDAR2003 | Chars74k |
|---|---|---|---|
| HOG+NN [18] | 2011 | 51.5 | 58 |
| GHOG [17] | 2013 | 76 | 62 |
| HOG+SVM | 2005 | 77 | 62 |
| Conv-Co-HOG [19] | 2014 | 81 | - |
| Stroke bank [1] | 2014 | 79.5 | 65.7 |
| DMSDR [15] | 2015 | 81.7 | 66.1 |
| SED [16] | 2014 | 82.0 | 67.1 |
| DSEDR [15] | 2015 | 82.6 | 71.8 |

**Table 2** Character recognition results on SVHN database

| Algorithm | Date | SVHN |
|---|---|---|
| HOG+SVM | 2005 | 80.01 |
| MTSM [15] | 2015 | 89.54 |
| K-MEANS [8] | 2011 | 90.6 |
| ConvNet/MS/L4/Padded [20] | 2012 | 91.55 |
| DMSDR | 2015 | 91.62 |
| DSEDR | 2015 | 92.12 |

the Conv-Co-HOG possess a little superiority than stroke bank methods, it is very complicated. From the Table 2 we can see that the DMSDR and DSEDR outperform HOG+SVM more than 13%. All of them utilize SVM as the classification method and the difference of the three methods lies in the feature representation. Although both the HOG+SVM and MTSM (Mixtures-of-Parts Tree-Structured Model) use HOG as descriptor, MTSM contains more structure information than HOG+SVM. Moreover, unlike DMSDR and MTSM which need the key points or parts labels, DSEDR learns the spatiality embedded codewords (stroke detectors) automatically.

## 4 Conclusion

In this paper, we have analyzed various feature representation methods under the framework of BoW model, including their motivations and mathematical representations. In addition, we have discussed their relations in theory, and empirically evaluated their performance. In the future, we will extend our comparison to include more advanced methods.

**Acknowledgements** This work is supported by National Natural Science Foundation of China under Grant No. 61401309, and No. 61501327, Natural Science Foundation of Tianjin under Grant No. 15JCQNJC01700, and Doctoral Fund of Tianjin Normal University under Grant No. 5RL134 and No. 52XB1405.

## References

1. S. Gao, C. Wang, B. Xiao, C. Shi , Z. Zhang, Stroke bank: a high-level representation for scene character recognition, in *International Conference on Pattern Recognition (ICPR)* (2014), pp. 2909–2913
2. Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, Action recognition using context-constrained linear coding. IEEE Signal Process. Lett. **19**(7), 439–442 (2012)
3. Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, Attribute regularization based human action recognition. IEEE Trans. Inf. Forensics Secur. **8**(10), 1600–1609 (2013)
4. L. Zheng, S. Wang, Q. Tian, Coupled binary embedding for large-scale image retrieval. IEEE Trans. Image Process. **23**(8), 3368–3380 (2014)
5. L. Zheng, S. Wang, Q. Tian, Fast image retrieval: query pruning and early termination. IEEE Trans. Multimedia **17**(5), 648–659 (2015)
6. T.E. de Campos, B.R. Babu, M. Varma, Character recognition in natural images, in *International Conference on Computer Vision and Applications* (2009), pp. 273–280
7. S.M. Lucas, A. Panaretos, L. Sosa, A. Tang, S. Wong, R. Young, ICDAR, robust reading competitions, in *International Conference on Document Analysis and Recognition* (2003), pp. 682–687
8. Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A.Y. Ng, Reading digits in natural images with unsupervised feature learning, in *NIPS Workshop Deep Learning Unsupervised Feature Learn* (2011) pp. 1–5
9. X. Bai, C. Yao, W. Liu, Strokelets: a learned multi-scale mid-level representation for scene text recognition. IEEE Trans. Image Process. **25**(6), 2789–2802 (2016)
10. S. Gao, C. Wang, B. Xiao, C. Shi, W. Zhou, Z. Zhang, Scene text recognition by learning co-occurrence of strokes based on spatiality embedded dictionary. IET Comput. Vision **9**(1), 138–148 (2015)
11. G. Csurka, C. Bray, C. Dance, L. Fan, Visual categorization with bags of keypoints, in *ECCV International Workshop Statistical Learning in Computer Vision*, vol. 1 (2004), pp. 1–22
12. J. Yang, K. Yu, Y. Gong, T. Huang, linear spatial pyramid matching using sparse coding for image classification, in *IEEE Conference on Computer Vision and Pattern Recognition* (2009), pp. 1794–1801
13. J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, Y. Gong, Locality-constrained linear coding for image classification, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2010), 3360–3367
14. H. Tatsuya, U. Yoshitaka, Y. Yuya, K. Yasuo, Discriminative spatial pyramid, in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 1617C1624 (2011)
15. C. Shi, S. Gao, M. Liu, C. Qi, C. Wang, B. Xiao, Stroke detector and structure based models for character recognition: a comparative study. IEEE Trans. Image Process. **24**(12), 4952–4964 (2015)
16. S. Gao, C. Wang, B. Xiao, C. Shi, W. Zhou, Z. Zhang, Learning co-occurrence strokes for scene character recognition based on spatiality embedded dictionary, in *IEEE International Conference on Image Processing (ICIP)*, 5956C5960 (2014)
17. C. Yi, X. Yang, Y. Tian, Feature representations for scene text character recognition: a comparative study, in *The 12th International Conference on Document Analysis and Recognition* (2013), pp. 907–911

18. K. Wang, B. Babenko, S. Belongie, End-to-end scene text recognition, in *International Conference on Computer Vision* (2011), pp. 1457–1464
19. B. Su, S. Lu, S. Tian, J.H. Lim, C.L. Tan, Character recognition in natural scene using convolutional co-occurrence HOG, in *International Conference on Pattern Recognition (ICPR)* (2014), pp. 2926–2931
20. P. Sermanet, S. Chintala, Y. LeCun, Convolutional neural networks applied to house numbers digit classification, *International Conference on Pattern Recognition (ICPR)* (2012), pp. 3288–3291

# Analysis of EEG Signal Evoked by Passive Movement and Motor Imagery

**Zhangliang Chen, Qilian Liang and Baoju Zhang**

**Abstract** In recent years, research based on the Motor Imagery (MI) and physical training of stroke rehabilitation therapy has become a hot topic. Among them, brain–computer interface (BCI)-based electrical stimulation (ES) and MI synchronously induced electroencephalo-graph (EEG) analysis is one of the research directions. This paper integrates MI with ES, and introduces the steady-state somatosensory evoked potential (SSSEP) on the MI, and then fuses features. The purpose is to achieve better accuracy, reducing training time, and enhancing target body imagine attention. We use ES and MI synchronous-evoked EEG analysis for rehabilitation training strategy in the process of synchronization. Based on the time–frequency diagram and brain topographic map, we make analysis on event-related desynchronization (ERD), time–frequency characteristic of the steady-state evoked potentials (SSEP), energy characteristics, and distribution regularity of brain regions of each task model. We apply the short-time Fourier transform (STFT), common spatial pattern (CSP), and filter bank common spatial pattern (FBCSP) to classify task pattern recognition on the basis of support vector machine. The results show that based on the lead less the FBCSP has certain superiority, the classification accuracy of ES versus ES&MI achieved 95% or more.

**Keywords** Stroke · Brain–computer interface · Functional electrical stimulation · Motor imagery · Event-related desynchronization · Common spatial pattern · Support vector machine

Z. Chen (✉) · Q. Liang · B. Zhang
Department of Electrical Engineering, University of Texas at Arlington, Arlington, TX 76019-0016, USA
e-mail: zhangliang.chen@mavs.uta.edu

Z. Chen · Q. Liang · B. Zhang
Tianjin Normal University, Tianjin 300387, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_89

## 1 Introduction

In recent years, the passive movement and motor imagery (MI) synchronously induced electroencephalo-graph (EEG) analysis has become a commonly used method for the treatment of stroke [1]. However, due to the synchronous process produced by different body motions induced by EEG physiological information and its limitations, the analysis methods of the specific changes of the EEG signals are unclear. Therefore, this paper will discuss the range of low electrical stimulation (does not produce movement of stimulation intensity, ES) exercise imagine EEG change at the same time, to seek the better rehabilitation training strategy in the process of synchronization.

In experiments we use noninvasive BCI system, electrodes cap collected EEG signals, in total of 64 lead lists [2]. The rest of this paper is organized as follows. In Sect. 2, we introduce procedure of experiment. In Sect. 3, we use time–frequency diagram and brain topographic map to analyze EEG signals. Section 4 is characteristics analysis of EEG signals. Section 5 is simulations of classification accuracy and Sect. 6 concludes this paper.

## 2 Experimental Scheme

In order to guarantee the success rate of experiments, experimenters are all trained. We collect five volunteer subjects of EEG signals. It includes three men and two women whom aged between 22 and 25 and all of them are right-handed. The whole experiment includes four paradigms. In the first paradigm we ask volunteers to image their right-hand clenched fists action and collect its EEG. In the second paradigm volunteers not only image fisting action but also are added ES on their right hands at the same time, and we collect their EEG. In the third paradigm we add ES on volunteers right hands only and collect their EEG. The fourth paradigm is a rest paradigm which means volunteers can relax and we collect their EEG. Each paradigm has 40 single repeat in 10 s, and the whole experiment has a total of 160 EEG data samples. Task model of four paradigms is same and shown in Fig. 1, divided into four periods, a total of 10 s.

First period is preparation period. There is a stationary white circle at the central of computer screen and lasting 2 s. It means experiment start, and volunteers are keeping relaxed state of no action. Second period is prompt period. In this period the white circle disappears and a red circle appears. Lasting 2 s which means prompting volunteers to prepare since imagine action thinking task is about to begin. For the second and the third paradigms, we give volunteers right hands ES at the same time. Third period is the imagination period. In this period red circle disappears but a right arrow appears and lasting 4 s. This signal asks volunteers to

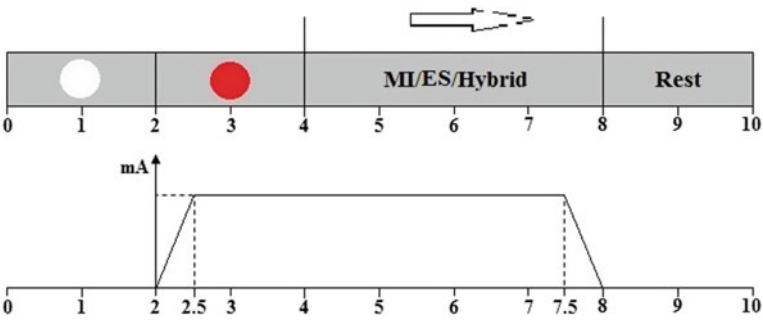



**Fig. 1** Task model diagram

focus on their right hands ES and imagine the action of their right hands clench fists. The fourth period is for recovery with the arrow disappears. A stationary word REST appears. Volunteers keep resting state [3]. Each paradigm of MI experiments includes 40 times single task and a total of 40 * 10 s = 400 s.

## 3 Motor Imagery Characteristic Analysis

### *3.1 Time–Frequency Diagram*

Figure 2 shows the average time–frequency diagram of experimental stage of the six volunteers in the C3, CZ, C4 lead lists in four experimental models. The horizontal axis is time and the vertical axis is frequency. It can be seen from the figure that the brain electrical signal evoked by the MI shows obvious event-related desynchronization (ERD) features and the obvious steady-state somatosensory

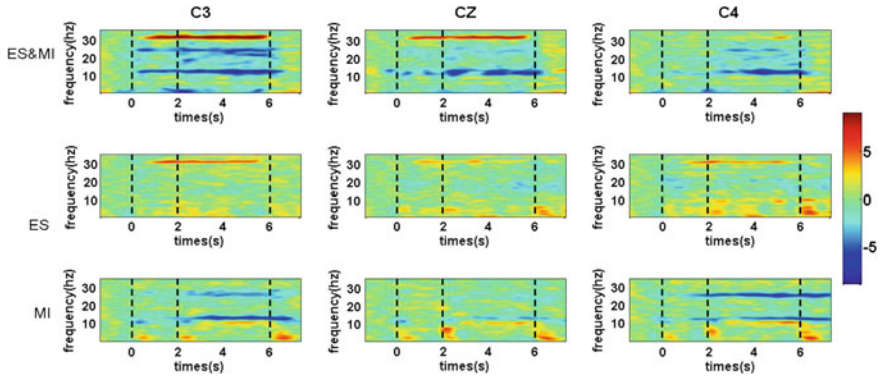



**Fig. 2** C3 C4 CZ lead lists time–frequency diagram; blue represents the energy drops; red represents the energy rise

evoked potential (SSSEP) characteristics of EEG signals evoked by the external electrical stimulation in the frequency domain [4, 5]. At the beginning of the 4 s MI, some significant energy attenuation phenomena occur in the 0.5 s, which appear in the alpha band (8–9 Hz) and the beta band (15–17 Hz). The ERD characteristic of alpha frequency band is more obvious and lasts about 3S time than the beta frequency band. When the volunteers are in a resting state, no ERD or SSSEP features are in the EEG.

It can be seen from the diagram that when volunteers imagine the right fist action the ERD characteristics of the corresponding C3 lead list is very obvious. CZ lead list which reflects adjacent brain regions reaction also appears ERD phenomenon but is not so obvious at lead list C3. When the volunteers imagine right fist action at the same time is added outside ES, C3 lead list appears SSSEP phenomenon at 31 Hz in addition to strong external ERD characteristics.

### *3.2 Brain Topographic Map*

Alpha band ERD phenomenon has the characteristics of good duration and intensity as well as more conducive to the follow-up observation and analysis, so we use alpha band ERD phenomenon for the main analysis.

We use window function to intercept ERD and SSSEP characteristics at obvious band of time–frequency diagram in order to select the best data for making brain topographic maps. We get ERD and SSSEP single-dimensional feature data of time–frequency within the window, and average each lead list separate time–frequency superimposed data, gaining feature data for brain map needed. Using the above method we make MI only, ES only, and ES&MI brain topographic maps of ERD and SSSEP characteristics as shown in Fig. 3.

From the brain topographic maps seen in the MI and ES experimental paradigms, it is obvious that the left side of the brain power attenuation is more obvious than the right side, with the obvious ERD and SSSEP contralateral dominant phenomenon [6].

## 4 MI Feature Extraction Based on Common Spatial Pattern

In this section, we use filter bank common spatial pattern (FBCSP) to analyze the characteristics of EEG. After EEG signal acquisition the whole 8–32 Hz is divided into six bands. The common spatial pattern (CSP) is used to extract the features and add the 1 Hz frequency doubling [7]. Then we use the feature selection algorithm to automatically select the recognition force of the dual band CSP and the corresponding characteristics.

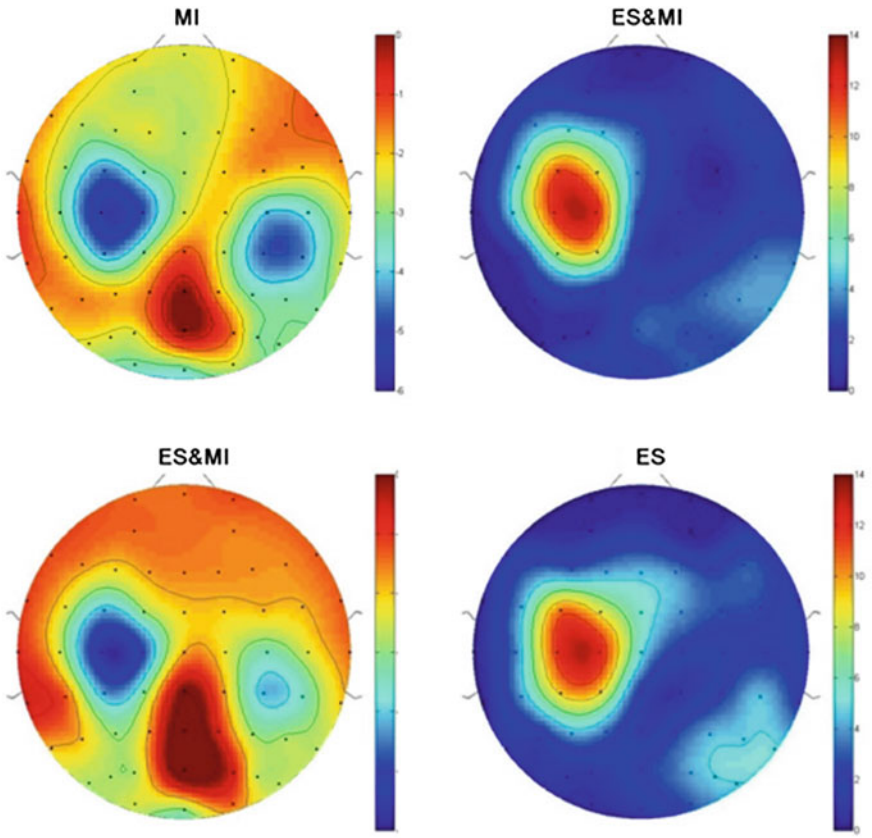



**Fig. 3** ERD and SSSEP brain topographic maps of three experimental paradigms

$$X_{CSP} = W^T * X, \tag{1}$$

where $X_{CSP}$ is obtained from original EEG signal $X$ after filtering, $W$ is the filter matrix, each column vector $w_j \in W^{N\times N}(j = 1, \ldots, N)$ is a filter, and $A = (W^{-1})^T$ is spatial pattern matrix, where each column $a_j \in A^{N\times N}(j = 1, \ldots, N)$ is a vector space model.

Assuming that the first-class matrix is represented by $X_l$, the second-class matrix is expressed as $X_r$, and the dimension is the lead number * sampling number. The spatial covariance matrix can be expressed as

$$R_l = \frac{X_l X_l^T}{trace(X_l X_l^T)} \qquad R_r = \frac{X_r X_r^T}{trace(X_r X_r^T)}, \tag{2}$$

where $T$ denotes transpose, and trace *trace(X)* is the sum of diagonal elements.

$X_l$ and $X_r$ can be linearly expressed as

$$X_l = [C_l C_c] \begin{bmatrix} S_l \\ S_c \end{bmatrix} \qquad X_r = [C_r C_c] \begin{bmatrix} S_r \\ S_c \end{bmatrix}, \tag{3}$$

where $S_l$ is the signal source for the first class, $S_r$ is the signal source for the second kinds of needles, and $C_l$ and $C_r$ are the corresponding feature vectors.

The purpose is to establish a spatial filters $F_l$ and $F_r$, which can be used to extract the components of the signal sources $S_l$ and $S_r$:

$$S_l = F_l * X_l \qquad S_r = F_r * X_r, \tag{4}$$

where $F_l$ and $F_r$ are the first and second types of action corresponding to the spatial filter, through the use of data training.

$R_l$ and $\mathrm{R_r}$ are the mean of covariance normalization::

$$\overline{R_l} = \frac{1}{N_l} \sum_{i=1}^{N_l} R_l(i) \qquad \overline{R_r} = \frac{1}{N_r} \sum_{i=1}^{N_r} R_r(i), \tag{5}$$

where $N$ is the number of experiments in the training data.

Then use the principal component analysis

$$\overline{R} = \overline{R_l} + \overline{R_r} = U_0 \Lambda U_0^T, \tag{6}$$

where $U_0$ is a feature vector and $\Lambda$ is the characteristic value.

Thus the whitening matrix is expressed as

$$P = \Lambda^{1/2} U_0^T. \tag{7}$$

Thus can be expressed as

$$Y_l = P \overline{R_l} P^T \qquad Y_r = P \overline{R_r} P^T. \tag{8}$$

Based on principal component analysis,

$$Y_l = U \Lambda_l U^T \qquad Y_r = U \Lambda_r U^T. \tag{9}$$

It can be shown that the eigenvalues of $Y_l$ and $Y_r$ have diagonal matrix $\Lambda_B = I$, where $I$ is the unitary matrix. So the final corresponding to the spatial filter $F_l$ and $Fr$ can be expressed as

$$F_l = U_l^T P \qquad F_r = U_r^T P. \tag{10}$$

## 5 Simulation

In this section, we use support vector machine (SVM) to classify feature vector of each training sample, and then draw the correct classification rate [8].

At the beginning we use all lead lists EEG data to classify feature vectors: the classification of ES versus REST as a result of the control group and the classification of ES&MI versus ES as the experimental group. Comparing these two results as shown in Table 1, we obtain more than 95% accuracy rate in ES&MI versus ES, which means our purpose is achieved. However, taking into account the full lead lists lead to patients discomfort, we use nine lead lists (C3, C4, CZ and their up and down lists) for classification as shown in Table 2.

We can see the accuracy rate of the classification of the nine lead lists. ES&MI versus ES has been greatly reduced. In order to improve the accuracy rate we use fast Fourier transform (FFT) to process EEG signal. After that we add amplitude at 31 Hz and frequency doubling at 62 Hz for feature extraction and classification. Each bandwidth is 1 Hz. Table 3 shows the classify accuracy rate.

Although the correct rate has improved, it is still lower than full lead lists. So finally on the basis of FFT we use FBCSP for feature extraction and classification. The accuracy is summarized in Table 4.

At this time the accuracy rate has reached more than 95%, infinitely close to the accuracy rate of the full lead lists.

**Table 1** CSP all lead lists feature extraction and classification accuracy rate

| CSP all lists classification accuracy rate | MI versus REST (%) | ES&MI versus ES (%) |
|---|---|---|
| Volunteer 1 | 100 | 97.5 |
| Volunteer 2 | 98.75 | 100 |
| Volunteer 3 | 100 | 100 |
| Volunteer 4 | 100 | 100 |
| Volunteer 5 | 100 | 80 |
| Average | 99.75 | 95.5 |

**Table 2** CSP 9 lead lists feature extraction and classification accuracy rate

| CSP 9 lists classification accuracy rate | MI versus REST (%) | ES&MI versus ES (%) |
|---|---|---|
| Volunteer 1 | 98.75 | 87.5 |
| Volunteer 2 | 88.75 | 87.5 |
| Volunteer 3 | 100 | 98.75 |
| Volunteer 4 | 98.75 | 93.75 |
| Volunteer 5 | 98.75 | 68.75 |
| Average | 97 | 87.25 |

**Table 3** CSP+FFT 9 lead lists feature extraction and classification accuracy rate

| CSP+FFT 9 lists classification accuracy rate | MI versus REST (%) | ES&MI versus ES (%) |
|---|---|---|
| Volunteer 1 | 98.75 | 90 |
| Volunteer 2 | 88.75 | 93.75 |
| Volunteer 3 | 100 | 96.25 |
| Volunteer 4 | 98.75 | 95 |
| Volunteer 5 | 98.75 | 83.75 |
| Average | 97 | 91.75 |

**Table 4** FBCSP+FFT 9 lead lists feature extraction and classification accuracy rate

| FBCSP+FFT 9 lists classification accuracy rate | MI versus REST (%) | ES&MI versus ES (%) |
|---|---|---|
| Volunteer 1 | 100 | 93.75 |
| Volunteer 2 | 91.25 | 98.75 |
| Volunteer 3 | 100 | 100 |
| Volunteer 4 | 98.75 | 95 |
| Volunteer 5 | 100 | 88.75 |
| Average | 98 | 95.25 |

## 6 Conclusions

First, we designed the movement pattern of MI, and the experimental type of ES was combined with MI, which constituted the four kinds of task paradigms. Second, we performed spatial filtering and denoising data preprocessing for the original EEG data. Third, we made qualitative analysis on the phenomenon of SSSEP induced by ES and the phenomenon of ERD induced by MI. Fourth, we used the CSP and FBCSP two classifications of spatial pattern algorithm add FFT algorithm. On the basis of that we combined CSP+FFT and FBCSP+FFT, two better feature extraction algorithms. Finally, we used SVM pattern recognition method to classify EEG data from fourtypes of tasks and compared the classification accuracy of MI versus REST with ES&MI versus ES. The classification accuracy is more than 95%. So we can conclude that MI&ES which can provide high classification accuracy on the basis of good patient comforts and convenient experiments is a better rehabilitation method in stroke rehabilitation strategy.

## References

1. G. Pfurtscheller, C. Brunner, A. Schlögl et al., Mu rhythm (de) synchronization and EEG single-trial classification of different motor imagery tasks. NeuroImage **31**(1), 153–159 (2006)
2. C. Sun, *Visual P300-Speller Induced ERP Study Based on Three Dimensional Coding Sequence*, Tianjin University (2011)
3. G. Pfurtscheller, C. Nerper, Motor imagery and direct brain-computer communication. Proc. IEEE **89**(7), 1123–1134 (2001)
4. M. Takahashi, K. Takeda, Y. Otaka et al., Event related desynchronization-modulated functional electrical stimulation system for stroke rehabilitation: a feasibility study. J. Neuroeng Rehabil. **9**(1), 1 (2012)
5. L. Salisbury, J. Shiels, I. Todd et al., A feasibility study to investigate the clinical application of functional electrical stimulation (FES), for dropped foot, during the sub-acute phase of stroke–A randomized controlled trial. Physiotherapy Theor. Pract **29**(1), 31–40 (2013)
6. W. Yi, *Research on EEG Feature Recognition Technique Based on Compound-limbs Motor Imagery* (Tianjin University, Tianjin, 2012)
7. X. Liu, *Movement Pattern Recognition and Trajectory Prediction Based on EMG-KJA Neural Musculoskeletal Model of Lower Limb*, Tianjin University (2012)
8. L. Jiao, *Functional Target Recognition and Classification* (Science Press, Beijing, 2010)
9. J.R. Wolpaw, Brain-computer interface: singals, method, and goals, in *Proceedings of the 1st international IEEE EMBS conference on Neural Engineering* (2003) pp. 2–22

# Part IX
# Digital Image & Video Processing

# Research on Fingerprint Image Enhancement Based on Improved Gabor Filtering Algorithm

**Xi Gong, Meijuan Yu and Yankai Liu**

**Abstract** With the gradual increase of accuracy requirement in the digital human identity authentication, biometric technology has been developing rapidly, where fingerprinting is most widely used. In this paper, according to the shortcoming in the actual fingerprint image acquisition, research on fingerprint image enhancement is based on improved Gabor filtering algorithm. Finally, by comparing the experiments, we prove that improved algorithm proposed in this paper can better achieve the minutiae enhance conducive accurate extraction of fingerprint features.

**Keywords** Image enhancement · Gabor filter · Fingerprint recognition

## 1 Introduction

With the continuous improvement of the degree of social information, identity authentication has gradually become an important part of people's lives. The traditional identity authentication is based on token authentication [1] and knowledge authentication [2], such as documents verification and secret key verification. But the highly digitization of society requests the higher level of accuracy and security for the identification, resulting in that the traditional methods of identification cannot meet this requirement. The public needs a more secure and reliable method of identity authentication that cannot be forged and cracked easily, which becomes the catalyst for the rapid development of the biometric technology.

X. Gong
Tianjin Normal University Jingu College, Tianjin 300387, China
e-mail: gx8000@126.com

M. Yu (✉) · Y. Liu
College of Computer and Information Engineering, Tianjin Normal University, Tianjin 300387, China
e-mail: amy_1121@163.com

Y. Liu
e-mail: ooaround@foxmail.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_90

Today, biometric technology has been widely used. The common recognition technology based on the biological properties [3–5] includes fingerprint recognition, vein recognition, iris recognition, DNA recognition, and voice recognition, in which the fingerprint recognition is the most widely used biometric technology because of its strong stability, accuracy, acquisition ability, and security. However, in the practical application, directly collected fingerprint images are generally with the noise and pollution, which means the quality is not so satisfactory. Therefore, before the feature extraction and recognition for the fingerprint image, a series of pretreatment need to be conducted. This paper is mainly about the improvement technology of fingerprint image.

## 2 The Improvement of Fingerprint Image Enhancement Technology

The extraction of minutiae in the fingerprint lies on the quality of fingerprint. After refinement of the fingerprint image, the ridge can be detected easily and the feature points can be extracted accurately. Fingerprint image enhancement is to deal with the imperfect image by a certain algorithm to make the ridge and feature clear, avoiding the false features to guarantee the accuracy and reliability of the feature extraction. There are many fingerprint enhancement algorithms, like Laplacian algorithm [6], the histogram equalization algorithm [7], and median filtering [8]. But these algorithms are not so ideal because they ignore the ridge orientation and frequency.

A. *Fingerprint Image Enhancement Algorithm Based on Gabor Filter*

Gabor function has good spatial and frequency domains [9]. To enhance the fingerprint image using Gabor function makes full use of the ridge direction. Direction filter is adopted to overcome the limitation of traditional image filtering algorithm, which can realize the intelligence enhancement according to the fingerprint ridge orientation, reaching a good effect of fingerprint enhancement. To filter the image by the real part of function gains the smoothing image effect. The imaginary part is an effective edge detection operator [10, 11].

Fingerprint is composed of the ridge line and valley line. Compared to the whole fingerprint image, the fingerprint texture changes greatly. But in the small region of fingerprint image, the ridge orientation and frequency characteristics are very close to each other. So the Gabor wavelet can be used for fingerprint image filtering enhancement.

The use of Gabor filtering can enhance the image intelligently from the aspects of fingerprint ridge orientation and the vertical orientation of ridge. In the direction of the fingerprint ridge, the ridge can be enhanced in the orientation field of one point to make up for the fracture and other defects of ridge in the image. In the vertical direction of the ridge, Gabor functions with the characteristics of interphase can conduct the oscillation enhancement to the image in frequency field.

Gabor filtering algorithm flow is as follows:

(a) To obtain the direction of fingerprint.
(b) To divide the fingerprint image into N * N blocks and calculate the frequency information of the ridge in each block.
(c) To conduct the Gabor filtering transform for each point in the fingerprint image.

B. *Research on Improved Algorithm of Gabor Filtering*

The fingerprint enhancement algorithm based on Gabor filtering can enhance the fingerprint image. But this approach is to divide the fingerprint image into several blocks and represents the direction and frequency of each pixel point with the direction and frequency of each block. So there exists a kind of block effect. In this paper, I will take the advantage of the method of least squares to get a balanced frequency $f$, which is the nearest value to each frequency. Then this balanced frequency will be used to enhance the image. Concrete steps are as follows:

(a) To divide the fingerprint image into 16*16 blocks without overlapping sub-blocks.
(b) To put the pixel $(i,j)$ in the sub-block as the center, sub-block fingerprint direction as the short axis. Accordingly, the size of a block is a rectangular area of 31*16.
(c) For each pixel $(i,j)$, to calculate the gray discrete signal X(k) in the perpendicular area according to the formula:

$$\mathrm{X(k)} = \frac{1}{w}\sum_{d=0}^{w-1} \mathrm{G(u}, v) \qquad \mathrm{k} = 0, 1, 2, \ \ldots, \ l-1 \tag{1}$$

$$\mathrm{u} = i + (d - \frac{w}{2})\cos\theta(i,j) + (k - \frac{l}{2})\sin\theta(i,j) \tag{2}$$

$$\mathrm{v} = j + (d - \frac{w}{2})\sin\theta(i,j) + (\frac{l}{2} - k)\cos\theta(i,j), \tag{3}$$

where $u$ and $v$ are the coordinate values in the rectangular window which is parallel to the point $(d,k)$ in the short axis and $\mathrm{G(u}, v)$ is the gray value of pixel $(\mathrm{u, v})$. Put the image retrieval gray value as the gray value of each point in the long axis direction of rectangular area, consisting of a discrete signal X(k) which is like a two-dimensional sine wave. Calculate the average value $\mathrm{A_0}$ of the sine wave peak distance A, namely the small ridge frequency $f_0$.

(d) To find the frequency $f_1, f_2, \ \ldots, f_n$ of local small block, the solving formula for filtering frequency $f$ is

$$(f - f_1) + (f - f_2), \ \ldots, \ (f - f_n) = 0. \tag{4}$$

## 3 Result and Analysis

For the original image, in this section, we will use Gabor to conduct the contrast experiment between the intelligent enhancement image and the enhancement image based on the improved algorithm. The original image and its histogram are as shown in Fig. 1. The fingerprint image enhancement effect based on Gabor filtering and gray-level histogram is as shown in Fig. 2. The fingerprint images enhancement effect based on an improved algorithm and gray-level histogram are as shown in Fig. 3.

We can see from Fig. 1 that the gray levels of original fingerprint image are very close to each other, namely the gray value of image is not uniform and the fingerprint patterns are not clear. The structure of image is fuzzy, which is going against with the extraction and recognition of fingerprint features. We can see from Fig. 2 that the patterns are clearer in the fingerprint image enhanced by Gabor

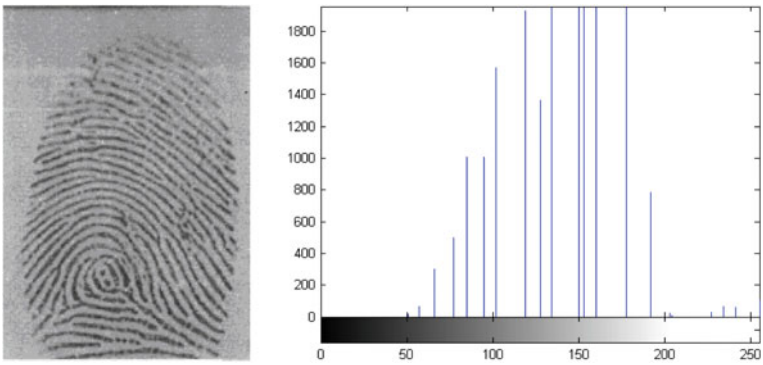



**Fig. 1** Original image and its histogram

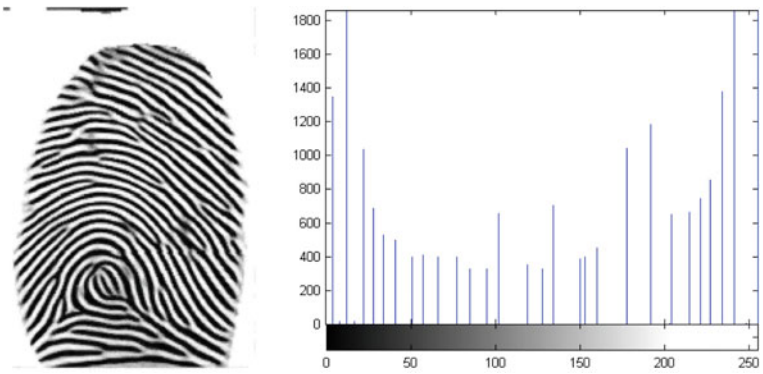



**Fig. 2** Fingerprint image enhancement effect based on improved algorithm and gray-level histogram

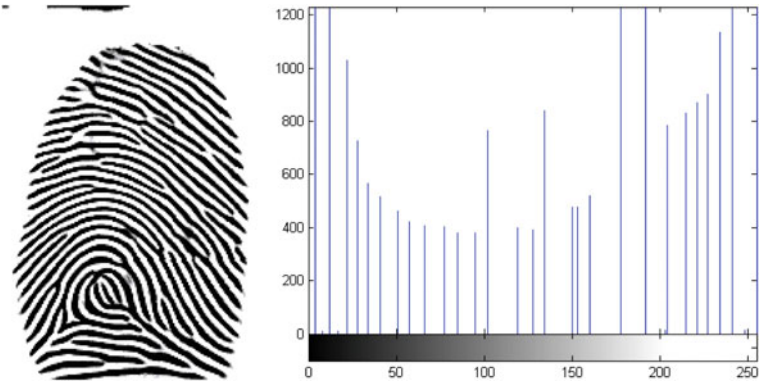



**Fig. 3** Fingerprint image enhancement effect based on improved algorithm and gray-level histogram

filtering. The distribution of image gray is more uniform. From Fig. 3, we can see that the details in the fingerprint image handled by the improved algorithm are clearer and the contrast ratio is improved. Compared with Fig. 2, the improved algorithm can better remove the block effect and the edge is more smooth after treatment.

## 4 Summary

This paper is mainly about the research on the fingerprint image enhancement technology based on Gabor filtering improved algorithm. First, the in-depth study of image enhancement algorithm based on Gabor function is conducted. This algorithm has a certain enhancement effect on the fingerprint image, making it clearer. But there exists block effect in this algorithm and it may produce some false ridge line and valley line. According to the shortcomings, we improve this algorithm and conduct the contrast experiment. The result shows that the image enhancement based on improved algorithm could highlight the details of the image and make the edge more smooth. The effect has been improved.

## References

1. C. Zhe, W. Yigang, Research and realization of the indentity authentication mechanism based on USBKey. Comput. Appl. Softw. **28**(2), 284–286 (2011)
2. W. Juan, H. Qi, Y. Fei, A user-centric indentity management and authentication system for mobile Internet. J. Shandong Univ. **47**(11), 12–17 (2012). (Science Edition)
3. P.K. Harmer, D.R. Reising, M.A. Temple, Classification selection for physical layer security augmentation in cognitive radio networks, in *IEEE International Conference on Communications, Budapest* (2013), pp. 2846–2851

4. Z. Li, K. Liu, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2014)
5. Z. Li, Y. Chen, NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. **2016**, 48 (2016)
6. M. Dusio, M.A. Olsena, C. Busch et al., Fingerprint sample quality assessment via ridge line count using Laplacian of Gaussian edge finding, in *2nd International Workshop on Biometrics and Forensics: 2nd International Workshop on Biometrics and Forensics (IWBF 2014)*, 27–28 March 2014, (Valletta, Malta, 2014), pp. 1–6
7. F.M. Hanoon, Contrast fingerprint enhancement based on histogram equalization followed by bit reduction of vector quantization. Int. J. Comput. Sci. Netw. Secur. IJCSNS 11(5), 116–123 (2011)
8. V. Sasikala, V.L. Prabha, A swarm intelligence based feature selection approach for efficient fake and real fingerprint classification using semisupervised learning. Int. J. Chem. Chem. Eng. **5**(1), 61–86 (2015)
9. X. Liu, Z. Li, Information-centric mobile ad hoc networks and content routing: a survey. Ad Hoc Netw. (2016). In Press, Corrected Proof
10. N. Moritsuka, K. Katsura, K. Matsuoka et al., Decadal sustainability of spatial distribution of soil properties in a paddy field as a fingerprint reflecting soil-forming factors and field management. Soil Sci. Plant Nutr. **61**(3), 516–527 (2015)
11. Z. Li, L. Song, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016). In Press, Corrected Proof

# Comparative Research on Different Color Systems in Three-Dimensional Color Gamut

**Zhendan, Liyan and Malingyun**

**Abstract** This paper introduces the Munsell color system and Pointer color system, and uses the algorithm to calculate the three-dimensional color gamut volume. In addition, we have tested a Toshiba display and got the data in three-dimensional color gamut, then hold the gamut that eyes can perceive as a denominator, and calculate the color gamut coverage respectively. This study provides some reference value for extending color gamut in three dimensions later.

**Keywords** Munsell color system · Pointer color system · The color gamut coverage

## 1 Introduction

In computer graphics, the color gamut is a complete subset of color [1]. Color gamut refers to the color range regions of the number that the equipment can express, namely the color range that all kinds of screen display; a printer or printing equipment can express. In the real world, the visible color spectrum makes up the largest color space, and the color space contains all the colors we can see.

The eye experiences the colorful world by the light, and human perceives the color of the object by the vision. To achieve this process requires three necessary factors, which are light, the object, and human vision. The paper studied the perceived color gamut based on the characteristics of the human eye, which we call it the color gamut that human eye can perceive. In this paper, we use this to evaluate the three-dimensional color gamut in different gamut color systems [2].

Zhendan (✉) · Liyan · Malingyun
Tianjin Normal University, Tianjin 300387, China
e-mail: zhendan0827@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_91

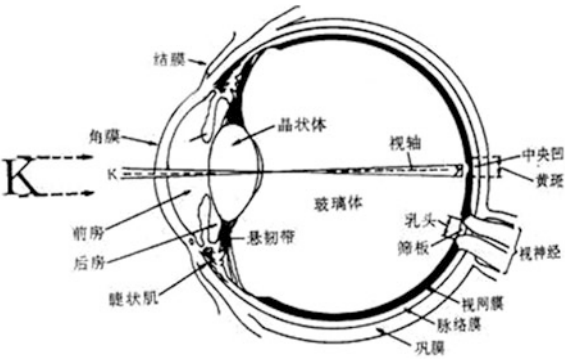



**Fig. 1** The human eye structure diagram

## 2 The Human Visual System

The human visual system is the world's best image processing system, but still not perfect. Human visual system to perceive the image is neither linear nor uniform, and it cannot perceive any change in the image. The human eye structure diagram is shown in Fig. 1.

### *2.1 Human Visual Cells*

Human visual cells include cone cells and rod cells. Rods are sensitive to light intensity, i.e., having high brightness sensitivity, and response to the wavelength of light is extremely low. The cones have strong selectivity toward different wavelengths of light, especially for red, green, and blue, but it does have this feature in bright situations.

### *2.2 Spectral Sensitivity of the Human Eye*

The human eyes have different sensitivities to various wavelengths of radiation; a large number of experiments about normal vision observer show that in a relatively bright environment, human vision is most sensitive to green of wavelength 550 nm: and In dark environments, the human visual is most sensitive to yellow-green of wavelength 500 nm.

In both cases, the human eye spectral sensitivity curve is shown in Fig. 2.

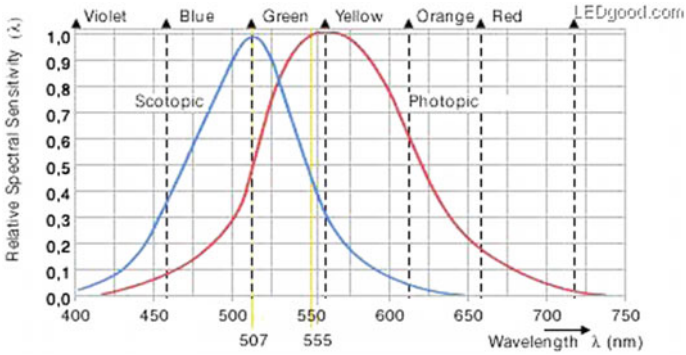



**Fig. 2** The human eye spectral sensitivity curve

## *2.3 Resolution of the Human Eye*

The resolution of the human eye is the human eye's ability to distinguish details on the scene. Generally, we call the minimum angle between the closest two black dots or white dots that human eye can distinguish on the objects is the limit angle of resolution. Its reciprocal is the resolution of the human eye. Factors that affect the resolution of the human eye contain light, color, relative contrast, and the velocity of the object and so on.

## *2.4 The Inertia of Human Eye*

Generally, the feeling of the human eye brightness changes lagging behind changes in actual brightness, and the persistence of vision characteristics collectively referred to as visual inertia. Visual inertia is one of the important characteristics of the human eye, which describes the relationship between subjective brightness and light effects of time.

# 3 Munsell Color System

Color system is divided into two categories: one is based on three color characteristics to classify, namely lightness, hue, and saturation; the other is based on the three primary colors, said that any color can be given in three colors mixed together by a certain percentage. The former is called a monochromatic classification color system, which is a composition of the standard color sample series, named after them in arranged sequentially.

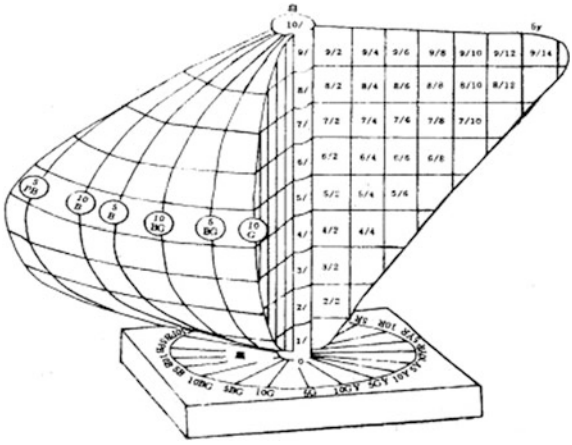


**Fig. 3** Munsell color stereoscopic

Alert Munsell, an artist, and as a professor of Normal University Art Institute Art Department in Massachusetts, wanted to create "a rational way to describe color", which will use decimal format rather than the name of the color. Then he can use it to teach his students. So the Munsell system is generated [3].

Munsell system is a color system established by Munsell of the twentieth century, who is an American artist. It uses a three-dimensional model to express all the features of variety colors, which contains brightness, hue, and saturation. It is the first one to split the brightness, hue, and saturation into the senses, which is separated into sensory uniform and independent dimensions, to describe colors in three-dimensional system.

Munsell color stereoscopic is shown in Fig. 3.

In the figure, the central axis represents the brightness level of neutral color in achromatic black and white series, black at the bottom and white at the top, called the brightness value of Munsell. It sets ideal white as 10, and ideal black as 0. The brightness values of Munsell range from 0 to 10 and divided into 11 visual equidistant levels.

We got the point cloud of Munsell color system in the CIELAB color space based on the data, and it is shown in Fig. 4.

In this paper, we use the alpha algorithm to calculate the three-dimensional color gamut volume of Munsell system [4], which evaluates to V = 797611.698. The color gamut map of Munsell in three dimensions is shown in Fig. 5.

## 4 Pointer Color System

The spectrum reflected from an object, which we saw in our real life, is very complicated, and not constituted by a single wavelength. M.R. Pointer made a series of experiments, and the result was published in the form of a paper in 1980,

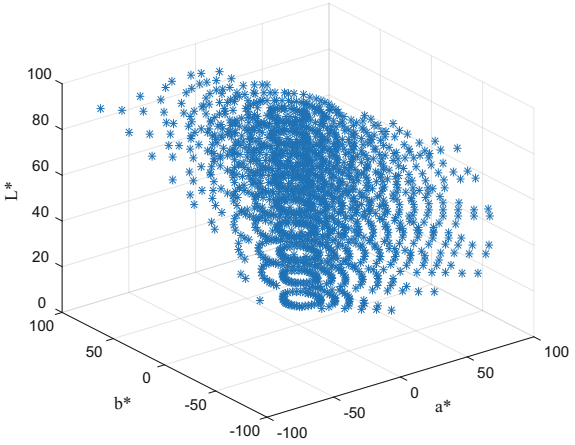



**Fig. 4** The point cloud of Munsell color system

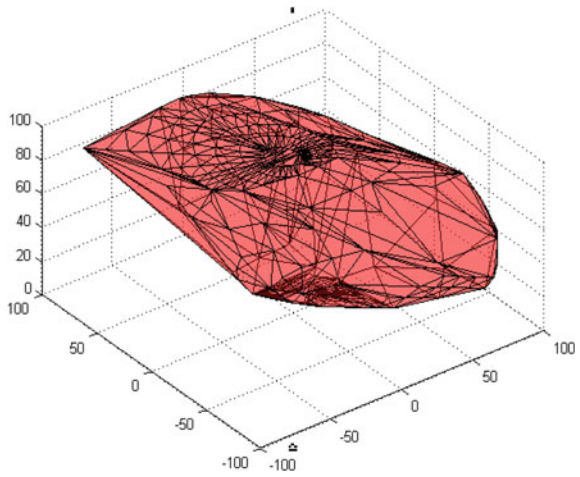



**Fig. 5** The color gamut map of Munsell in three dimensions

named "The Gamut of Real Surface Colours" [5]. Mr. Pointer obtained the color samples of 4089 real objects and got their color coordinate in the xy space. In brief, most of the natural colors that human eye can see are in this range. It is shown in Fig. 6.

The purpose of display technology is to achieve the reduction of the true color of the object. Pointer gamut represents the largest color gamut of target color gamut [6–9]. It also said that we do not need to implement every color of the visible spectrum, and we only need to restore the color inside Pointer color gamut, so Pointer gamut has very important practical significance.

The point cloud of Pointer color system in the CIELAB color space is shown in Fig. 7.

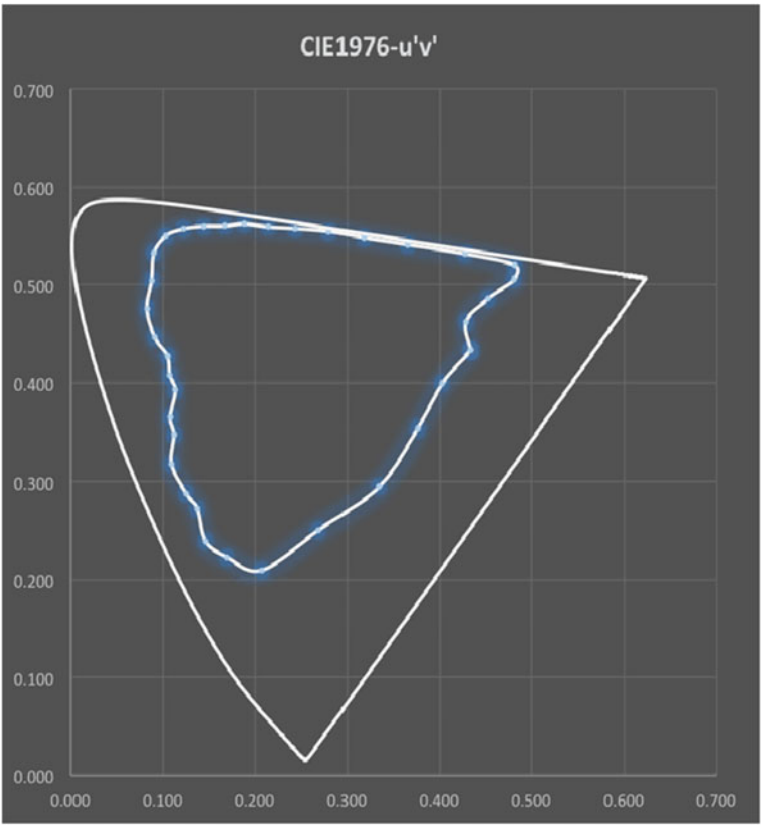



**Fig. 6** The color gamut of Pointer in two dimensions

**Fig. 7** The point cloud

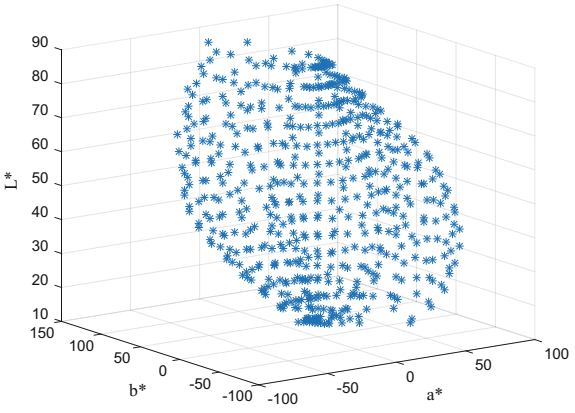
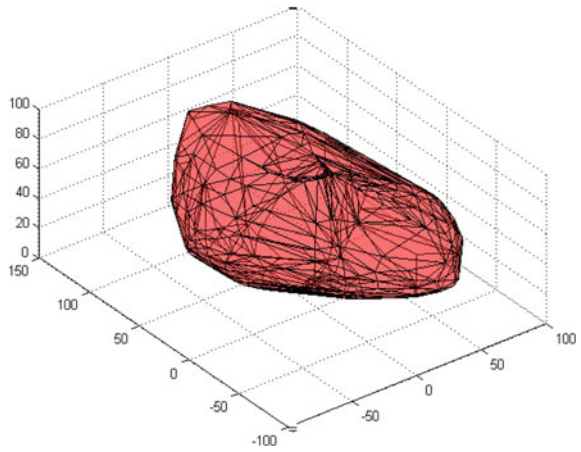


**Fig. 8** The color gamut map




We also use the alpha algorithm to calculate the three-dimensional color gamut volume of Pointer system, which evaluates to V = 662312.824. The color gamut map of Pointer in three dimensions is shown in Fig. 8.

## 5 Experiments

We got to test the data of Toshiba display, and then use the color gamut coverage to describe the color effect of the display quantitatively [1]. The formula of color gamut coverage is as follows:

$$P = \frac{V_e}{V} \times 100\%. \tag{1}$$

In this formula, P represents the color gamut coverage, $V_e$ represents the color gamut of the tested display, and V represents the visually perceptible three-dimensional color gamut.

Based on the data of each system, we plotted the three-dimensional color gamut coverage figure of Pointer system, Munsell system, and the Toshiba display by MATLAB. We take Munsell system and Pointer system as examples, and the figures are shown in Figs. 9 and 10.

According to three-dimensional color gamut coverage map, we obtained the color gamut and the color gamut coverage of Pointer, Munsell, and Toshiba (HDMI) display calculated by the algorithm [10, 11]. The calculated results are shown in Table 1.

As can be seen from the table, the color gamut coverage of the display measured in three-dimensional color space is 0.292, which is larger than the Pointer and smaller than the Munsell.

**Fig. 9** Pointer in spectrum

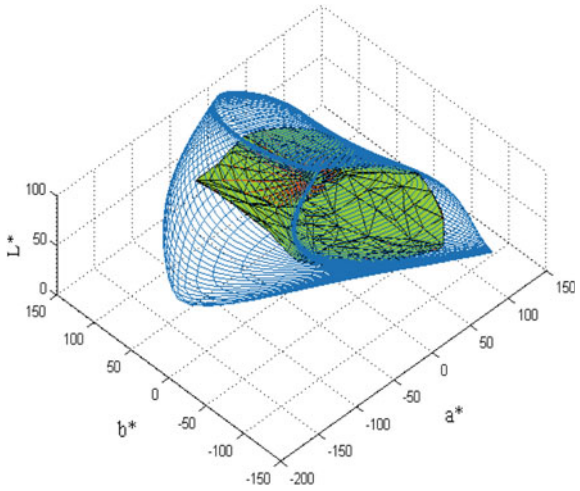



**Fig. 10** Munsell in spectrum

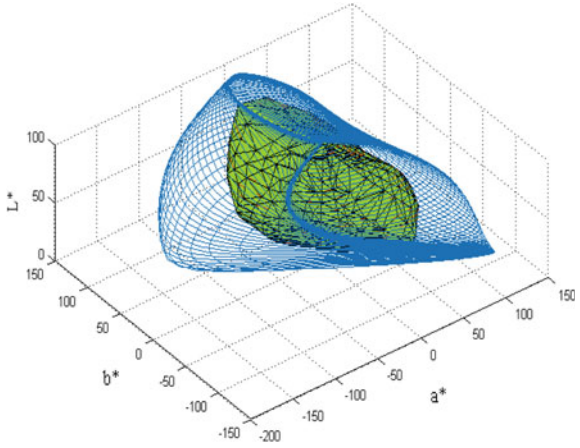



**Table 1** The color gamut volume and color gamut coverage

| Color gamut data points | Color gamut volume (V) | Color gamut coverage |
|---|---|---|
| Munsell Lab | 797611.698 | 0.347 |
| Pointer Lab | 662312.824 | 0.288 |
| The test data of Toshiba (HDMI) | 671729.166 | 0.292 |

## 6 Conclusions

In this paper, we calculated the three-dimensional color gamut volume of Pointer, Munsell, and the tested display. Moreover, we calculated the three-dimensional color gamut coverage of two color systems and Toshiba display, respectively, which provides a clear contrast. This study shows a certain reference value for the later widen gamut research of display.

## References

1. Q. Liu, Color Gamut and Color Gamut Coverage. Telev. Technol. **11**, 49–52 (2004)
2. Y. Li, X. Yan, G. Li, Research on three-dimensional gamut model of high definition television. Acta Scientiarum Naturalium Universitatis Nankaiensis. **44**(1), 30–35 (2011)
3. A.S. Burns, B.J. Cohen, N.E. Kuznetsov, Munsell color system in fundamental color space. Color Res. Appl. **15**(1), 29–51 (1990)
4. J. Sun, X. Yan, G. Li, The fast three-dimensional color gamut volume algorithm based on the alpha form. Telev. Technol. **38**(21), 29–31 (2014)
5. M.R. Pointer, Gamut of real surface colours. Color Res. Appl. **5**(3), 145–155 (1980)
6. R. Miwa, Y. Manabe, N. Yata, 5–5 Color gamut expansion method of 3D projector using multibands projection, in *Ite Winter Annual Convention* (2013), pp. 5–5
7. J.R. Manders, L. Qian, High efficiency and ultra-wide color gamut quantum dot LEDs for next generation displays. J. Soc. Inform. Display **23**(11), 523–528 (2015)
8. G. Braun, K. Spaulding, Method for evaluating the color gamut and quantization characteristics of output-referred extended-gamut color encodings, in *IS&T/SID Tenth Color Imaging Conference*, New York (2002), pp. 99–105
9. D.S. Park, S.Y. Choi, H.Y. Lee, et al., A new wide-gamut RGB primary set and efficient color encoding methods for ultra-high definition television(UHDTV). Proc. IEEE **101**(1), 18–30 (2013)
10. A. Willert, M. Flaspohler, A.C. Hubler, Calculating a color gamut border using a grid approach. conference on colour in graphics, imaging, and vision. Soc. Imaging Sci. Technol. **2004**(1), 156–160 (2004)
11. S. Kuang, X. Wang, Z. Wang, The tetrahedral meshing algorithm to describe the color gamut based on regional segmentation. Packag. Eng. **35**(5), 126–130 (2014)

# A Projected Algorithm Based on the Convex Hull of the Triangle in Three-Dimensional Color Space

**Yan Li, Dan Zhen and Lingyun Ma**

**Abstract** With the rapid development of display technology, modern display equipment future development goal is diversified and widen the color gamut Miwa et al. (2013) *ITE Winter Annual Convention*, pp. 5–5: [1], Manders, Qian (2015) J Soc Inform Display 23(11), 523–528: [2]. In order to find a unified approach to evaluate different types of displays, more and more scholars tend to the three-dimensional color space to evaluate the color effect on the display. However, due to the three-dimensional color, gamut of display shows irregular shape, so the calculation is relatively complicated, and it always needs a long time. In this paper, the convex hull projection algorithm of three-dimensional color gamut based on the triangular surface has been proposed; it has been greatly improved computation time.

**Keywords** Display technology · Triangular surface · Convex hull projection · Three-dimensional color gamut · Volume calculations

## 1 Introduction

At present, the research about color gamut of the device mainly includes the fields of display and color printing. Studies have shown that, compared with the traditional two-dimensional color gamut, the color gamut of a display device in a three-dimensional color space can be more comprehensive to compare and analyze the color characteristics of the display [3]. Regarding the quantification of display three-dimensional color gamut, the key is to calculate the volume of the color gamut.

资助项目:天津市高等学校科技发展基金 **(20140719).**

Y. Li · D. Zhen (✉) · L. Ma
Tianjin Normal University, Tianjin 300387, China
e-mail: zhendan0827@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_92

In recent years, more and more researchers have studied the three-dimensional space gamut; there are some three-dimensional color gamut volume calculation methods that have been proposed. Andreas Willert et al. calculated the gamut volume of color printing equipment in the CMY and CMYK color space, and gamut volumes are divided into a series of sheets by the brightness, and then mesh on each sheet to calculate the size of the color gamut [4]. Kuang et al. proposed tetrahedral meshing algorithm based on space region segmentation to achieve the reproduction of display device gamut in three dimensions and obtain the color gamut volume [5]. Yan et al. used alpha form algorithms extending in three-dimensional space, by selecting the appropriate radius to calculate the three-dimensional color gamut volume of the display in CIELUV color space [6]. Bangyong Sun et al. divided gamut boundary surface into small triangles based on convex hull triangulation technology, take the center of the surface and connected the triangle to form tetrahedron, by calculating the sum of all tetrahedrons to obtain the gamut volume [7]. These gamut volume calculation methods were surface reconstructed with three-dimensional gamut data points of display, and most algorithms have fewer sampling points and poor accuracy. While in three-dimensional color space to evaluate the gamut, it requires the data of volume without the need for surface reconstruction of data points.

Through analyzing the color gamut boundary, points of the tested monitor in the CIELAB show that the color gamut boundary shows the feature of big middle and small side. In order to calculate the color gamut volume, we proposed a projected algorithm based on the convex hull of the triangle to calculate the three-dimensional color gamut volume of the display.

## 2 The Projected Algorithm Based on the Convex Hull of the Triangle

Step of the algorithm is as follows:

(1) To construct the convex hull of triangles based on a three-dimensional color gamut boundary points

According to the convex shell model, which is one of space point set mathematical models, the convex hull of spatial point set P is a minimum convex set containing the point set [8]. The convex of point set P represents all convex combinations of three-dimensional point set P. Its formula is shown in Eq. (1):

$$\varphi(P) = \{ \sum_{i=1}^{n} \alpha_i p_i | p_i \in P, \alpha_i \geq 0, \sum_{i=1}^{n} \alpha_i = 1, n = 1, 2, 3, \cdots \}. \quad (1)$$

In the above formula, α and n are constants. According to Euler's formula, if the internal point set P has n vertices, the maximum number of edge in the point set is

$3n - 6$, and the maximum number of plane is $2n - 4$. In the three dimensions, the complexity of the convex hull of arbitrary N points is $O(n)$.

The algorithm will take a random incremental method to construct the convex hull of color gamut boundary point set. The algorithm procedure is as follows, in which P indicates the color space gamut boundary points, $P_i$ represents the $i_{th}$ sample points of the gamut, and $\phi(P)$ represents the convex hull of color gamut boundary points.

① Elected four color gamut sample points that are not coplanar in gamut boundary points form a tetrahedron, and seen it as an initial gamut convex hull $\phi(P_4)$;
② Determine the direction of each surface in the initial gamut convex hull, making the other gamut sample points, which are outside the convex hull, into their right-handed direction.
③ Arrange the remaining color gamut sample points with $P_5$, $P_6$, $P_7$,…,$P_n$. Then processed every color gamut sample points individually, at the same time maintain the gamut convex hull;
④ Depending on the state of the color gamut sample point $P_m$, which is to be processed, to determine how to deal. If the color gamut sample point $P_m$ falls inside the convex hull or at the boundary of the convex hull $\phi(P_{m-1})$, this represents $\phi(P_m) = \phi(P_{m-1})$, otherwise turn to step ⑤ to update color gamut convex hull. If the color gamut sample point $P_m$ falls outside the convex hull $\phi(P_m)$, then searches through the color gamut convex hull $\phi(P_m)$, look for the horizon and the plane of the sample point $P_m$ on the color gamut convex hull $\phi(P_{m-1})$.
⑤ Delete all planes that the gamut sample point $P_m$ constitutes in the visible area of the color gamut convex hull $\phi(P_{m-1})$; and along the horizon to connect it with the color gamut sample point $P_m$ to form a new plane, while maintaining the direction of the plane, and then added the sample point $P_m$ to the color gamut convex hull $\phi(P_{m-1})$; thus a new color gamut convex hull $\phi(P_m)$ is formed.
⑥ Repeat the steps ④–⑤, until processed all the color gamut sample points.

In the above algorithm, we used the concept of the ground plane and the visible area of the convex hull. For the color gamut of sample point $P_m$, the visible area of the color gamut convex hull $\phi(P_{m-1})$ means that, if the sample point $P_m$ falls outside the color gamut convex hull $\phi(P_{m-1})$, then from the position of the point $P_m$ to view the color gamut convex hull $\phi(P_{m-1})$, the visible planes on the surface of the convex hull $\phi(P_{m-1})$ formed the connected region. Connect the edges on the color gamut convex hull, which surrounded the connected region, to form a broken line. This line is called the horizon of the color gamut sample point $P_m$ on the color gamut convex hull $\phi(P_{m-1})$.

By this method, we can get the color gamut convex hull composed of the color gamut boundary points of the display in three-dimensional space. In this paper, we

tested the color gamut boundary points of the Toshiba display in three-dimensional space, the point cloud and the convex hull of the display is shown in Figs. 1 and 2.

(2) By the projection to constitute the pentahedron and then calculate the color gamut volume.

Seen from the figure, the color gamut surface of the convex hull obtained by this algorithm is a triangular mesh model. This article will calculate the volume of the color gamut convex hull by projection.

The specific process of the projected algorithm is as follows: First, select the projected plane, and select the plane $a^*b^*$ to herein as the projected plane, to facilitate understanding; we take one color gamut convex hull for example to calculate. Suppose one of the triangle convex hulls is $\phi(P_m)$, which was projected onto the plane $a^*b^*$, so that it can form a projected boundary of the convex hull in the plane $a^*b^*$, then connected the projected boundary and its convex hull shell, thus the pentahedron is formed. Its principle is shown in Fig. 3.

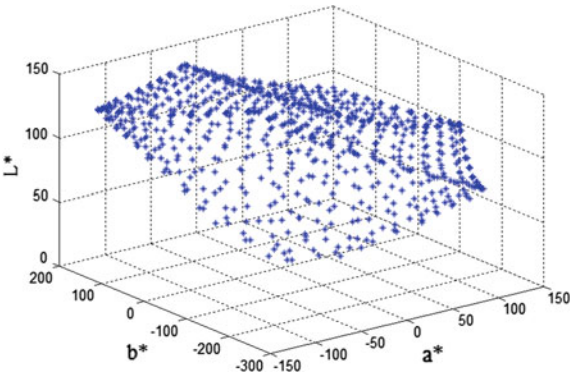



**Fig. 1** The point cloud of the display

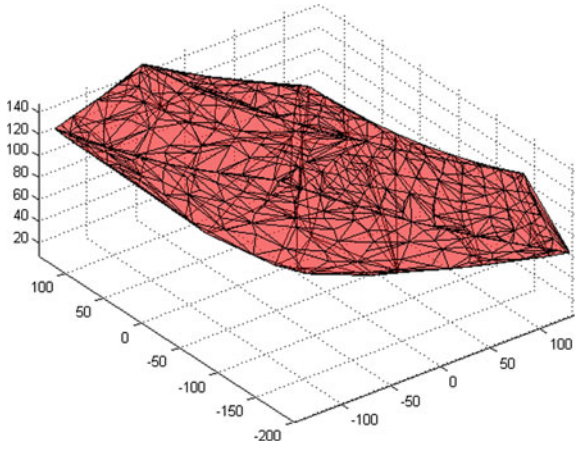



**Fig. 2** The convex hull of the display

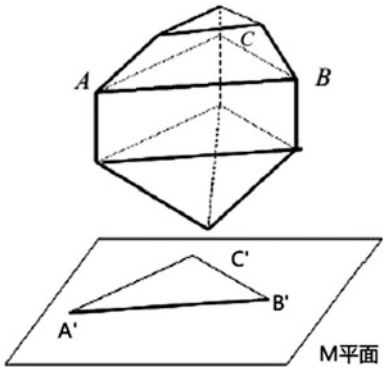



**Fig. 3** The projection of the convex hull

By this method, the volume of the convex hull can be divided into three parts to solve, namely the upper triangular mesh surface $M_U$, the lower triangular mesh surface $M_D$, and the triangular mesh face $M_C$ coincided with the prism side, since the plane $M_C$ in the middle is perpendicular to the plane M, so the volume of the convex hull is

$$V = V_U - V_D. \tag{2}$$

For such a triangular mesh patch, composed of the projected plane to form a pentahedron, it is shown in Fig. 4.

The volume of the pentahedron V (ABCA′B′C′) can be calculated by the sum of three tetrahedrons' volume to obtain, respectively: V(AA′B′C′), V(ABCB′), and V (ACA′C′). Seen by the analytic geometry, tetrahedron volume can be obtained by the vector product of the edge; the calculation is shown in formula (3):

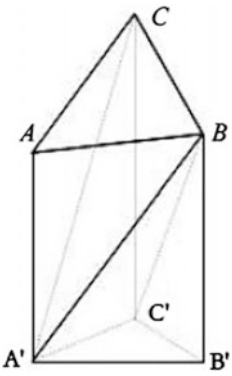



**Fig. 4** The pentahedron

$$V(AA'B'C') = \frac{1}{6} \times \begin{vmatrix} 1 & 1 & 1 & 1 \\ a_1^* & a_2^* & a_3^* & a_4^* \\ b_1^* & b_2^* & b_3^* & b_4^* \\ L_1^* & L_2^* & L_3^* & L_4^* \end{vmatrix}. \tag{3}$$

In this formula, ($a_1^*$, $b_1^*$, $L_1^*$), ($a_2^*$, $b_2^*$, $L_2^*$), ($a_3^*$, $b_3^*$, $L_3^*$), and ($a_4^*$, $b_4^*$, $L_4^*$) are spatial coordinates of four vertices of the tetrahedron. The volume of the pentahedron is shown in formula (4):

$$V(ABCA'B'C') = V(AA'B'C') + V(ABCB') + V(ACA'C'). \tag{4}$$

At the same time, judging a tetrahedral edge vector is a right-handed or left-handed to determine the triangle on the convex hull is upper triangular mesh surface or lower triangle mesh surface. Take any point of the triangle, and projected it to the plane, then connect this point and the triangle to form a tetrahedron. When the edge vector of the tetrahedron is right-handed, it belongs to upper triangular surface; on the contrary, it belongs to lower triangular surface.

Finally, the volume of convex hull in three dimensions is determined as follows:

$$V(\varphi(P)) = \sum_{i}^{m} V(\Delta_i) - \sum_{j}^{n} V(\Delta_j). \tag{5}$$

In this formula, V(ϕ(P)) represents the final calculated volume, M represents the number of pentahedron constituted by the upper triangular surface, and N represents the number of pentahedron constituted by the lower triangular surface.

Through this algorithm, the final calculated three-dimensional color gamut volume of this display is V = 837474.168.

Pyramid segmented method is one of the basic methods of three-dimensional space calculation of the volume of data points [9]. In this paper, we use pyramid segmented method to calculate the color gamut volume, and compared the calculating time. The results show that the algorithm of this paper processed the data faster.

## 3 Conclusion

When the display device color gamut is analyzed in three-dimensional space, it requires a lot of experimental test data to analyze and evaluate the color characteristics of the display [10]. Color gamut volume is considered to be one of the indexes to evaluate the color characteristics of the display; high computational complexity limits its practical application. When testing a large gamut of data points, the projected algorithm that proposed in this paper can guarantee the accuracy of the results at the same time greatly reduce the time used in the calculation.

## References


1. R. Miwa, Y. Manabe, N. Yata, 5–5 color gamut expansion method of 3D projector using multibands projection, in *ITE Winter Annual Convention* (2013), pp. 5–5
2. J.R. Manders, L. Qian, High efficiency and ultra-wide color gamut quantum dot LEDs for next generation displays. J. Soc. Inform. Display **23**(11), 523–528 (2015)
3. Q. Liu, Color gamut and color gamut coverage. Telev. Technol. **11**, 49–52 (2004)
4. A. Willert, M. Flaspohler, A.C. Hubler, calculating a color gamut border using a grid approach. conference on colour in graphics, imaging, and vision. Soc. Imaging Sci. Technol. **2004**(1), 156–160 (2004)
5. S. Kuang, X. Wang, Z. Wang, The tetrahedral meshing algorithm to describe the color gamut based on regional segmentation. Packaging Eng. **35**(5), 126–130 (2014)
6. J. Sun, X. Yan, GuiLing Li, The fast algorithm to calculate the color gamut volume in three-dimension based on the alpha form. Telev. Technol. **38**(21), 29–31 (2014)
7. B. Sun, H. Liu, W. Li et al., A color gamut description algorithm for liquid crystal displays in CIELAB space. Sci. World J. **2014**(1), 671964 (2014)
8. K. Deshpande, P.J. Green, M.R. Pointer, Metrics for comparing and analyzing two colour gamuts. submitted for publication, in *Visualization of Expanded Printing Gamuts Using 3-Dimensional Convex Hulls*, ed. by K. Guyler, Proceedings of TAGA 2000, pp 696–708
9. M. Zhang, J. Wen, Z. Zhang, Measure point cloud registration based on the volume of inter-surface triangular pyramid. China Mech. Eng. **2**, 175–178 (2010)
10. Y.H. Luo, M.H. Lin, Three-dimensional color gamut visualization of digital output device based on icc profile. Appl. Mech. Mater. **262**, 36–39 (2012)

# A Multi-scale Image Registration Algorithm Based on Wavelet Transform

**Qingfeng Sun and Jixiang Zhang**

**Abstract** Image registration has been an attractive research area of image processing; it has wider applications in many fields such as pattern recognition, image fusion, computer vision, and other practical problems. Due to the multi-resolution features, wavelet becomes the focus of research new algorithm and improves the traditional matching algorithm. Based on the study of traditional Harris corner detection algorithm, combined with wavelet transform, proposed an image registration algorithm based on the wavelet edge detection and Harris corner detection.

**Keywords** Image registration · Wavelet transform · Corner detection

## 1 Introduction

Image registration is the align process in space of overlaying two or more images of the same scene taken at different times, from different viewpoints, or by different sensors [1]. It is the key step of image processing technology; it has a wide range of applications in computer vision, pattern recognition, remote sensing image processing, medical image processing, automatic navigation, artificial intelligence, etc. It is also used in image communication field, but interference management has been a severe problem in wireless communication [2]. We need to choose a stable method to enhance anti-interference ability and reduce the amount of computation.

Image registration has a very wide range of methodology, based on the use of the image information; they can be divided into the method based on image gray-level information and image feature-based registration method. Among them, compared

Q. Sun (✉) · J. Zhang
School of Automation and Electronic Engineering, Tianjin University of Technology and Education, Tianjin 300222, China
e-mail: sunqife@126.com

Q. Sun
Department of Electronic Engineering, Anhui Technical College of Mechanical and Electrical Engineering, Wuhu 241000, Anhui, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_93

with the method use image intensity information to registration, feature-based image registration algorithm has less computation, good robustness, adaptability of geometric distortion, etc.; it is the research focus.

In recent years, combine the traditional registration algorithm with multi-scale technology to get more efficient, more accurate, more reliable registration algorithm is a popular research direction; this paper combines wavelet transform and Harris corner detection; make the classic Harris corner detection algorithm which has a multi-scale feature and enhance its anti-noise ability.

## 2 Harris Corner Detection

Harris corner detector is based on the local autocorrelation function of a signal, where the local autocorrelation function measures the local changes of the signal with patches shifted by a small amount in different directions. A discrete predecessor of the Harris detector was presented by Moravec [3], where the discreteness refers to the shifting of the patches.

Given a shift $(\Delta \mathrm{x}, \Delta y)$ and a point $(\mathrm{x}, y)$, autocorrelation function is defined as

$$\mathrm{c}(\mathrm{x}, y) = \sum_{W} [I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2, \tag{1}$$

where $I(\cdot, \cdot)$ denotes the image function and $(x_i, y_i)$ are the points in the window W (Gaussian) centered on $(x, y)$.

The shifted image is approximated by a Taylor expansion truncated to the first-order terms,

$$I(\mathrm{x_i} + \Delta x, y_i + \Delta y) \approx I(x_i, y_i) + \left[\mathrm{I_x}(x_i, y_i) I_y(x_i, y_i)\right] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \tag{2}$$

where $I_{\mathrm{x}}(\cdot, \cdot)$ and $I_y(\cdot, \cdot)$ denote the partial derivatives in x and y, respectively.

Substituting approximation equation into $\mathrm{c}(x, y)$,

$$\begin{aligned} \mathrm{c}(\mathrm{x}, y) &= \sum_{W} [I(x_i, y_i) - I(x_i + \Delta x, y_i + \Delta y)]^2 \\ &= \sum_{W} \left\{ I(x_i, y_i) - I(x_i - y_i) - \left[I_x(x_i, y_i) I_y(x_i, y_i)\right] \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix} \right\}^2 \\ &= [\Delta x, \Delta y] M(x, y) \begin{bmatrix} \Delta x \\ \Delta y \end{bmatrix}, \end{aligned} \tag{3}$$

where matrix M(x, y) captures the intensity structure of the local neighborhood.

Let $\lambda_1$ and $\lambda_2$ be the eigenvalues of matrix M(x, y). The eigenvalues form a rotationally invariant description. There are three cases to be considered:

1. If both eigenvalues are small, the local autocorrelation function is flat (i.e., little change in M(x, y) in any direction); the windowed image region is of approximately constant intensity.
2. If one eigenvalue is high and the other low, so the local autocorrelation function is ridge-shaped, then only local shifts in one direction (along the ridge) cause little change in M(x, y) and significant change in the orthogonal direction; this indicates an edge.
3. If both eigenvalues are high, so the local autocorrelation functions sharply peaked, then shifts in any direction will result in a significant increase; this indicates a corner.

In order to avoid the eigenvalues of a matrix M, it can be used Tr(M) and Det(M) to replace and indirectly:

$$\begin{aligned} Tr(M) &= \lambda_1 + \lambda_2 = A + B \\ Det(M) &= \lambda_1 \lambda_2 = AB - C^2. \end{aligned} \tag{4}$$

Harris corner responses function with the following:

$$R = Det - KTr^2, \tag{5}$$

where K is a constant factor, usually between 0.03 and 0.15. Only when the image pixel value R is greater than a certain threshold, and in eight directions around is local maximum value, consider the point is a corner.

The Harris corner detector is a popular interest point detector due to its strong invariance to the influence of camera position and illumination variation [4].

However, it is sensitive to noise, the corner prone to drift at different image resolutions. In addition, the traditional Harris corner detection operators are not scale invariant. However, the image feature points often appear in different scales, and each corner of the scale information is unknown.

Generally, in the larger scale, corner is not easy to accurate positioning, but it can be better to eliminate false alarms and detect real corners. Conversely, in smaller scale, the location of corner is more accurate, but the proportion of false detection will increase. Therefore, it seems necessary and practical to use multi-scale technology for feature extraction.

## 3 Wavelet Transform-Based Image Registration

If scale as the camera lens, then, from big to small changes in scale, equivalent to the camera is from far and near close to the target. Large-scale space corresponds to observe object under the far lens, reflecting the general overview of objectives;

small-scale space corresponds to observe target in the close lens, and the details of the part can be observed. Therefore, scale changes from big to small can achieve the goal observation from coarse to fine, which is multi-scale (or multi-resolution) thought.

1. The Basic Theory of Wavelet Transform

One-dimensional continuous wavelet transform definition is

$$W_f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x)\overline{\psi\left(\frac{x-b}{a}\right)}\,dx, \tag{6}$$

where a is the scale parameter (corresponding to the frequency information), b is the shift factor (Corresponding to the space-time information), $\overline{\psi(x)}$ is the wavelet function, and $\overline{\psi(x)}$ denotes the complex conjugate of $\psi(\mathrm{x})$. If a discretization, let $a = 2^k$, and binary wavelet transform can be obtained as

$$W_{2^k}(b) = \frac{1}{\sqrt{2^k}} \int_{-\infty}^{+\infty} f(x)\overline{\psi\left(\frac{x-b}{2^k}\right)}\,dx. \tag{7}$$

If a, b are all discretization, let $b = n2^k$, then get discrete wavelet transform:

$$W_{2^k} f(n) = \frac{1}{\sqrt{2^k}} \int_{-\infty}^{+\infty} f(x)\overline{\psi(2^{-k}x - n)}\,dx. \tag{8}$$

Two-dimensional wavelet transform has a similar definition. In image processing, discrete wavelet is often used. Compared with Fourier transform, wavelet transform has time–frequency localization and multi-resolution features can effectively extract required information from a given signal. Through dilation and translation operations, it can refine the function and signal.

2. Wavelet Transform-Based Image Registration

Generally speaking, image processing is based on wavelet transform, most built on the wavelet decomposition and reconstruction, the wavelet transform decomposition of the image automatically to form a pyramid structure; the image information is divided into low- and high-frequency part of detail parts, and it is easy for image feature extraction and utilization [5–8].

The stationary wavelet transform (SWT) is designed to shift-invariant discrete signal analysis. In a decomposition step, SWT will produce two signals of the same length. These two signals are called approximate coefficients and detail coefficients. Unlike the discrete wavelet transform (DWT), these coefficients are not down-sampled. In contrast, decomposition filters are up-sampled before the next

step of decomposition. The dyadic decomposition is often used, and the algorithm is often called "à trous" algorithm In this paper, wavelet-based image registration algorithm is the choice of stationary wavelet transform.

Combined with multi-scale characteristic of wavelet transform and the Harris corner detection, get a kind of multi-scale image registration algorithm based on feature point, and procedure is as follows (Fig. 1):

(1) Decompose reference image and recent image by SWT (choose sym4 wavelet); it will produce four series coefficients in every decomposition level and the size of them is the same with the original signal; denote A for approximate coefficient, H for horizontal high-frequency coefficient, V for vertical high-frequency coefficient, and D for diagonal direction high-frequency coefficients.
(2) After the wavelet decomposition, use horizontal and vertical direction high-frequency coefficient (H, V) to get the modulus image:

$$\begin{aligned} &\text{for i} = 1\text{: scale;} \\ &\quad \text{Mod}(:,:,\text{i}) = \text{sqrt}(\text{V}(:,:,\text{i}).\text{^}2 + \text{H}(:,:,\text{i}).\text{^}2); \end{aligned} \tag{9}$$

(3) To avoid the error caused by the selection of constant factor K, adopt modified Harris corner response function:

$$R = \left(I_x^2 * I_y^2 - I_{xy}^2\right) / \left(I_x^2 + I_y^2 + eps\right), \tag{10}$$

extracted feature point from modulus image that get from the previous step;
(4) According to the correlation criterion to determine the matching point;
(5) Determine the transformation relationship between two images.

## 4 Simulation Results and Analysis

Use Lena map to analyze the proposed algorithm. Figure 2 shows two images; one is reference image and the other one is an image wait to registration. Figure 3 shows the match results of the Lena character image at different scales; we can see (a) (scale = 1) have no error matching pair, Figure (b) (scale = 2) have only one error pair of matching, and the stability is well, in addition, the repeatability of match points under two scales is high.

Figure 4 shows that the match results using based on SURF (Speeded Up Robust Features) operator [9], which have high speed and high stability, have been wide recognized in feature detection and image registration. Compared with the method upon in this paper, we can see that the match point pairs in two algorithms have high degree of coincidence.

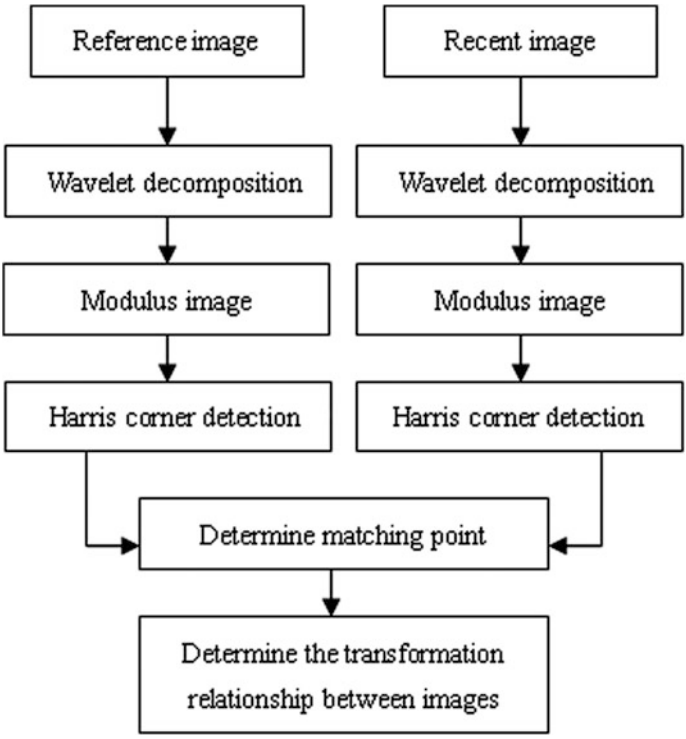



**Fig. 1** Framework for wavelet transform-based image registration methodology

**Fig. 2** Reference image

**(a)** scale = 1

**(b)** scale = 2

**Fig. 3** The match results in different scales using the method proposed

**Fig. 4** The match results based on SURF operator

## 5 Conclusion

Aim at the scale invariance and anti-noise insufficiency of traditional Harris corner detection, based on maintain its advantages in rotate, zoom, and so on; combine with the wavelet multi-scale technology; and propose an image registration algorithm based on wavelet transform. The experimental results show that the method proposed has high match accuracy at different scales.

**Acknowledgements** Project found: 1. Young teachers development and support program of Anhui Technical College of Mechanical and Electrical Engineering (project number: 2015yjzr028); 2. Quality engineering project in Anhui province: specialty of industrial robot (project number: 2015tszy080); 3. Virtual simulation experimental teaching center for industrial robot (project number:2016xnzx007).

## References

1. B. Zitova, J. Flusser, Image registration methods: a survey. Image Vis. Comput. **21**, 977–1000 (2003)
2. Z. Li et al., Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.02.009
3. C. Harris, M.J. Satephens, A combined corner and edge detector, in *4th Alvey Vision Conference* (1988), pp. 147–152
4. C. Schimid, R. Mohr, C. Bauckhage, Evaluation of interest point detectors. Int. J. Comput. Vis. **37**(2), 151–172 (2000)
5. J. Le Moigne, W.J. Campbell, R.F. Cromp, An automated parallel image registration technique based on the correlation of wavelet features. IEEE Trans. Geosci. Remote Sens. **40**(8), 1849–1864 (2002)
6. J. You, P.A. Bhattacharya, Wavelet-based Coarse-to-fine Image Matching Scheme in a parallel virtual machine environment. IEEE Trans. Image Process. **9**(9), 1547–1559 (2004)
7. Y. Ding, W. Li, Medical image registration algorithm based on the SAM of multi-scale Harris corner. J. Image Graph. **15**(12), 1762–1768 (2010)
8. Z. Fan, C. Wang, High accuracy subpixel image registration based on phase-only correlation of Harris strength. Appl. Res. Comput. **28**(2), 788–791 (2011)
9. H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speed up robust features, in *European Conference on Computer Vision* (2006), pp. 404–417

# Camera Calibration Based on New Lens Distortion Model

**You Zhai, Xiwei Guo and Deliang Liu**

**Abstract** In the classic two-step technique based camera calibration methods, pinhole parameters and lens distortion parameters are estimated together. The calibration results are affected by the coupling of these parameters. Image pixel coordinate expressed lens distortion models are derived, which are modifications to existing image physical coordinate expressed lens distortion models and reduce the coupling between linear parameters and distortion parameters. Camera calibration method is proposed based on the new distortion models and used to calibrate binocular vision cameras with a three-dimensional target. The experiment results show that compared with traditional camera calibration method, the proposed method is feasible and can efficiently improve calibration accuracy.

**Keywords** Camera calibration · Distortion model · Parameter decoupling · Binocular vision

## 1 Introduction

Camera calibration is the process of camera modeling, which is the foundation and premise of vision measurement. There are two kinds of parameters describing camera model: pinhole model parameters and lens distortion parameters. The pinhole model parameters consist of intrinsic parameters and extrinsic parameters. The intrinsic parameters include focal length, aspect ratio, principal points and skew factor, which depict the perspective projection transformation from 3D points in the scene to 2D points in the image. The extrinsic parameters give the position and orientation of the camera coordinate system with respect to a world (reference) coordinate system. Due to errors arising from lens design, manufacturing and assembling, the image coordinates deviate from the ideal image coordinates. The lens distortion parameters describe the distortion of the lens and give the trans-

Y. Zhai (✉) · X. Guo · D. Liu
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: youyou1952@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_94

formation from distorted image coordinates to ideal image coordinates. There are many lens distortion models and most of them model lens distortion in image physical coordinate system, such as in [1–3].

Pinhole parameters and lens distortion parameters are usually estimated together. Due to the coupling of these parameters, the calibration results are only optimal for the calibration targets and are valid to minor scene depth variation. Carlos et al. proposed methods decoupling lens distortion from pinhole model by nonmetric lens distortion correction in [4–6]. Nonmetric lens distortion correction is carried out directly in the image pixel coordinate and therefore we have to deal with lens distortion model in image pixel coordinate (pixel-LDMs). In the process of nonmetric lens distortion calibration, lens distortion models in image physical coordinate are directly used as pixel-LDMs, such as in [7], which implicitly assumes square pixels (namely, aspect ratio to be 1). But in fact, due to manufacturing errors, there is no square pixel. Therefore, non-one aspect ratio should be taken into account. In short, lens distortion model in image pixel coordinate is not consistent with distortion model in image physical coordinate.

In order to solve the problem mentioned above, a new camera calibration method based on pixel-LDMs is proposed. We first analyze the image formation process and general models in pixel coordinate system are obtained. Then, combining Faugeras's calibration method in [8, 9] and pixel-LDMs, a two-step technique based calibration method is proposed.

## 2 Pixel Coordinate Expressed Lens Distortion Model

### *2.1 Camera Modeling*

Camera modeling is usually broken into four steps, as is detailed in the following list.

First, change from world coordinate system $\mathrm{X_w} = (x_\mathrm{w}, y_\mathrm{w}, z_\mathrm{w})^\mathrm{T}$ to camera coordinate system $\mathrm{X_c} = (x, y, z)^\mathrm{T}$, as shown in Eq. (1). The transformation consists of a 3 × 3 rotation matrix $R$ and a 3 × 1 translation vector $T$.

$$\mathrm{X_c} = R\mathrm{X_w} + T \tag{1}$$

Second, project 3D points $\mathrm{X_c}$ on the image plane obtaining 2D points $(x_u, y_u)^\mathrm{T}$ in image physical coordinate system using a perspective transformation, as shown in Eq. (2), where $f$ is focal length.

$$x_u = f\frac{x}{z} \quad y_u = f\frac{y}{z} \tag{2}$$

Third, change undistorted image physical coordinate system $(x_u, y_u)^{\mathrm{T}}$ to distorted image physical coordinate system $(x_d, y_d)^{\mathrm{T}}$ by distortion model expressed by undistorted image points, as shown in Eq. (3). When calibrating a camera, usually lens distortion model expressed by distorted image points is used, as shown in Eq. (4). Function $G$ and function $L$ is lens distortion model respectively.

$$(x_d, y_d)^{\mathrm{T}} = G(x_u, y_u) \tag{3}$$

$$(x_u, y_u)^{\mathrm{T}} = L(x_d, y_d) \tag{4}$$

Fourth, change from image physical coordinate system $(x_d, y_d)^{\mathrm{T}}$ to image pixel coordinate system $(u_d, v_d)^{\mathrm{T}}$, as shown in Eq. (5), where $d_x$ and $d_y$ is the size of each CCD pixel along horizontal and vertical direction separately. They are usually provided by manufacturer and $\lambda = d_y/d_x$ represents the ratio of the size of CCD pixel in horizontal and vertical direction, called aspect ratio. Usually it is close to 1, but due to manufacturing errors it is not equal to 1. $(u_0, v_0)$ is intersection between optical axis and image plane, called principal point. It usually coincides with image distortion center.

$$u_d = \frac{x_d}{d_x} + u_0 \quad v_d = \frac{y_d}{d_y} + v_0 \tag{5}$$

## 2.2 Derivation of the New Lens Distortion Model

There are many distortion models that can be used in the third step of camera modeling as shown in [1–5]. These distortion models are all *expressed* by image physical coordinate. In nonmetric lens distortion correction, lines and points in the image are used to build projective invariant constraints to estimate the image distortion and camera calibration in not a necessary. The correction is done directly in image pixel coordinate system. This means that pixel coordinate expressed lens distortion model is needed to model lens distortion. In [6–8], lens distortion models expressed by image physical coordinate system are adopted as lens distortion models in image pixel coordinate system directly, which neglects the effect of non-one aspect ratio $\lambda \neq 1$. Pixel coordinate expressed lens distortion model is derived using distortion model represented by undistorted image point.

Rearranging Eq. (5) and replacing $u_d - u_0$ with $\bar{u}_d$, the following Eq. (6) are obtained.

$$(\bar{u}_d, \bar{v}_d)^T = \begin{bmatrix} 1/d_x & 0 \\ 0 & 1/d_y \end{bmatrix} (x_d, y_d)^T \tag{6}$$

Taking Eq. (3) into Eq. (6), Eq. (7) can be get:

$$(\bar{u}_d, \bar{v}_d)^T = \begin{bmatrix} 1/d_x & 0 \\ 0 & 1/d_y \end{bmatrix} G(x_u, y_u) \tag{7}$$

where $G(x_u, y_u)$ represents function about $(x_u, y_u)^{\mathrm{T}}$ and $\psi$ represents lens distortion parameter space. The common used distortion models, such as radial and tangential distortion model, polynomial distortion model, division model and rational function distortion model, are all rational polynomials of $(x_u, y_u)^{\mathrm{T}}$. Equation (8) gives a uniform expression of those model.

$$G(x_u, y_u) = \begin{bmatrix} \left(\Psi, \left[x_u, y_u, \sum x_u^i, \sum y_u^j, \sum x_u^p y_u^q\right]\right) \\ \left(\Psi, \left[x_u, y_u, \sum x_u^i, \sum y_u^j, \sum x_u^p y_u^q\right]\right) \end{bmatrix} \tag{8}$$

Taking Eq. (8) into Eq. (7), the following Eq. (9) is obtained.

$$(\bar{u}_d, \bar{v}_d)^T = \begin{bmatrix} \frac{1}{d_x}\left(\Psi, \left[x_u, y_u, \sum x_u^i, \sum y_u^j, \sum x_u^p y_u^q\right]\right) \\ \frac{1}{d_y}\left(\Psi, \left[x_u, y_u, \sum x_u^i, \sum y_u^j, \sum x_u^p y_u^q\right]\right) \end{bmatrix} \tag{9}$$

Making the following mathematical transformation, substitute $x_u/d_x = u - u_0$ and $y_u/d_y = v - v_0$ with $\bar{u}$ and $\bar{v}$. The process is shown in Eq. (10) and finally Eq. (11) can be obtained.

$$(\bar{u}_d, \bar{v}_d)^T = \begin{bmatrix} \frac{1}{d_x}\left(\Psi, \left[d_x\frac{x_u}{d_x}, d_y\frac{y_u}{d_y}, \sum d_x^i\frac{x_u^i}{d_x^i}, \sum d_y^j\frac{y_u^j}{d_y^j}, \sum d_x^p\frac{x_u^p}{d_x^p}d_y^q\frac{y_u^q}{d_y^q}\right]\right) \\ \frac{1}{d_y}\left(\Psi, \left[d_x\frac{x_u}{d_x}, d_y\frac{y_u}{d_y}, \sum d_x^i\frac{x_u^i}{d_x^i}, \sum d_y^j\frac{y_u^j}{d_y^j}, \sum d_x^p\frac{x_u^p}{d_x^p}d_y^q\frac{y_u^q}{d_y^q}\right]\right) \end{bmatrix} \tag{10}$$

$$(\bar{u}_d, \bar{v}_d)^T = \begin{bmatrix} \frac{1}{d_x}\left(\Psi, \left[d_x\bar{u}, d_y\bar{v}, \sum d_x^i\bar{u}^i, \sum d_y^j\bar{v}^j, \sum d_x^p\bar{u}^p d_y^q\bar{v}^q\right]\right) \\ \frac{1}{d_y}\left(\Psi, \left[d_x\bar{u}, d_y\bar{v}, \sum d_x^i\bar{u}^i, \sum d_y^j\bar{v}^j, \sum d_x^p\bar{u}^p d_y^q\bar{v}^q\right]\right) \end{bmatrix} \tag{11}$$

Because $G_x$ and $G_y$ represent rational polynomial of $(x_u, y_u)^{\mathrm{T}}$, $1/d_x$, $1/d_y$, $d_x$ and $d_y$ can be coupled with coefficient of rational polynomial. Meanwhile taking $\lambda = d_y/d_x$ into Eq. (11), the following equation can be obtained.

$$(\bar{u}_d, \bar{v}_d)^T = \begin{bmatrix} \left(\Psi, \left[\bar{u}, \lambda\bar{v}, \sum \bar{u}^i, \lambda\sum \bar{v}^j, \sum \bar{u}^p\bar{v}^q\right]\right) \\ \left(\Psi, \left[\frac{1}{\lambda}\bar{u}, \bar{v}, \frac{1}{\lambda}\sum \bar{u}^i, \sum \bar{v}^j, \sum \bar{u}^p\bar{v}^q\right]\right) \end{bmatrix} \tag{12}$$

Similarly, we can get general distortion model expressed by distorted image pixel coordinate as shown in Eq. (13).

$$(\bar{u}, \bar{v})^T = \begin{bmatrix} \left(\Phi, \left[\bar{u}_d, \lambda\bar{v}_d, \sum \bar{u}_d^i, \lambda \sum \bar{v}_d^j, \sum \bar{u}_d^p \bar{v}_d^q\right]\right) \\ \left(\Phi, \left[\frac{1}{\lambda}\bar{u}_d, \bar{v}_d, \frac{1}{\lambda}\sum \bar{u}_d^i, \sum \bar{v}_d^j, \sum \bar{u}_d^p \bar{v}_d^q\right]\right) \end{bmatrix} \tag{13}$$

Equations (12) and (13) give general image pixel coordinate expressed lens distortion model. According to Eqs. (12) and (13), we can easily rewrite image physical coordinate expressed lens distortion models to obtain their correspondent image pixel coordinate expressed lens distortion models, which are often used in nonmetric distortion calibration. For example, Eqs. (14) and (15) are distorted image pixel coordinate expressed radial and tangential distortion model and division model, where the following substitution is used: $k_1 d_x^2 = k_1$, $p_1 d_x = p_1$, $p_2 d_x = p_2$, $s_1 d_x = s_1$, $s_2 d_x = s_2$.

$$\begin{cases} \bar{u} = \bar{u}_d\left(1 + k_1\left(\bar{u}_d^2 + \lambda^2\bar{v}_d^2\right)\right) + p_1\left(3\bar{u}_d^2 + \lambda^2\bar{v}_d^2\right) + 2p_2\lambda\bar{u}_d\bar{v}_d + s_1\left(\bar{u}_d^2 + \lambda^2\bar{v}_d^2\right) \\ \bar{v} = \bar{v}_d\left(1 + k_1\left(\bar{u}_d^2 + \lambda^2\bar{v}_d^2\right)\right) + p_2\left(\frac{\bar{u}_d^2}{\lambda} + 3\lambda\bar{v}_d^2\right) + 2p_1\bar{u}_d\bar{v}_d + s_2\left(\frac{\bar{u}_d^2}{\lambda} + \lambda\bar{v}_d^2\right) \end{cases} \tag{14}$$

$$\begin{cases} \bar{u} = \dfrac{\bar{u}_d}{1 + k_1\left(\bar{u}_d^2 + \lambda^2\bar{v}_d^2\right)} \\ \bar{v} = \dfrac{\bar{v}}{1 + k_1\left(\bar{u}_d^2 + \lambda^2\bar{v}_d^2\right)} \end{cases} \tag{15}$$

A special case is logarithmic fish eye lenses distortion model, as shown in Eq. (16), which is not a rational function of $(x_d, y_d)^T$, but the substitution shown in the first part of Eq. (13) can be also applied. Its correspondent lens distortion model represented by distorted image pixel coordinate is shown in Eq. (17), where $s = s/d_x$, $\rho = \rho \times d_x$.

$$r_u = s \log\left(1 + \rho r_d\right) \tag{16}$$

$$\sqrt{\bar{u}^2 + \lambda^2\bar{v}^2} = s \log\left(1 + \rho\sqrt{\bar{u}_d^2 + \lambda^2\bar{v}_d^2}\right) \tag{17}$$

By observing Eqs. (12) and (13), it can be inferred that the difference between image pixel coordinate expressed lens distortion model and image physical coordinate expressed lens distortion model is caused by aspect ratio $\lambda$. The effect of $\lambda$ to lens distortion is taken into account by explicitly including it in our general image pixel coordinate expressed lens distortion model, which can deal with both $\lambda = 1$ and $\lambda \neq 1$ cases. With the improvement of CCD manufacturing technique, the square pixel can be well guaranteed. This means $\lambda$ is close to 1, but taking $\lambda \neq 1$ explicitly into account will definitely improve the accuracy of lens distortion model.

Besides, it further reveals the interrelationship between intrinsic parameters and lens distortion, which is useful for reduce the effect of the intrinsic parameters to lens distortion model parameters.

## 3 Camera Calibration

A 3D target based camera calibration method is proposed based on the new lens distortion model. After 3D scene control points $X_w = (x_w, y_w, z_w)^T$ and their correspondent imaging points $(u_d, v_d)^T$ are obtained, the Faugeras's method and the new distortion model are combined together to calibrate camera intrinsic, extrinsic and distortion parameters.

### *3.1 Pinhole Model Calibration*

Camera pinhole model include intrinsic and extrinsic parameters and the pinhole model are calibrated by Faugeras's method as shown in the following equations, for more detail in [8].

$$M = \frac{1}{\sqrt{m_{31}^2 + m_{32}^2 + m_{33}^2}} M \tag{18}$$

$$\begin{cases} u_0 = m_1 m_3^T \\ v_0 = m_2 m_3^T \end{cases}, \begin{cases} \alpha_x = \sqrt{m_1 m_1^T - u_0^2} \\ \alpha_y = \sqrt{m_2 m_2^T - v_0^2} \end{cases} \tag{19}$$

$$\begin{cases} [r_{11} \quad r_{12} \quad r_{13}] = \text{sgn} \times \frac{m_1 - u_0 m_3}{\alpha_x} \\ [r_{21} \quad r_{22} \quad r_{23}] = \text{sgn} \times \frac{m_2 - v_0 m_3}{\alpha_y} \\ [r_{31} \quad r_{32} \quad r_{33}] = \text{sgn} \times m_3 \end{cases}, \begin{cases} t_x = \text{sgn} \times \dfrac{m_{14} - u_0 m_{34}}{\alpha_x} \\ t_y = \text{sgn} \times \dfrac{m_{24} - v_0 m_{34}}{\alpha_y} \\ t_z = \text{sgn} \times m_{34} \end{cases} \tag{20}$$

$\boldsymbol{m}_i$ (i = 1, 2, 3) represents the *i*th row of projection matrix $\boldsymbol{M}$; $m_{jk}$ represents the *j*th row and the *k*th column (j = 1, 2, 3, k = 1, 2, 3, 4) of projection matrix $\boldsymbol{M}$; sign is sign function and the sign of the function depends on the camera's position relative to the z axis of the world coordinate system. If the camera locates in the positive direction of the z axis, the sign of the function is positive, otherwise sign is negative. Then SVD (Singular Value Decomposition) can be used to improve accuracy. The Euler angle (a, b, c) can be decomposed from the elements ($r_{ij}$) of the rotation matrix.

### 3.2 Lens Distortion Model Calibration

As mentioned above, the pinhole model parameters can be calibrated by Faugeras's method. The initial value of pinhole camera parameters can be used to calibrate the lens distortion model. When calibrating lens distortion model, the pixel coordinate expressed radial and tangential distortion model is used, as shown in Eq. (14). The procedure to compute the lens distortion model is shown below.

First, least square method is used to estimate the initial value of the lens distortion parameters. Reproject the 3D scene control points with initial pinhole camera parameters and the reprojection image points ($u_i$, $v_i$) can be obtained. Then compute the errors between reprojection image points ($u_i$, $v_i$) and the actual image points ($u_{di}$, $v_{di}$) according to Eq. (21).

$$\begin{cases} \delta_{ui} = u_i - u_{di} = u_0 + \alpha_x \frac{r_{11}x_i^w + r_{12}y_i^w + r_{13}z_i^w + t_x}{r_{31}x_i^w + r_{32}y_i^w + r_{33}z_i^w + t_z} - u_{di} \\ \delta_{vi} = v_i - v_{di} = v_0 + \alpha_y \frac{r_{21}x_i^w + r_{22}y_i^w + r_{23}z_i^w + t_y}{r_{31}x_i^w + r_{32}y_i^w + r_{33}z_i^w + t_z} - v_{di} \end{cases} \tag{21}$$

Rearrange Eqs. (14) and (22) is obtained, where ($\delta_{ui}$, $\delta_{vi}$) is related with distortion parameters and $\bar{u}_{di} = u_{di} - u_0$, $\bar{v}_{di} = v_{di} - v_0$

$$\begin{cases} \delta_{ui} = k_1 \bar{u}_{di}\left(\bar{u}_{di}^2 + \lambda^2 \bar{v}_{di}^2\right) + p_1(3\bar{u}_{di}^2 + \lambda^2 \bar{v}_{di}^2) + 2p_2\lambda \bar{u}_{di}\bar{v}_{di} + s_1\left(\bar{u}_{di}^2 + \lambda^2 \bar{v}_{di}^2\right) \\ \delta_{vi} = k_1 \bar{v}_{di}\left(\bar{u}_{di}^2 + \lambda^2 \bar{v}_{di}^2\right) + p_2(\frac{\bar{u}_{di}^2}{\lambda} + 3\lambda \bar{v}_{di}^2) + 2p_1 \bar{u}_{di}\bar{v}_{di} + s_2\left(\frac{\bar{u}_{di}^2}{\lambda} + \lambda \bar{v}_{di}^2\right) \end{cases} \tag{22}$$

According to Eqs. (21) and (22), two linear equations about the lens distortion parameters can be obtained. When N (N > 3) pairs of such equations are computed, least square method can be used to estimate the initial value of the lens distortion parameters. When computing, the value of ($u_0$, $v_0$) and $\lambda = \alpha_y/\alpha_x$ is obtained from Eq. (19).

Second, fixing the initial value of the pinhole parameters, LM algorithm can be used to improve the accuracy of the lens distortion parameters. The object function for LM optimization is shown in Eq. (23).

$$f(k_1, p_1, p_2, s_1, s_2) = (D_{ui}, D_{vi}) \tag{23}$$

Where ($D_{ui}$, $D_{vi}$) can be computed according to Eq. (24).

$$\begin{cases} D_{u_{di}} = \alpha_x \frac{r_{11}x_i^w + r_{12}y_i^w + r_{13}z_i^w + t_x}{r_{31}x_i^w + r_{32}y_i^w + r_{33}z_i^w + t_z} - \bar{u}_i \\ D_{v_{di}} = \alpha_y \frac{r_{21}x_i^w + r_{22}y_i^w + r_{23}z_i^w + t_y}{r_{31}x_i^w + r_{32}y_i^w + r_{33}z_i^w + t_z} - \bar{v}_i \\ \bar{u}_i = \bar{u}_{di}\left(1 + k_1\left(\bar{u}_{di}^2 + \lambda^2 \bar{v}_{di}^2\right)\right) + p_1(3\bar{u}_{di}^2 + \lambda^2 \bar{v}_{di}^2) + 2p_2\lambda \bar{u}_{di}\bar{v}_{di} + s_1\left(\bar{u}_{di}^2 + \lambda^2 \bar{v}_{di}^2\right) \\ \bar{v}_i = \bar{v}_{di}\left(1 + k_1\left(\bar{u}_{di}^2 + \lambda^2 \bar{v}_{di}^2\right)\right) + p_2(\frac{\bar{u}_{di}^2}{\lambda} + 3\lambda \bar{v}_{di}^2) + 2p_1 \bar{u}_{di}\bar{v}_{di} + s_2\left(\frac{\bar{u}_{di}^2}{\lambda} + \lambda \bar{v}_{di}^2\right) \end{cases} \tag{24}$$

Third, LM algorithm is used to optimize all the parameters including intrinsic, extrinsic and lens distortion parameters, and the object function is shown in Eq. (25).

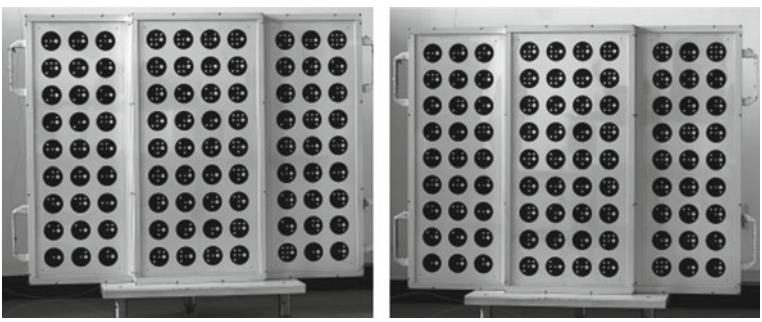
$$f\left(\alpha_x, \alpha_y, u_0, v_0, a, b, c, t_x, t_y, t_z, k_1, p_1, p_2, s_1, s_2\right) = (D_{ui}, D_{vi}) \tag{25}$$

## 4 Experimental Results and Analysis

Experiments are carried out to verify the feasibility and accuracy of the proposed method and the proposed method is compared with several classic calibration methods, such as Faugeras's method in [8], Weng's method in [10]. The proposed method is used to calibrate binocular vision cameras and the stereo images of a 3D calibration target captured by the binocular vision system are shown in Fig. 1. The target contains 90 coded markers and the 3D position of the markers is known. Coded marker can be used to accelerate the recognition procedure of the 3D target.

The value of each parameter cannot reflect the accuracy of the method. Hence, reprojection errors and 3D measurement errors are used to evaluate different methods.

Table 1 and Table 2 show the reprojection errors for calibration results of left and right camera respectively. Reprojection errors reflect the consistence between calibration results and the actual 3D to 2D mapping. From Tables 1 and 2, the reprojection errors of the proposed method are smallest and the mean and standard deviation are both less than 0.2 pixels.



**(a)** left **(b)** right

**Fig. 1** Stereo images of 3D calibration target

**Table 1** Reprojection errors of the left camera

| Params\Method | | Proposed | Fagera's | Weng |
|---|---|---|---|---|
| RES | Mean/pixel | 0.1527 | 0.1796 | 0.1824 |
| | Std/pixel | 0.1070 | 0.1218 | 0.1088 |

**Table 2** Reprojection errors of the right camera

| Params\Method | | Proposed | Fagera's | Weng |
|---|---|---|---|---|
| RES | Mean/pixel | 0.1447 | 0.2518 | 0.1951 |
| | Std/pixel | 0.1096 | 0.1606 | 0.1170 |

**Table 3** 3D measurement error

| Result\Method | Proposed | Faugera's | Weng |
|---|---|---|---|
| Mean/mm | 0.1821 | 0.3858 | 0.2022 |
| Std/mm | 0.1115 | 0.2616 | 0.1283 |

3D coordinates can be computed using binocular cameras and the 3D measurement results are shown in Table 3. The 3D measurement errors of the proposed method are also smallest in the three methods.

From the results above, it can be referred that the accuracy of the proposed method is better than the other two methods and the proposed method is feasible.

## 5 Conclusion

In this paper, a camera calibration method based on 3D calibration target is proposed, which utilizes a new lens distortion model. The new model is expressed in image pixel coordinates. It is derived from the image physical coordinate expressed lens distortion models. The effect of $\lambda$ to lens distortion is taken into account by explicitly including it in our general image pixel coordinate expressed lens distortion model, which can deal with both $\lambda = 1$ and $\lambda \neq 1$ cases. With the new model, a two-step camera calibration method is proposed. First, Faugera's method is used to estimate the pinhole model. Second, least square method and LM method is used to compute the lens distortion model. Experiment results show that the proposed method is feasible and are high in accuracy.

## References

1. R. Hartley, S. Kang, Parameter-free radial distortion correction with center of distortion estimation. IEEE Trans. Pattern Anal. Mach. Intell. **29**(8), 1309–1321 (2007)
2. J. Wang, F. Shi, J. Zhang, Y. Liu, A new calibration model of camera lens distortion. Pattern Recogn. **41**, 607–615 (2008)
3. J.P. Tardif, P. Sturm, M. Trudeau, S. Roy, Calibration of cameras with radially symmetric distortion. IEEE Trans. Pattern Anal. Mach. Intell. **31**(9), 1552–1566 (2009)
4. C. Ricolfe-Viala, A.-J. Sanchez-Salmeron, A. Valera, Efficient lens distortion correction for decoupling in calibration of wide angle lens cameras. IEEE Sens. J. **13**(2), 854–863 (2013)
5. C. Ricolfe-Viala, A.-J. Sánchez-Salmerón, Robust metric calibration of non-linear camera lens distortion. Pattern Recogn. **43**, 1688–1699 (2010)

6. C. Ricolfe-Viala, A.-J. Sánchez-Salmerón, Using the camera pin-hole model restrictions to calibrate the lens distortion model. Opt. Laser Technol. **43**, 996–1005 (2011)
7. A. Wang, T. Qiu, L. Shao, A simple method of radial distortion correction with centre of distortion estimation. J. Math. Imaging Vis. **35**, 165–172 (2009)
8. O.D. Faugeras, G. Toscani, The calibration problem for stereo, in *Proceedings of the IEEE Computer Vision and Pattern Recognition* (1986), pp. 15–20
9. O.D. Faugeras, G. Toscani, Camera calibration for 3D computer vision, in *Proceedings of International Workshop Industrial Application of Machine Vision and Machine Intelligence* (1987), pp. 240–247
10. J.Y. Weng, P. Cohen, M. Herniou, Camera calibration with distortion models and accuracy evaluation. IEEE Trans. Pattern Anal. Mach. Intell. **14**(10), 965–980 (1992)

# A Wavelet Transform-Based Image Mosaic Algorithm

**Qingfeng Sun and Hao Yang**

**Abstract** From the basis of studying in SIFT (Scale-Invariant Feature Transform) operator and wavelet transform, a multi-scale image mosaic algorithm is proposed. The proposed mosaic algorithm maintains the advantage that SIFT operator is invariant to scaling and rotation transformations, combining with the wavelet transform image fusion method, it enhance the anti-noise ability, maintains a certain stability to brightness change, improved the joint effect.

**Keywords** SIFT operator · Image mosaic · Wavelet transform

## 1 Introduction

Image mosaic is a integral image technology, through a series of operations include space registration, image transform, resampling, and image fusion, split two or more image which have overlapping portions into a large seamless image [1].

Image mosaic has a very wide range of applications, such as computer vision, resolution enhancement, medical imaging, and so on. Commonly, mosaic involves 3 steps: image preprocessing, image registration, image fusion. Image registration is one of the most important and foundation step which use to estimate the geometric transformation relationship of a set of images. Then, reproject the image in a big background merging both the overlapping portions and no overlap ones. The next step is to the discontinuities in the boundary between two images.

Q. Sun (✉) · H. Yang
Department of Electronic Engineering, Anhui Technical College of Mechanical and Electrical Engineering, Wuhu, Anhui, China
e-mail: sunqife@126.com

Q. Sun
School of Automation and Electronic Engineering, Tianjin University of Technology and Education, Tianjin, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_95

In recent years, many researchers proposed a lot of methods in image registration and fusion. Compared with many traditional registration algorithms, feature-based method such as SIFT and SURF detectors proved to be more efficient, more accurate, more reliable registration algorithm. This paper combines wavelet transform and SIFT operator, proposed an image mosaic algorithm.

## 2 Image Registration Based on SIFT

Image registration is the align process in space of over laying two or more images of the same scene taken at different times, from different view points, or by different sensors [1]. Image registration based on SIFT first finds the corresponding feature points of the overlapping parts of the two images, then calculated the transformation relation between the images by using these features, and then transformed the two images into the same coordinate system.

SIFT operator, called the scale-invariant feature transform, proposed by Lowe [2, 3], which is an image local feature descriptor based on the scale space theory. SIFT feature points on the image have invariance to scale and rotation, scaling, strong noise resistance characteristics.

1. Feature detection

Using Gauss's function as the convolution kernel to construct the scale space function.

Take $I$(x, y) to be input image, two-dimensional Gauss kernel function is defined as

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\left(x^2+y^2\right)/2\sigma^2} \tag{1}$$

Then build Gauss Pyramid:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \tag{2}$$

In order to detect the location key points in the scale space effectively, Lowe proposed to construct DoG (Difference of Gaussian) Pyramid by Gauss differential convolution (Fig. 1).

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma) \tag{3}$$

After establishing DoG scale space, in order to find its extreme points, compare each of the sampling points with all the 26 adjacent points of it, to ensure that in the scale space and the 2D image space are detected extreme point, then choose the point when it is the smallest or largest one. The points selected in this step are as the candidate points.

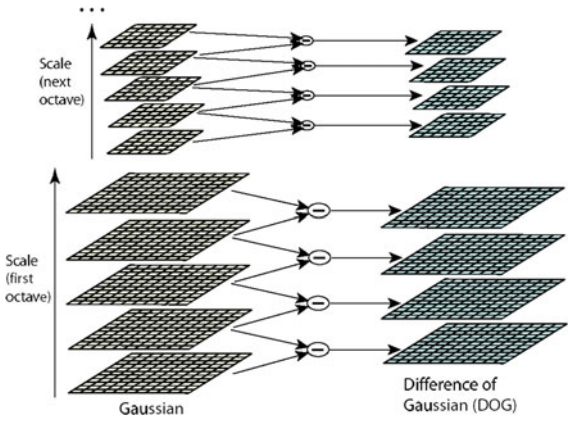



**Fig. 1** Scale space construction of SIFT operator

2. Feature point description

For each feature point $(x, y)$, its gradient $m(x, y)$ and direction $\theta(x, y)$ are calculated as follows:

$$\begin{aligned} m(x,y) &= \sqrt{(L(x+1,y)-L(x-1,y))^2+(L(x,y+1)-L(x,y-1))^2} \\ \theta(x,y) &= \tan^{-1}(L(x,y+1)-L(x,y-1))/(L(x+1,y)-L(x-1,y)) \end{aligned} \tag{4}$$

Among them, the L value is the size of each feature point.

Feature points as the center, take the 16 * 16 window, divided into 4 * 4 sub regions, each region through the gradient histogram statistics 8 directions, formation a 128 dimensional feature vector.

3. Feature point matching

SIFT operator feature points on the image have scale and rotation, scaling invariance, strong noise resistance characteristics. The feature points extracted by the traditional SIFT operator are invariant to rotation, scaling, translation, light, and so on. However, the SIFT operator is 128 dimensional, it has large computational complexity. Aiming to overcome the short comes of the high dimension of SIFT operator, the domestic and foreign research scholars try to improve it, in order to keep the good characteristics of the original operator, as far as possible to reduce its dimension, reduce the amount of computation. In these studies, PCA-SIFT [4] and GLOH [5] are more representative, each has its own characteristics, and are using PCA (Component Analysis Principal) technology to achieve dimension reduction processing.

In this paper, we use PCA technology to reduce the dimension to get a fast result. Then use the ratio of the nearest neighbor Euclidean distance feature points and the next nearest neighbor Euclidean distance of the feature point to get the matching points. Finally, use RANSAC (random sample consensus) algorithm [6] to remove

the false matching points. There have a special application in railway station that need image communication. The rapid growth of train speed make the packet data communication mode based on TCP/IP unable to deal with communication handover and other problems [7]. So we need to reduce the amount of computation.

## 3 Wavelet Transform-Based Image Fusion

Any pair of adjacent image cannot be exactly the same in the acquisition conditions. Therefore, for the same image, some characteristics in the two images showed not exactly the same. Image mosaic gap appear in the image when one image region transitions to another region, some relevant characteristics of the image has been changed by the jump. Image fusion is to make the splicing gap between the image not obvious, stitching more natural. Direct registration between the images will have a clear joint, at this time, it need to be smooth at the seams, that is image fusion. Image fusion is generally divided into three levels: Pixel level, feature level, symbol decision level. There are many image fusion methods in practical application, such as weighted pixel fusion method, hierarchical fusion method and time domain fusion, frequency domain fusion. In this paper, we choose wavelet based multi-scale decomposition and fusion.

1. Wavelet transform

One-dimensional continuous wavelet transform is as follows:

$$W_f(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{+\infty} f(x)\overline{\psi\left(\frac{x-b}{a}\right)}\,dx \tag{5}$$

where, the definition of wavelet transform is let a shift wavelet function $\overline{\psi(x)}$ inner product with signal f(x) which to need be analyzed in different scale among the formula, a is scale parameter, b is shift factor.

Wavelet transform applied in the field of image processing, mainly because of its decomposition and reconstruction. Through wavelet transform, we can get image information in different scales, high-frequency details and low-frequency overview, its convenient to deal with separately. In addition, through the reconstruction, we can get the image reappear. Especially after the propose of Mallat algorithm [8], wavelet transform began to be one of the most useful and quickly method in image processing.

2. Image fusion use wavelet transform

In recent years, with the development of wavelet theory and its application, wavelet multi-resolution decomposition is applied to image fusion. The wavelet transform-based method has the following advantages:

(1) Reconstruction ability well, ensure the signal without information loss and redundant information in the decomposition process;
(2) Image is decomposed into a average images and details image, representing the different structure of the image, and is easy to extract the structure and detail information of the original image.
(3) With a fast algorithm, its easy and fast to realization.

From the result of SIFT match, we can get the overlap part (AB) of two original image (image A and image B) and get a match image (C), then, use wavelet transform decomposition each images, for the low-frequency coefficients, using the method of weighted mean to fusion, for the high-frequency coefficients, use absolute value fusion rules, then using inverse wavelet transform to reconstructed the image, get mosaic image.

## 4 Simulation Results and Analysis

Mosaic algorithm steps:

First, feature extraction. Use SIFT operator to extract the feature points of two images (image A and image B), then, adopt nearest neighbor (NN) algorithm to match the image, use RANSAC method to eliminate the error matching point pair.

Second, image registration. Calculate the transform parameters between two images, determine the transformation relationship between the images, get a preliminary stitching result.

Finally, use wavelet transform fusion method to eliminate the mosaic trace.

Use two maps to analyze the proposed algorithm, Fig. 2 gives two images taken at different angles. Figure 3 shows the match results using SIFT descriptor, get 54 pairs of matching points, the stability is well, almost no error matching pair. In Fig. 4, we can see the final mosaic result using wavelet transform method, the edge treatment effect is better.

**Fig. 2** Two original images

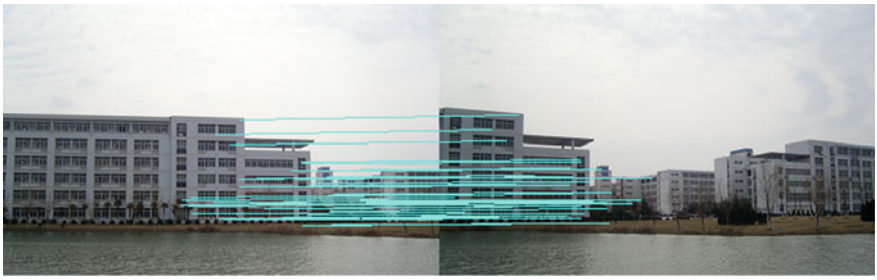


**Fig. 3** The match result based on SIFT operator

**Fig. 4** Image mosaic result

## 5 Conclusion

Combine SIFT-based image registration method with wavelet multi-scale technology, proposed an image mosaic algorithm. The experimental results show that the method proposed has high speed and good effect.

**Acknowledgements** Project fund: 1. Young teachers development and support program (project number: 2015yjzr028); 2. Anhui Quality engineering project:specialty of industrial robot (project number: 2015tszy080); 3. Virtual simulation experimental teaching center for industrial robot (project number:2016xnzx007).

## References

1. B. Zitova, J. Flusser, Image registration methods: a survey. Image Vis. Comput. **21**, 977–1000 (2003)
2. D.G. Lowe, Object recognition from local scale-invariant features, in *International Conference on Computer Vision*, vol. 2 (1999), pp. 1150–1157

3. D.G. Lowe, Distinctive image features from scale-invariant key points. Int. J. Comput. Vis. **60** (2), 91–110 (2004)
4. Y Ke, R. Sukthankar, PCA-SIFT: a more distinctive representation for local image descriptors. Comput. Vis. Pattern Recogn. 511–517 (2004)
5. K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell. **27**(10), 1615–1630 (2005)
6. M. Fischler, R. Bolles, Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography. Commun. ACM **24**, 381–385 (1981)
7. Z. Li et al., NDN-GSM-R: a novel high-speed railway communication system via Named Data Networking. EURASIP J. Wirel. Commun. Netw. **2016**, 48 (2016)
8. S. Mallat, A compact multiresolution representation: the wavelet model, in *Proceedings of IEEE Workshop Computer Vision* (1987), pp. 2–7

# An Effective SURE-Based Wiener Filter for Image Denoising

**Xiaobo Zhang**

**Abstract** The Wiener filter is widely used for image denoising. Although good performance can be obtained, the parameters such as the sizes of window and etc. still limit the performances of algorithm. To deal with this problem, the Stein's unbiased risk estimate (SURE) approach has been used in related methods. Although the proposed method employs the SURE strategy, unlike previous approaches the interscale and intrascale information of wavelet coefficient is adopted to improve the Wiener filter. Compared to the related state-of-the-art wavelet-based algorithms, the proposed method shows good performance but very simplicity.

**Keywords** Image denoising · Wiener filter · Stein's unbiased risk estimate (SURE)

## 1 Introduction

Noise removal for images has received considerable attention in last several decades. The improved Wiener filter methods have produced substantially outstanding performances. For example, under the frame of Wiener filter, Dabov et al. [1] proposed a collaborative image denoising algorithm by patch matching and sparse 3D transform (BM3D); Zhang et al. developed two-stage image denoising scheme by exploiting principal component analysis with local pixel grouping (LPG-PCA) [2]. Although these methods obtain excellent performances, their implementations are a little complex. At this time, inspired by anisotropic diffusion (AD) method [3], the multiple-step local Wiener filter algorithm (MSLWF) [4] was designed based on wavelet transform. Although the MSLWF obtains the desired results, it did not take advantage of parent coefficient so that the performance is limited. For this, this paper

X. Zhang (✉)
Institute of Graphics and Image Processing, Xianyang Normal University, Xianyang 712000, China
e-mail: zhangxiaobo9876@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_96

presents a new design of Wiener filter. The stein unbiased risk estimate (SURE) is used to determine the form of new threshold function (SURE-Wiener). To effectively and fairly demonstrate the performance of SURE-Wiener, as with MSLWF, the proposed method operates in dual tree complex wavelet transform (DT-CWT) domain since its approximate shift invariance and better directional selectivity [5]. Although the SURE strategy has been used to improve the Wiener filter [6–8], the novelty of the proposed method lies in that the interscale and intrascale information of wavelet coefficient is fully employed in the design of algorithm.

The rest of this paper is organized as follows. Section 2 presents the proposed denoising algorithm. Section 3 presents experimental results. Section 4 concludes the paper.

## 2 The Proposed Method

In this section, the denoising of an image contaminated by white Gaussian noise with zero mean and variance $\sigma_n^2$ will be done. Since DT-CWT is used, in each subband the complex wavelet coefficient of noisy image can be formulated as:

$$y_{i,j} = x_{i,j} + n_{i,j} \tag{1}$$

where $y_{i,j}$ is the noisy wavelet coefficient, $x_{i,j}$ is the true coefficient satisfying conditionally independent zero-mean Gaussian distribution and the variances are independent and identically distributed and local strongly correlated random variables, and $n_{i,j}$ can also be approximately considered as the noise which is Gaussian distribution with zero mean and variance $\sigma_n^2$ and independent of $x_{i,j}$. So, as with [4], it is possible that restoring the clean wavelet coefficients by employing Wiener filter. In the following, the proposed algorithm is described in detail.

### *2.1 Shrinkage Function Using Intrascale and Interscale Information of Wavelet Coefficient*

In order to use the information of neighboring coefficients in the process of denoising, one uses the following Wiener shrinkage:

$$\hat{y}_{i,j} = \omega_{i,j} y_{i,j} \tag{2}$$

where

$$\omega_{i,j} = \frac{\max(q_{i,j} - \lambda\sigma_n^2, 0)}{\max(q_{i,j} - \lambda\sigma_n^2, 0) + \lambda\sigma_n^2} \tag{3}$$

Equation (3) is taken from [7, 8], the $\lambda$ is used to tune denoising amount. The $q_{i,j}$ is obtained by averaging the squared values of coefficients in a window centered at $(i,j)$, according to the following way:

$$q_{i,j} = \frac{1}{(2R+1)^2} \sum_{k1=-R}^{R} \sum_{l1=-R}^{R} y_{i-k1,j-l1}^2 \tag{4}$$

In the following, for ease of notation, one arranges the noisy wavelet coefficient matrix with $y_{i,j}$ and $x_{i,j}(1 \le i \le M, 1 \le j \le N)$ into the vector form with $y_s$ and $x_s(1 \le s \le MN)$, respectively. So, the subscript $i,j$ of all symbols is substituted with $s$.

As with [8], the thresholding function incorporating intrascale information is expressed as:

$$\psi(y_s) = \alpha y_s + \beta \hat{y}_s \tag{5}$$

where $\alpha$ and $\beta$ are parameters to be set in the test.

Now, let $y_s^p$ denote the value of the parent coefficient corresponding to the noisy coefficient $y_s$. As with [9], the parent coefficient function

$$f(y_s^p) = e^{-(y_s^p)^2/(2T^2)} \tag{6}$$

is used as a discriminator between High Signal-to-Noise Ratio (SNR) wavelet coefficients and low SNR wavelet coefficients. $T$ is set to $\sqrt{6}\sigma$. By joining the interscale predictor (6) and intrascale estimator (5), the new Wiener filter denoising function is expressed as:

$$\theta(y_s) = e^{-(y_s^p)^2/(2T^2)}(a_1 y_s + a_2 \hat{y}_s) + (1 - e^{-(y_s^p)^2/(2T^2)})(a_3 y_s + a_4 \hat{y}_s) \tag{7}$$

where linear parameters $a_k(1 \le k \le 4)$ are solved for by minimizing the mean square error (MSE) estimate, the detail descriptions are presented in Sect. 2.3. In (7), let $\varphi_1(y_s) = y_s e^{-(y_s^p)^2/(2T^2)}$, $\varphi_2(y_s) = \hat{y}_s e^{-(y_s^p)^2/(2T^2)}$, $\varphi_3(y_s) = y_s(1 - e^{-(y_s^p)^2/(2T^2)})$ and $\varphi_4(y_s) = \hat{y}_s(1 - e^{-(y_s^p)^2/(2T^2)})$, and then (7) can be reformulated as:

$$\theta(y_s) = \sum_{k=1}^{4} a_k \varphi_k(y_s) \tag{8}$$

## 2.2 Denoising Algorithm

The denoising algorithm is as follows:

(1) Perform a $J$ level DT-CWT to the original noisy image.

(2) For each subband (except the low-pass residual), compute the interscale predictor $\{y_s^p\}$ by expanding the parent by a factor two. At the same time, in order to increase the homogeneity inside the regions of similar coefficients, Gaussian smoothings are used in wavelet parent coefficients and in output of Wiener filter of wavelet coefficients using (2).
(3) Compute the estimated value $\theta(y_s)$ according to (8).
(4) Reconstruct the denoised image from the processed subband and the low-pass residual.

## 2.3 Choice of Parameters and Analysis of Performance

The proposed new Wiener filter (SURE-Wiener) (8) is similar in form to the thresholding function in SURE-LET in [9]. So, as with [9], the SURE approach is used to determine the weight factor of the proposed thresholding function.

In practice, the following random variable:

$$\tilde{\varepsilon} = \frac{1}{MN}\sum_{s=1}^{MN}\left(\theta^2(y_s) - 2y_s\theta(y_s) + 2\sigma^2\theta'(y_s)\right) \tag{9}$$

is used to represent an unbiased estimator of the MSE. Introducing (8) into the estimate of the MSE and performing differentiations over $a_k$, obtaining for all $k$

$$\sum_{l=1}^{4}\underbrace{\langle\varphi_k(y)\varphi_l(y)\rangle}_{Q_{k,l}}a_l - \underbrace{\langle y\varphi_k(y) - \sigma^2\varphi_k'(y)\rangle}_{c_k} = 0 \tag{10}$$

These equations can be summarized in matrix form as $Qa = c$, where $a = [a_1, a_2, a_3, a_4]^T$ and $c = [c_1, c_2, c_3, c_4]^T$ are vectors of size $K \times 1$, and $Q = [Q_{k,l}]_{1\le k,l\le 4}$ is a matrix of size $4 \times 4$. This linear system is solved for $a$ by

$$a = Q^{-1}c \tag{11}$$

which lets the thresholding function adaptive to the different subbands.

In practice, as with [8], the $\theta'(y_s)$ is computed as

$$\theta'(y_s) = e^{-(y_s^p)^2/(2T^2)}(a_1 + a_2\omega_s) + (1 - e^{-(y_s^p)^2/(2T^2)})(a_3 + a_4\omega_s) \tag{12}$$

In above, $(\varphi_2(y_s))' = \omega_s e^{-(y_s^p)^2/(2T^2)}$ and $(\varphi_4(y_s))' = \omega_s(1 - e^{-(y_s^p)^2/(2T^2)})$. In the test, $\lambda$ is set to 1.33 by hand. And the sizes of square window are taken as 7 × 7, 3 × 3, 3 × 3, 3 × 3, 3 × 3 in turn from the finest scale to the coarsest scale ($J = 5$) also by hand.

## 3 Experimental Results

In this section, the results of two closed methods such as UWT SURE-LET [10] and SURE-NeighShrink [6] are used to compare. The result of the state-of-the-art MSLWF [4] employing Wiener filter are also presented to verify the effectiveness of the proposed method. One has tested stand 8-bit grayscale images such as Lena and Barbara (size 512 × 512), corrupted by simulated additive Gaussian noise at three different power levels $\sigma \in [15, 20, 25]$. The parameters for all kinds of methods have been set in term of the values given by their respective authors in the corresponding papers. These comparisons were reliable because that the various authors have provided their respective Matlab source codes on the related websites.

In the test, assume that the noise variance is known for all methods. The PSNR measure is used to evaluate the performance of all methods. Let $\tau$, $\upsilon$ and $\eta$ denote the original, the denoised image and the number of pixels, respectively. The PSNR in decibels is given by

$$\mathrm{PSNR} = 10\log_{10}\left(255^2 / \left(\frac{1}{\eta}\sum_{\gamma}(\tau_\gamma - \upsilon_\gamma)^2\right)\right) \tag{13}$$

Table 1 summarizes the results obtained. It can be seen that the proposed method gets higher PSNRs compared to MSLWF, SURE-NeighShrink and UWT SURE-LET.

The computational burden of the filters is measured by CPU time of obtaining optimal PSNR provided those methods are denoising the same image on the same computer. From Table 2, it can be seen that the proposed SURE-Wiener takes the least time in three SURE methods. Although the computational time with SURE-Wiener is more than that of MSLWF by 1.19 s, the high PSNRs obtained by the proposed filter justify this amount of the high computational burden.

**Table 1** The PSNRs (in dB) presented by the different noise levels

| σ | 15 | 20 | 25 | 15 | 20 | 25 |
|---|---|---|---|---|---|---|
| Method | Lena | | | Barbara | | |
| MSLWF | 33.68 | 32.42 | 31.44 | 31.72 | 30.18 | 29.01 |
| SURE-NeighShrink | 33.58 | 32.30 | 31.25 | 31.63 | 30.07 | 28.88 |
| UWT SURE-LET | 33.33 | 32.08 | 31.11 | 30.15 | 28.44 | 27.17 |
| SURE-Wiener | **33.79** | **32.54** | **31.55** | **31.72** | **30.18** | **29.02** |

**Table 2** Computational burdens (second) for different denoising methods

| Method | UWT SURE-LET | SURE-NeighShrink | MSLWF | SURE-Wiener |
|---|---|---|---|---|
| Unite of time | 7.75 | 48.70 | 1.92 | 3.11 |


**Fig. 1** Comparison of the restoration results from the different methods. Zoom into file for a better view. **a** Original Lena image; **b** noisy image ($\sigma = 25$); **c**–**f** shows restored Lena images using MSLWF, SURE-NeighShrink, UWT SURE-LET and proposed SURE-Wiener, respectively

In addition, it is noted that the proposed method does not sacrifice the visual performance of denoised images. Figure 1 shows the comparison between several methods when applied to the noisy Lena image with noise standard deviation 25. It can be seen that the results with UWT SURE-LET are worst. And other methods achieve similar results. The details such as fairs and so on are preserved well. Although the visual differences are very subtle, it still can be found that there exist more artifacts in the results of MSLWF and SURE-NeighShrink in the smooth regions compared to the proposed method.

## 4 Conclusion

In this paper, an effective image denoising algorithm is proposed. It is based on an integration of the local Wiener filter, DT-CWT and the interscale and intrascale information of wavelet coefficient. This method can be considered as supplementary of the Wiener filter method by employing SURE strategy. The tests show the proposed method obtains the desired performance compared to related algorithms.

**Acknowledgements** This work is partially supported by National Nature Science Foundation of China (Grant No. 61401383) and Natural Science Foundation of Xianyang Normal University (Grant No. 14XSYK006) and Qinglan Talent Program of Xianyang Normal University (Grant No. XSYQL201503).

## References

1. K. Dabov, A. Foi, V. Katkovnik, K. Egiazarian, Image denoising by sparse 3D transform-domain collaborative filtering. IEEE Trans. Image Process. **16**(8), 2080–2095 (2007)
2. L. Zhang, W. Dong, D. Zhang, G. Shi, Two-stage image denoising by principal component analysis with local pixel grouping. Pattern Recogn. **43**(4), 1531–1549 (2010)
3. P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion. IEEE Trans. Pattern Anal. Mach. Intell. **12**(7), 629–639 (1990)
4. X. Zhang, X. Feng, Multiple-step local Wiener filter with proper stopping in wavelet domain. J. Vis. Commun. Image R **25**(2), 254–262 (2014)
5. I.W. Selesnick, R.G. Baraniuk, N.C. Kingsbury, The dual-tree complex wavelet transform. IEEE Signal Process. Mag. **22**(6), 123–151 (2005)
6. D. Zhou, W. Cheng, Image denoising with an optimal threshold and neighbouring window. Pattern Recogn. Lett. **29**(11), 1694–1697 (2008)
7. X. Zhang, The SURE-LET approach using hybrid thresholding function for image denoising. Comput, Electr. Eng. (2016) (in press)
8. X. Zhang, Image denoising using local Wiener filter and its method noise. Opt. Int. J. Light Electron. Opt. **127**(17), 6821–6828 (2016)
9. F. Luisier, T. Blu, M. Unser, A new SURE approach to image denoising: inter-scale orthonormal wavelet thresholding. IEEE Trans. Image Process. **16**(3), 593–606 (2007)
10. T. Blu, F. Luisier, The SURE-LET approach to image denoising. IEEE Trans. Image Process. **16**(11), 2778–2786 (2007)

# Part X
# Circuit Processing System & System Design

# Simulation Analysis and Improved Design of the Control System of a Certain Missile

**Xiwei Guo, Shen Zhao, Peng He and Jianhua Xie**

**Abstract** To validate the performance of a certain type of anti-tank missile under altiplano environment, the operate principle of control system is analyzed first, and then the simulation model based on Simulink is established. The stability and dynamic performance of the missile is analyzed through the curve of flight and attack angle. Toward the instability of control system under altiplano environment, the emendation network is designed, and it is testified that the stability of control system is improved.

**Keywords** Anti-tank missile · Control system · Simulation analysis

## 1 Introduction

Certain type of anti-tank missile is mainly used to attack the armor target and other fortifications in plain area. It is crucial to analyze and validate the working performance of the control system, to the demand of combat mission at high altitudes [1]. The truth data of the performance can be acquired by flight test, but the test method involves the expense limitation and environment condition restriction, so it is impossible to obtain the all required data by launching mass expensive missile, nor to acquire the true data in different test conditions [2]. While the method by computer simulation take the advantage of expense/cost saving comparatively [3]. The paper adopts the simulation method to analyze the working performance of the anti-tank missile control system and proposes the improving method to the existent problem.

X. Guo (✉) · S. Zhao · P. He · J. Xie
Department of Missile Engineering, Ordnance Engineering College, Shijiazhuang, China
e-mail: xixinting@sohu.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_97

## 2 Modeling of the Control System

The control and guide equipment of the missile implement the guidance trajectory in three-point method. The shooter manipulates the elevating and traversing mechanism continually after the missile launched, to control cross-mark on the aiming mirror aim at the target. In the flight course, the shooter must observe flight state of the missile and guide it to the target. The guide equipment measures the angular deviation of the missile with the line-of-sight, and brings forward the control instructions, which is used to guide and control the missile to the line-of-sight. The course of guidance is equivalent to the process of measuring and eliminating the angular deviation [4–7].

As the key components of the whole system, the control and guide equipment of certain type of missile is constitutive of three main mechanisms (goniometer, control box, and the missile body), including nine functional modules (angle measurement block, amplitude limitation block, range transform block, overshoot network block, gravity compensating instruction block, nonlinear control block, missile body, movement block and the view angle transform bock) [8]. The framework of the functional modules is denoted in Fig. 1.

The flight trajectory model of the controlled missile can be established based on the framework. Some blocks could be simplified considering the complexity of the true system. According to analysis measure of the control system, the missile body in pitch channel is simplified into second-order mode, and the transfer functions of the missile body and attack angle can be obtained. The transfer function of the input instruction to the inclination angle change of the flight trajectory is derived as [8]

$$W_K^{\dot{\Theta}}(s) = \frac{k_K^{\dot{\Theta}}\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \tag{1}$$

The block shown in Eq. (1) is named as the missile body block commonly. The transfer function of attack angle $\alpha$ with the input is

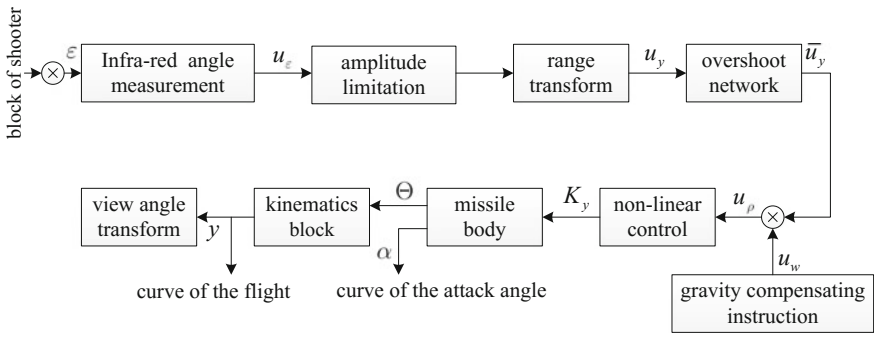



**Fig. 1** Framework of the control and guide equipment

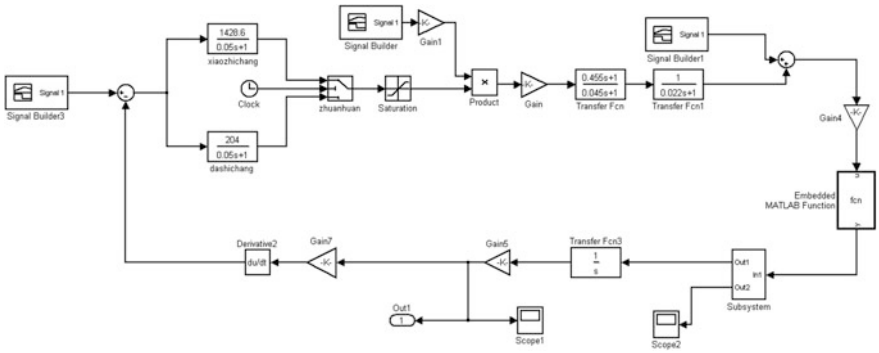



**Fig. 2** Simulation model of the whole system

$$W_K^{\alpha}(s) = \frac{k_K^{\alpha}\omega_n^2}{s^2 + 2\xi\omega_n s + \omega_n^2} \tag{2}$$

where, $\omega_n = \sqrt{a_2 + a_1 a_4}$ is the inherent frequency, $k_K^{\dot{\Theta}} \approx \frac{a_3 a_4}{a_2 + a_1 a_4}$ and $k_K^{\alpha} \approx \frac{a_3}{a_2 + a_1 a_4}$ are amplification coefficients, and $\xi = \frac{a_1 + a_4}{2\sqrt{a_2 + a_1 a_4}}$ is the relative damping coefficients, respectively [9]. Further, parameters $a_1$, $a_2$, …, $a_5$ are the dynamical coefficients of the pitching movement, with unit $s^{-1}$.

The simulation model of the whole system based on Simulink can be implemented in the form shown in Fig. 2.

## 3 Simulation and Analysis

According to the index demands such as the running speed of the system and precision, the fourth-order Runge–Kutta method is adopted to calculate the integrator. The simulation step and initial value are setup, and the simulation result of the curve of flight and attack angle are analyzed further.

### *3.1 Characteristic of the Flight*

The flight course is simulated, with altitude of 4500 m and stationary target distance at 2500 m. The curve of the flight track on vertical plane, or the position of the center of gravity vary with time is exhibited in Fig. 3.

According to the simulation result, with the current altitude and parameter setup, the missile falls down at 8.41 s. Besides, the vibration of the curve shows that the

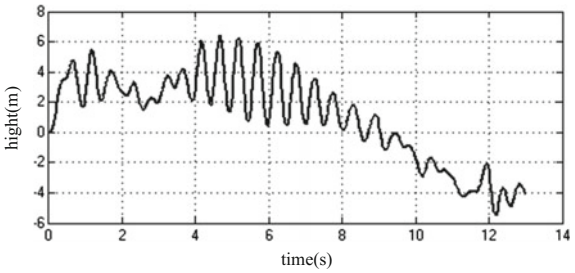



**Fig. 3** Curve of the flight on vertical plane

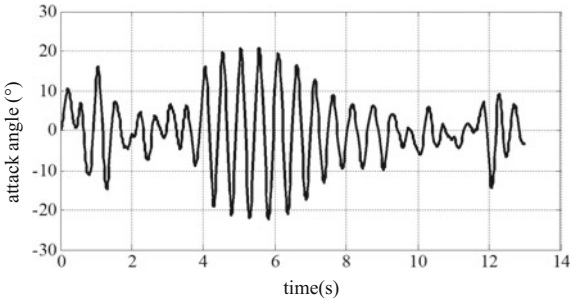



**Fig. 4** Curve of the attack angle

flight stability is badly within the whole 13 s simulation course. The vibration is violent especially after the simulation runs after 4 s, and the height diversity shocks with 5 m maximal. It can be concluded that the missile is probably out of control with such flight fluctuation.

## *3.2 Characteristics of the Attack Angle*

The curve of the attack angle throughout the simulation process, shown in Fig. 4, is investigated further to analyze the cause of falling down. It is visible that the attack angle behaves a divergent trend, especially after the simulation runs after 4 s. The maximum of the divergency peaks at ±20°, so the stability of the missile is destroyed in the simulation process.

## *3.3 Analysis on the Performance*

With the simulation result, it is crucial to investigate the impact on the stability of the control system induced by the atmosphere density change for the high altitude. The stability of the control system is turned into the stability of an equivalent

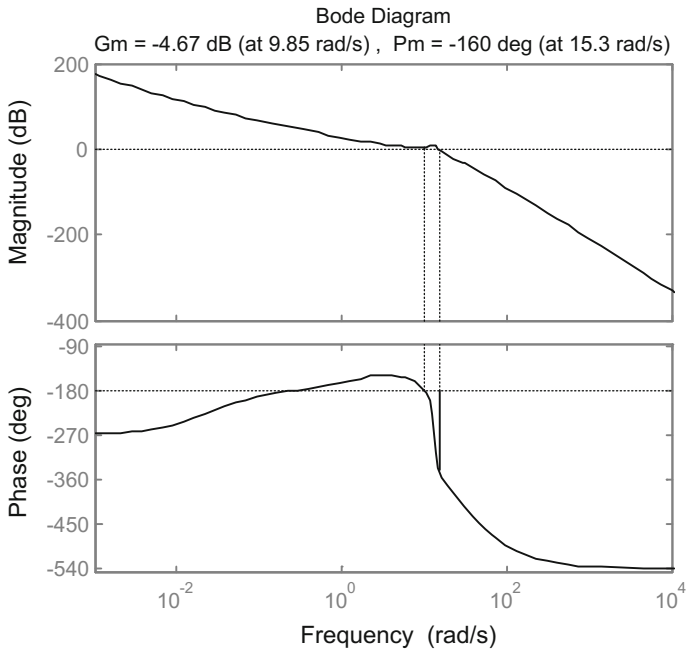



**Fig. 5** Open-loop bode diagram of the control system within 3 s

minimum phase systems [10]. With the contract method, it is considered that the atmosphere density decreases by a factor of 0.6 compared with the plain area.

Applying the atmosphere density variety and other condition steady, according to the framework of the control system, it can be derived that the open-loop gain

$$\begin{aligned} G_k(s) &= [W_e(s) \cdot W_D(s) + G(s)] \cdot W_K^{\dot{\Theta}}(s) \cdot W_{\Upsilon}(s) \\ &= \left[ \frac{204}{0.05s+1} \frac{0.455s+1}{1.5(0.045s+1)(0.022s+1)} + \frac{3.31s+0.0678}{s^2} \right] \frac{23.11}{s^2+1.93s+121.13} \frac{1}{s^2} \\ &\approx \frac{0.0037s^3 + 1430.4s^2 + 3152s + 76.49}{s^3(0.05s+1)(0.045s+1)(0.022s+1)(s^2+1.93s+121.13)} \end{aligned} \tag{3}$$

The corresponding open-loop bode diagram can be calculated, as shown in Fig. 5. Where the amplitude across frequency is $\omega_c = 13.685$ rad/s, the gain stability margin is $h = 0.342$, the phase crossover frequency is $\omega_g = 8.683$ rad/s and the phase stability margin is $\gamma = -160.144°$, respectively.

To the basic stability criterion of the minimum phase systems, the amplitude across frequency $\omega_c$ is larger than the phase crossover frequency is $\omega_g$, which illuminates that the original control system is unstable at high altitudes. The amplitude of the vibration is so distinct and time of the vibration is excessive, which

is also a validation of the unstability. Meanwhile, low crossover frequency induces that the adjusting time of the system is so long, which bring that the response to the instruction is not in time adequately.

## 4 Design Improvement

With the simulation result and analysis of the control system, it can be concluded that the correcting block of the original anti-tank missile is designed for the plain area, and not suited for the high altitudes condition. The missile may fall down at high altitudes in the flight process, with the outside cause of the working conditions change. The atmosphere density decreases heavily with higher altitude, which make the inflect of the control object (or the transfer function of the missile body) $W_K^{\Theta}$. Along with the altitude grow up to 4500 m, the atmosphere density decreases by 0.6 with the plain area, and inducing the five coefficients $a_1$, $a_2$, $a_4$, $\omega_n$ and $\xi$ to be change. As these coefficients play an key value in the stability of the missile, their continuously change bring in the unstable and falling down of the missile.

### *4.1 Lag–Lead Correction Design with the Bode Diagram*

According to experimental data of the classical control theory [11], the lead correction method affords the positive angle, conducing higher cut-off frequency, while the phase margin it offers only in amount of 40–60°. The lag correction method makes the cut-off frequency lower, with cut-off frequency losing in exchange with a lager phase margin. So the problem cannot be solved by only adopting the lead or lag correction. The only way is to impose the lead and lag correction jointly to achieve the desire goal, which means that a lager phase margin and more stable of the system.

The transfer function of the lag–lead correction network is

$$G_c(s) = KG_{c1}(s) \times G_{c2}(s) = K\frac{1+T_1 s}{1+\beta T_1 s} \cdot \frac{1+T_2 s}{1+\alpha T_2 s} \tag{4}$$

In the network designing, the tune parameters at the feature points should be optimized, and then the correction coefficients in flight course could be established by linear interpolation method. The designed targets are included below.

(1) To guarantee the stability margin $h$ of the system, the phase margin $\gamma$ must be no less than 30°.
(2) To avoid the resonation of the missile body, the cut-off frequency of the closed-loop must be lower than a factor of 0.3 by the natural frequency of the missile body, which means $\omega_c \leq 14.13$ rad/s.
(3) To prevent the falling down of the missile in flight course, it pursues a smaller system overshoot.

According to the classical automatic control theory and the designing step of the phase lag correction based on Bode diagram, the transfer function of correction is particular as

$$G_c(s) = KG_{c1}(s) \times G_{c2}(s) = \frac{1}{1.65} \cdot \frac{0.8333s+1}{7.083s+1} \cdot \frac{1.232s+1}{0.1225s+1} \tag{5}$$

## 4.2 Frequency Domain Analysis Validation Method with the Bode Diagram

The transfer function of correction in Eq. (5) can be substituted into the original control system loop, and the transfer function of the whole system can be derived as

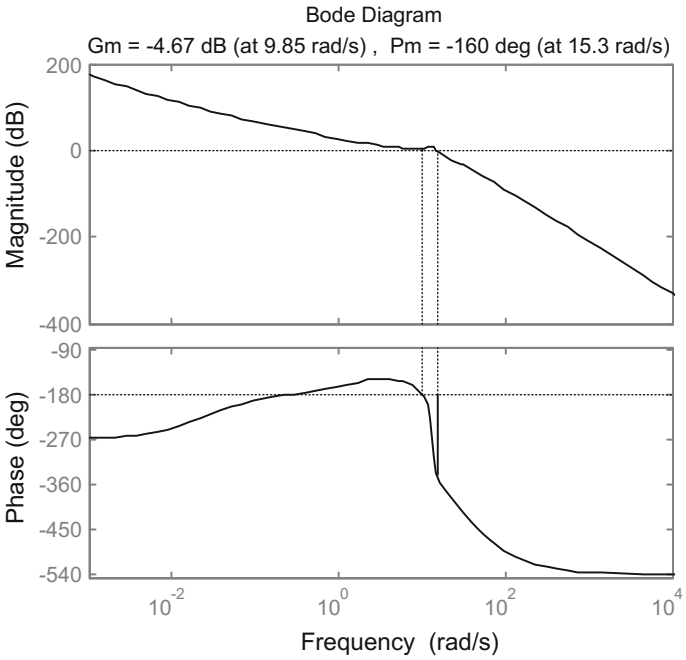



**Fig. 6** Open-loop bode diagram of the control system within 3 s

$$
\begin{aligned}
G_k(s) &= [W_\varepsilon(s)\cdot G_c(s)+G(s)]\cdot W_K^{\dot{\Theta}}(s)\cdot W_Y(s) \\
&= [\frac{204}{0.05s+1}\cdot\frac{1}{1.65}\cdot\frac{0.8333s+1}{7.083s+1}\cdot\frac{1.232s+1}{0.1225s+1}+\frac{3.31s+0.0678}{s^2}]\cdot\frac{23.11}{s^2+1.93s+121.13}\cdot\frac{1}{s^2} \\
&\approx 23.11\cdot\frac{127.98s^3+84.66s^2+150.91s+3.31}{s^3(0.05s+1)(7.083s+1)(0.1225s+1)(s^2+1.93s+121.13)}
\end{aligned}
\tag{6}
$$

To analyze the meliorated transfer function in frequency domain, the Bode diagram of the open-loop system is depicted in Fig. 6. In the figure, M is the gain margin of open-loop (in unit dB), and P is the phase margin (in unit degree).

It shows in the Fig. 6 that the amplitude across frequency $\omega_c \leq 3.1191$ rad/s, the phase stability margin $\gamma = 46.2074°$, the phase crossover frequency is $\omega_g = 8.4006$ rad/s and the phase stability margin is $\gamma = 46.2074°$, respectively.

According to the index, the phase stability margin satisfied the common criterion ($30° \leq \gamma \leq 70°$) in engineering. Besides, the phase crossover frequency is lower than the demanded target (14.13 rad/s) ultimately. Above all, it fulfills the goal of stability for the missile flight at high altitudes by means of lag–lead correction design with the Bode diagram.

## 5 Conclusion

Based on establishing the model of the control system of the missile, the stability and the dynamic characteristic of the control system are analyzed under altiplano environment, which draws the conclusion that the control system is unstable at high altitudes. The atmosphere density decreases with altitude higher, which changes the transfer function of the control system and is not able to afford control demand. A method based on lag–lead correction is proposed to improve on the control system to be accommodated at high altitudes. The analysis and validation based on Bode diagram shows that the meliorated control system is stable under altiplano environment.

## References


1. Boli Zhu, Hao Chang, Zhanyuan Qi, Study of precision improving for spinning anti-tank missile's guided system. J. Syst. Simul. **7**(19), 3361–3364 (2007)
2. Wei Jiang, Guocai Dong, Discussion on the anti-tank missile simulation experiment. Technol. Found. Natl. Def. **7**, 55–56 (2007)
3. Jianwei Chen, Fu Huimin, Dynamic consistency test method for air-to-air missile simulation results. J. Syst. Simul. **20**(19), 5121–5124 (2008)
4. Z. Li, K. Liu, Y. Zhao et al. MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)

5. Z. Li, Y. Chen, H. Shi et al. NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. **48**, 1–5 (2016)
6. X. Liu, Z. Li, P. Yang et al. Information-centric mobile ad hoc networks and content routing: a survey. Ad Hoc Netw (2016). doi:10.1016/j.adhoc.2016.04.005
7. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.02.009
8. X. Li, Q. Fang, *Dynamic of Winged Missile* (Northwestern Polytechnical University Press, 2004)
9. Yimin Wen, Xu Duosheng, The analysis on dynamical stability of winged missile. Aerosp. Control **1**, 41–47 (2001)
10. Z. Koruba, L. Nocon, Programmed control of the flat track anti-tank guided missile, in *Proceedings of the 15th International Carpathian Control Conference* (2014), pp. 237–242
11. Hua Geng, Geng Yang, Algebraic loop problems in simulation of control systems and the methods to avoid it. Electric Mach. Control **10**(6), 632–635 (2006)

# DDRII SDRAM Memory Controller Interface Design and Application Based on Virtex-5 FPGA

**Binfei Li, Jun Liu and Fudong Zhou**

**Abstract** For the currently the most widely used data storage memory is DDRII SDRAM which is used in the high-speed, high-precision, and high-memory depth of the data storage and communication system, using ISE software and calling the IP core of Xilinx to create MCB, adopting the Verilog HDL to achieve a common DDRII SDRAM controller interface for universal design and application based on Xilinx's FPGA chip and industrial standard. This paper deeply analyses the working principle of DDRII SDRAM, focusing on the read operation, write operation and refresh operation and validates the design of accuracy and stability in the latest Xilinx Virtex-5 series of FPGA platform.

**Keywords** DDRII · Controller · IP core · ChipScope pro

## 1 Introduction

In recent years, due to the FPGA (Field Programmable Gate Array) in the field of digital circuit design powerful parallel, high-speed, reusable, site design, programming, verification, modification, analysis characteristic has been the rapid development and application. With the high-speed data communication and storage requirements of growing, processor and interface connection peripheral speed continues to improve and the traditional static memory has been difficult to meet the requirements. DDRII SDRAM is a kind of new structure, high-speed, low-cost, high-data throughput of the static random access memory based on embedded system and FPGA's high-speed storage system has been the urgent demand, but in a high-speed memory controller interface design more difficult. This paper using

B. Li (✉) · J. Liu · F. Zhou
Tibet Vocational Technical College, Lhasa 850000, Xizang, China
e-mail: xiaofei_83@sohu.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_98

Xilinx ISE12.2 software and the latest Virtex-5 series of 601 development board, through research on the working principle of DDRII SDRAM, using Verilog hardware language implementation of DDRII SDRAM controller interface design, and at the operating frequency 333.3 MHz of the hardware system to verify the write and read out the data consistency and stability, simplify the design cycle, reduce the design cost.

## 2 SDRAM DDRII Overview

SDRAM DDRII is the second generation of double-data-rate synchronous dynamic access memory. Its English full name is: Double-Data-Rate Two Synchronous Dynamic Random Access Memory. The I/O DDRII buffer uses a technique called "Pumping double" to transfer data to the rising and falling edges of the clock signal. The core frequency of DDRII memory unit is prefetching queue equivalent frequency of one-fourth (and not 1/2), which requires a 4-bit-deep in without changing the memory unit itself, DDRII can effectively achieve the DDR data transmission speed which is twice that of [1]. Its main features are in the rising and falling edge of the clock and data transmission; DDRII SDRAM read and write supports burst access, starting from the selected location, you can choose four or eight burst length, the DDRII SDRAM read and write begins with a registered activation command information, registration information after the completion of the read and write operations; allows the column strobe; commands the first clock cycle in the inserted strobe command, column gate command in additional delay AL (Additive Latency) behind remain effective, which can improve the utilization rate of the memory bus driver; support offline calibration of OCD (Off-Chip Driver), which reads the voltage/write operation is consistent by adjusting the pull-up resistor value and pull-down resistor value, thereby reducing the inclination of DQ and DQ_N, in order to improve the integrity and consistency of the signal by 1.8 V working voltage. And 0.9 V pull in voltage, voltage lower than DDR SDRAM, thereby reducing the overall power consumption and heat generation; provide a built-in end resistance ODT (On Die Termination), can improve the signal integrity and increase the timing margin, reduce the multiple reflection signals, and DDR SDRAM compared to reduce to prevent the use of data terminal reflection signal a lot of end resistance; used FBGA package, with better electrical performance and heat dissipation performance, provides a good guarantee for the stability of the controller.

## 3 SDRAM DDRII Working Way

SDRAM DDRII must first be based on the actual application of the initial operation, the DDRII of the work mode and timing parameters, etc., and then can be read/write operation. DDRII using burst mode data transmission and continuous

transmission involves the storage unit (column) number is burst length, transmission when given only the starting address and burst length, can automatically on the back of a corresponding number of columns for read/write operations. Since the DDRII uses the capacity to store data information, in order to save the internal data, every time it is to perform a refresh command for each line. DDRII internal storage unit is managed by bank, according to the size of the capacity is generally divided into 4 or 8 bank [2]. Each bank is divided into rows and columns, the column address strobe signal CAS (Column Address Strobe) select rows and columns, bank width is memory chip data width. At work, allowing only read/write operations to open, if you want to read/write operations to other lines in the current bank shall be first with a precharge ordered the closure of the current line, and then activated command opened for read/write operations. Send the active command to give bank address and the row address, select line to be opened up and wait for a certain time interval and then to the column address. Because of the high frequency of DDRII, the data window is very narrow, in order to ensure that the data acquisition is accurate, DDRII uses the differential signal DQ, DQ_N to collect data. Write operation, DQ signal emitted by the controller, DQ signal and data window central alignment; read operation, DQ signal issued by the DDRII Memory, DQ signal and data window edge aligned, the controller receives a DQ signal to DQ signal and data window phase offset 90 degrees, make the DQ signal and data window center align [3].

## 4 DDRII Controller Interface Design

According to the working mode and characteristics of the DDRII SDRAM memory, combined with the actual usage, the DDRII controller designed in this paper mainly realizes the following functions: automatic sending activation, refresh, a precharge command, users only need to send write/read command; correctly implement the burst length is 8 read/write operations; the DQ of the data channel for the DDRII and 16 bit, automatically complete the initialization of DDRII; data input/output is 128; working frequency 333.3 Hz [4].

## 5 DDRII Controller Structure

The DDRII controller is mainly composed of a control unit and a MCB, and the control unit is valid after the initial signal (Calib_done) of the MCB is received, and the user's write/read instruction is executed normally. In order to facilitate the

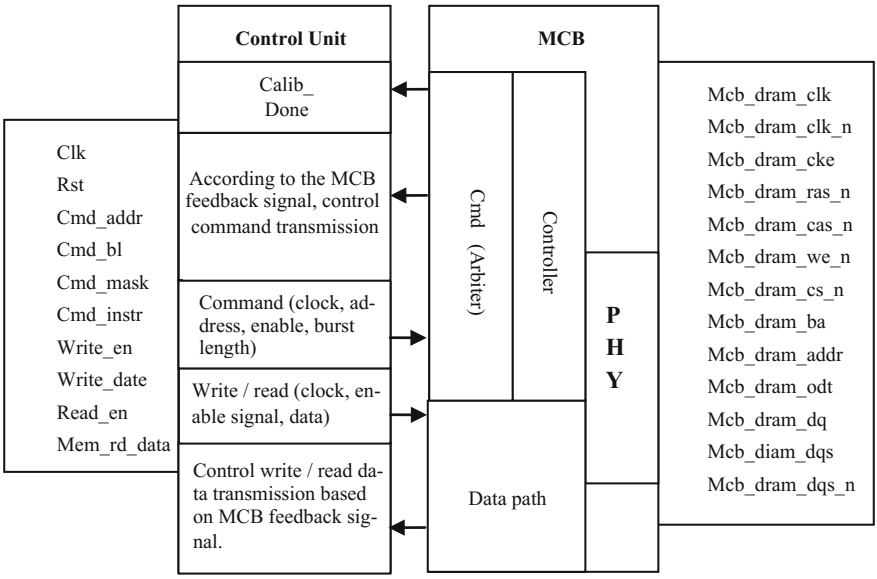



**Fig. 1** DDRII controller structure

control, this design adopts the automatic refresh mode. At every clock cycle refresh controller sends the refresh command, performs a refresh command, the controller will temporarily stop the operation signal end user, wait until after the completion of the operation, the controller will again enable signal operation, the user can continue to send the read/write command refresh. The command input port of the MCB depth 4 FIFO, data input and output ports were depth for 64 FIFO [5] control unit according to the feedback signal of the FIFO control write/read instructions and data sent correctly, ensure the continuity of the data and address. MCB according to the users write/read instructions, the data will be written in accordance with the work of DDRII or read out SDRAM DDRII. The structure of DDRII controller is shown in Fig. 1.

## 6 DDRII Controller Function

In the system after power controller is first performed is the initialization process, mainly to complete configuration of DDRII SDRAM mode register and extended mode register, of DDRII perform global precharge, indicating signal (Calib_done)

done effective, then performs a read/write operation. In the process of reading/writing, if you encounter a refresh request, the priority of the refresh command is higher than the read/write operation; so the first implementation of the refresh command. If the refresh operation conflicts with read/write operation, the controller will notify the user to stop sending read/write commands, and wait for the existing read/write operation to complete, and then refresh the operation.

When the initial indication signal effectively, the user writes which can efficiently, the control unit according to the feedback signal of the MCB, the data is written first, write FIFO, after reaching the set burst length, start the write command and the address, the data is written to DDRII; user read enable effective control unit according to the feedback signal of the MCB, first issue read command and address. Data is then entered into read FIFO, control unit according to the MCB feedback signal readout data. Write/read operation, the controller first according to the mapping of the row addresses to determine whether the operation is active, if not activated and in the bank to transmit activation command sends the column address and the read write command; if not activated and in the same bank, first to send precharge ordered the closure of the open line, followed by sending an activation command and read and write command; if it has been activated, sent directly to the column address. In the implementation of the read operation, the read command issued, according to the DQ signal to receive the read data back, the dual rate of 16 bit data into a single rate of 128 bit data to the user port.

## 7 Verification of FPGA Controller for DDRII

We used Spartan company's Xilinx 601 development board to verify the design of the DDRII controller. The development board has a model for the FPGA XC6SLX16 chip, models for the DDRII E1116AEBG chip, and the compiler tool for ISE12.2 Xilinx. The basic principle is to call MIG Memory (Interface Generator Xilinx) 3.5 of the IP core to produce Controller Block Memory (MCB) [6]. Based on this MCB, the compiler control unit for DDRII to read/write, self-refresh and other operations. Through the simulation and on-chip logic analyzer chip scope

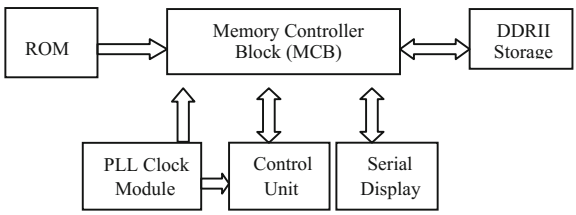



**Fig. 2** System block diagram

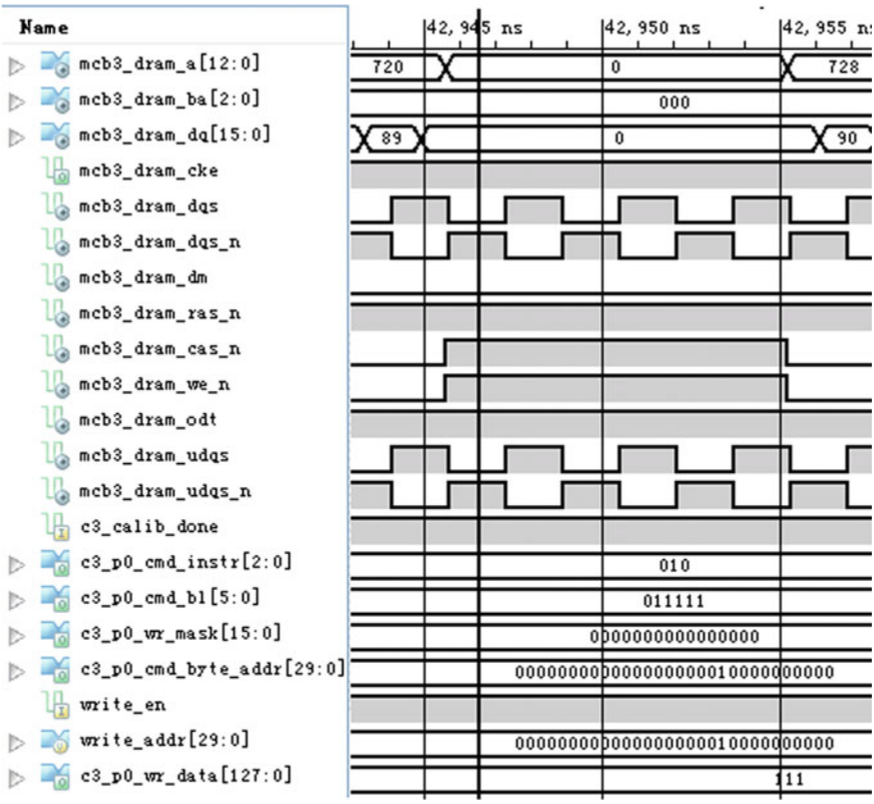



**Fig. 3** Read emulation timing

Pro) to write and read timing calibration and detection of write and read data consistency, and through the serial port to read data display in computer design verification of correctness.

Based on verification system, by the control unit sends the write instruction, the ROM data readout and with the write command sent to the MCB, MCB according to the DDRII operating mode will be a corresponding data according to the address of the corresponding to 8 for burst length is stored into the DDRII, and then by the control unit issued the read instruction, according to the address of the corresponding and burst length from the DDRII read data and transmission to serial (UART). The verification system block diagram is shown in Fig. 2.

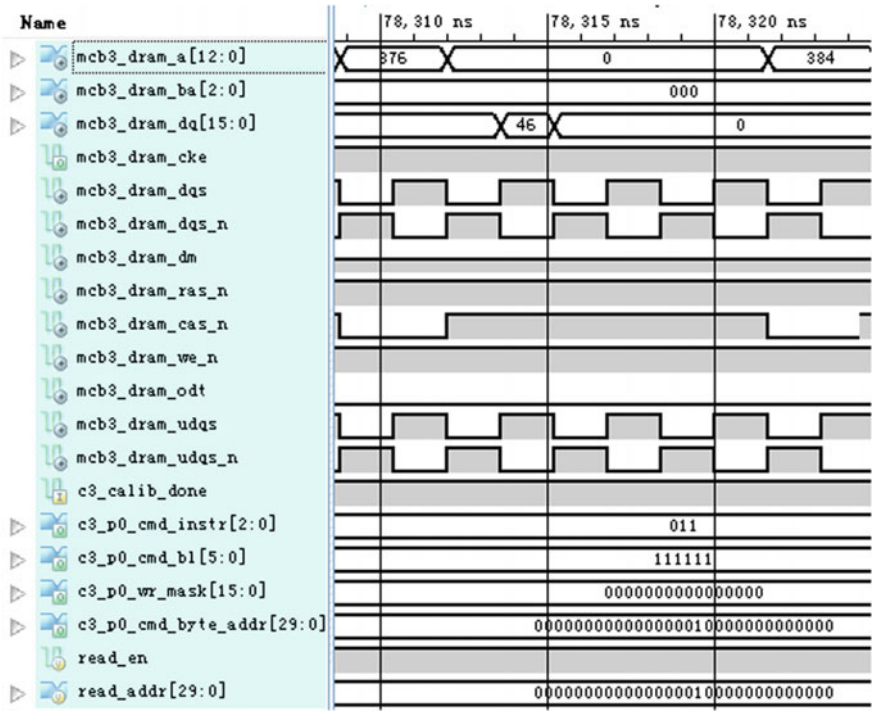



**Fig. 4** Write emulation timing

Waveform simulation: using ISE simulation software to write/read execution timing simulation and write timing as (Fig. 3), read timing as (Fig. 4) is shown, and figure C3 C3_Calib_done done on behalf of the initialization signal port, the port of C3 C3_Calib_done done above port on behalf of DDRII SDRAM, below the done C3 C3_Calib_done end mouth on behalf of the user port [7].

We used ISE to generate executable bit file, the bit file is downloaded to the development board on, using ISE of on-chip logic analyzer chip scope Pro) of instructions and data port for real-time scanning, get the write operation (Fig. 5) and read the real-time status of the operation (Fig. 6) corresponding to the data port and port instruction values [8]. By observing the discovery, write/read data port for the ROM in the 1–128 consecutive data [9].

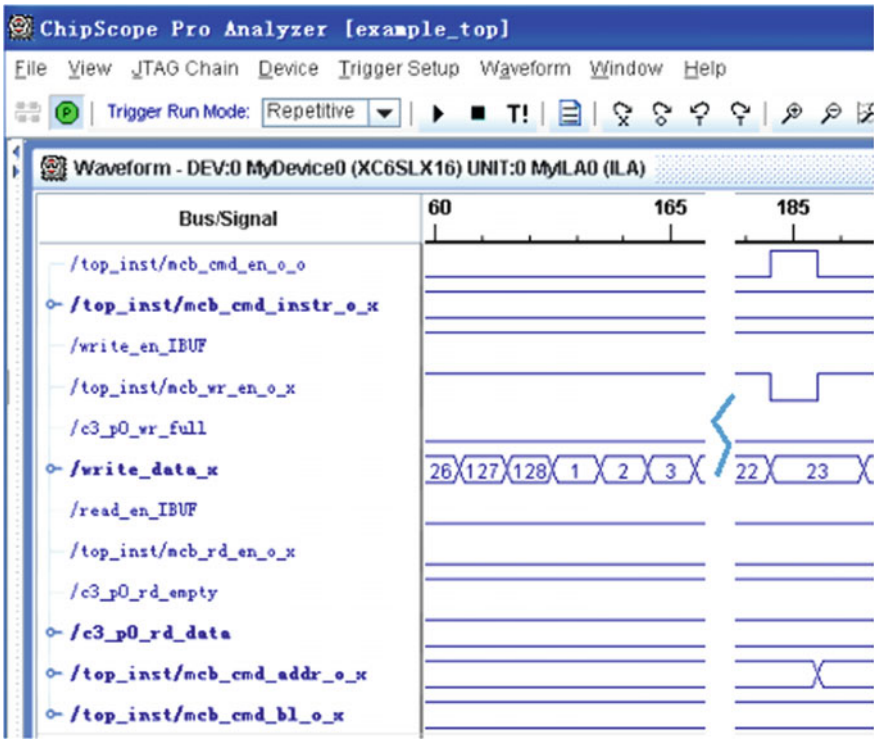



**Fig. 5** Write operation

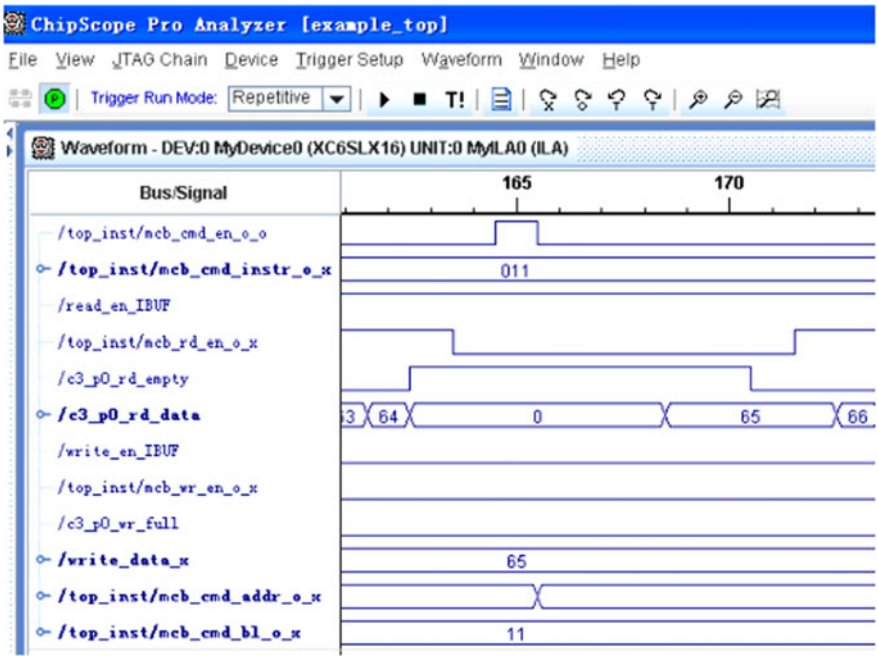



**Fig. 6** Read operation

## 8 Conclusion

In this paper, the design of DDRII SDRAM controller interface, through the FPGA verification, the correct timing relationship, and in the development board through the on-chip logic analyzer is written and read the data, they are the same, in the clock frequency 333.3 MHz still can correct operation, namely specifier synthetic DDRII controller design requirements, and to achieve the good stability, shorten the development cycle, in favor of DDRII SDRAM based on FPGA high-speed data storage system in the practical application.

## References

1. JEDEC Solid State Technology Association. JESD79–2A, DDRII SDRAM SPECIFICATION. JEDEC STANDARD, May 2007
2. Elapid Corporation.1G Bits DDRII SDRAM Datasheet (2008), http://www.elpida.com. Accessed Sept 2008
3. Z. Fan, Design and verification of SDRAM. DDRII controller (Xi'an: Xi'an Electronic and Science University, 2009)
4. W. Yang, J. Huang, Z. Wang, SDRAM.DDR in high speed data acquisition system in the application and design of [J]. Autom. Inf. (8) (2006)
5. Xilinx Corporation. Spartan6-FPGA Memory Controller User Guide (2010), http://www.xilinx.com. Accessed Apr 2010
6. Xilinx Corporation. Spartan6-FPGA Memory Interface Solutions (2010), http://www.xilinx.com. Accessed Apr 2010
7. Hynix DDRII SDRAM Device Operation & Timing Diagram (2007)
8. Olivier Despaux DDR SDRAM Controller Using Virtex-4 Devices. XILINX.2006 (2)
9. L. Zhuo et al., NDN-GSM-R: a novel high-speed railway communication system via Named Data Networking. EURASIP J. Wirel. Commun. Netw. **2016**, 48 (2016)

# Decoupling Control Methods for Spinning Missiles

**Chang-an Wang, Wei Wu, Sheng-bing Shi, Yong-chao Chen and Dan Fang**

**Abstract** Decoupling control is one key component during the design of autopilot for spinning missiles. However, in addition to the strong dynamic coupling effect, its parameters vary very fast over a large range, which induces great challenge to the design of autopilot. To address this issue, the classic complex summation decoupling control method and the newly developed robust gain-scheduling control method are both studied in detail, which are then applied to the design of autopilot of the spinning missile. The results from comparative studies show that the complex summation decoupling control method yields poor tracking performance and robustness; while the robust gain-scheduling control method performs much better during the whole trajectory. It is indicates that the robust gain-scheduling control method is very applicable to the spinning missile with strong dynamic coupling, and large fast time-varying parameters.

**Keywords** Spinning missiles · Robust control · Gain-scheduling · Complex summation decoupling

## 1 Introduction

The spinning of missiles brings many benefits, which has been widely adopted in modern weapon systems. However, the spinning may induce serious dynamic coupling between pitch and yaw channels, which would greatly deteriorate the stability as well as the performances of missiles. Therefore, it is necessary to consider the decoupling problem during the design of the control system of spinning missiles [1, 2].

C. Wang · W. Wu (✉) · S. Shi
Baicheng Ordnance Test Center, Baicheng 137001, China
e-mail: wuwei19781116@163.com

W. Wu · Y. Chen · D. Fang
Shijiazhuang Mechanical Engineering College, Shijiazhuang 050003, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_99

Due to the dynamic parameters of spinning missiles vary fast over a large range. In this case, the traditional decoupling control methods including the complex summation method developed for the time-invariant or slowly varying systems are inapplicable. Especially, when considering the control of the whole trajectory of spinning missiles, the decoupling controller designed according to the conventional gain-scheduling strategy can hardly work [3, 4]. In contrast, aiming at dealing with uncertainties and the fast time-varying characteristic of the system, the $H$ robust gain-scheduling control method based on the linear parameter varying (LPV) system has been demonstrated to be highly effective [5]. However, it is very difficult to solve the controller when there are lots of dependent parameters due to the unaffordable intensive computational cost. To address this issue, it has been proposed to select some important parameters that have large impact on the LPV system and construct a partial parameter dependent Lyapunov function (PPDLF) to obtain the controller [6, 7].

In this paper, the well-recognized complex summation decoupling autopilot with gain scheduling and the newly developed robust gain-scheduling control design method mentioned above are respectively presented in detail, which are then employed to the design of decoupling controller of a spinning guided missile with strong dynamic coupling and fast time-varying parameters in a wide range. Simulations on some specific set points as well as the whole trajectory for the two approaches are both conducted, of which the results are compared to verify the relative merits of both approaches.

## 2 Linear Dynamic Equations of the Spinning Missile

A typical linear model of a spinning missile is described as below.

$$\begin{bmatrix} \beta \\ \dot{\alpha} \\ \dot{\vartheta} \\ \ddot{\psi} \end{bmatrix} = \begin{bmatrix} -c_{N\alpha} & 0 & 0 & -1 \\ 0 & -c_{N\alpha} & 1 & 0 \\ -c_{mq\alpha} I_p & c_{m\alpha} & -c_{mq} & -I_p \\ -c_{m\alpha} & -c_{mq\alpha} I_p & I_p & -c_{mq} \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \\ \vartheta \\ \dot{\psi} \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \\ c_{m\delta} & 0 \\ 0 & c_{m\delta} \end{bmatrix} \begin{bmatrix} \delta_y \\ \delta_z \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} a_y \\ a_z \\ \ddot{\vartheta} \\ \ddot{\psi} \end{bmatrix} = \begin{bmatrix} -c_{N\alpha} V & 0 & 0 & 0 \\ 0 & -c_{N\alpha} V & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \\ \dot{\vartheta} \\ \dot{\psi} \end{bmatrix} \quad (2)$$

where $a_y$ and $a_z$ are the acceleration outputs in pitch and yaw channels, respectively; and the state variable $x = [\beta \quad \alpha \quad \vartheta \quad \dot{\psi}]^T$ represent the sideslip angle, angle of attack, sideslip angle and pitch angular velocity speed of the missile in the non-spinning coordinates [8]; $[\delta_y \quad \delta_z]^T$ are the ideal commands of rudder angles of the missile in the non-spinning coordinates.

## 3 The Complex Summation Decoupling Method

From Eq. (1), it is noticed that the system is a symmetric coupling system [5]. Therefore, the system can be described in the form of complex summation as below by defining $\xi = \beta + i\alpha$ , $\mu = \vartheta + i\psi$ and $\delta = \delta_y + i\delta_z$.

$$\begin{bmatrix} \xi \\ \dot{\mu} \end{bmatrix} = \begin{bmatrix} -c_{N\alpha} & i \\ -c_{mp\alpha} I_p - ic_{m\alpha} & -c_{mq} + iI_p \end{bmatrix} \begin{bmatrix} \xi \\ \mu \end{bmatrix} + \begin{bmatrix} 0 \\ c_{m\delta} \end{bmatrix} \delta$$
$$\begin{bmatrix} a/V \\ \mu \end{bmatrix} = \begin{bmatrix} -c_{N\alpha} & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \xi \\ \mu \end{bmatrix} \tag{3}$$

Considering that the overload autopilot has been widely applied to missiles, Eq. (3) can be further expressed as

$$\dot{x} = \begin{bmatrix} -c_{N\alpha} & i \\ -c_{mp\alpha} I_p - ic_{m\alpha} & -c_{mq} + iI_p \end{bmatrix} x + \begin{bmatrix} 0 \\ c_{m\delta} \end{bmatrix} u$$
$$y = \begin{bmatrix} -c_{N\alpha} & 0 \\ 0 & 1 \end{bmatrix} x \tag{4}$$

The actual inputs of the control system are the acceleration commands $\boldsymbol{a}_c$ from the guidance system and the control variables are

$$u = -Fx + k_{dc} a_c \tag{5}$$

where $a_c = -a_{cz} + ia_{cy}$ are the lateral acceleration commands with the complex summation form, and $k_{dc}$ is the adjustment coefficient of the closed-loop gain and is used to adjust the steady-state error of the system.

Let $A = \begin{bmatrix} -c_{N\alpha} & i \\ -c_{mp\alpha} I_p - ic_{m\alpha} & -c_{mq} + iI_p \end{bmatrix}$, $B = \begin{bmatrix} 0 \\ c_{m\delta} \end{bmatrix}$, $C = \begin{bmatrix} -c_{N\alpha} & 0 \\ 0 & 1 \end{bmatrix}$, and substitute Eq. (5) into Eq. (4) yield

$$\begin{aligned} \dot{x} &= (A - BF)x + Bk_{dc} a_c \\ y &= Cx \end{aligned} \tag{6}$$

For the linear time-invariant system in Eq. (4), the necessary and sufficient condition that all poles of the system can be arbitrarily placed by the state feedback method is that the plant is controllability. Since usually controllability matrix $[\boldsymbol{B}\ \boldsymbol{AB}]$ is non-singular matrix, the system is controllable. The characteristic polynomial of the closed-loop system is

$$\det[sI-(A-BF)]=s^2+p_1s+p_2 \tag{7}$$

Substituting $\boldsymbol{A}$, $\boldsymbol{B}$ and $\boldsymbol{F}$ into Eq. (7), one obtains

$$\begin{cases} p_1=c_{N\alpha}-c_{m\delta}f_1+c_{mq}-iI_p+c_{m\delta}f_2 \\ p_2=c_{N\alpha}\left[c_{mq}-iI_p+c_{m\delta}f_2\right]+\left[-c_{m\alpha}-iI_p+c_{m\delta}f_1\right] \end{cases} \tag{8}$$

Denoting the expected poles as $\lambda_1^*$ and $\lambda_2^*$, which often can be selected as a pair of conjugate complex numbers according to the requirements of the system response, one has

$$\begin{cases} p_1=-\left(\lambda_1^*+\lambda_2^*\right) \\ p_2=\lambda_1^*\lambda_2^* \end{cases} \tag{9}$$

## 4 Robust Gain-Scheduling Control Method

The control system framework based on the $H$ robust control theory is defined in Fig. 1, where $K(\theta)$ is the controller, $G(\theta)$ is the plant, and $\theta=\theta(t)$ is parameter dependent vector of the time-varying system.

Based on the above system architecture definition, a typical LPV system with disturbance for the spinning missile is constructed in Eq. (11).

$$G:\begin{cases} \dot{x}=A(\theta)x+B_1(\theta)w+B_2(\theta)u \\ z=C_1(\theta)x+D_{11}(\theta)w+D_{12}(\theta)u \\ y=C_2(\theta)x+D_{21}(\theta)w \end{cases} \tag{10}$$

Usually, it is assumed that the matrixes $A(\theta)$, $B_1(\theta)$ are affine linear functions of the parameter-dependent vector. Generally, an output feedback controller $K(\theta)$ with dependent parameters is designed for the plant $G(\theta)$. Similarly, $K(\theta)$ is assumed to be an affine linear function of dependent parameters $\theta$. Then the state space expression of the controller is

$$K(\theta):\begin{cases} \dot{x}_K=A_K(\theta)x_K+B_K(\theta)y \\ u=C_K(\theta)x_K+D_K(\theta)y \end{cases} \tag{11}$$

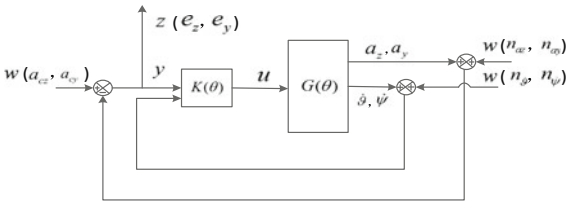



**Fig. 1** The basic framework of $H$ robust control for spinning missile

The order of controller $K(\theta)$ is generally considered as the same as that of the control plant. The reason why it is called gain-scheduling control is that the parameter matrixes $A_K(\theta)$, $B_K(\theta)$, $C_K(\theta)$ and $D_K(\theta)$ will be updated online according to the dependent parameters obtained in real time. Substituting Eq. (12) into Eq. (11), one obtains the closed-loop system below.

$$G_{cl}(\theta): \begin{cases} \dot{x}_{cl} = A_{cl}(\theta)x_{cl} + B_{cl}(\theta)w \\ z = C_{cl}(\theta)x_{cl} + D_{cl}(\theta)w \end{cases} \tag{12}$$

where $x_{cl}^T = \begin{bmatrix} x^T & x_K^T \end{bmatrix}$, $A_{cl}(\theta) = A_0(\theta) + \tilde{B}(\theta)\Omega(\theta)\tilde{C}(\theta)$, $B_{cl}(\theta) = B_0(\theta) + \tilde{B}(\theta)\Omega(\theta)\tilde{D}_{21}(\theta)$, $C_{cl}(\theta) = C_0(\theta) + \tilde{D}_{12}(\theta)\Omega(\theta)\tilde{C}(\theta)$, $D_{cl}(\theta) = D_{11}(\theta) + \tilde{D}_{12}(\theta)\Omega(\theta)\tilde{D}_{21}(\theta)$, $A_0(\theta) = \begin{bmatrix} A(\theta) & 0 \\ 0 & 0_{n\times n} \end{bmatrix} B_0(\theta) = \begin{bmatrix} B_1(\theta) \\ 0 \end{bmatrix}$, $C_0(\theta) = \begin{bmatrix} C_1(\theta) & 0 \end{bmatrix}$, $\tilde{B}(\theta) = \begin{bmatrix} 0 & B_2(\theta) \\ I_n & 0 \end{bmatrix}$, $\tilde{C}(\theta) = \begin{bmatrix} 0 & I_n \\ C_2(\theta) & 0 \end{bmatrix}$, $\tilde{D}_{12}(\theta) = \begin{bmatrix} 0 & D_{12}(\theta) \end{bmatrix}$ and $\tilde{D}_{21}(\theta) = \begin{bmatrix} 0 \\ D_{21}(\theta) \end{bmatrix}$.

The design task of robust gain-scheduling control is to design a time-varying controller $K(\theta)$ for a time-varying system $G(\theta)$ to make the closed-loop system $G_{cl}(\theta)$ asymptotically stable and satisfy $\int_0^t z^T z d\tau \le \gamma^2 \int_0^t w^T w d\tau$ for arbitrary bounded energy disturbances, where $\theta$ varies with time. In this paper, the robust gain-scheduling method based on PPDLF is employed to make a compromise between computational cost and robustness. In order to employ the mature linear matrix inequality (LMI) method in $H$ control problems to solve the controller, two theorems are introduced below.

**Lemma 1** *The system shown in* (12) *is asymptotically stable and has the* $L_2$ *norm boundary* $\gamma$ *if and only if there exists a parameter dependent symmetric and derivable matrix* $X_{cl}(\theta) > 0$, *such that the matrix inequalities below are satisfied on* $(\theta, \theta) \in H \times H_d$.

$$\begin{bmatrix} A_{cl}^T(\theta)X_{cl}(\theta) + X_{cl}(\theta)A_{cl}(\theta) + \dot{X}_{cl}(\theta) & X_{cl}(\theta)B_{cl}(\theta) & C_{cl}^T(\theta) \\ B_{cl}^T(\theta)X_{cl}(\theta) & -\gamma I & D_{cl}^T(\theta) \\ C_{cl}(\theta) & D_{cl}(\theta) & -\gamma I \end{bmatrix} < 0 \tag{13}$$

where $H$ and $H_d$ are two multidimensional rectangular spaces corresponding to dependent parameters and their derivatives, respectively. Defining control matrix

$$\Omega(\theta) = \begin{bmatrix} A_K(\theta) & B_K(\theta) \\ C_K(\theta) & D_K(\theta) \end{bmatrix} \tag{14}$$

Since $X_{cl}(\theta)$ and $\Omega(\theta)$ are nonlinear, the method based on the LIM is no longer applicable. Lemma 2 will present the method to solve $X_{cl}(\theta)$.

**Lemma 2** *The system is asymptotically stable and has the $L_2$ norm boundary $\gamma$ if and only if there exist parameter dependent symmetric and derivable matrixes $X(\theta)$ and $Y(\theta)$, such that matrix inequalities below are satisfied on $(\theta,\theta)\in H\times H_d$.*

$$\begin{bmatrix} N_X(\theta) & 0 \\ 0 & I \end{bmatrix}^T \left[\begin{array}{cc|c} A(\theta)^T X(\theta)+X(\theta)A(\theta)+\dot{X}(\theta) & X(\theta)B_1(\theta) & C_1(\theta)^T \\ B_1(\theta)^T X(\theta) & -\gamma I & D_{11}(\theta)^T \\ \hline C_1(\theta) & D_{11}(\theta) & -\gamma I \end{array}\right] \begin{bmatrix} N_X(\theta) & 0 \\ 0 & I \end{bmatrix} < 0$$

$$\begin{bmatrix} N_Y(\theta) & 0 \\ 0 & I \end{bmatrix}^T \left[\begin{array}{cc|c} A(\theta)Y(\theta)+Y(\theta)A(\theta)^T-\dot{Y}(\theta) & Y(\theta)C_1(\theta)^T & B_1(\theta) \\ C_1(\theta)Y(\theta) & -\gamma I & D_{11}(\theta) \\ \hline B_1(\theta)^T & D_{11}(\theta)^T & -\gamma I \end{array}\right] \begin{bmatrix} N_Y(\theta) & 0 \\ 0 & I \end{bmatrix} < 0 \tag{15}$$

$$\begin{bmatrix} X(\theta) & I \\ I & Y(\theta) \end{bmatrix} > 0 \tag{16}$$

*where $N_X(\theta)$ and $N_Y(\theta)$ are any basic zero space of matrix of $\begin{bmatrix} C_2(\theta) & D_{21}(\theta) \end{bmatrix}$ and $\begin{bmatrix} B_2(\theta)^T & D_{12}(\theta)^T \end{bmatrix}$, respectively.*

Assume that $X(\theta)$ and $Y(\theta)$ are considered as the affine functions of dependent parameters. Then a parameter-dependent Lyapunov function (PDLF) can be built. $\dot{X}(\theta)$ and $\dot{Y}(\theta)$ in (15) can be expressed as

$$\begin{cases} \dot{X}(\theta)=\sum\limits_{i=1}^{N}\theta_i X_i = X(\theta)-X_0 \\ \dot{Y}(\theta)=\sum\limits_{i=1}^{N}\theta_i Y_i = Y(\theta)-Y_0 \end{cases} \tag{17}$$

According to the convex theory of linear matrix inequalities, it is only required that the matrix inequalities (15) are satisfied at the vertices of the convex set. In general, the infinite linear matrix inequality (LMI) constraints can be transformed into finite LMI constraints by meshing the parameter space, which are then solved by the LMI convex optimization. Once $X(\theta)$ and $Y(\theta)$ are determined, the Lyapunov function $x_{cl}^T X_{cl}(\theta)x_{cl}$ can be constructed, where $X_{cl}(\theta)$ is expressed as

$$X_{cl}(\theta)=\begin{bmatrix} X(\theta) & X(\theta)-Y^{-1}(\theta) \\ X(\theta)-Y^{-1}(\theta) & X(\theta)-Y^{-1}(\theta) \end{bmatrix} \tag{18}$$

Further, one has

$$\dot{X}_{cl}(\theta)=\begin{bmatrix} \dot{X}(\theta) & \dot{X}(\theta)+Y^{-1}(\theta)\dot{Y}(\theta)Y^{-1}(\theta) \\ \dot{X}(\theta)+Y^{-1}(\theta)\dot{Y}(\theta)Y^{-1}(\theta) & \dot{X}(\theta)+Y^{-1}(\theta)\dot{Y}(\theta)Y^{-1}(\theta) \end{bmatrix} \tag{19}$$

Substituting (17) and (18) into Lemma 1, one has

$$\Psi_{X_{cl}}(\theta) + Q^T(\theta)\Omega^T(\theta)P_{X_{cl}}(\theta) + P_{X_{cl}}^T(\theta)\Omega(\theta)Q(\theta) < 0 \tag{20}$$

where

$$\Psi_{X_{cl}}(\theta) = \begin{bmatrix} A_0^T(\theta)X_{cl}(\theta) + X_{cl}(\theta)A_0(\theta) + \dot{X}_{cl}(\theta) & X_{cl}(\theta)B_0(\theta) & C_0^T(\theta) \\ B_0^T(\theta)X_{cl}(\theta) & -\gamma I & D_{11}^T(\theta) \\ C_0(\theta) & D_{11}(\theta) & -\gamma I \end{bmatrix}$$

$$P_{X_{cl}}(\theta) = \begin{bmatrix} \tilde{B}^T(\theta)X_{cl}(\theta) & 0 & \tilde{D}_{12}^T(\theta) \end{bmatrix}, Q(\theta) = \begin{bmatrix} \tilde{C}(\theta) & \tilde{D}_{21}(\theta) & 0 \end{bmatrix}$$

In the LMI (20), the variables to be solved are the parameter matrix of controller $\Omega(\theta)$ that are dependent on the time-varying parameters $\theta$. If the $\theta$ can be obtained in real time, other matrixes at this moment can be determined according to $\theta$ of the system.

## 5 Simulation Analysis and Comparison

### *5.1 Simulation on Discrete Points*

In this section, the complex summation decoupling control method (denoted as CSD) and the PPDLF robust gain-scheduling control method (denoted as PPDLF) are applied to the autopilot design of spinning missile, of which the simulation results are compared to verify the effectiveness and relative merits of the two approaches. Simulations are conducted on the highest point of trajectory.

Due to the different aerodynamic characteristics, the available overloads of the spinning missile at the two discrete points are relatively different. In the initial and terminal phases of the trajectory, the overloads are relatively large. However, in the middle of the trajectory, the available overload is significantly reduced, due to the high altitude, low velocity and low dynamic pressure of missile. Therefore, the amplitudes of the step input signal are set as 50 m/s$^2$ and 0.2 m/s$^2$, respectively. The step responses considering $\pm 10\%$ variations of the aerodynamic coefficients by the two control approaches are illustrated in Fig. 2.

Clearly, although the coupling between the two channels are complete eliminated by the CSD method (i.e., the overload command in the pitching channel did not cause any output overload in the yaw channel), when considering uncertainties, the command tracking performance of the decoupled overload autopilot is poor with large static error. While, the PPDLF method has good control performances (decoupling, overshot, and steady-state error etc.), which is more robust to the parameter uncertainties compared to the CSD method.

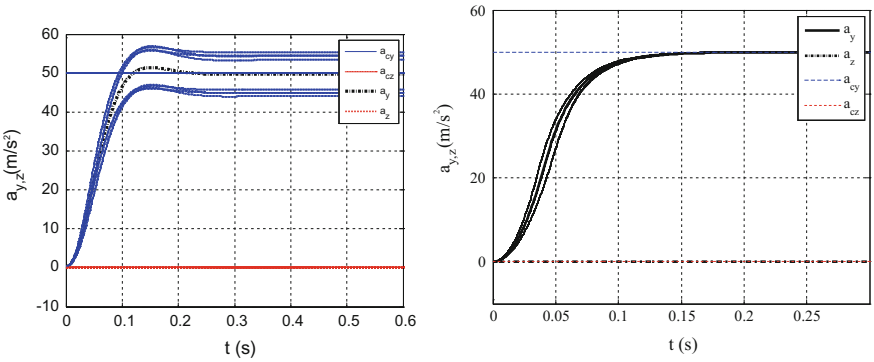



**Fig. 2** Step responses (*left* CSD; *right* PPDLF)

## 5.2 *Simulation of the Whole Trajectory*

The good performances of controller at the discrete points cannot ensure the good performance during the whole flight especially when the system parameters vary fast in a wide range. Therefore, it is necessary to carry out the closed-loop full trajectory simulation of the spinning missile. For the CSD autopilot, simulation is conducted by calling the altitude-speed-speed gain-scheduling table. According to the available overload of missile, the amplitude of the step signal is 1 m/s2 during the whole trajectory. The responses of the overload and required angle of attack of the two approaches are shown in Figs. 4 and 5, respectively.

From Fig. 3, it is noticed that during about 0–25 s (initial phase) and 12–140 s (terminal phase) of the flight, the overload autopilot by the CSD method performs well in tracking the overload command. Meanwhile, good decoupling effect is observed. However, during about 50–100 s, the control performance of the CSD method degrades greatly. The interpretation is that with the increase of the flight

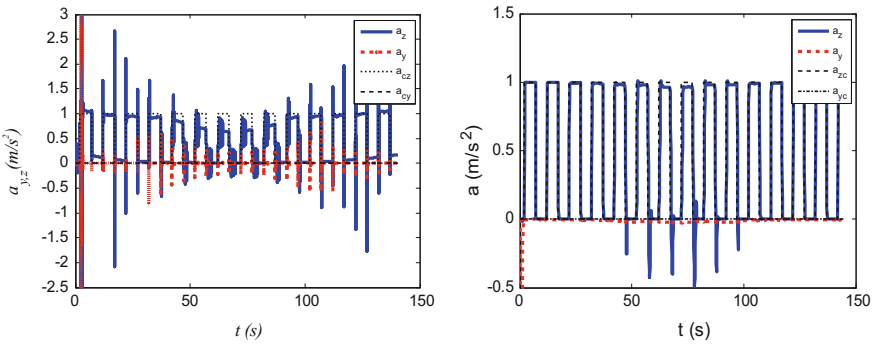



**Fig. 3** Step response during the whole trajectory (*left* CSD; *right* 1PPDLF)

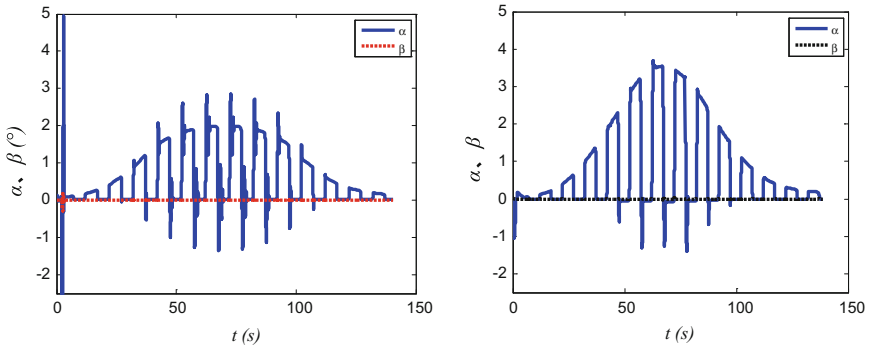



**Fig. 4** The required angle of attack during the whole trajectory (*left* CSD; *right* PPDLF)

height, the flight velocity and the dynamic pressure decrease greatly, which deteriorates the maneuvering ability of missile. Therefore, to provide the same overload, larger angle of attack is required during this phase (see the left in Fig. 5). During this flight phase, the overload autopilot based on CSD method of the spinning missile cannot track the command well. For the PPDLF method, it is found that generally the missile has good decoupling and dynamic characteristics, and high tracking accuracy, except that during the initial moment and engine off moment, some small output fluctuations are observed. Compared to the CSD controller, during 50–100 s with low aerodynamic efficiency, the PPDLF controller can generate larger angle of attack quickly so as to track the overload command as much as possible (see the right in Fig. 4). Therefore, the effectiveness and advantages of the PPDLF based robust gain-scheduling controller in dealing with systems with fast time-varying parameters in a wide range and the coupling effect is demonstrated, compared to the interpolation based gain-scheduling controller.

## 6 Conclusions

In this paper, two decoupling control methods for spinning missiles are introduced and compared. Extended from the traditional overload autopilot, the complex summation decoupling autopilot shows good performances in command tracking and decoupling when the aerodynamic efficiency is high. However, it performs poor when the parameters vary fast and aerodynamic efficiency is low. In contrast, the PPDLF based robust variable gain-scheduling control always performs well, exhibiting high robustness and good command tracking and decoupling abilities during the whole trajectory flight, which is considered to be an effective control method with extensive application prospects.

## References

1. K. Li, S. Yang, L. Zhao, Stability of spinning missiles with an acceleration autopilot. J. Guid. Control Dyn. **35**(3), 774–786 (2012)
2. P.W. Fortescue, E.M. Belo, Control decoupling analysis for gyroscopic effects in rolling missiles. J. Guid. Control Dyn. **12**(6), 798–805 (1989)
3. W.J. Rugh, J.S. Shamma, Research on gain scheduling. Automatica **36**, 1401–1425 (2000)
4. D.J. Leith, W.E. Leithead, Survey of gain-scheduling analysis and design. Int. J. Control **73** (11), 1001–1025 (2000)
5. A.J. Calise, M. Sharma, J.E. Corban, Adaptive autopilot design for guided munitions. J. Guid. Control Dyn. **23** (5), pp. 837–843 (2000)
6. S. Theodoulis, G. Duc, Missile autopilot design: gain-scheduling and the gap metric. J. Guid. Control Dyn. **32**(3), 986–996 (2009)
7. S.M. Shahruz, S. Behtash, Design of controllers for linear parameter-varying systems by gain scheduling technique. J. Math. Anal. Appl. **168**, 195–217 (1992)
8. Zarchan P. *Tactical and Strategic Missile Guidance* (American Institute of Aeronautics & Astronautics Inc, (6), 1990), p. 555
9. G. Becker, A. Packard, Robust performance of linear parametrically varying systems using parametrically-dependent linear feedback. Syst. Control Lett. **23**, 205–215 (1994)

# A Method of Antenna Impedance Matching Based on Vector Fitting

**Lingling Tan, Yunpeng Wang and Guizhen Yu**

**Abstract** Based on the principle of vector fitting and the theory of impedance matching, a method of antenna impedance matching is proposed for electromagnetic systems. A method of fitting of measured frequency domain responses named vector fitting is proposed. The *S* parameter obtained at certain frequency is transformed to *Z* parameter, then method of Smith chart is used to achieve impedance matching of the corresponding *Z* parameter at certain frequency. Finally, the method is used for frequency of 315 MHz antenna impedance matching. The simulation results verify the effectiveness of vector fitting in impedance matching of radio frequency (RF) domain.

**Keywords** Impedance matching · Smith chart · 315 MHz · Vector fitting · Signal fitting

## 1 Introduction

The high frequency behavior of antenna is characterized by its impedance characteristics, of which the input impedance is meaning to realize the antenna impedance matching with circuit output impedance. When the antenna impedance matches with the circuit output impedance, all the current will flow into the antenna from circuit without any being reflected back. Usually, the output impedance of circuit is designed in 50 Ω [1]. In order to make the antenna and the connected

L. Tan · Y. Wang · G. Yu (✉)
School of Transportation Science and Engineering, Beihang University, Beijing 100191, China
e-mail: yugz@buaa.edu.cn

L. Tan
e-mail: tanlingling@buaa.edu.cn

Y. Wang
e-mail: ypwang@buaa.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*, Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_100

circuit match with each other, the input impedance of antenna should be designed to equal with the output impedance of circuit.

The signal integrity and electrical properties should be taken into account in PCB board design and IC packaging design. In electronic systems, such as antenna, the voltages and currents of every port are closely related with the frequency and it is difficult to establish exact model of the systems [2]. Therefore, signal fitting of measured frequency domain responses is important for electromagnetic magnetic compatibility (EMC) simulation in electromagnetic systems [3]. The fitting is based on calculating a rational function which described a certain wide-band frequency-dependent black box model. The method of signal fitting named vector fitting was widely applied to the admittance parameters and scattering parameters [4] of the complicated signal to establish the models corresponding to rational functions [5].

This paper presents the method of antenna impedance matching based on vector fitting. First, theory of vector fitting for achieving approximation of the sampled scatting parameters of the antenna which needs to be matched was introduced. An application example of the antenna working at frequency of 315 MHz was taken to demonstrate how to realize its impedance matching with circuit output impedance of 50 Ω using computer simulation tool of Smith chart [6].

## 2 Method of Antenna Impedance Matching

### *2.1 The Theory of Vector Fitting*

In a passive system, the transfer function of the measured system is in the following [7],

$$H(s) = \sum_{n=1}^{N} \frac{r_n}{s - p_n} + d + sh + s^2 e + \ \ldots \tag{1}$$

If the non-linearity of Eq. (1) is ignored, it can be written into Eq. (2)

$$H(s) \approx \sum_{n=1}^{N} \frac{r_n}{s - p_n} + d + sh \tag{2}$$

where $d$, $h$, $e$ is constant, the poles $\{p_n\}$ and residues $\{r_n\}$ are either real quantities or come in complex conjugate pairs.

In order to obtain the approximated transfer function of the measured system, an initial poles vector $\overline{P}_n = (\overline{p}_1, \overline{p}_2, \ldots, \overline{p}_n)$ is set to construct function (3) [8]

$$\delta(s) = \sum_{n=1}^{N} \frac{\tilde{r}_n}{s - \overline{p}_n} + 1 \tag{3}$$

where $\delta(s)$ is forced to approach unity at very high frequency. Multiplying Eq. (3) with $H(s)$, we get

$$\left(\sum_{n=1}^{N} \frac{r_n}{s - \overline{p}_n} + d + sh\right) \approx \left(\sum_{n=1}^{N} \frac{\tilde{r}_n}{s - \overline{p}_n} + 1\right) H(s) \tag{4}$$

Equation (4) can be rewritten as

$$\left(\sum_{n=1}^{N} \frac{r_n}{s - \overline{p}_n} + d + sh\right) - \left(\sum_{n=1}^{N} \frac{\tilde{r}_n}{s - \overline{p}_n}\right) H(s) \approx H(s) \tag{5}$$

For a given sampled frequency point $s_k$, we get [8]

$$A_k x = b_k \tag{6}$$

where

$$A_k = \left[\frac{1}{s_k - \overline{p}_1} \quad \ldots \quad \frac{1}{s_k - \overline{p}_N} \quad 1 \quad s_k \quad \frac{-H(s_k)}{s_k - \overline{p}_1} \quad \ldots \quad \frac{-H(s_k)}{s_k - \overline{p}_N}\right] \tag{7}$$

$$x = [r_1 \quad \ldots \quad r_N \quad d \quad h \quad \tilde{r}_1 \quad \ldots \quad \tilde{r}_N]^T, \quad b_k = H(s_k) \tag{8}$$

After $x$ is solved as a least squares problem, Eq. (4) can be rewritten into a rational function [8]

$$h \frac{\prod_{n=1}^{N+1} (s - z_n)}{\prod_{n=1}^{N} (s - \overline{p}_n)} = \frac{\prod_{n=1}^{N} (s - \tilde{z}_n)}{\prod_{n=1}^{N} (s - \overline{p}_n)} H(s) \tag{9}$$

Thus,

$$H(s) = h \frac{\prod_{n=1}^{N} (s - z_n)}{\prod_{n=1}^{N} (s - \tilde{z}_n)} \tag{10}$$

In this way, another poles vector $\tilde{Z}_n = (\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_n)$ is solved. Then the new poles $\tilde{Z}_n$ should be substituted into Eq. (4). Repeating the procedure of Eqs. (4)–(10) until the ultimate poles $\overline{P}_\infty = (\overline{p}_1, \overline{p}_2, \ldots, \overline{p}_n)$ makes the measured system stable, the ultimate residues $R_n = (r_1, r_2, \ldots, r_n)$, $d$ an $h$ are obtained through solving Eq. (6) [8]. The approximated system is achieved by technique of pole relocation, in which the poles are improved iteratively [9].

Equation (2) can also be written into the form of state space [10]

$$H(s) = C(sI - A)^{-1}B + D + sE, \tag{11}$$

where $H(s)$ is the transfer function of the system, $(sI - A)^{-1}$ is the characteristic matrix, $A$, $B$, $C$, $D$, and $E$ is coefficient matrix.

## 2.2 The Theory of Antenna Impedance Matching

$S$ parameter is taken to describe the two-port network system through the reflecting signal to its port and signal transmitted from one port to another port, its description is as follows, (Fig. 1)

$$\begin{bmatrix} b_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \tag{12}$$

where $a_k$ is the incident wave of the $k$th port, $b_k$ is the reflection wave of the $k$th port, $S_{11}$ is the voltage reflection coefficient of port 1 when port 2 is matched, $S_{12}$ is the reverse voltage gain when port 1 is matched, $S_{21}$ is the forward voltage gain when port 2 is matched, $S_{22}$ is the voltage reflection coefficient of the port 2 when port 1 is matched [11].

The input impedance $Z_{in} = R + jX$ of the antenna determines the matching condition between the antenna and the transmitter or receiver; it is used to realize the impedance between the source and the load to achieve maximum radiated power [12]. It is difficult to measure the input impedance $Z_{in}$ at high frequency, while there is equation for the conversion between $S_{11}$ and $Z_{in}$ when port 2 is matched [13], which is shown in Eqs. (13) and (14). Smith chart is a curve coordinate diagram to solve the impedance matching problem, through which the normalized impedance value, the admittance value, reflection coefficient [14] at a certain point can be read.

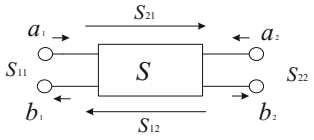



**Fig. 1** The description of $S$ parameter

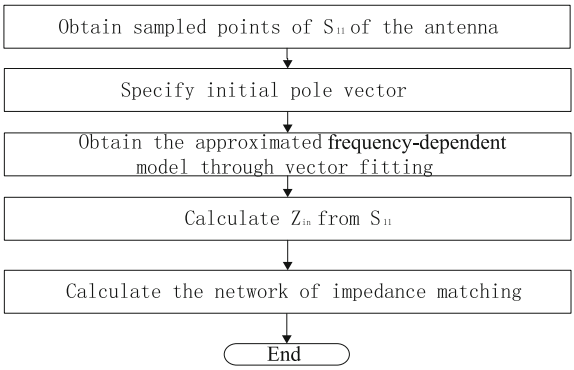



**Fig. 2** The flow diagram of antenna impedance matching based on vector fitting

$$s_{11} = \frac{Z_{in} - Z_0}{Z_{in} + Z_0} \tag{13}$$

$$Z_{in} = \frac{1 + S_{11}}{1 - S_{11}} Z_0 \tag{14}$$

where $Z_0$ is the characteristic impedance, usually which is 50 Ω, $Z_{in}$ is the input impedance.

For an antenna system, the following flow diagram demonstrates the method of its impedance matching based on vector fitting (Fig. 2).

## 3 Application Example

In this section, a design example is provided to illustrate the design procedure of antenna impedance matching using vector fitting. We consider the impedance matching of an antenna worked at 315 MHz frequency, and the sampled frequency domain responses $S_{11}$ of the antenna is taken to establish the models of the antenna. The frequency range of the sampled signal is [300 MHz, 1 GHz] with 0.7 MHz as the interval frequency, the amplitude characteristics are shown in Fig. 3.

A rational approximation of the sampled $S_{11}$ is calculated by fitting in the frequency range 300 MHz–1 GHz using approximation of 500 poles, the final amplitude characteristics of the fitted signal are shown in Fig. 4.

Figure 4 shows the amplitude characteristics of the approximated data, seen from it, the resulting fitting of $S_{11}$ coincides with the sampled frequency domain responses at most of the frequency domain, from the results of which the value of the fitted $S_{11}$ at 315 M frequency can be calculated by Eq. (1). Finally, we get

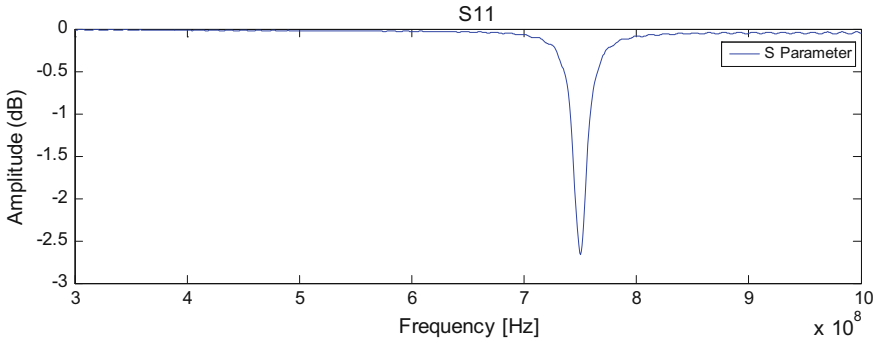



**Fig. 3** The amplitude characteristics of $S_{11}$ of the antenna

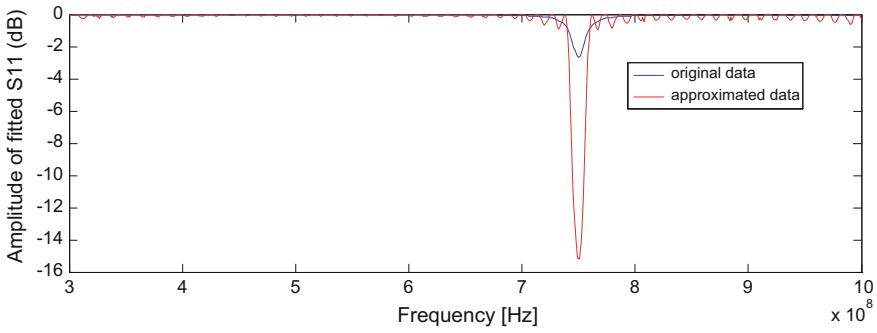



**Fig. 4** The amplitude characteristics of the approximated data

$$s_{11}@315\,MHZ = -0.8805 + 0.4255i \tag{15}$$

As was pointed out in Sect. 2.2, $S_{11}$ can be converted to $Z_{in}$ through Eq. (14), thus

$$Z_{in}@315\,MHZ = 0.5874 + 11.4464i \quad with\, Z_0 = 50\,\Omega \tag{16}$$

In practical application, Smith chart is taken to carry out the impedance matching work, the objective is to make the whole circuit resonate at 315 MHz after increasing a matching circuit. According to the L-shaped matching circuit with a 48.3 pF capacitor connected in series with a 47 nH inductor, shown in Fig. 5, the impedance of the antenna connected in series with the matching circuit nearly comes to 50 Ω, as shown in Fig. 6.

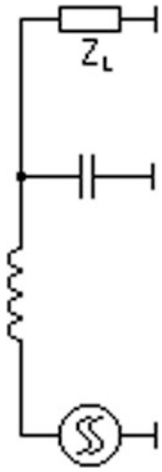



**Fig. 5** The matching circuit

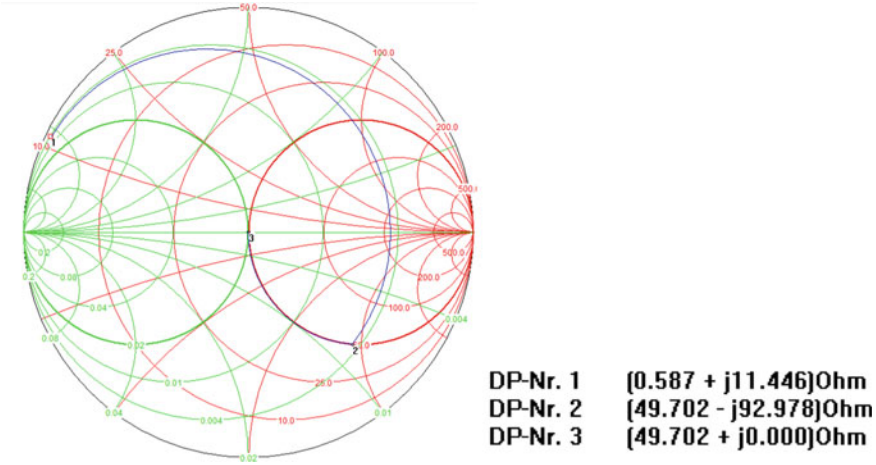



**Fig. 6** The Smith chart of the impedance matching

## 4 Conclusion

This paper has extended the method of vector fitting into application of antenna impedance matching, which has been achieved through fitting of measured frequency domain responses and Smith chart has been used to achieve impedance matching of the corresponding $Z$ parameter at frequency of 315 MHz. The simulation results verify the effectiveness of vector fitting in impedance matching of antenna design.

**Acknowledgements** This research was funded partially by the National Science Foundation of China under Grant #61371076.

## References

1. Y. Tsut sumi, H. Kahaya, K. Yoshida, Design and performance of an electrically small slot loop antenna with a miniaturized superconducting matching circuit. IEEE Trans. Appl. Supercond. **15**(2), 1020–1023 (2005)
2. R. Pintelon, P. Guillaume, Y. Rolain, J. Schoukens, H.V. Hamme, Parametric identification of transfer functions in the frequency domain—a survey. IEEE Trans. Autom. Control **39**(11), 2245–2260 (1994)
3. Z. Qi, H. Yu, P. Liu, S.X.D. Tan, L. He, Wide band passive multiport model order reduction and realization of RLCM circuits. IEEE Trans. Comput. Aided Des. Integr. Circuits Syst. (TCAD) **25**(8), 1496–1509 (2006)
4. B. Gustavsen, A. Semlyen, Simulation of transmission line transients using vector fitting and modal decomposition. IEEE Trans. Power Deliv. **13**(2), 605–614 (1998)
5. B. Gustavsen, Improving the pole relocating properties of vector fitting. IEEE Trans. Power Deliv. **21**(3), 1587–1592 (2006)
6. Y. Wu, Y. Zhang, Y. Liu et al., Theory of the spherical generalized Smith chart. Microw. Opt. Technol. Lett. **51**(1), 95–97 (2009)
7. A.O. Soysal, A. Semlyen, Practical transfer function estimation and its application to transformers. JEEE Trans. PWRD **8**(3), 1627–1637 (1993)
8. B. Gustavsen, A. Semlyen, Rational approximation of frequency domain responses by vector fitting. IEEE Trans. Power Deliv. **14**(3), 1052–1061 (1999)
9. D. Deschrijver, M. Mrozowski, T. Dhaene, D. De Zutter, Macromodeling of multiport systems using a fast implementation of the vector fitting method. IEEE Microw. Wireless Compon. Lett. **18**(6), 383–385 (2008)
10. B. Gustavsen, Computer code for rational approximation of frequency dependent admittance matrices. IEEE Trans. Power Deliv. **17**(4), 1093–1098 (2002)
11. A. Frickey, Conversions between S, Z, Y, h, ABCD, and T parameters which are valid for complex source and load impedances. IEEE Trans. Microw. Theory Tech. **42**(2), 205–211 (1994)
12. C.A. Balanis, Antenna theory: analysis and design. New York, 2005
13. R. Ludwig, P. Bretchko, RF circuit design: theory and applications, in *Proceedings of the Prentice-Hall* (2000)
14. P.H. Smith, Electronic applications of the Smith chart, in *Proceedings of the IEEE* (1995)

# Hydraulic Equipment Detection System Design on Certain Launching Device

**Li Hongru, Xu Baohua and Ye Peng**

**Abstract** In this paper, one hydraulic equipment detection system on certain launching device has been established applying data acquisition, virtual instrument, and signal analysis technologies. This detection system can detect pressure and flow parameters on-line, and has such advantages as such convenient operation and high accuracy.

**Keywords** Hydraulic equipment · Data acquisition · Virtual instrument · LabVIEW

## 1 Introduction

Hydraulic equipments, vital components of certain launching device, have such functions as launching pad and erecting arm's lifting-falling and locking-unlocking. The hydraulic equipments have relative high failure rate because of the bad working conditions as transportation vibration and extreme temperature. There is no specific detection system of hydraulic equipment on this launching device, so the technicians have to utilize simple instruments or their experiences to detect the failures when hydraulic equipments disabled, which can no longer meet the needs of modern warfare. Therefore, research on hydraulic equipment detection system for this launching device has significant meanings in performance measurement, failure detection and equipment maintenance.

L. Hongru · X. Baohua · Y. Peng (✉)
Department 4, Mechanical Engineering College, Shijiazhuang 050003, China
e-mail: yepeng_oec@sina.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_101

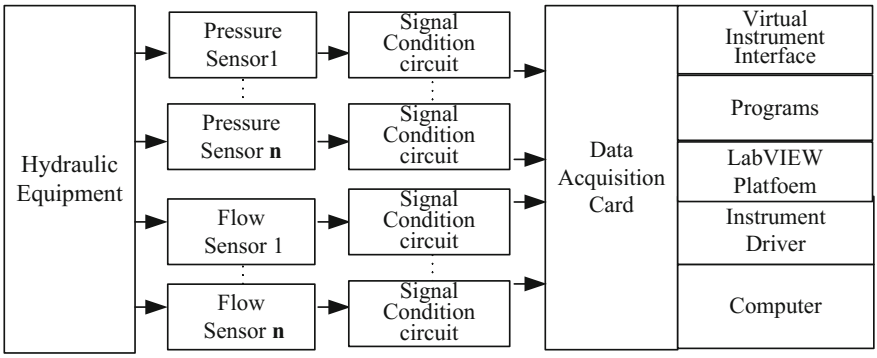



**Fig. 1** System general structure

## 2 General Design

Theoretically, the dynamic signals representing hydraulic equipment's functional state include pressure, flow, vibration, temperature signals, and oil contaminated level. Generally considering hydraulic equipment's bad working conditions and online detection [1–9], pressure and flow signals are adopted to be measured in this paper. Conditional instruments cannot meet the needs of detection due to the complex procedures, so one hydraulic equipment detection system based on virtual instrument has been established with general structure shown in Fig. 1.

The hardware design is developed as PC-DAQ virtual instrument structure composed of sensors, signal condition circuit, data acquisition card and computer. The hardware design is developed based on LabVIEW platform.

## 3 Hardware Design

### *3.1 Sensor*

Sensor is the first link in state signal acquisition of the detection system, and the measuring errors from sensors cannot be compensated in later links, so sensors' quality is vital to the accuracy of the whole system. According to general design, pressure and flow sensors are adopted to measure the pressure and flow of hydraulic equipments in this paper.

Pressure sensor: the hydraulic equipment detection system can measure steady and dynamic pressures at the same time. HM801 pressure sensor is adopted, which is a pressure-resistant sensor with diffused-silicon sensitive chip made with stainless steel isolation diaphragm. HM801 pressure sensor has such advantages as high sensitivity, high accuracy, fast frequency response, simple structure, small capacity,

and light weight. Its major parameters are: measuring range 0–16 Mpa, accuracy level 0.1%, and output voltage 2–10 V.

Flow sensor: LWGY turbine flow sensor is adopted, because the hydraulic oil has relative high pressure and relative small flow. Its major parameters are: measuring range 0.25–1.2 $m^3$/h, rated working pressure $\leq$25 MPa, accuracy level 0.5%, and output current 4–20 mA.

### *3.2 Signal Condition Circuit*

Signal condition circuit amplifies, filters, isolates, and linearizes the output signals from the sensors and transmitters, and transforms them into signals which is easy to be acquired by data acquisition device.

Pressure sensor output signal condition: this detection system filters the high-frequency noise from HM801 pressure sensor's output voltage by low-pass second-order filter circuit.

Flow sensor output signal condition: this detection system transforms current into 2–10 V signals using 500 Ω sample resistance from HM801 pressure sensor's output current, and filters the signal by low-pass second-order filter circuit.

### *3.3 Data Acquisition Card*

PCI-6221 data acquisition card from NI Company is adopted, which is one multiplex card based on standard computer interface with 16-way single-end analog input channels, 8-way difference input channels, 250 KS/s sample rate, ±200 mV to ±10 V input range and 68-pin port.

## 4 Software Design

In this detection system, software is programmed in modularization. All modules are linked and positioned after each is programmed, compiled and debugged. In top-down design, the detection system software can be divided into such function modules as user logging and verification module, system initiation module, data acquisition module, data management module, data analysis and process module, alert module and interactive interface, shown in Fig. 2.

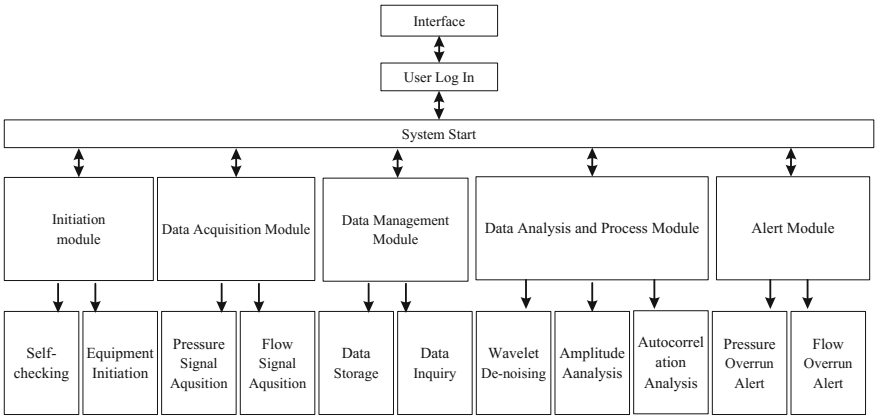



**Fig. 2** Software function modules

## *4.1 User Logging and Verification Module*

This module can verify the logged users and prevent illegal users entering system.

## *4.2 System Initiation Module*

This module can be divided into two sub-modules as system self-checking and equipment initiation. This module can check the hardware's state, initialize data acquisition card, load card configuration from system configuration files such as hardware address and device number.

## *4.3 Data Acquisition Module*

This module is the basis of the whole detection system, the most frequently used module and the precondition of data storage, analysis and process. SubVI in PCI-6221 data acquisition card can control and acquire data. In this module, acquisition parameters must be configured such as channel, sample time, input way, AD mode, start and stop, display, and alert set-up.

### 4.4 Data Management Module

This module can inquire storage and history data. The database is Microsoft Access, and can be accessed by LabSQL.

The acquired pressure and flow data can be stored in Access database. Database tables must be established in Access first; then acquired data can be scheduled into the tables by LabSQL in time order.

Data inquiry can call history data which are stored in Access in certain searching conditions. In this system, the specific searching condition is the data saving moment, and it is unique in timeline.

### 4.5 Data Analysis and Process Module

Data analysis and process is to eliminate the noises in signals and transform the signals into other forms easy to be analyzed and obtain useful information. This module can be divided into three sub-modules such as wavelet de-noising, amplitude analysis and autocorrelation analysis. This module can analyze real-time data and history data. Wavelet de-noising utilizes wavelet to de-noise the useless parts in the signal. Amplitude analysis is to analyze signals' amplitude parameters such as peak value, effective value and peak value index. Autocorrelation analysis is to determine signals' nature, detect the periodic signals in random ones.

## 5 Conclusion

Based on certain launching device, one hydraulic equipment detection system has been established applying virtual instrument technology, system's hardware and software have been designed and developed in this paper. Experiments indicate this detection system has such advantages as complete functions, good interactive interface and convenient operation, and can meet the needs of hydraulic equipments detection on certain launching device.

## References

1. W. Qinghua, *Research on fault diagnosis for hydraulic system based on fractal theory [D]* (Dalian Maritime University, Dalian, 2004)
2. L. Qiong, W. Quanche, F. Daliang, Analysis on general signal conditioning in data acquisition system [J]. J. Zhanjiang Normal College **27**(3), 130–133 (2006)
3. L. Zhongshen, A quick design method provided for the high order Butterworth low pass electrical filter [J]. Ind. Instrum. Autom. **4**, 27–29 (2007)

4. Z. Handong, J. Wenzhi, The database access technology based on LabVIEW in data acquisition system [J]. Ind. Instrum. Autom. **4**, 63–66 (2009)
5. C. Jing'an, L. Kejian, P. Dehui, Study and implementation on program method of LabVIEW with MATLAB [J]. Comput. Meas. Control **16**(5), 737–739 (2008)
6. Z. Li, K. Liu, Y. Zhao et al., MaPIT: An enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)
7. Z. Li, Y. Chen, H. Shi et al., NDN-GSM-R: a novel high-speed railway communication system via Named data networking. EURASIP J. Wireless Commun. Netw. **48**, 1–5 (2016)
8. X. Liu, Z. Li, P. Yang et al., Information-centric mobile ad hoc networks and content routing: a survey. Ad Hoc Netw. (2016), http://dx.doi.org/10.1016/j.adhoc.2016.04.005
9. Z. Li, L. Song, H. Shi, Approaching the capacity of k-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016). doi:10.1016/j.adhoc.2016.02.009

# The Arithmetic Research Based on the Probability Matching of Low Sampling Rate of Satellite Navigation Map

**Yankai Liu and Meijuan Yu**

**Abstract** Map-matching is a process of matching the track point of GPS to the digital map. The existing map-matching algorithm is based on the high sampling rate so that the algorithm has a low precise matching rate when sampling interval time increases. For that reason, this paper came up a special map-matching algorithm aiming at GPS track point with a low sampling rate. By considering many aspects like the geometry of road network structure, topological structure, mutual influence between the neighboring points and so on, the algorithm improves the accuracy of matching and determines the best matching results of GPS tracking points by probability calculation. Finally, it has been proved by experiment that the algorithm has good run time and accurate matching result.

**Keywords** A low sampling rate · Map-matching · GPS navigation · Probability matching

## 1 Introduction

As a result of mature development of the Internet technology, the smart city can be built and develop quickly, of which intelligence transportation is indispensable. The construction of intelligence transportation includes several areas: vehicle navigation, traffic flow analysis, and satellite positioning ETC which has not been intensively studied. All these mentioned application programs are based on the track. Its core steps are involved in GPS to accurately position the GPS track data of vehicles on the road [1], in other words, the map-matching.

Y. Liu · M. Yu (✉)
College of Computer and Information Engineering,
Tianjin Normal University, Tianjin 300387, China
e-mail: amy_1121@163.com

Y. Liu
e-mail: liuyankai@mail.tjnu.edu.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_102

Typical GPS track points data is a serious of sequential track points. Each GPS point consists of latitude, longitude and time-stamp information. However, on account of the limitation of GPS itself, the sampling and measuring process of GPS data and the return or accept process of the measuring data will have possible errors, which further lead to inaccurate GPS data [2]. Therefore, the original data need to be processed and then be used on the road network, that is to say, the map-matching.

## 2 Application Conditions of the Algorithm

With the development of science and technology, the number of any travel navigation system has increased sharply, such as a GPS embedded PAD and smart phone. Due to the spread of these devices, a large number of track point data can be available. But in the practice of lour life, only a low sampling rate (e.g., a sample point every 2 min) of GPS can be got because of energy consumption, cost consumption and so on. For example, there are more than 60,000 taxis in Beijing and most of them are equipped with GPS [3, 4]. Usually the taxi drivers drive on the road. In order to save energy consumption, the time intervals of their passing the GPS point is bound to increase, which leads to lower sampling rate of GPS track data [5, 6].

However, at present, the algorithm of map-matching is only for processing GPS data with high sampling rate (usually 10–30 s every one track point) [7]. When they use points with low sampling rate as their data, the matching error is over 50% [8–10]. Therefore, in view of the track point with a low sampling point, the paper will put forward an improved algorithm of map-matching. Besides, the low sampling rate here means collecting one track point every 1.5 min and more [11, 12].

## 3 The Algorithm Design

Map-matching system of GPS navigation based on a low sampling rate consists of three parts: the preparation of candidate point, the analysis with the times and spatial factors and the result matching.

### *3.1 The Preparation of Candidate Point*

The algorithm will give full consideration to the geometric structure of road network, so as to calculate the candidate point of the track point. It needs two steps to achieve this goal. First, finding out the possible section of the track point, in other words, the candidate point. Second, calculating the candidate point in the section by making use of the point-to-curve in the present geometric map-matching algorithm.

### 3.2 Spatial Analysis

In this step, we need to make the most of the geometry and topology information of road network to evaluate the candidate point getting from the first step. In this paper, geometry information is represented by the observation density and topology information is represented by transmission probability.

### 3.3 Time Analysis

In most cases, the algorithm can find out the best candidate point through spatial analysis so that choosing the true path $P_i$ among the candidate path $\{P_{i-1}, P_i \ldots P_n\}$. However, there is a special kind of situation that cannot be solved by the spatial analysis. As shown in Fig. 1.

In the above figure, a thick yellow line stands for highway and the thin blue line represents the common roads. The two roads are very close, so if we use spatial analysis to calculate the candidate point of $P_{i-1}$ and $P_i$, the results of the algorithm of two roads may be same. But if the average speed from $P_{i-1}$ to $P_i$ is 85 km/h, the two track points can be match on the highway because of road speed limits. Therefore, it needs the time analysis of track points.

Firstly, the algorithm needs to calculate the average speed $\bar{v}$ from $P_{i-1}$ to $P_i$, the formula is as follows.

$$\bar{v}_{(i-1,t)\to(i,s)} = \frac{\sum_{u=1}^{k} l_u}{\Delta t_{i-1\to i}} \tag{1}$$

The candidate point of $P_{i-1}$ is $C_{i-1}$, the candidate point of $P_i$ is $C_i$, the shortest path from $C_{i-1}$ to $C_i$ is a series of sections $\left[e_1^{'}, e_2^{'} \ldots e_k^{'}\right]$. Therefore, the $l_u = e_u^{'} \cdot l$ in formula, that is $l_u$, is the length of $e_u^{'}$, the member means the shortest length from $C_{i-1}$ to $C_i$, the denominator $\Delta t_{i-1\to i}$ means the time intervals from $P_{i-1}$ to $P_i$.

The algorithm thinks that every section $e_u^{'}$ in the road network has its own speed constraint value $e_u^{'} \cdot v$. This paper will use cosine calculation to describe the

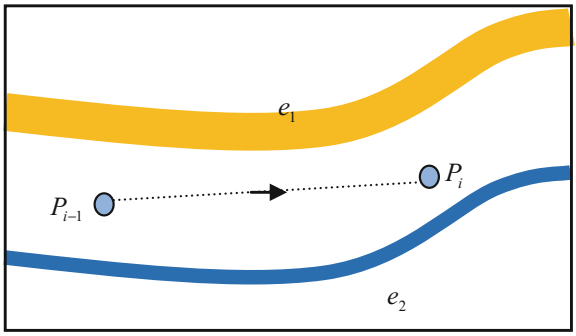



**Fig. 1** Influence factors of track points matching—time speed information

similarity between the average speed from $C_{i-1}$ to $C_i$ and the section constraint value $e_u^{'} \cdot v$. Therefore, the time analysis using time and speed information is defined as follows.

$$F_t\left(C_{i-1}^t \to C_i^s\right) = \frac{\sum_{u=1}^{k}\left(e_u^{'} \cdot v \times \bar{v}_{(i-1,\,t)\to(i,\,s)}\right)}{\sqrt{\sum_{u=1}^{k}\left(e_u^{'} \cdot v\right)^2} \times \sqrt{\sum_{u=1}^{k} \bar{v}^2_{(i-1,\,t)\to(i,\,s)}}} \tag{2}$$

### 3.4 The Result Matching

After spatial matching, the paper can find out a candidate graph $G^{'}\left(V^{'}, T^{'}\right)$ to its given track sequence $T: P_1 \to P_2 \to \cdots \to P_n$. $V^{'}$ is the candidate point of the track point, $T^{'}$ is the side represented by the shortest path between two adjacent candidate points.

Every candidate point in the candidate graph $G^{'}$ can be described by $N\left(C_i^s\right)$. Every side can be described by $V\left(C_{i-1}^t \to C_i^s\right)$ and $F_t\left(C_{i-1}^t \to C_i^s\right)$. To sum up, the paper can define the probability function of the map-matching as follows.

$$F\left(c_{i-1}^t \to c_i^s\right) = N\left(c_i^s\right) \times V\left(c_{i-1}^t \to c_i^s\right) \times F_t\left(c_{i-1}^t \to c_i^s\right), 2 \le i \le n \tag{3}$$

At last we can get a candidate path collection from the whole track (T)—$P_c$ $P: C_1^{S_1} \to C_2^{S_2} \to \cdots \to C_n^{S_n}$. If calculating each path's value $F(p_c)$, $F(P_c) = \sum_{i=2}^{n} F(c_{i-1}^{S_{i-1}} \to C_i^{S_i})$, the maximum will be the final matching path.

## 4 The Experimental Results and Analysis of the Algorithm

Based on the algorithm proposed in this paper, using the true data of road network and track network, the experiment can be designed and made successfully. The experimental data includes road network data and track data. Using the road network in Beijing and downloading from the open source OpenStreetMap, the road network contains 112 sections and 74 road network sites. The track data came from the true Shenzhou taxi GPS track data provided by Shenzhou Taxi Company. The formats are shown in Fig. 2.

```
1  116.400469 40.005635 13.27 357.80 2015-10-16 07:01:08
2  116.408297 40.022195 0.00 0.00 2015-10-16 07:03:07
3  116.408306 40.022177 0.00 0.00 2015-10-16 07:05:03
4  116.431571 40.021983 1.78 90.80 2015-10-16 07:09:25
5  116.436195 40.021727 6.88 102.70 2015-10-16 07:11:18
```

**Fig. 2** GPS track data

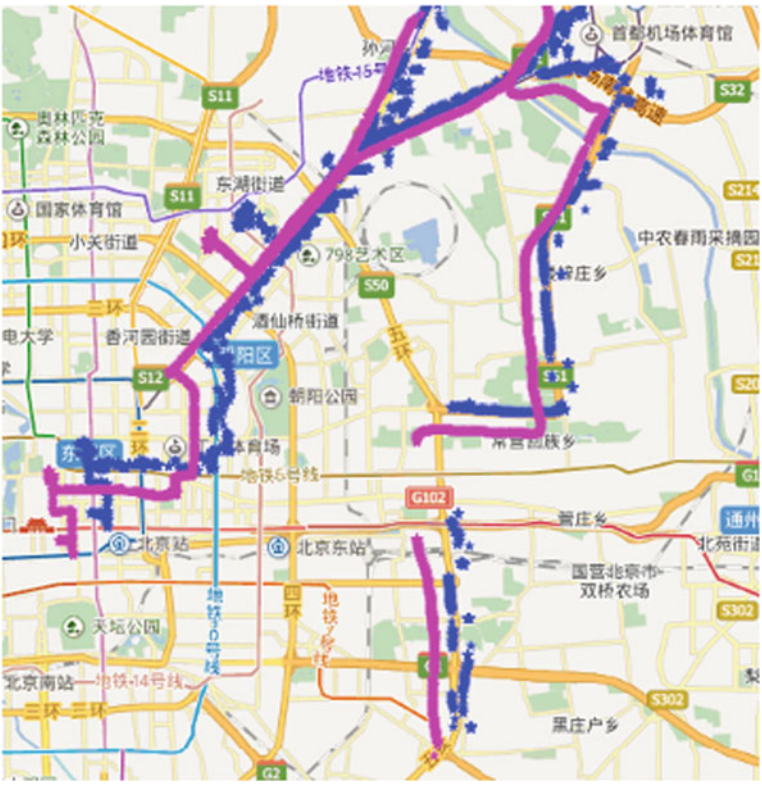



**Fig. 3** The matching effects

The first column of data represents the longitude of the driving point, the second column of data represents the latitude, the third represents the speed of the vehicle, the fourth represents the point of view and the last represents the time-stamp. The overall effects of the final matching of the experiment are shown in Fig. 3. The purple line is the track point sequence of the effect figure before matching and the blue line is the track point sequence after matching.

(a) Matching precision. In the experiment, there are about 1089 track points matching correctly on the roads, namely, the algorithm in this case can make the matching accuracy reach above 80%.
(b) Matching time. In this case, the track point n is 1354 and the section of road network m is 112. According to the theoretical calculation complexity formula, the algorithm operation time in this experiment is 3.2 min.

## 5 Conclusion

In view of the limitation of energy and resources in real life, the actual sampling intervals of getting GPS track points are large, while the existing map-matching algorithm is aimed at a high sampling rate. Therefore, the paper came up a special map-matching algorithm aiming at GPS track point with a low sampling rate. Based on probability matching, fully considers the geometric structure of road network (observation probability), topological structure (transmission probability), and the time speed information of vehicles (spatial analysis), the algorithm calculates every track point and determines the best matching point according to the results of probability calculation. At last, the paper verifies the matching precision of algorithm and time complexity through experimental analysis with actual data.

## References

1. P. Misra, P. Enge, Global positioning system: signals, measurements, and performance [M], (Ganga-Jamuna Press, 2011)
2. A.M. Pinto, A.P. Moreira, P.G. Costa, A localization method based on map-matching and particle swarm optimization [J]. J. Intel. Robotic Syst. **77**(2), 313–326 (2015)
3. K.W. Anshi, The method research of GPS data track [J]. Sci. Technol. Eng. **15**(11), 125–130 (2015)
4. H. Fan, The estimation of reasonable number of taxi and realization way in megacity—taking beijing as an example [J]. Res. Econ. Manag. **36**(8), 91–94 (2015)
5. X. Liu, Z. Li, Information-centric mobile ad hoc networks and content routing: a survey [J]. Ad Hoc Networks, In Press, Corrected Proof, 2016, Accessed 19 April 2016
6. Z. Li, Y. Chen, NDN-GSM-R: a novel high-speed railway communication system via named data networking [J]. EURASIP J. Wireless Commun. Netw. **2016**, 48 (2016)
7. Y. Zheng, L. Capra, O. Wolfson, H. Yang, Urban computing: concepts, methodologies, and applications. ACM Trans. Intel. Syst. Technol. **5**(3) (2014)
8. A.U. Peker, O. Tosun, T. Acarman, Particle filter vehicle localization and map-matching using map topology [C]. Intel. Veh. Symp. (IV), 2011 IEEE, 248–253 (2011)
9. S. Lamy-Perbal, N. Guenard, M. Boukallel et al., A HMM map-matching approach enhancing indoor positioning performances of an inertial measurement system [C], in *International Conference on Indoor Positioning and Indoor Navigation. IEEE*, (2015)
10. L. Giovannin, A novel map-matching procedure for low-sampling GPS data with applications to traffic flow analysis [J]. Universita Di Bologna (2011)
11. Z. Li, L. Song, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme [J]. Ad Hoc Networks, In Press, Corrected Proof, 2016, Accessed 2 March 2016
12. Z. Li, K. Liu, MaPIT: an enhanced pending interest table for NDN with mapping bloom filter [J]. IEEE Commun. Lett. **18**(11), 1915–1918 (2014)

# Study on the Simulation and Training System State Transition Method of the Complex Weapon Equipment on Operation-Oriented

**Peng Liu, Silong Zheng and Baohua Wei**

**Abstract** Aiming at the question which is hard to describe the simulation and training system state transition, the state transition method is put forward based on the FSM. The paper introduces the conception of the FSM, describes the system state process transition, establishes the state transition modeling, and then introduces the application by this method. The application shows that this method is suitable for the system which has finite states; the software code which adopts this method has good reusable, high efficient design and easy maintenance cost.

**Keywords** Weapon equipment · Simulation and training · State transition

## 1 Introduction

The state process of the simulation and training system is to describe the simulation and training process through abstracting the operation sequence or the operation rules of the weapon equipment and modeling the realistic and manageable state transition set based on the operation time and the logic. From the view of data flow, the state process transition is the process from one steady state to another steady state. There are two factors to decide the system state transition: one is the current system state, and the other is the event which causes the change of the system state.

The weapon equipment is more complex, the description of the system state transition is more difficult. The concrete manifestation is that: (1) The large system has the complex time, the logic relation and many keys; (2) One key has different meanings on different cases; (3) In order to accomplish one subject operation, some

P. Liu (✉) · S. Zheng (✉) · B. Wei (✉)
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: baty2012lp@126.com

S. Zheng
e-mail: zhengsilong2016@126.com

B. Wei
e-mail: weibaohua2016@126.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_103

keys need to be combined input. So it is a difficult question to adopt an appropriate method to describe the system state transition in the field of the simulation and training system.

Now, the "if … then" method is adopted more for the system state transition. The method is simple, but has more shortcomings: (1) The program code is complex and lengthy; (2) The coupling phenomenon of the logic relation is serious; (3) It has bad module and hard debugging; (4) It is fit for the system state transition in the simple logic. So the system state process transition method is put forward based on the finite state machine. This method could describe the system transition process simply and clearly.

## 2 The Basic Concept of the FSM

The Finite State Machine is engineering application of the finite automatic machine and has been widely used in computer science, information coding theory and so on. It mainly describes the process behavior that one system or one object is transited from different states. Its mathematics concept is as follows:

The FSM M is an ordered set including five objects: $M = \{S, I, O, f, g\}$

| | |
|---|---|
| $S = \{S_1, S_2, \ldots, S_n\}$ | is the state set; |
| $O = \{O_1, O_2, \ldots, O_n\}$ | is the output set; |
| $I = \{I_1, I_2, \ldots, I_n\}$ | is the input set; |
| $f$ | is the function from $I \times S$ to $S_k$; |
| $g$ | is the function from $I \times S$ to $O_k$; |

From the view of the complex time and the logic relation, it could be expressed as follows:

$$S_{i+1} = f(I_i, S_i), \quad O_i = g(I_i, S_i)$$

The working principle is that: the state set $S$ and the input set $I$ is known. Once the current input $I_i$ and the current state $S_i$ are confirmed, the next state $S_{i+1}$ and the output $O_i$ are confirmed on the effect of $f$ and $g$.

The core of the finite state machine is one object has one certain state at one certain moment. The transition between two states needs certain condition. When receiving one input (event), one output is confirmed and the transition happens.

So the finite state machine has distinct relation. The system which could apply the finite state machine also needs several conditions: (1) the system has several states which are finite; (2) the current state of the system is confirmed only by the last state and the input; (3) the system states are relatively stable and can keep invariable when no input. Above all, the finite state machine could describe the simulation and training system state process on operation-oriented effectively.

# 3 The System State Process Transition Method Based on the FSM

## *3.1 The Description of the System State Process*

The object of study for the simulation and training system on operation-oriented is the weapon equipment. It emphasizes the operation sequence and the operation flow rather than the description of the electric, dynamic or machinery. In general the state process transition is shown in Fig. 1.

(1) In certain stable state, the trainers press some buttons or switches at the simulating panel;
(2) The system acquires the state change of the buttons or switches and saves it;
(3) The system analyses the state change, judges the logic and output the result;
(4) The output displays at the simulating panel (For example, some indicator light is on or off, or the instrument panels changes, or the display changes), then the system is at relatively stable state.

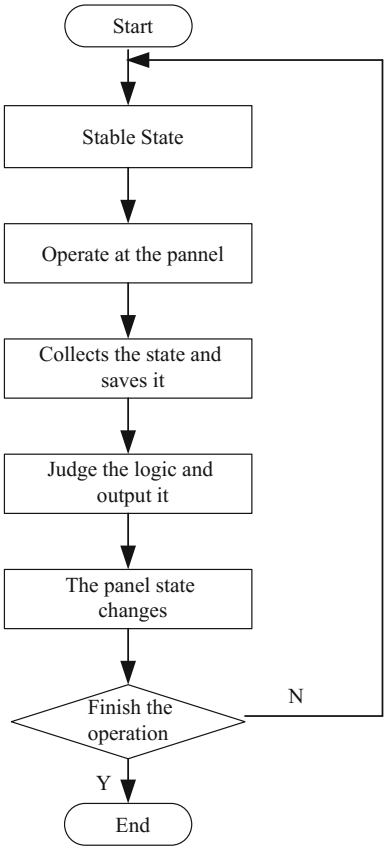



**Fig. 1** The system state process transition flow diagram

(5) The trainers operate continuously. The system state process operates according to the 1–4 until the end of the operation.

## 3.2 The Realization of the System State Process Transition Based on the FSM

1. **The system process transition model**
   Taking example by the structure of the AI expert system, the system process transition model is designed based on the FSM which is shown in Fig. 2. The model is comprised of state database, inference machine, knowledge database and the human–machine interactive interface.

   (1) The human–machine interactive interface
   The human–machine interactive interface is the interface that the system exchanges with the outside. The interface has two tasks: getting the input (event) and outputting the next state.
   From the view of design, there are two channels to getting the output: one is the change of the buttons or keys when the trainers operate at the simulating panel, the other is to receive the training interactive information from other subsystem. The output displays the conclusion (system state) from the inference machine.

   (2) The state database
   The state database saves the finite state of the simulation and training process. According to the operation sequence of the weapon equipment, the system state process can be divided into several states. Every state has several basic elements: the state name, the time logical sequence and the state phenomenon. The state phenomenon which describes the states of the lights, keys or switches at some certain state is the core of the state database.

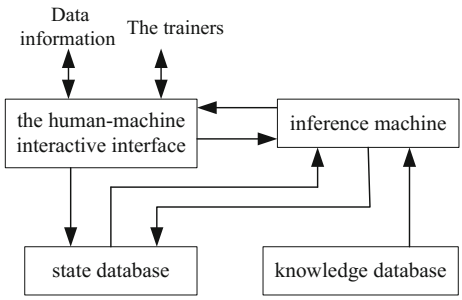



**Fig. 2** The structure of the state process transition model

(3) The knowledge database
The knowledge database saves the elements which are to drive the system state process transition. Its elements includes: the last system state, the current system state, the next system state, the event and so on. When the inference machine gets the current system state and input (event), read the content of the knowledge database, and then the next system state could be confirmed. The knowledge database derives the content from the operation sequence of the weapon equipment and the logic, and saves it according to the structure of the knowledge database.
The knowledge database adopts the mode of the state transition table. The state transition table shows the state transition logically. The basic pattern is shown that the current state (B) and the event (Y) confirm the next state (D). It is shown in Table 1. The state transition table focuses on the event, the current and the next state. This pattern has more feasibility and operability in software design.

(4) The inference machine
The inference machine implement. Through receiving the input (event) or the information from Ethernet, its main task is to search the state database and the knowledge database, choose the matching rule in the knowledge database, execute the rule to transmit the system state and output the result.

2. **Design thoughts**

(1) The description of the operation sequence
Detail the operation flow of the weapon equipment and fill in the corresponding the operation flow table. The operation flow table is used to describe the operation logic of the weapon equipment in overall processes. It should be detailed that could describe the logical relationship between the keys and the lights clearly.

(2) The confirmation of the system state division and the event
The system state is divided and confirmed according to the operation flow table. Obtain the input event and make the event as the system state division sign. The system state should be complete which includes all the states of the keys, switches, lights and so on.

(3) Design the state database and the knowledge database
Design the structure of the state database and the knowledge database. Fill in the system state and the input event into the corresponding database according to the definition of the database.

The state process transition flow design is shown in Fig. 3.

**Table 1** The state transition table

| The state transition table | | | |
|---|---|---|---|
| | State A | State B | State C |
| Event X | … | … | … |
| Event Y | … | State D | … |
| Event Z | … | … | … |

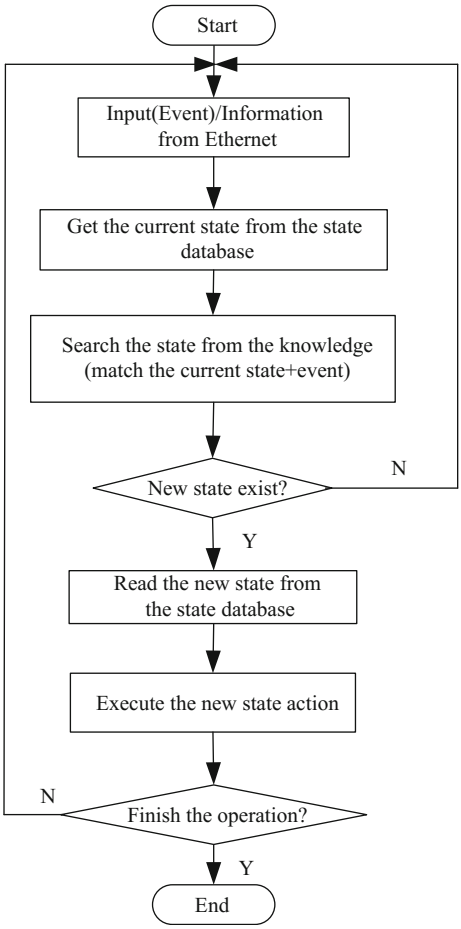



**Fig. 3** The state process transition flow diagram

# 4 The Application of the Simulation and Training System State Transition Method of Air-Defense Missile Based on the FSM

## *4.1 The Basic Structure of the Simulation and Training System*

The simulation and training system for an air-defense missile is composed of the control unit, S vehicle simulation subsystem and the F vehicle simulation subsystem. The system is established adopting by semi-physical mode. "Industrial control computer + data acquisition card + signal conditioning circuit + operation panel" is adopted in hardware design. VC++ is adopted to realize the control, communication and transition in software design. Vega is adopted to simulate the radar

display. MS Access is adopted to structure the database. Winsock is adopted to communicate with each subsystem.

## 4.2 Design and Realization

1. **The operation and training module flow design**
   The operation and training module is the core module of the software. Its task is to acquire the hardware state, judge the state transition and display the state. The UML use case diagram is shown in Fig. 4.
2. **The design of the hardware acquisition and control mode**
   S vehicle simulation subsystem and the F vehicle simulation subsystem have hundreds of keys, switches, and the lights. If adopting the interrupt mode, there will be so many interrupt that would effect the normal operation of the training process. The inquiry mode is adopted because of its easiness of the exploitation debugging and dispense with the state protection.
   When setting the inquiring cycle, one of the core targets is to make sure that every change of the hardware should be acquired timely and accurately. If the inquiring mode is too short, the invalid time of CPU is too much; The CPU work efficiency would be slow down, even the system works abnormally. If the inquiring mode is too long, there will be several operation times in one inquiring time. That case would lead to the inaccuracy of state acquisition result. Tests

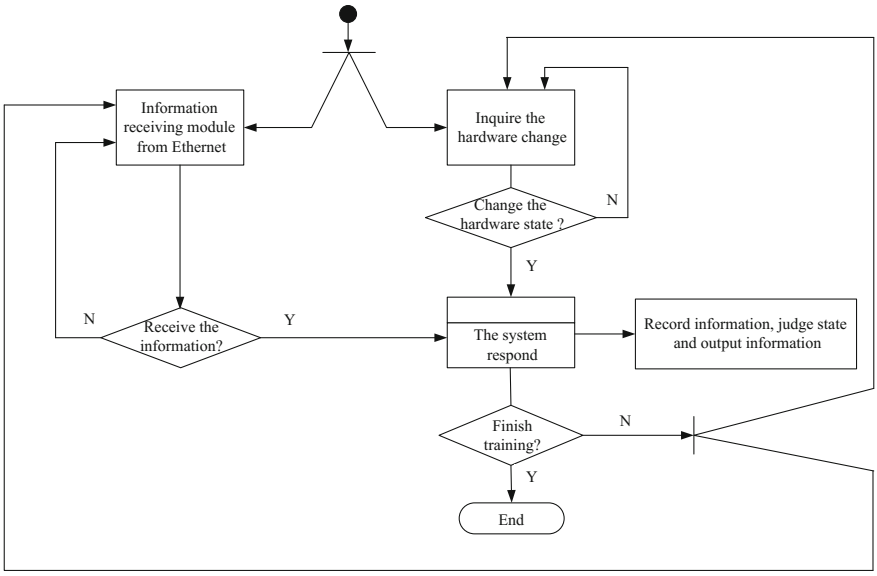



**Fig. 4** The operation training module based on the UML use case diagram

**Fig. 5** The equipment cabinet of the F vehicle subsystem

have shown that 100 ms is rational as the inquiring cycle which could fulfill the CPU work ability and ensure the accuracy of the inquiring state.

3. **The establishment of the state database**

   (1) Classify and describe the state of all the keys, switches, and the lights. BOOL type is adopted to describe the keys and the lights. Enumeration type is adopted to describe the band switches. Float type is adopted to describe the change of the instrument hand.
   (2) Design the structure which includes the entire variable. This structure could describe all the state of the hardware in the operation panel in any case.
   (3) Divide the simulation and training process state. Confirm the finite states. Every system state is filled in one system state table. The state database is composed of the entire system state Table

4. **The establishment of the knowledge database**
   According to the operation sequence and the system finite state, confirm the event which leads to the state transition, and then fill the system finite states and state transition event in the state transition table which composes the knowledge database.
5. **System implementation**
   The system state transition method based on the FSM is applied to the software design of the simulation and training system of air-defense missile based on the FSM. The equipment cabinet of the F vehicle subsystem is as follows (Fig. 5).

## 5 Conclusion

(1) The task interface between the equipment engineer and the software engineer is very distinct. It is the pipelined software designing mode which has high designing efficiency.
(2) The software code has good reusable which could apply to the any software design of the simulation and training system. Maintain the software easily and design it efficiently.
(3) This method is suitable for the system which has finite states. If the system has too many states, the state database would be expanded rapidly. It needs to be researched intensively how to solve this question.

## References

1. T. Xiao, Y. Zhang, J. Chen, *Introductory Theory of the System Simulation* (Tsinghua University Press, Beijing, 2000)
2. X. Liu, K. Huang, X. Zhu, The application of finite state machine in the CGF's behavior modeling. J. Syst. Simul. **9**(5), 663–665 (2001)
3. R. Zhu, X. Xie, A. Yu, J. Tu, Research on process model in virtual operating training system based on FSM. Microcomput. Inf. **28**, 153–154 (2009)
4. X. Liu, J. Pan, Finite state machine and its application to CGF of naval combat simulation. Comput. Simul. **8**, 24–27 (2007)

# Research on the Formal Representation of ATML Documents

**Shuyi Fan, Huixia Jiang, Baohua Wei and Wanming Liu**

**Abstract** In the process of constructing the ATS, developing of test program is an important and time consuming job. The traditional development approach of test program is in manual way. It makes the development cost very high, development cycle very long and portability very poor. Next, the automatic generation of test program will be a trend. Based on the study of ATML document characteristics, an approach to convert ATML documents into formal representation is proposed. First, the way of data type conversion is studied. Then based on the characteristics of entity documents, the formal way of entity conversion is studied. Finally, the method of converting the test description document into the formal representation of process control is studied. The effectiveness of the proposed approach is demonstrated by an example.

**Keywords** ATML · Formal representation · Test program automatic generation

## 1 Introduction

The purpose of test program automatic generation is converting ATML (Automatic Test Markup Language) test description documents and related documents into equivalent source code which can be compiled. Currently, research on automatic conversion of ATML documents to test programs is still less. Only some conversion methods between other format documents are studied in [1–4].

In addition to the test procedure information contained in test description document and contents in other documents are static definition or explanation information. This information can be directly converted into compiled source code.

S. Fan · H. Jiang · B. Wei
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: 316925561@qq.com

W. Liu (✉)
Hebei Normal University, Shijiazhuang, China
e-mail: lwmwlh@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_104

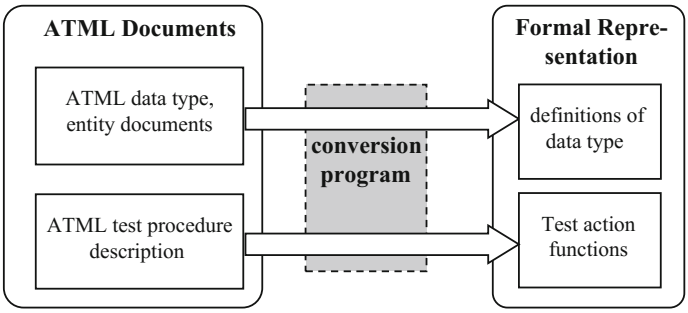



**Fig. 1** Conversion progress

The test procedure information contained in test description document need to be converted into a mid-formal format first, and then into source code that can be compiled by COTS (commercial off-the-shelf) compiler.

Depending on the contents of the ATML documents, there are three cases that converting the ATML document into formal representation. First is the conversion of data type. In ATML documents, the data types are mainly exist in Common.xsd. Second is the conversion of entity type description documents, such as UUT description of the document, etc. The three parts are divided into three independent modules, which can be configured to execute one or more modules. It should be noted that the data type conversion is done before the resource analysis and match in order to provide the desired data structure. The conversion progress is depicted in Fig. 1.

## 2 Conversion of Data Type in ATML Documents

Definitions of data type in ATML mainly contained in common elements, internal models, and signal definitions, such as STDBSC.xsd and STDTSFLib.xsd. The data type can be used by other ATML documents directly.

The primitive data types defined in XML Schema is directly referenced in ATML. Some of the primitive data types can be represented directly in high level language, others should be defined in customized data type, such as struct, union, etc.

In addition, we need to take extra action to some data type, such as data type normalizedString in XML Schema. All carriage return, line feed, and tab character should be replaced by a single space. We can define all of the string data type as string*, then call API function of the free or commercial XML parse program, such as MSXML, TinyXML in [5], XML Spy in [6] and Schematron in [7]. The API function can return correct string data. But we still need to manually write program for some special data type, such as dateTime.

In ATML documents, there are many abstract data type and the specific implementation data type derived from the abstract data type. As we know, many

high level language have not abstract data type. So we should use struct, abstract class, or union to represent these abstract data type. If ATML codes directly reference the derived data type, struct data type should be used. If ATML codes reference the abstract data type, abstract class should be used. If the ATML abstract type needs to be passed as a parameter in the COM interface function, the union should be used [7].

## 3 Conversion of Entity in ATML Documents

The entity documents mentioned here refers to the ATML XML Schema documents that descript parts of hardware or configurations of ATS. These documents include Instrument description, UUT description, test configuration, test adapter, test platform, and test results.

Each entity document can be defined as a class. The fields of the class depend on the contents of the entity document. Next, we take the UUT description document as an example to introduce the conversion of entity description document. The structure of UUT description document is depicted in Fig. 2. UUT description documentation refers to data type HardwareItemDescription in HardwareCommon (the prefix hc: in Fig. 2). It also refers to data type NonBlankString, ItemDescription, Document and attributes group DocumentRootAttributes in Common (the prefix c: in Fig. 2). The UUT description document can be converted into high level programming source code a

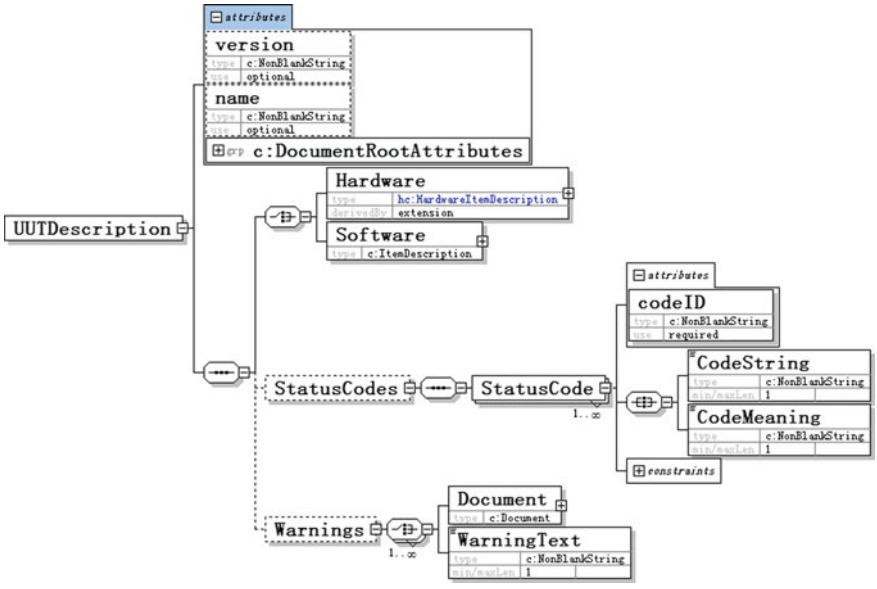



**Fig. 2** Structure of UUTDescription document

```
#include <Common.h>
#include <HardwareCommon.h>
class UUTDescription{
  public:
        char* version;
        ……
        c_ItemDescription SoftWare;
        tpWarings Warings;
        ……
private:
        typedef union{
        ……
        tpWarings* next;
        } tpWarings;
        ……
}
```

The principles and methods of conversion for other entity document are similar to this. In addition, the structure to save the test results should be defined for document TestResult.xsd. Generally, for the purpose of performance considerations and operation convenient, these results are stored in the database. It can be converted into XML format that meets the requirement of ATML.

## 4 Conversion of ATML Test Description Document

Different from the previous introduction of the data type and entity description document, the ATML test description document contains test flow and test action. Test flow should be converted into control structure of programming language, including flow control statements and functions, etc. Test action should be converted into customized formal statements which control test instrument, switch or adapter. Therefore, the focus of conversion of ATML test description document is the division of test actions, usage of control statements and the calling method of driver, including test instrument, switch, or other test resource. The major part of test description document is the element DetailedTestInformation whose conversion methods and principles will be discussed in detail below. The structure of element DetailedTestInformation is shown in Fig. 3.

### *4.1 td:EntryPoints*

The element identifies the test groups or actions that can be used as entry points. For each entry point, test conditions are completely stated and are not dependent on

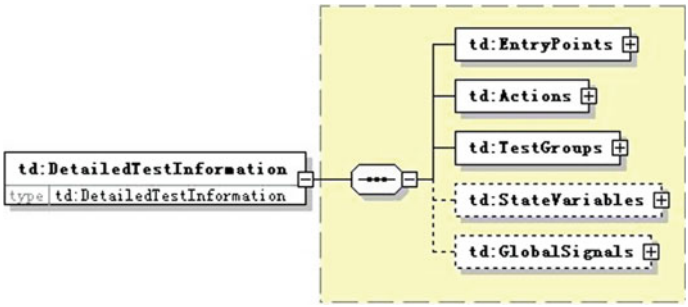



**Fig. 3** Structure of element DetailedTestInformation

previous tests or setups. The element can contain multiple entry points. When the element contains multiple entry points, the switch statement can be used to determine the entry point function the user select.

## *4.2 td:Actions*

The Action elements of type Test used to detect and isolate UUT faults. A typical test verifies a characteristic of the UUT by (1) applying one or more stimuli, (2) measuring the UUT response to these stimuli, and (3) comparing the response with limits or expected values that define the acceptable UUT operation. When measurement results are within limits or are equal to the expected values, the test returns a "Passed" outcome. When limits are exceeded or the expected values are not matched, the test returns a "Failed" outcome and, if applicable, a qualifier indicating which limit was exceeded (e.g., "Low" or "High"). When a test was not completed, it returns an "Aborted" outcome. The qualifier of that outcome may provide additional information regarding the cause of the abort, for example, "Manual Intervention," "Timeout," "Instrument Error," and so on.

The Action elements of type SessionAction represent individual nontest actions to be performed on the UUT. When successfully completed, session actions return a "Done" outcome. When not completed, they return an "Aborted" outcome. The qualifier of that outcome may provide additional information regarding the cause of the abort. The structure of element Action and its conversion is shown in Fig. 4.

In the left part of Fig. 4, element td:Parameters describe the parameters of the action and shall specify their values. Element td:LocalSignals identify local signals, shared by multiple operations specified for the current action. Element td:Conditions specify the preconditions and postconditions for the action. The preconditions are used to ensure the Action can be executed only if these conditions are satisfied. The postconditions indicate the conditions that Action can achieve after successful execution. Element td:Behavior is the most important element in Action. It describes the behavior of the action in detail.

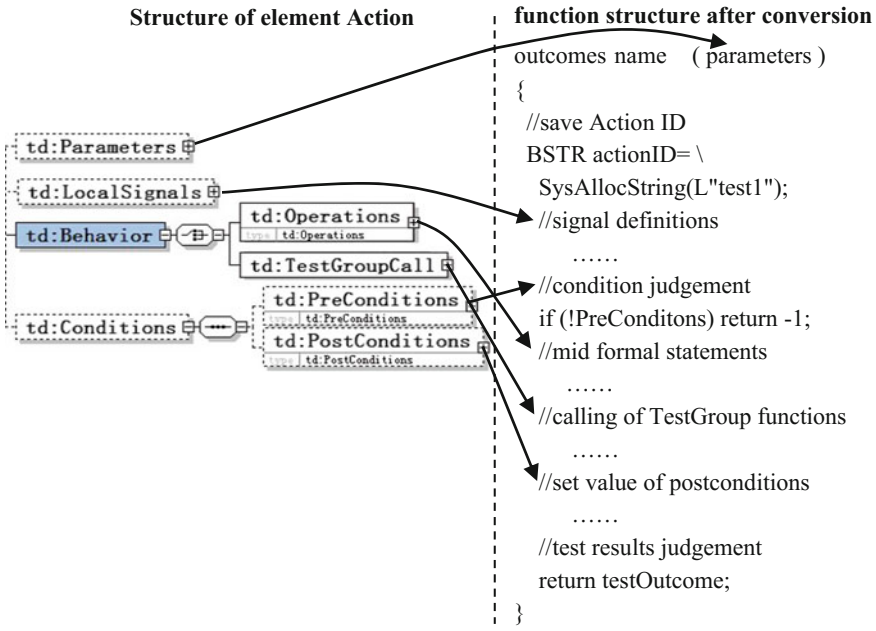



**Fig. 4** Structure of element Action and its conversion

## *4.3 td:TestGroups*

The element td:TestGroups contains one or more td:TestGroup elements, and each Td:TestGroup element can contain one or more Action/TestGroup. Execution order, parameters, preconditions, postconditions, initial operation, and termination operation of Actions can be specified in element TestGroup. One TestGroup can verify multiple characteristics of UUT and get one outcome. When performing UUT fault detection, it will return "passed" or "failed." When performing fault isolation, it will return "passed" or a certain kind of "failed." We can isolate the fault according to the detailed description of "failed." There are five types of TestGroup. They are TestGroupDiagnosticModel, TestGroupParallel, TestGroupSequence, TestGroupSerial, and TestGroupUnspecifiedOrder.

TestGroupDiagnosticModel describes a test group where the order of execution is determined by a diagnostic reasoner in accordance with a diagnostic model. TestGroupParallel describes a test group where actions can be executed in parallel, depending on resource availability. The execution of the test group finishes after all of actions finish. TestGroupSequence describes a test group where actions are executed sequentially. The order of execution depends on the outcomes returned by the tests. Each step in the sequence represents of the execution of an action (test or session action) or a test group. This mode of representation for test sequencing is commonly known as a "fault tree." TestGroupSerial describes a test

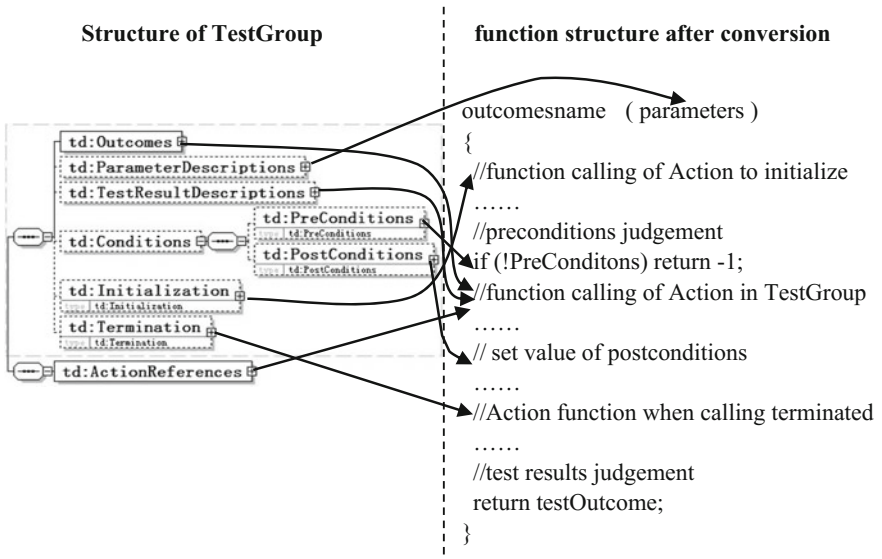



**Fig. 5** Transformation of TestGroup

group where actions are executed sequentially. The order of execution is fixed. TestGroupUnspecifiedOrder describes a test group where execution is serial and can be in any order.

The representation and conversion of TestGroup are similar to Action. The corresponding relationship between the structure of the element TestGroup and the converted formal code is shown in Fig. 5.

The execution order of the Action contained in the element td:ActionReferences at the left side of Fig. 6 is related to the type of TestGroup. And it also determines the execution order of the corresponding Action function. The content of element td:Outcomes and td:TestResultDescription combine with the judgment process of test resultsand Outcome that get from calling of Action function. If the Outcome value of a test is more than two, it will be converted to "switch" structure, otherwise it will be converted to "if…else…" structure.

## 5 Example of Conversion

The conversion of entity document is simple. So the example of entity will not be given below. We will give a conversion example of test description document. Figure 7 shows an example of test description of test description document. There are ten Actions and two TestGroups. The relationship of Action and TestGroup is shown in Fig. 6.

**Actions and Test Groups**

td:Actions

td:Action (10)

| | = xsi:type | = ID | = name |
|---|---|---|---|
| 1 | td:Test | test1 | V_CC_to_GND_Resistance_Test |
| 2 | td:SessionAction | sa2 | Apply_UUT_DC_Power |
| 3 | td:Test | test3 | V_O_AC_Voltage_Test |
| 4 | td:Test | test4 | V_C_AC_Voltage_Test |
| 5 | td:Test | test5 | V_C_DC_Voltage_Test |
| 6 | td:Test | test6 | V_E_DC_Voltage_Test |
| 7 | td:Test | test7 | V_B_DC_Voltage_Test |
| 8 | td:Test | test8 | VBE_Test_on_V_C_Fail_High |
| 9 | td:Test | test9 | VBE_Test_on_V_C_Fail_Low |
| 10 | td:SessionAction | sa1 | Remove_UUT_DC_Power |

td:TestGroups

td:TestGroup (2)

| | = xsi:type | = ID | = name |
|---|---|---|---|
| 1 | td:TestGroupSequence | seq1 | FaultDetectionAndIsolation |
| 2 | td:TestGroupSequence | seq2 | VBE_Tests |

**Fig. 6** Transformation of TestGroup

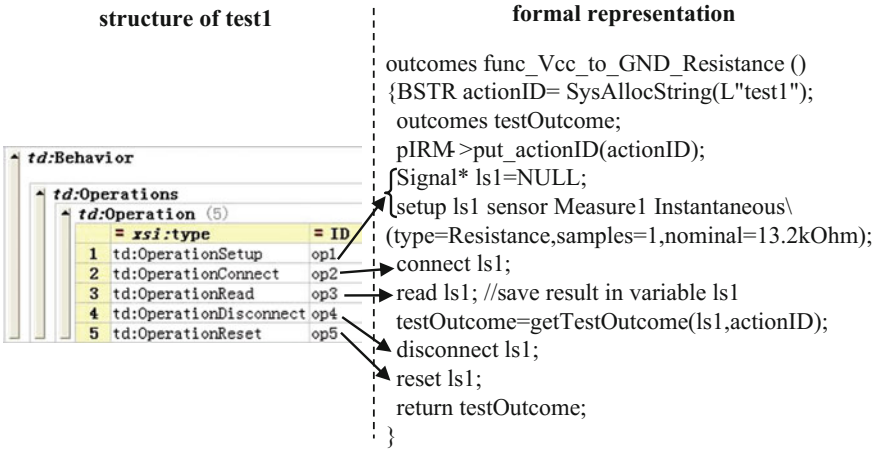



**Fig. 7** Document structure of test1 and FTDL codes

In Fig. 7, the Action with ID = "test1," name = "Vcc_to_GND_Resistance" is the first Action. This Action is used to measure the Vcc resistance to ground. It contains five Operations. These Operations are used to define measurement signal (resistance), connect the measurement signal, measure and reset the measurement signal. According to the principle and method mentioned last section, the structure of "test1" and the converted formal code is shown in Fig. 7.

In Fig. 8, function getTestOutcome (ls1, test1) is a customized public function. It can compare the test value with Upper and lower limits. And then return the comparison result. The parameter "test1" is used to find out the Upper and lower limits.

## 6 Summary

In this paper, we discuss the conversion method to formal representation of entity document and test description document in ATML. The entities in entity documents can be converted into suitable data types. The Actions and testGroups in test description document can be converted into suitable formal statements. The effectiveness of the proposed method is demonstrated by an example. The formal conversion method proposed in this paper will lay a solid foundation for test program automatic generation based on ATML documents.

## References

1. D. Ordonez Camacho, K. Mens, APPAREIL: a tool for building automated program translators using annotated grammars, in *23rd IEEE/ACM International Conference on Automated Software Engineering*, May 2008, pp. 489–490
2. Y. Wang, M. Wu, Case studies on translation of RTPA specifications into Java programs, in *IEEE CCECE 2002*, August 2002, pp. 675–680
3. R.C. Waters, Program translation via abstraction and reimplementation. IEEE Trans. Softw. Eng. **14**(8), 1207–1228 (1988)
4. K. Yasumatsu, N. Doi, SPiCE: a system for translating smalltalk programs into a C environment. IEEE Trans. Softw. Eng. **21**(11), 902–912 (1995)
5. Z. Li, K. Liu, Y. Zhao et al., MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)
6. Z. Li, Y. Chen, H. Shi et al., NDN-GSM-R: a novel high-speed railway communication system vis named data networking. EURASIP J. Wirel. Commun. Netw. **48**, 1–5 (2016)
7. X. Liu, Z. Li, P. Yang et al., Information-centric mobile ad hoc networks and content routing: a survey. Ad hoc Netw. (2016). doi:10.1016/j.adhoc.2016.04.005

# Whole Design of Anti-tank Missile Equipment Maintaining Training System

**Jianhua Xie, You Li, Bo Dong and Peng He**

**Abstract** This paper aims at the development of maintenance training system of anti-tank missile equipment. It uses technique of large-scale real-time circuit simulation, technique of virtual maintenance and half substance simulation to present the whole structure of maintenance training system of anti-tank missile equipment and design the model of layered architecture of federation based on HLA/RTI to support the whole structure. It can supply firmly base for developing anti-tank missile equipment maintaining training system. The whole design of maintenance training system and layered architecture of federate model have better ideas for similar system.

**Keywords** Maintaining training · Whole structure · Layered architecture of federation

## 1 Introduction

According to anti-tank missile equipment maintaining supply, a lot of work is testing, maintaining and changing for electronic circuit, electronic component and mechanical component. Then the knowledge of electronic circuit, maintaining skill and operation level of maintenance man have important real effects according to protection missile equipment in good condition and to improve operating life and effectiveness of weapon system.

And now, there is no relative complete maintaining training equipment for anti-tank missile equipment. Maintenance man can only do maintaining training

J. Xie (✉) · P. He
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: rainmanbenben@163.com

Y. Li
Information Technology College, Nanchang University, Nanchang, China

B. Dong
Bengbu Armor College, Bengbu, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_105

depends on the drawing and real equipment, it has more limitation. This paper aims the demand and presents the method of design and implement for anti-tank missile equipment maintaining system.

The anti-tank missile equipment maintaining system is a distributed system based on HLA/BOM. The system adopts technique of computer simulation, technique of virtual maintenance, technique of large-scale electronic simulation, technique of distributed interactive simulation to build. The system can do maintaining training and operation training in the realistic virtual training environment, it includes electronic fault maintaining training, mechanic fault training and operation training. It can cover the training content of multiform model of anti-tank missile equipment.

## 2 Whole Structure of Anti-tank Missile Equipment Maintaining Training System

The whole structure of anti-tank missile equipment maintaining training system is a distributed interactive system based on network bus. Figure 1 shows the composition structure of anti-tank missile equipment maintaining training system [1].

The system includes master computer of master control board, simulation computer of training terminal, man–machine interaction, etc. The master computer implements man–machine interaction for the trainer and the system, implements to set fault [3], set system, monitor training process, monitor simulation, etc. Training terminal can divides to four types according to functions. The first type is electronic

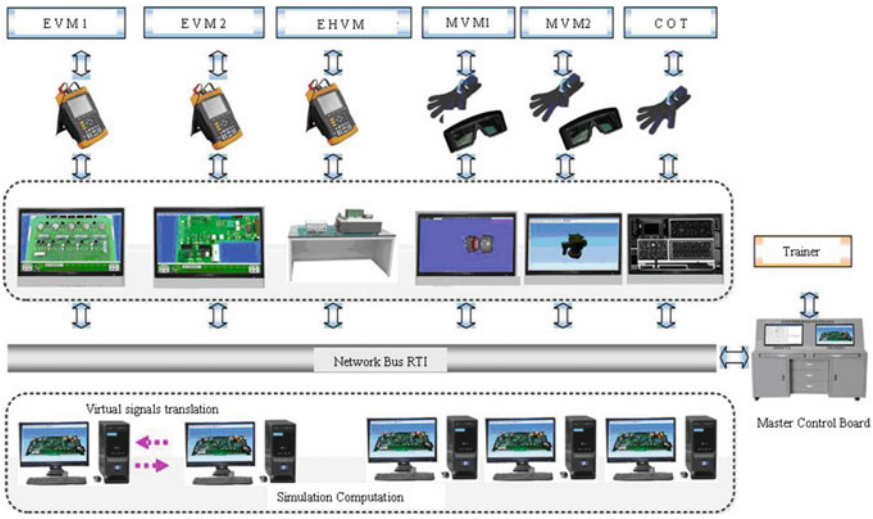



**Fig. 1** Structure of anti-tank missile equipment maintaining training system

fault virtual maintaining training terminal, we call it EVM. The second type is electronic fault half substance maintaining training terminal, we call it EHVM. The third type is mechanic fault virtual maintaining training terminal, we call it MVM. The fourth type is complex operation virtual training terminal, we call it COT. Every training terminal is an agent instance, it implements virtual vision and working process simulation for electronic fault, mechanic fault, complex operation training of anti-tank missile equipment. The training man uses interactive equipments to measure and debug for stated electronic fault, to department and maintain for mechanic fault and to do operation complex training.

## *2.1 Electronic Fault Virtual Maintaining Training Terminal*

According to electronic fault virtual maintaining training terminal, simulation computer can do action with virtual maintaining operation according to translation relation of virtual signals and circuit simulation. It can simulate real working state. Every simulation node can realize distributed interaction through BOM. It can compose distributed simulation network to simulate the relation and signals translation of each circuit model of anti-tank missile equipment, then it can simulate whole working states of anti-tank missile equipment. According to simulation computer of simulation node, it uses Spice3f5 descriptive language as circuit descriptive language of anti-tank missile equipment for circuit models. After set virtual fault by trainer, it can form fault descriptive language according to circuit descriptive language of anti-tank missile equipment to modify circuit descriptive language of anti-tank missile equipment corresponding, and then it can simulate fault state of anti-tank missile equipment.

Figure 2 shows the basic principle and process of circuit fault virtual maintenance of anti-tank missile equipment [1, 2].

When the virtual signals from former class are coming to original class, it forms virtual signal description according to characters of signal at first, then it do simulation with Spice3f5 according to circuit description of anti-tank missile equipment, and then it gets signal characters of every circuit nodes to form virtual output, after that it can translate the output to back class to calculate [4]. When the maintenance man is doing virtual maintaining operations, the operations of circuit device repairing and changing done by him can transform to description language of maintenance process, and then it can modify the circuit description language to simulate process of virtual maintenance. Spice3f5 simulating engine can simulate with circuit description of anti-tank missile equipment, then it can simulate circuit working state after maintenance, and then training man can test and measure it until he eliminates the fault.

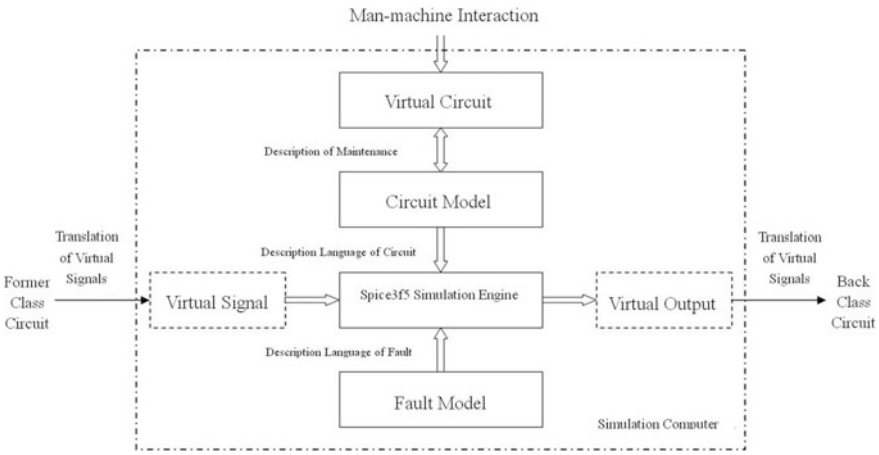



**Fig. 2** Basic principle and process of circuit fault virtual maintenance of anti-tank missile equipment

## *2.2 Electronic Fault Half Substance Maintaining Training Terminal*

According to electronic fault half substance maintaining training terminal, it is supplement relationship with electronic fault virtual maintaining training. It can be used to solve maintenance questions of circuit board during realistic circuit simulation cannot simulate in electronic fault virtual maintaining. According to complex circuit board of missile equipment, realistic circuit simulation software cannot finish its working process. Electronic fault half substance maintaining training terminal adopts electronic signal generator, high voltage transfer card, waveform generator board, etc. In the environment of VC++, it designs special API software package. It can finish half substance simulation of circuit board of missile equipment, it can use oscilloscope and universal meter to measure the test points and fault debugging for maintenance operations. It can solve maintenance training questions much better fitting-in with electronic fault virtual maintaining for electronic fault of anti-tank missile equipment.

## *2.3 Mechanic Fault Virtual Maintaining Training Terminal*

According to mechanic fault virtual maintaining training terminal, simulation computer constructs maintenance model machine of mechanic fault virtual maintaining. Mechanic fault maintenance model machine is facing to task to design. It has definite geometric and function truth degree similar with physical model machine. It has sport and physical characters of space, time and freedom restraint. It

can realize to simulate mechanic fault maintaining process combining with virtual reality.

Mechanic fault maintenance model machine of anti-tank missile equipment includes modelling of geometric characters, modelling of physical characters, modelling of interaction characters and modelling of behaviour characters. Modelling of geometric characters is not only real external appearance, but also it can show the structure and restraint relationship of each parts of anti-tank missile equipment. According to modelling of interaction characters, it needs to define interactive device and interactive modes. According to modelling of behaviour characters, mechanic fault maintenance model machine shows characters of unrestrained transition and rotation. Demounting/assembling simulation of mechanic components is hot point for research of virtual maintenance. But it has universal problem of complex modelling methods. The virtual maintenance training system presents VDN of anti-tank missile equipment based on theory of synchronizing network. It can describe demount action relation, sequence planning and other maintenance information of maintenance process. VDN uses relation of demount logic as first description object, it shows variation functions with demount time and resource, it do not increase elements of the net, and so it reduces the complex degree of modelling, and then it solves the question better during demount modelling.

### *2.4 Complex Operation Virtual Training Terminal*

According to complex operation virtual training terminal, it constructs operation training model with accuracy to size and realistic external appearance of certain gravis type anti-tank missile. The model was constructed by software of AutoCAD and Multigen Creator. According to investigation for real operation rules, it realizes virtual whole process operation training of certain gravis type anti-tank missile in the development environment of VC++. It can simulate each operation of certain gravis type anti-tank missile. It can use the fingers of operator and touch screen to directly operate on–off, rotary knob, etc., it has realistic effect.

## 3 Architecture Design of Maintenance Training System

### *3.1 Whole Structure*

According to HLA/RTI simulation standard, structure characters of anti-tank missile equipment, working principle of every parts and analysis of process of maintenance training, architecture of anti-tank missile equipment based on HLA adopts topological structure based on RTI showed by Fig. 3.

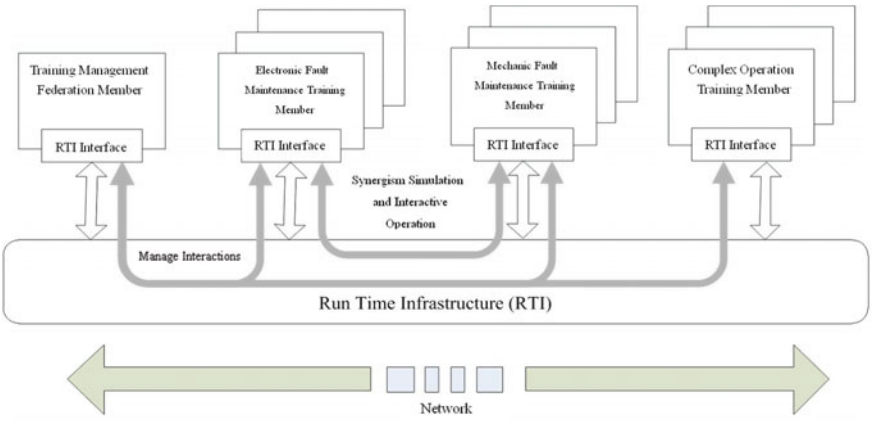



**Fig. 3** Architecture of anti-tank missile equipment based on HLA

In the structure, RTI has the effect of distributed operation system. All the simulation interactions of federation members are through RTI. RTI manages all the federation, it includes putting in and exit of members, timing of members, time pushing, refreshing and mirroring of object class, delivering and receiving of interactive class, it includes every interactions.

Tint arrow shows the designed member interface according to RTI between every federation members and RTI. It is finished through callback service of RTI Ambassador and Federate Ambassado. It is explicit to see in the codes of the system. It combines two function classes in common condition (mainly RTI service, callback service itself is method structure of the given class. The programmer only needs to imply the method.)

Deep color arrow between every federation members shows the detail content of interactive information. It is defined through analysis detail working principle of every components of the equipment and through the process of corresponding maintenance training and operation training. The information mixes together in RTI service, it is not to shows in the codes of the system. The interactive content can be divided to two classes, the first class is the management information between federation management member with maintenance member and operation member, the second class is the interactive information between different maintenance members.

## *3.2 Layering Structure of Maintenance Training Federation Member*

In the maintenance training system federation, according to each federation member, especially structure and function of federation member responding for

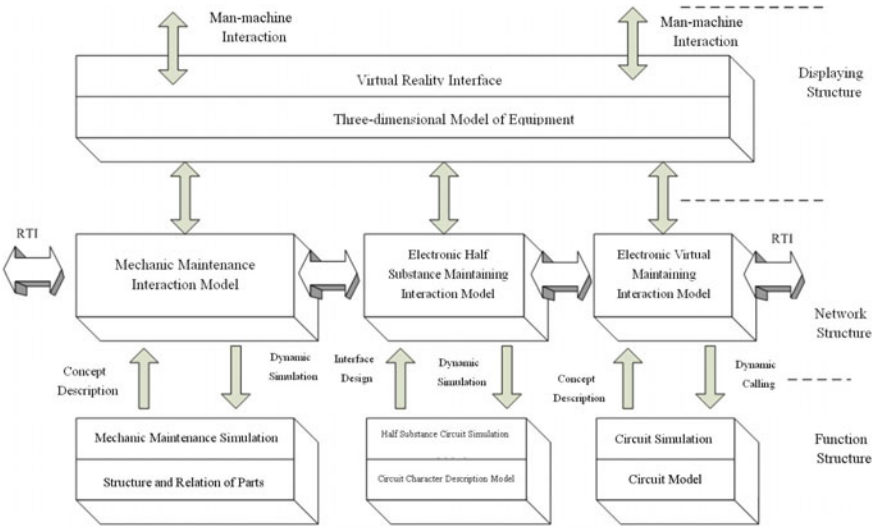



**Fig. 4** Layered architecture of maintenance federate

maintenance work is an important research. The system presents a kind of layered architecture of federate (LAF) as reference model of design of federation member of anti-tank missile equipment maintaining training system.

Layered architecture of federate of anti-tank missile equipment maintaining training system is showed by Fig. 4. There are function structure, network structure and display structure from low to high. Function structure is the base structure, it can simulate simulation object, it can simulate the process of working and maintenance of anti-tank missile equipment and it is the base of maintenance training. Function structure is mathematic model of counted in common, it can supply data for network structure and displaying structure, and it can receive control information from upper structure. Interaction information from upper structure can dynamically call function structure to finish simulation.

Network structure is the middle structure, it contacts with function structure and displaying structure, it is abstract description of function structure and defining every impossible interaction modes during the system operating. Network structure's most important function is to realize the data communication with other federation members based on HLA/RTI.

Displaying structure is the top structure, it displays all the vision effect of the virtual maintenance training system. It should include three-dimensional model of all components of the equipment and half substance measuring nodes. It is the maintenance interface of the system. Operating man realizes man–machine interactions through the displaying structure.

According to layered architecture of federate is the layered modelling method. Table 1 shows the corresponding relation between layered architecture and model.

**Table 1** Layered architecture model

| Architecture | Model | Modelling method |
|---|---|---|
| Function | Simulation model | Field tool |
| Network | Interaction model | Concept description, OMT, BOM |
| Displaying | Three-dimensional model; half substance measuring node | 3D Max, creator |

## 4 Conclusion

The paper presents design of the whole architecture and its implementing method with the background of developing anti-tank missile equipment maintaining training system. And then the paper presents a kind of layered architecture of federate to solve the problem of the information interactions of maintenance training system. It can supply firmly base for developing anti-tank missile equipment maintaining training system. The whole design of maintenance training system and layered architecture of federate model have better ideas for similar system.

## References

1. J. Xie, J. Ma, X. He, Design and implement of circuit virtual maintenance system of missile equipment, in *The 6th System Modeling and Simulation high grade Forum of China* (2011), pp. 313–316
2. Y. Yang, J. Zhao, W. Wu, Design of circuit simulation software based on spice. [J] Comput. Simul. **24**(5), 2390–2394 (2007)
3. Z. Li, K. Liu, Y. Zhao et al., MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)
4. Z. Li, Y. Chen, H. Shi et al., NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wireless Commun. Netw. **48**, 1–5 (2016)

# Analysis on Asynchronous Start Permanent Magnet Synchronous Motor Cogging Torque Optimization Based on Equivalent Magnetic Motive Force

**Chen Wang, Qingfeng Sun, Guanghua Cao and Jian Zeng**

**Abstract** Asynchronous Start Permanent Magnet Synchronous Motor has slots at both sides, so the formation mechanism of its cogging torque is more complex. To avoid influence on analyzing cogging torque by rotor slots, this paper propose a method of making the magnetic motive force generated by rotor bar equivalent to magnetic motive force generated by the permanent magnets, obtaining the analytical expression formula of such motor's cogging torque, to analyze the impact of the rotor tooth width, rotor slots number and rotor teeth shape, and other rotor parameters on motor cogging torque, and gives the method of reducing asynchronous start permanent magnet synchronous motor cogging torque; uses the effectiveness of the method proposed by finite element analysis verification and conducts the comparative analysis of the motor's performance before and after optimization. The results showed that selecting the appropriate rotor tooth width and rotor slots number can effectively weaken the asynchronous start permanent magnet synchronous motor cogging torque.

**Keywords** Asynchronous start permanent magnet motor · Cogging torque · Rotor tooth width · Magnetic motive force

C. Wang (✉) · Q. Sun (✉)
Department of Electrical Engineering, Anhui Technical College of Mechanical and Electrical Engineering, 241000 Wuhu, Anhui, China
e-mail: wangchen1071@163.com

Q. Sun
e-mail: sunqife@126.com

G. Cao · J. Zeng
School of Electrical Engineering and Automation, Jiangxi University of Science and Technology, 341000 Ganzhou Jiangxi, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_106

## 1 Introduction

With the rapid development of high-performance permanent magnetic materials, permanent magnet motors have been widely used in industrial robots, elevators, and other industrial places thanks to such advantages as simple rotor structure, high power density, high power factor, etc. [1–4]. However, there is mutual interaction between the magnetic field generated by rotor permanent magnets of the permanent magnet motor and the stator slot, generating cogging torque. And the generation of cogging torque will result in greater torque ripple, causing motor vibration and noise. So how to reduce the cogging torque has become one of the problems focused by relevant experts and scholars of permanent magnet motor direction.

Literature 5 uses the method of finite element analysis to analyze the influence on permanent magnet operating point by the magnetic motive force of stator and rotor. The results showed that different loading conditions and initial position of the rotor have great influence on demagnetization of permanent magnet; Literature 6 established 2 Pole 1.5 kW finite element analysis model of asynchronous start permanent magnet synchronous motor to study the effect on cogging torque by pole arc coefficient, the stator auxiliary slot, and other parameters, but it has a long calculation period and amount of calculation; Based on the consideration of saturation, Literature 7 uses the finite element to analyze the influence of the motor's stator and rotor on cogging torque, but the method is not easy to obtain the relationship between the parameters and cogging torque; Literature 8 proposes a new method for weakening built-in permanent magnet motor cogging torque. The studies on cogging torque from abovementioned Literatures all focus on the unilateral slotted permanent magnet motor, using the motor equivalent gap length method to study the relational expression between structural parameters and cogging torque, having proposed using the stator chute, rotor skewing, change pole arc coefficient and other methods to weaken cogging torque of permanent magnet motor.

Asynchronous start permanent magnet synchronous motor is a permanent magnet motor with special structure. There are squirrel-cage bars on its rotor in order to improve the starting torque of the motor. The squirrel-cage bars and stator windings make slots on both sides of the stator and rotor, and the bilateral slotting makes the motor's air gap more complex. The analysis method of common unilateral slot cogging torque is difficult to analyze and calculate such motor's cogging torque. Literature 15 derived the equivalent gap length expression of the asynchronous start permanent magnet synchronous motor using analytical method, and then deduced the cogging torque expression, proposed the method of reducing cogging torque of such motor. But the cogging torque expression proposed by this method is complex, which is difficult to obtain a clear physical concept.

This paper makes the rotor magnetic motive force equivalent to the distribution magnetic motive force corresponding to stator magnetic motive force, which has avoided the complex calculations of air gap effective length, deduced the cogging torque analytic expression of asynchronous start permanent magnet synchronous

motors, given out a clear relationship between the rotor and stator structural parameters and cogging torque, analyzed the influence of on motor cogging torque by rotor tooth width, the rotor slot number and size of the rotor slots, proposed the new methods of changing the rotor tooth width and rotor eccentricity to weaken asynchronous start permanent magnet synchronous motor cogging torque, using the method of finite element analysis to verify validity of this method.

## 2 Analytical Analysis of Cogging Torque

The method of analytical analysis is to determine the relationship between the stator and rotor parameters and the motor cogging torque. The following assumptions are made without considering its accurate calculation: (1) silicon steel permeability approaches infinity; (2) the centerline position the specified permanent magnetic pole is $\theta = 0$; (3) the slot shape of both stator and rotor is rectangular; and (4) the relative position angle between stator and rotor is $\alpha$.

Cogging torque is the reluctance torque generated by the relative movement between rotor permanent magnet and the cogging between stator and rotor. As for asynchronous start permanent magnet synchronous motors, permanent magnet is on the rotor, when relative motion is generated between stator and rotor, the energy inside magnetic steel remain unchanged, so we just need to consider the energy changes in air gap magnetic field, and the cogging torque expression can be expressed as below:

$$T_{cog} = -\frac{\partial W}{\partial \alpha} = -\frac{\partial}{\partial \alpha}\left[\frac{1}{2\mu_0}\int_V B^2(\theta,\alpha)dV\right]. \quad (1)$$

This paper adopted the asynchronous start permanent magnet synchronous motor with tangential structure as shown in Fig. 1. Analyze such motor's cogging torque, and radial motor as well as such motor with other structural type have similar analysis method.

As for the asynchronous start permanent magnet synchronous motor with tangential structure, the rotor teeth are evenly distributed under each pole. Since there is no need to calculate the cogging torque, the magnetic flux leakage at rotor slot can be ignored, and the approximately equivalent magnetic flux runs through the entire rotor teeth. The distribution of remanence $Br(\theta)$ of permanent magnet is shown in Fig. 2: in the figure, $t_0$ represents rotor tooth width, $\theta r$ represents the rotor slot width, and Br is the residual magnetic density of the permanent magnet. This model is the rotor with four rotor slots at each pole, when the rotor at each pole is n, the square wave number between $\pi/2p$ and $-\pi/2p$ in the figure is n.

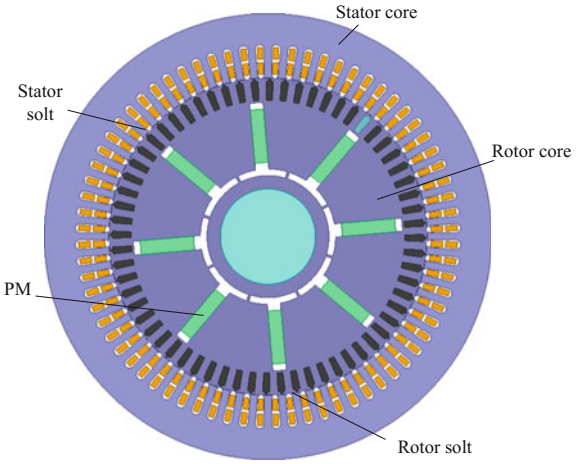



**Fig. 1** Diagram of asynchronous start permanent magnet synchronous motor

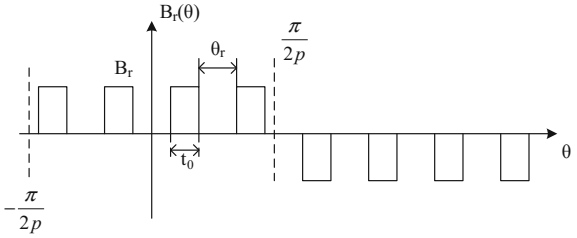



**Fig. 2** Distribution map of $B_r(\theta)$

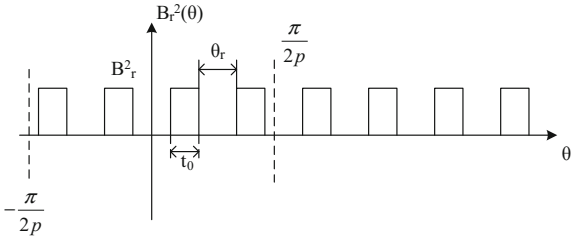



**Fig. 3** Distribution map of $B_r^2(\theta)$

According the distribution map of $B_r(\theta)$ under each pair of pole, the distribution map of $B_r^2(\theta)$ under each pair of pole can be obtained as shown in Fig. 3.

According to the distribution map of $B_r^2(\theta)$ along the circumferential direction of the air gap, its Fourier expansions can be obtained as below:

$$B_r^2(\theta) = B_{r0} + \sum_{n=1}^{\infty} B_{rn} \cos nQ_2\theta. \tag{2}$$

where various Fourier decomposition coefficients are:

$$B_{r0} = \frac{Q_2}{\pi}\int_0^{\frac{\pi}{Q_2}} B_r^2(\varepsilon)d\varepsilon = \frac{Q_2}{\pi}\int_0^{\frac{t_0}{2}} B_r^2 d\varepsilon = \frac{Q_2}{\pi}\frac{t_0}{2}B_r^2. \tag{3}$$

$$\begin{aligned} B_{rn} &= \frac{2Q_2}{\pi}\int_0^{\frac{\pi}{Q_2}} B_r^2(\varepsilon)\cos nQ_2\varepsilon d\varepsilon = \frac{2Q_2}{\pi}\int_0^{\frac{t_0}{2}} B_r^2\cos nQ_2\varepsilon d\varepsilon \\ &= \frac{2B_r^2Q_2}{\pi}\frac{1}{nQ_2}\sin nQ_2\frac{t_0}{2} = \frac{2B_r^2}{n\pi}\sin\frac{t_0}{t_2}n\pi. \end{aligned} \tag{4}$$

Since the air permeability is nearly equal to that of permanent magnets, so the rotor tooth of asynchronous start permanent magnet synchronous motor can be made equivalent to that of a permanent magnet, this move can make the bilateral asynchronous start permanent magnet synchronous motor equivalent to a surface mounted permanent magnet motor. Its expansion can be expressed as below with reference to the Fourier decomposition of equivalent gap:

$$\left[\frac{h_m}{h_m + g(\theta,\alpha)}\right]^2 = G_0 + \sum_{n=1}^{\infty} G_n \cos nQ_1(\theta+\alpha). \tag{5}$$

The cogging torque expression of tangential asynchronous start permanent magnet synchronous motor can be obtained as below:

$$T_{cog} = \frac{\pi Q_1 L_{Fe}}{4\mu_0}(R_2^2 - R_1^2)\sum_{n=1}^{\infty} nG_n B_{r\frac{nQ_1}{Q_2}}\sin nQ_1\alpha. \tag{6}$$

In the above formulas, $Q_1$ represents stator slots, $Q_2$ represents the number of rotor slots, $t_2$ represents rotor tooth pitch, $L_{Fe}$ represents axial length of the motor; $R_1$ and $R_2$ are the inner and outer radius of the stator and rotor, n is the integer that makes $nQ_1/2p$ an integer.

## 3 Influence on Cogging Torque by Rotor Side Parameters

### *3.1 Influence of Number of Rotor Slots*

Adopt the 72-slot 8-pole asynchronous start permanent magnet synchronous motor as an example to study rotor side parameters' influence on cogging torque, as shown in Fig. 1. Its main dimensions of the structure are shown in Table 1.

According Fourier decomposition formula of air gap flux density, rotor slot will affect the harmonic amplitude increase of rotor magnetic motive force, thereby affecting changes in harmonic content of the air gap flux density. When the rotor slot number changes the influence of harmonic wave number that generate cogging

**Table 1** Main dimension parameters of motor 1

| | | | |
|---|---|---|---|
| Number of poles | 8 | Width of permanent magnet/mm | 48 |
| Number of stator slots | 72 | Thickness of permanent magnet/mm | 12 |
| Outer diameter of stator/mm | 400 | Lamination factor | 0.97 |
| Outer diameter of rotor/mm | 283.6 | Outer diameter of rotor/mm | 85 |
| Core lengthen/mm | 225 | Rated voltage/V | 380 |

torque can is negligible, but the change of rotor number will affect the generation of harmonic wave magnetic motive force of rotor teeth. Based on the foregoing, the asynchronous start permanent magnet synchronous motor rotor teeth may be equivalent to a surface mounted permanent magnet. Magnetic motive force refers to the change in energy of permanent magnets, which will inevitably lead to changes in the size of cogging torque.

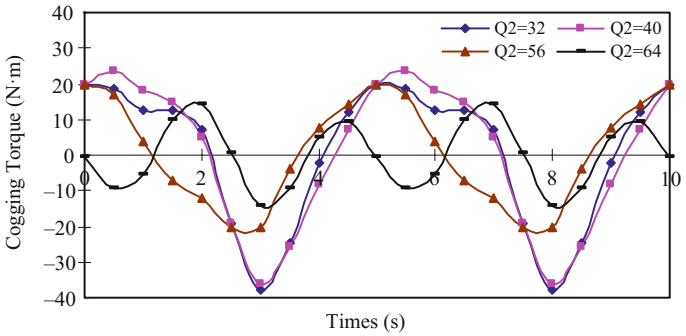
Using finite element analysis, conduct a parametric analysis of the impact on cogging torque by the number of rotor slots, obtaining the waveform of cogging torque with different number of rotor slots as shown in Fig. 2.

Figure 4 shows that when the rotor slot number changes, cogging torque's cycle number and amplitude undergo some changes. When the number of rotor slots is 32 and 40, respectively, the two cogging torques close to each other; when the number of rotor slots is 24 and 48, respectively, the two cogging torques close to each other; and cogging torques with 56 slots and 64 slots nearly equal to each other. And the corresponding relationship between the amplitude value of the cogging torque is $Q_2$ (24, 48) greater than $Q_2$ (32, 40) greater than $Q_2$ (56, 64).

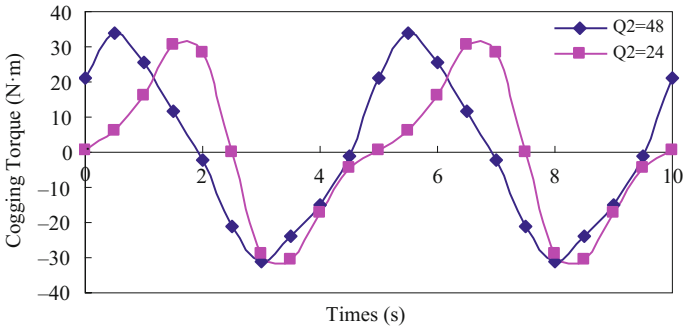
The change cycle number of cogging torque is defined as $N_{cog}$, and $N_q$ is the least common multiple of $N_{cog}$ and $Q_2 2p$. According to the analytical formula of cogging torque, the change cycle of motor 1' cogging torque is 2. Table 2 shows the comparison of $N_q$, $N_q 2p/Q_2$, and cogging torque amplitude value.

In conjunction with Fig. 4 and Table 2, the following conclusions can be drawn: when $N_q 2p/Q_2$ is larger, the generated amplitude value of cogging torque is smaller; when $N_q 2p/Q_2$ is smaller, the generated amplitude value of cogging torque is greater. It can be seen further from Table 2 that, when $Q_2 = (24, 48)$, $N_q 2p/Q_2 = 3$; when $Q_2 = (32, 40)$, $N_q 2p/Q_2 = 9$; when $Q_2 = (56, 64)$, $N_q 2p/Q_2 = 9$. The causes of this phenomenon is that larger $N_q 2p/Q_2$ results in lower harmonic content of cogging torque teeth flux density that generates cogging torque, so the cogging torque is small.

Further explanation of this phenomenon is that rotor tooth are equivalent to the surface mounted permanent magnet, number of rotor slots affect the magnetic motive force at rotor tooth, and according to the analytic formula of cogging torque, n-th harmonic amplitude value of cogging torque is related to $nQ_1/2p$–th harmonic amplitude value of $B_r 2(\theta)$ and number of harmonic wave of equivalent air gap. When $N_q 2p/Q_2$ is larger, corresponding harmonic wave number of $\left[\frac{h_m}{h_m + g(\theta, \alpha)}\right]^2$ is higher, and the cogging torque generated is smaller.




(a) Number of Rotor Slots is 32, 40, 56, 64




(b) Number of Rotor Slots is 24 and 48

**Fig. 4** Cogging torque with different number of rotor slots **a** number of rotor slots is 32, 40, 56, 64 **b** number of rotor slots is 24 and 48

**Table 2** Comparison of cogging torque, $N_q$ $N_qp/Q_2$ with different number of rotor slots

| $Q_2$ | Nq | $Nq2p/Q_2$ | Cogging torque value (N m) |
|---|---|---|---|
| 24 | 9 | 3 | 31 |
| 32 | 36 | 9 | 43 |
| 40 | 45 | 9 | 42 |
| 48 | 18 | 9 | 34 |
| 56 | 63 | 9 | 20 |
| 64 | 72 | 9 | 16 |

## *3.2 Influence of the Rotor Tooth Width*

According to Formulas 2, 4, 6, $nQ_1/Q_2$—th Fourier decomposition coefficients of $B_r2(\theta)$ have a certain effect on the amplitude value of the cogging torque. Make the $nQ_1/Q_2$—th Fourier decomposition coefficients of $B_r2(\theta)$ 0

$$\sin\frac{t_0}{t_2}\frac{nQ_1}{Q_2}\pi = 0. \tag{7}$$

$$\frac{t_0}{t_2}\frac{nQ_1}{Q_2}\pi = k\pi, \quad \text{k is integer.} \tag{8}$$

According to Formula 8, we can know that when the rotor teeth width and the rotor tooth pitch satisfies $t_2/t_0 = nQ_2/kQ_1$, i.e., when $t_2 = nQ_2/kt_0Q_1$, the n-th and kn-th harmonic amplitude value of $B_r2(\theta)$ is 0, the n-th and kn-th harmonic wave of cogging torque is also zero, which can effectively weaken the asynchronous start permanent magnet synchronous motor's cogging torque.

Take the dimension of asynchronous start permanent magnet synchronous motor from Table 1 as an example. As for the 8-pole 72-slot asynchronous start permanent magnet synchronous motor, the select number of rotor slots is 64. According to formula 8, when $t_2/t_0 = 9/8$, namely when the rotor tooth width is 5°, the cogging torque of such motors can be significantly weakened, as shown in Fig. 5.

Figure 5 shows that, with original rotor tooth width, the motor cogging torque is 15 N m, when the rotor tooth width is changed to 5°, cogging torque amplitude value of the motor is reduced to about 9.5 N m, and the cogging torque is significantly weakened.

To further verify the validity of this method, the prototype 2 is selected as example to be analyzed, the main sizes of the motor are shown in Table 3.

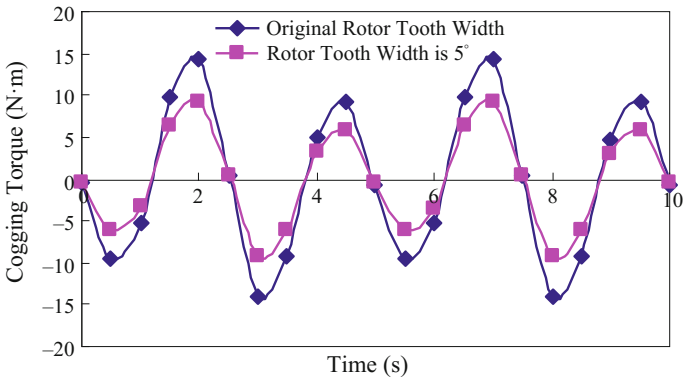



**Fig. 5** Cogging torque changed by stator tooth width change

**Table 3** Main size parameters of motor 2

| | | | |
|---|---|---|---|
| Number of poles | 4 | Width of permanent magnet/mm | 39 |
| Number of stator slot | 24 | Thickness of permanent magnet/mm | 3 |
| Outer diameter of stator/mm | 120 | Laminated coefficient | 0.97 |
| Outer diameter of rotor/mm | 74 | Inner diameter of rotor/mm | 25 |
| Core length/mm | 65 | Rated voltage/V | 220 |

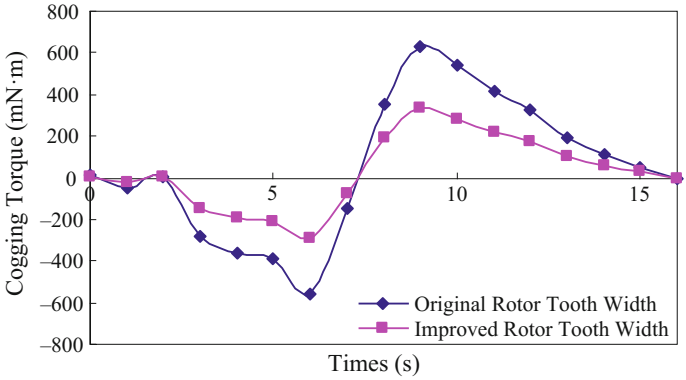



**Fig. 6** Cogging torque changed by stator tooth width change

The shape of permanent magnet used by Motor 2 is V-shaped structure, and number of stator slots is 24. According to Formulas 7 and 8, when the rotor tooth width satisfies $t_2/t_0 = 3/4$, cogging torque of the motor can be significantly weakened. After the stator tooth width is improved, the motor cogging torque waveform diagram is shown as Fig. 6.

Figure 6 shows that, before improvement, the motor cogging torque is approximately 1.5 N m, and after the rotor tooth width is improved, the motor cogging torque is reduced to 0.5 N m or so. Cogging torque has been significantly weakened, thus the motor performance is improved.

## 4 Influence of Rotor Side Parameters on Motor Electromagnetic Properties

It can be seen from the above conclusion that changes in relevant rotor side parameters can effectively weaken the asynchronous start permanent magnet synchronous motor's cogging torque. But changing the rotor side parameters may have certain influence on the motor's electromagnetic properties. Using the time step finite element analysis method, this paper calculates and compares prototype's no-loaded back electromotive force (BEMF) and rated load current before and after optimization, which are shown in Table 4, respectively.

**Table 4** No-loaded back electromotive force before and after optimization

| Main sizes of the motor | | BEMF | Rated current |
|---|---|---|---|
| The motor | Being improved | 213.5 | 43.15 |
| | Changing rotor tooth width | 211.6 | 41.26 |
| | Curves with rotor teeth pairing | 209.3 | 40.85 |

It can be seen from Table 4, after the optimized motor size is adopted, the prototype's unloaded back electromotive force basically remains unchanged, while the changes generated by rated current is substantially negligible. The method proposed in this paper can effectively weaken the asynchronous start permanent magnet synchronous motor's cogging torque under the premise of maintaining motor rated performance.

## 5 Conclusion

On the premise of analyzing the mechanism asynchronous start permanent magnet synchronous motor cogging torque generation, a new method of analyzing asynchronous start permanent magnet synchronous motor cogging torque is proposed. Make the built-in rotor permanent magnet equivalent to magnetic motive force generated by surface mount permanent magnets, and analyze the influence of number of rotor slots, rotor tooth width and rotor shape on motor cogging torque, obtaining the following conclusions:

(1) When larger Nq2p/Q2 is selected, can make asynchronous start permanent magnet synchronous motor's cogging torque smaller;
(2) Select the appropriate rotor tooth width can effectively weaken motor cogging torque;
(3) Eccentric distance of the rotor has a certain impact on the asynchronous start permanent magnet synchronous motor's cogging torque.

## References

1. J.C. Urresty, J.R. Riba, L. Romeral et al., A simple 2-D finite-element geometry for analyzing surface-mounted synchronous machines with skewed rotor magnets. IEEE Trans. Magn. **46** (11), 3948–3954 (2010)
2. W. Chen, C. Guanghua, C. Dong, A new method for weakening built-in permanent magnet synchronous motor cogging torque. Micro Motor **43**(1), 9–12 (2015) (in Chinese)
3. R. Tang, *Modern Permanent Magnet Theory and Design* (China Machine Press, Beijing, 2005) (in Chinese)
4. L. Jinglin, W. Li, Analysis of influence on rare earth permanent magnet chute motor performance by magnetizing type of magnetic steel. Rare Earth **35**(1), 16–22 (2014) (in Chinese)
5. X. Tang, X. Wang, Y. Li et al., Research on permanent magnet demagnetization during the starting process of asynchronous start permanent magnet synchronous motor. Proc. CSEE **35** (4), 962–970 (2015) (in Chinese)
6. P. Guo, C. Shi, L. Wang, Research on asynchronous start permanent magnet synchronous motor cogging torque. Micro Mot. **46**(5), 21–26 (2013) (in Chinese)
7. M.F. Hsieh Y.H. Yeh, Rotor eccentricity effect on cogging torque of pm generators for small wind turbines. IEEE Trans. Magn. **49**(5), 1897–1900 (2013)

8. L. Yansheng, D. Manfeng, F. Xin, Analytical calculation of air-gap magnetic field and cogging torque of surface mounted permanent magnet motor. Micro Mot. **40**(12), 9–15 (2012) (in Chinese)
9. S. Huang, T. Liu, H. Ouyang, Method of permanent magnet motor cogging torque weakening based on notch deviation. Trans China Electrotechnical Soc. **28**(3), 100–105 (2013) (in Chinese)

# Roll Attitude Solving Algorithm of Projectile with Geomagnetic Field Sensors

**Qingwei Guo, Yongchao Chen, Xieen Song and Lei Zhang**

**Abstract** The traditional attitude measurement methods have many constraints when they are applied on the traditional ammunition. The advantage of geomagnetic technology has yielded that magnetic sensors are smaller, steadier, and more sensitive when they are contributing in the attitude measurement system. The paper gives solving algorithm with the information not only about building of the solving model but the actual processing. First, the basic assumptions and the reference frame are given, so as the local geomagnetism analysis. Then transformation of coordinate system is carried out and the roll angle is worked out by the range analysis. From the experiment, the results demonstrate that the algorithm is a sample method and it can reach a high level that error angle is less than 5° (3σ). The algorithm has the well foreground and potential application in the future traditional ammunition improvement.

**Keywords** Geomagnetic · Roll attitude solving algorithm · Traditional ammunition

## 1 Introduction

Trajectory correction projectiles are simplified guidance weapons which cost less but have high precision. It is the development trend of traditional ammunition. Timely measurements of projectile roll attitude is a pivotal technology which is directly associated with the realization of trajectory correction. Now there are plenty of methods to measure projectile roll attitude [1]. For example, gyroscope can conduct the mission but with the disadvantage that the error cumulates along the time and the zero point moves. Comparing with missile and air vehicle, traditional ammunition has the special characters that they have smaller volume and they are more easily affected and less expensive [2]. Also the bad environment, which goes

Q. Guo (✉) · Y. Chen · X. Song · L. Zhang
Shijiazhuang Mechanical Engineering College, Shijiazhuang, China
e-mail: gqingwei@sina.cn



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_107

with high temperature, heavy strike and much vibration, limits the choices. The precise measurement system is desiderated urgently. Recent advance in magnetic technologies have yielded that magnetic sensors are smaller, steadier, and more sensitive when they are contributing in the roll attitude measurement system [3, 4].

While the advantages and significant features of the magnetic field, the day/light and all-weather capability, non-emissive, passive, and so on, have been conscious and the technologies provide the practice, the applications become popular. The magnetic sensors have widely applied in the aviation and aerospace engineering, not only the satellites, but the space-probe and the airplanes [4, 5]. The development is the widely usage in the navigation which includes the attitude measurement, the guidance, and the control [6–8]. The sensors assist with the raw data for the roll attitude measurement with respect to geomagnetic field.

This dissertation provides a new way to solve the roll attitude of projectile with the geomagnetic field. The model was given and the algorithm was conducted. From the experiment data, the roll attitude solving algorithm was proved working well and having a good precision.

## 2 Modeling of Roll Attitude Measurement System

Target movement is relative, which is said actually to perform in a reference coordinate system. The necessary coordinate system should be established to analyze the projectile motion and to describe the flight attitude of projectile integrally. The common coordinate systems used to research the flight attitude of projectile are ground coordinate system, launching coordinate system $Ox_ty_tz_t$, projectile coordinate system $Ox_py_pz_p$, and quasi-projectile coordinate system $Ox_p'y_p'z_p'$, as shown in Fig. 1.

Geomagnetic field varies with different positions and different time. But these changes are regular and can be measured. The strength of local geomagnetic field can be obtained by local latitude and longitude and other related geomagnetic elements, except few abnormal places.

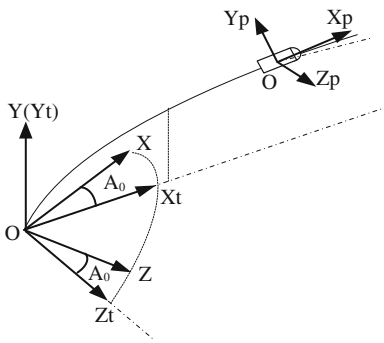



**Fig. 1** Sketch map of coordinate system

When the launching site is confirmed, the latitude and longitude are acquired, combing with other known parameters (altitude, time and so on), and the geomagnetic field strength $B_0 = [\,B_{x0} \quad B_{y0} \quad B_{z0}\,]$ of launching site can be given by the IGRF model or the software Geomag70.

## 3 Geometric Algorithm of Roll Attitude

The geomagnetic sensors are installed along the projectile coordinate axes. The strength of magnetic field $B_n$ along three axes can be measured during the flight. The strength of ground geomagnetic field $B_0$ is acquired by latitude and longitude. And the relationship equations of $B_n$ and $B_0$ are established by the coordinate transformation, namely the attitude matrix. The roll angle is solved with the known conditions of $B_0$, $\varphi_0$ and $B_n$.

The angle between ground coordinate system and launching coordinate system is $A_0$ based on the known conditions, and the transform matrix is

$$C_0 = \begin{bmatrix} \cos \mathrm{A}_0 & 0 & -\sin \mathrm{A}_0 \\ 0 & 1 & 0 \\ \sin \mathrm{A}_0 & 0 & \cos \mathrm{A}_0 \end{bmatrix} \tag{1}$$

Transformation relationship equation of geomagnetic field strength is

$$B_t = C_0 * B_0 \tag{2}$$

Transformation between Ground coordinate system and launching coordinate system can be denoted by attitude angle $\gamma$, $\varphi$, $\psi$ according to Euler theorem, namely the projectile coordinate system can be gained by launching coordinate system from the three coordinate system, where the Euler angle is $\gamma$, $\varphi$, then the attitude matrix are gained as follow,

$$C_{\mathrm{n}} = \begin{bmatrix} \cos\varphi \cos\psi & \sin\varphi & -\cos\varphi \sin\psi \\ -\sin\varphi \cos\psi \cos\gamma + \sin\psi \sin\gamma & \cos\varphi \cos\gamma & \sin\varphi \sin\psi \cos\gamma + \cos\psi \sin\gamma \\ \sin\varphi \cos\psi \sin\gamma + \sin\psi \cos\gamma & -\cos\varphi \sin\gamma & -\sin\varphi \sin\psi \sin\gamma + \cos\psi \cos\gamma \end{bmatrix} \tag{3}$$

The transformation of geomagnetic field strength between the launching coordinate system and projectile body coordinate system is,

$$\begin{bmatrix} B_{xt} \\ B_{yt} \\ B_{zt} \end{bmatrix} = C_n^{-1} * \begin{bmatrix} B_{xn} \\ B_{yn} \\ B_{zn} \end{bmatrix} \tag{4}$$

Using the attitude transform matrix, $\psi \approx 0$ from the basic suppose, then $C_n$ can be simplified and the Eq. (4) is unfolded,

$$\begin{cases} B_{xn}* \cos\varphi - B_{yn}* \sin\varphi \cos\gamma + B_{zn}* \sin\varphi \sin\gamma = B_{xt} \\ B_{xn}* \sin\varphi + B_{yn}* \cos\varphi \cos\gamma - B_{zn}* \cos\varphi \text{in}\gamma = B_{yt} \\ B_{yn}* \sin\gamma - B_{zn}* \cos\gamma = B_{zt} \end{cases} \tag{5}$$

By solving the Eq. (5), $\gamma_0$ get

$$\gamma_0 = 2* \arctan\left(\frac{B_{yn} \pm \sqrt{B_{yn}^2 + B_{zn}^2 - B_{zt}^2}}{B_{zn} + B_{zt}}\right) \tag{6}$$

The roll angle $\gamma \in [0, 2\pi]$, if $\gamma_0 > 0$, $\gamma \in [0, \pi]$; if $\gamma_0 < 0$, $\gamma = 2\pi + \gamma_{1,2}$, $\gamma \in [\pi, 2\pi]$.

In Fig. 2, $B_p^{'}$ is the projection of geomagnetic field on the plane $Oy_p^{'}z_p^{'}$. Supposing the angle between angle of geomagnetic field in the projectile coordinate system and the axis $Oy_p^{'}$ is $\gamma_0$, then $|\tan\gamma_0| = \left|B_{zt}/\sqrt{B_{yn}^2 + B_{zn}^2 - B_{zt}^2}\right|$, the value of $\gamma_0$ is associated with three components of $B_t$. If $z_t > 0$, $\gamma_0 \in [0, \pi]$, if $z_t < 0$, $\gamma_0 \in [\pi, 2\pi]$. When $y_t > 0$, $B_p^{'}$ is above the axis, $\gamma_0 = 2\pi + \arctan(B_{zt}/\sqrt{B_{yn}^2 + B_{zn}^2 - B_{zt}^2})$ $(B_{zt} > 0)$ or $\arctan(B_{zt}/\sqrt{B_{yn}^2 + B_{zn}^2 - B_{zt}^2})$ $(B_{zt} < 0)$. When $y_t < 0$, $B_p^{'}$ is below the axis, namely $\pi - \arctan(B_{zt}/\sqrt{B_{yn}^2 + B_{zn}^2 - B_{zt}^2})$, as shown in Table 1.

On the basis of measured values by the geomagnetic sensors, where $\gamma_0 \in [0, \pi/2]$, if $B_{yn} > 0, B_{zn} > 0$, then $\gamma \in [0, \gamma_0] \cup [3\pi/2 + \gamma_0, 2\pi]$, if $B_{yn} > 0$, $B_{zn} < 0$, $\gamma \in [\gamma_0, \pi/2 + \gamma_0]$, if $B_{yn} < 0$, $B_{zn} < 0$, $\gamma \in [\pi/2 + \varphi_0, \pi + \gamma_0]$, if $B_{yn} < 0$, $B_{zn} > 0$, $\gamma \in [\pi + \gamma_0, 3\pi/2 + \gamma_0]$. Similarly, when $\gamma_0 \in [\pi/2, \pi]$, $[\pi, 3\pi/2]$, and $[3\pi/2, 2\pi]$, the range of roll angle $\gamma$ can be obtained, as shown in Table 2. The roll angle is resolved by Eq. (6) combined with the angle range.

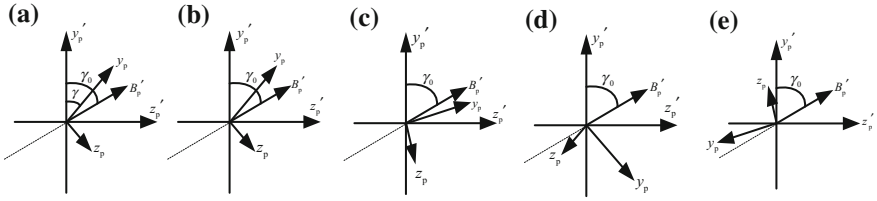



**Fig. 2** Sketch map of roll angle

**Table 1** Range of $\gamma_0$ in different condition

| $B_t$ | $B_{yt} > 0$, $B_{zt} > 0$ | $B_{yt} > 0$, $B_{zt} < 0$ | $B_{yt} < 0$, $B_{zt} > 0$ | $B_{yt} < 0$, $B_{zt} < 0$ |
|---|---|---|---|---|
| $\gamma_0$ | $[0, \pi/2]$ | $[3\pi/2, 2\pi]$ | $[\pi, 3\pi/2]$ | $[\pi/2, \pi]$ |

**Table 2** Value range of $\gamma_0$ in different condition

| $B_n$<br>$\gamma$<br>$\gamma_0$ | $B_{yn}>0$, $B_{zn}>0$ | $B_{yn}>0$, $B_{zn}<0$ | $B_{yn}<0$, $B_{zn}<0$ | $B_{yn}<0$, $B_{zn}>0$ |
|---|---|---|---|---|
| $[\pi/2, \pi]$ | $[\gamma_0-\pi/2, \gamma_0]$ | $[\gamma_0, \gamma_0+\pi/2]$ | $[\gamma_0+\pi/2, \gamma_0+\pi]$ | $[\gamma_0+\pi, 2\pi]\cup$<br>$[0, \gamma_0-\pi/2]$ |
| $[\pi, 3\pi/2]$ | $[\gamma_0-\pi/2, \gamma_0]$ | $[\gamma_0, \gamma_0+\pi/2]$ | $[\gamma_0+\pi/2, 2\pi]\cup$<br>$[0, \gamma_0-\pi]$ | $[\gamma_0-\pi, \gamma_0-\pi/2]$ |
| $[3\pi/2, 2\pi]$ | $[\gamma_0-\pi/2,\ \gamma_0]$ | $[\gamma_0, 2\pi]\cup$<br>$[0, \gamma_0-3\pi/2]$ | $[\gamma_0-3\pi/2, \gamma_0-\pi]$ | $[\gamma_0-\pi, \gamma_0-\pi/2]$ |

From the processing of roll angle solving, the method has certain precondition and constraint condition. The necessary and sufficient condition, in which the Eq. (6) has solution, is $\Delta = B_{\mathrm{yn}}^2 + B_{zn}^2 - B_{zt}^2 \geq 0$. The projection of geomagnetic field on the direction of axis $z_t$ is constant and in the same plane with the geomagnetic sensors on the projectile cross-section. For $\sqrt{B_{\mathrm{yn}}^2 + B_{zn}^2 - B_{zt}^2}$ is the denominator in the processing of solution, the equation has no solution when $B_{\mathrm{yn}}^2 + B_{zn}^2 - B_{zt}^2 = 0$. But the probability that these points appears is very little on every cycle during the processing of measuring, thus the solving can be done very well.

## 4 Test and Analysis of Results

The trial site is in Northeast China. The ground geomagnetic parameters which are solved by the model of IGRF11 are: $B_0 = [26424.5, -47410.6, 4259.2]$ nT. Then the strength of geomagnetic field in the launching coordinate system is as follows:

$$B_t = C_0 * B_0 = \begin{pmatrix} \cos 35 & 0 & -\sin 35 \\ 0 & 1 & 0 \\ \sin 35 & 0 & \cos 35 \end{pmatrix} * \begin{pmatrix} 26424.5 \\ -47410.6 \\ -4259.2 \end{pmatrix} = \begin{pmatrix} 24088.6 \\ -47410.6 \\ 11667.5 \end{pmatrix} \quad (7)$$

According the analysis $B_t$ is approximately a constant, then $B_{zt} = 11667.5$ nT. The real-time strength $B_n$ can be gained from projectile-borne geomagnetic sensors measurement at any time interval throughout the flight. And azimuth angle is $\varphi_0 = 35°$, height above sea level is 1426 m. Other important parameter is the initial yaw angle $\psi_0 = 314.3°$(from the due north direction in the clockwise direction).

As shown in Fig. 3, the curves are about the components of geomagnetic field strength measured by the strap-on geomagnetic sensors. As known, the projectiles rotate with a slow variation frequency throughout the flight. The data result in a sinusoidal curve of nearly constant frequency, as shown in Fig. 3b, d. The variation

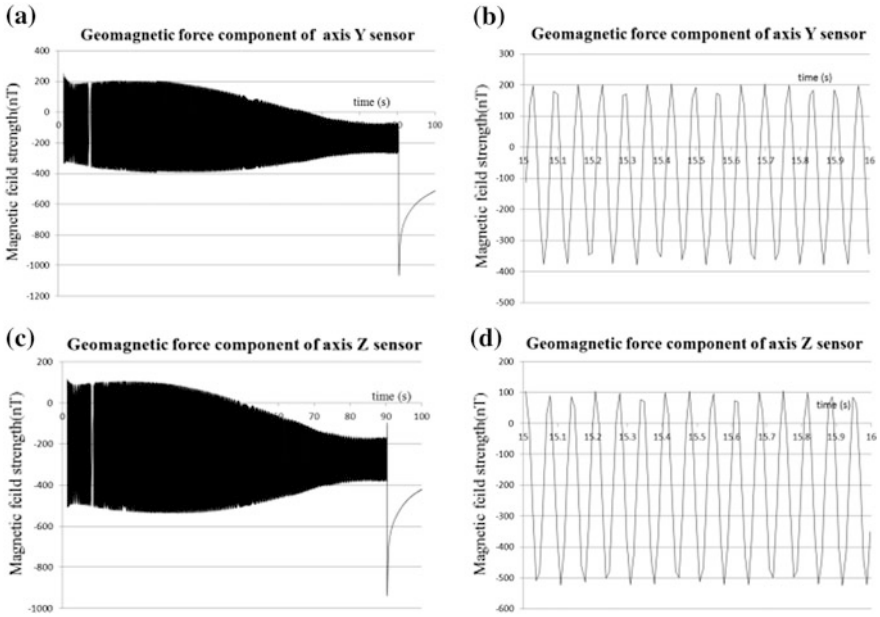



**Fig. 3** Geomagnetic force component of magnetic resistance sensor

changes periodically and the changes are very small in different cycles. Therefore, the experiment results are in accord with the theory analysis.

The initial flight data includes the extraneous influences which are commonly composed of magnetic resistance sensor errors, geomagnetic environment interference errors, and other application errors. There is no doubt that the initial data should be well compensation. In accordance with the empirical statistics data, the online automatic calibration method is designed to improve the precision and the real-time capability. The actual processing data is shown in the Fig. 4.

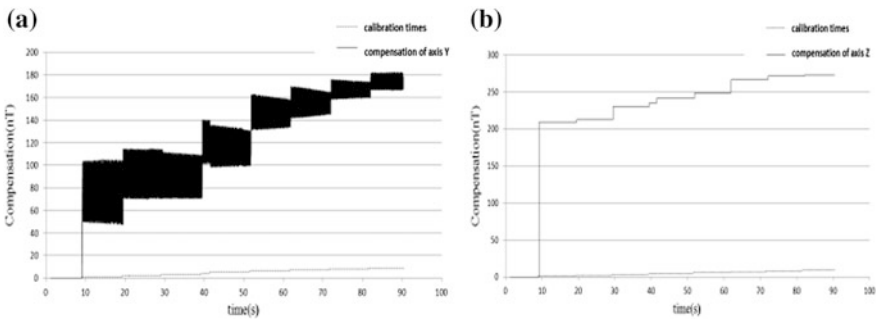



**Fig. 4** Compensation of axis Y and axis Z magnetic field strength

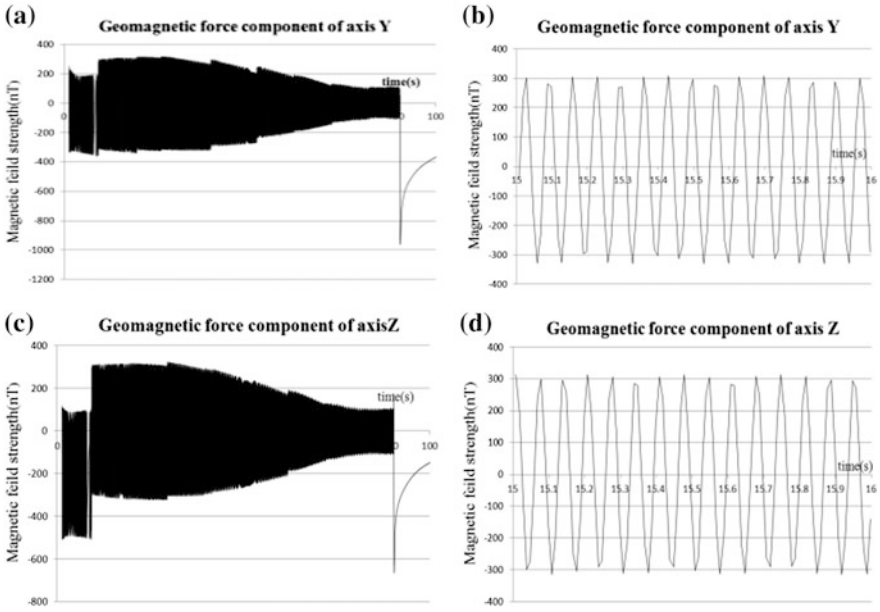



**Fig. 5** Geomagnetic force components of axis Y and axis Z

Integrated the compensation to the raw sensors data, the precise magnetic field data is gained, shown by Fig. 5.

On the basis of the precise earth magnetic field data of actual flight, it can be concluded that the roll angle is measured based on the above geometric algorithm. Substituting the data of the earth magnetic field strength components (axis Y, axis Z) into Eqs. (5) and (6) and comparing the actual condition with Tables 1 and 2, the results lead to the following graphs.

Figure 6 covers the variation of roll angle with time on the full flight. In the full flight time, the roll angle solving with geometric algorithm performs well and proves a good change rule. The resolving result remains better consistency, no high wave in the full process. From the graph of Fig. 6, there is a diverse curve in the 1950 s, where is the moment that the projectile is flying over vertex of the trajectory and the slope angle of trajectory changes too acutely to precise online automatic calibration of the strap-on magnetic sensors. Then the detail views are provided in the Figs. 7 and 8.

In Fig. 7, there are two different curves of different phases, which are gained from up trajectory (a, 15.36–15.56 s) and descent trajectory (b, 46.70–46.92 s), respectively. Those two curves suggest that solving roll angle increases from 0° up to 360° smoothly and periodically and remains good consistency.

Figure 8 shows a single cycle of solving roll angle (near 50.10 s). The period is about 0.067 s and the angle changes from 2.2° to 352.6°. Supposing that the red beeline is the theoretical curve, which joins the first point and the end point in a

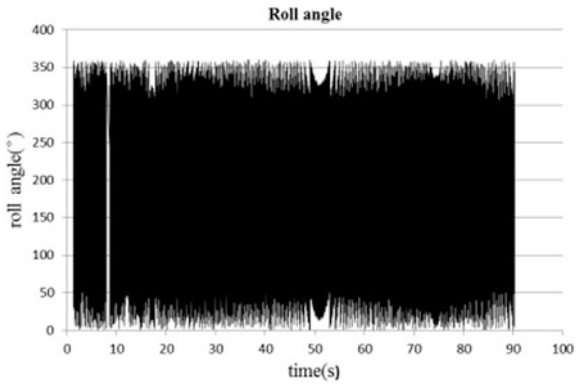



**Fig. 6** Result of roll angle solving (throughout the flight)

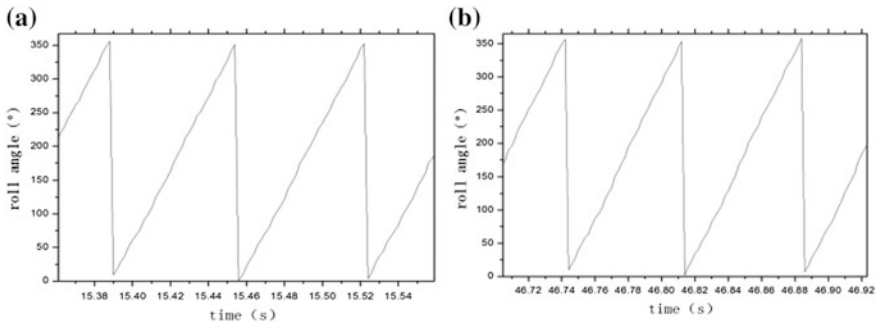
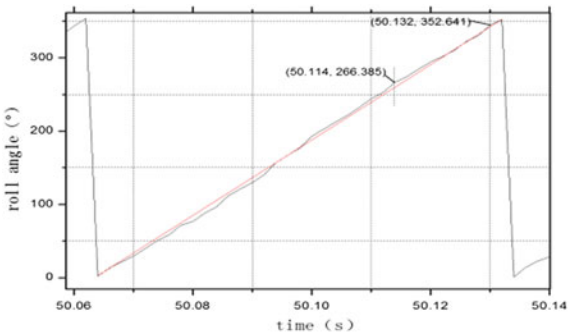



**Fig. 7** Results of roll angle solving

**Fig. 8** Detail views of roll angle solving




period, the difference between the actual curve and the theoretical curve is the resolving error. Consequently, from the comparison in Fig. 8, the largest error (at 50.114 s) is 4.7° and the integrated average error is less than 5°. The solving precision can reach a relatively high level.

The roll angle changes periodically and the period are similar to that of geomagnetic field strength, which is in accord with the roll attitude of the projectile. With the precise geomagnetic field data, the roll angle solving can reach a high level that error angle is less than 5° (3σ). The algorithm is proved to be a useful method to provide the roll attitude.

## 5 Conclusion

A methodology for the roll angle of the spinning projectile relative to geomagnetic field has been formulated. The mathematical model is established by solving the attitude matrix and the real-time parameters are calculated by measuring geomagnetic field on the foundation of known factors. The real-time roll angle is acquired from the analysis of experiment data, which has great practical value. The geomagnetic field has itself limitation, it is necessary to consider the magnetic anomaly interference and the influence of no solution which result from the small angle in the direction of geomagnetic field. It is very essential to research on the data filter processing and error analysis and compensation, which will prompt the attitude measurement with geomagnetic field to exert the greater military application value. The methodology will provide an all-weather, day/night angular measurement capability for spinning projectiles, and will be applied to the new programs in the near future.

## References

1. C. Hong-song, C. Guo-guang, Investigating roll attitude detection method of the flying ammunition. J. Projectiles Rocket. Missiles Guidance **24**(1), 141–143 (2004)
2. W. Guang-long, Z. Jing, Z. Wen-dong, Earth magnetic field sensor and its application in flight body attitude measurement. J. Beijing Inst. Technol. **3**(1), 361–364 (1999)
3. LV. Qing-li, X.-D. Gao, X.-M. Wang, Q. Qu, The research on real-time calculation of projectile's roll angle based on geomagnetic detection. J. Projectiles Rocket. Missiles Guidance **31**(3), 30–31 (2011)
4. T. Harkins, D. Hepner, Magsonde (patent pending): a device for making angular measurements on spinning projectiles using magnetic sensors, in *Proceedings of SPIE. US: SPIE* (2000), pp. 60–67
5. S. Lian-yan, Ya. Shu-xing, Z. Xia-qing, Analyzing of MR/GPS guiding applied in rotating rocket powered missile. J. Projectiles Rocket. Missiles Guidance **26**(2), 1145–1147 (2006)
6. G. Creamer, The Hessi magnetic attitude control system, in *AIAA Guidance, Navigation, and Control Conference*, Portland, Oregon, Aug 1999
7. F. Santoni, P. Tortora, Magnetic attitude determination and control of inertial pointing small momentum bias spacecraft, in *Proceedings of the AAS/AIAA Space Flight Mechanics Meeting*, Santa Barbara (CA, 2001)
8. F. Goldenberg, Geomagnetic navigation beyond magnetic compass. Pos. Locat. Navig. Symp. **4**(25), 684–694 (2006)

9. Z. Li, K. Liu, Y. Zhao, et al., MaPIT: an enhanced pending interest table for NDN with mapping bloom filter. IEEE Commun. Lett. **18**(11), 1915–1918 (2011)
10. Z. Li, Y. Chen, H. Shi, et al., NDN-GSM-R: a novel high-speed railway communication system via named data networking. EURASIP J. Wirel. Commun. Netw. **48**, 1–5 (2016)
11. X. Liu, Z. Li, P. Yang, et al., Information-centric mobile ad hoc networks and content routing: a survey. Ad Hoc Netw. (2016), http://dx.doi.org/10.1016/j.adhoc.2016.04.005
12. Z. Li, L. Song, H. Shi, Approaching the capacity of K-user MIMO interference channel with interference counteraction scheme. Ad Hoc Netw. (2016). 10.1016/j.adhoc.2016.02.009

# Design and Analysis of a Novel Structure of Electromagnetic Metamaterial with Negative Permeability and Permittivity

**Cheng Gu and Xiu Zhang**

**Abstract** Different from the traditional electromagnetic material, the electromagnetic metamaterial has tremendous application potential in antenna designing, superlens and stealth fields because of its negative properties in permeability and/or permittivity. In this paper, a novel structure of electromagnetic metamaterial with middle-frequency band is designed and analyzed. The unit of this novel structure is composed of a copper wire and a resonator which are distributed in both sides of a substrate respectively. In this case, the resonators realize the negative permeability, and the copper wire realizes the negative permittivity. The simulation results indicate that the effective refraction index is negative from 360 to 450 MHz. The advantage of this structure is that its unit is compact and easy to form artificially periodic array.

## 1 Introduction

In 1968, Veselago first proposed the concept of the material with negative permittivity and permeability which different from the common electromagnetic material, unfortunately, this theoretical concept did not receive the recognition from the scientific community because this material can't be found in nature [1]. Until 1996, British scientist Pendry [2] creatively proposed periodic array wire can realize the negative value of permittivity. In 1999, he proposed periodic arrangement of opening resonant ring can achieve negative magnetic permeability [3]. Since then, this electromagnetic metamaterial received wide attention from all over the world because of its special properties such as reversed Doppler Effect, reversed Snell Refraction, reversed Cerenkov Radiation and so on [4–8].

C. Gu · X. Zhang (✉)
College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin, China
e-mail: zhang210@126.com

C. Gu · X. Zhang
Tianjin Key Laboratory of Wireless Mobile Communication and Wireless Power Transmission, Tianjin Normal University, Tianjin, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_108

In 2006, the research team in Purdue University designed the electromagnetic metamaterial cloak at microwave frequencies to achieve stealth for the metal cylinder [5]. In 2012, Cao et al. designed the broadband periodic antenna using split-ring resonator structure to enhance the gain of the antenna [6]. In addition, the researchers [7–10] implement the superlens using electromagnetic metamaterial to improve the imaging quality. Recently, the researchers [11, 12] applied the electromagnetic metamaterial in the wireless power transfer system to improve its performance, because this special material can focus the electromagnetic field.

In this paper, a novel structure of electromagnetic metamaterial is designed and analyzed. In order to apply it in a wireless power transfer system, the operation frequency of the metamaterial should be megaHz.

## 2 Theoretical Foundation of Electromagnetic Material

According to the polarity of the permittivity $\varepsilon$ and permeability $\mu$, the material can be divided into four categories as shown in Fig. 1. If it have two negative parameters, the electromagnetic metamaterial is also called Left-handed material (LHM), while the traditional electromagnetic material named Right-handed material (RHM).

To analysis the performance of the left-handed material, the theoretic foundation is still based on Maxwell equations:

$$\begin{cases} \nabla \times \vec{H} = \vec{J} + \mathrm{j}\omega\vec{D} \\ \nabla \times \vec{E} = -\mathrm{j}\omega\vec{B} \\ \nabla \cdot \vec{B} = 0 \\ \nabla \cdot \vec{D} = \rho \end{cases} \tag{1}$$

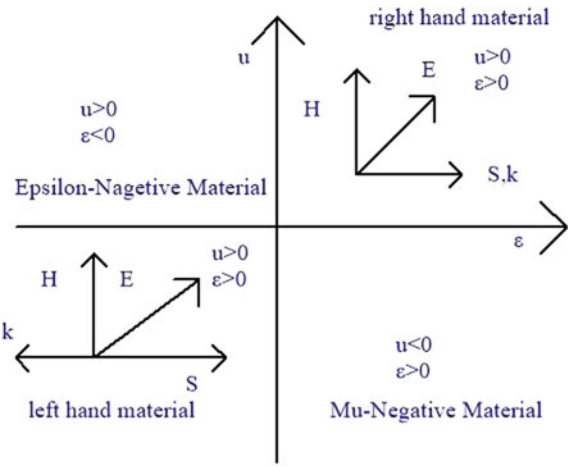



**Fig. 1** Classification of the materials according the polarity of the permittivity $\varepsilon$ and permeability $\mu$

The constitutive relation of medium is

$$\begin{cases} \vec{D} = \varepsilon \vec{E} \\ \vec{B} = \mu \vec{H} \end{cases} \tag{2}$$

For a homogeneous one-dimensional flat plate material with a thickness of $d$ ($kd$<<1), the transmission matrix can be expressed as

$$T = \begin{pmatrix} \cos(nkd) & -\frac{z}{k}\sin(nkd) \\ \frac{k}{z}\sin(nkd) & \cos(nkd) \end{pmatrix} \tag{3}$$

Then the $S$ parameters can be deduced as

$$S_{21} = S_{12} = \frac{1}{\cos(nkd) - \frac{i}{2}\left(Z + \frac{1}{Z}\right)\sin(nkd)} \tag{4}$$

$$S_{11} = S_{22} = \frac{i}{2}\left(\frac{1}{Z} - Z\right)\sin(nkd) \tag{5}$$

Thus, the refractive index $n$ and impedance $Z$ can be expressed as:

$$n = \frac{1}{kd}\cos^{-1}\left[\frac{1 - S_{11}^2 + S_{21}^2}{2S_{21}}\right] \tag{6}$$

$$Z = \sqrt{\frac{(1 + S_{11})^2 - S_{21}^2}{(1 - S_{11})^2 - S_{21}^2}} \tag{7}$$

The permittivity and permeability of the material can be obtained using the following relationships:

$$\varepsilon = \frac{n}{Z}, \quad \mu = nZ \tag{8}$$

## 3 Design of a Novel Structure of Left-Handed Material

With further study and continuous improvement of left-handed materials, the design of left-handed materials requires a higher degree of integration, wider frequency band and lower loss. Furthermore, the unit structure also needs to be easy to manufacture. While the existing left-handed materials are almost difficult to meet these requirements at the same time, in this paper, a novel structure of the left-handed material is designed and analyzed.

**Fig. 2** Layout of the novel resonator

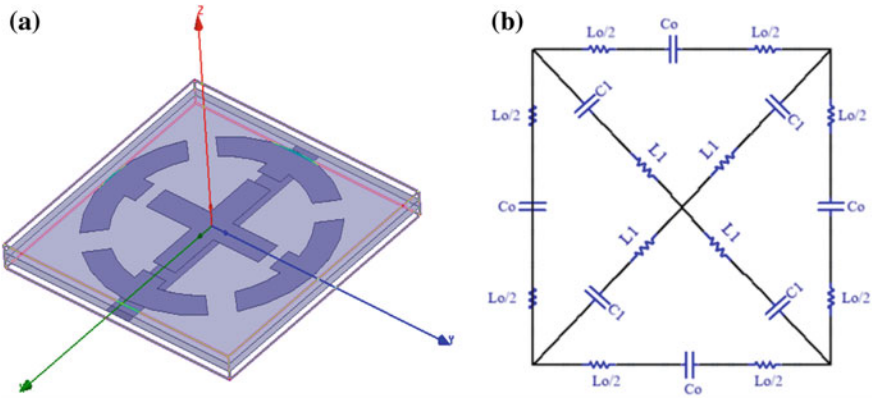



**Fig. 3** **a** 3D model of the novel resonator; **b** The equivalent circuit of the novel resonator

In this paper, the copper wires are used to realize the capacitance and inductance in parallel branch circuit. But the capacitance and inductance in series branch are realized by a novel structure of resonant unit. The ichnography of novel resonator is shown in Fig. 2, which consists of multiple split-ring and a cross-wire inside.

The 3D model of the novel resonator is shown in Fig. 3a, and its equivalent circuit is shown in Fig. 3b. The model is symmetrical in which the copper wires form the inductance and the opening size determine the value of the capacitance. In the equivalent circuit, L1 and L0 represent the inductance of the circle arc and the cross structure respectively; C1 and C0 are the capacitance of the opening on the circle and the cross structure.

In this model, the FR4 material is used as substrate material which its relative dielectric constant is 4.4, and loss angle tangent is 0.02, and the thickness is 0.25 mm. The thickness of the copper wire is 0.017 mm. The length of metal rod is 10 cm, and the width is 1 cm. The width of each opening is 1 cm, and the outer radius and inner radius of split metal ring are 9 cm and 8 cm, respectively. The line width of cross type is 1 cm, and the length of each opening between them is 1 cm too.

## 4 Simulation Results

The novel unit structure is simulated using software HFSS. Its S parameters are obtained as shown in Fig. 4. From the results, it can be seen that there are two subsidence bands in vicinity of 350 and 450 MHz. the maximum value of S11 in this band is –18 dB, which indicates the novel structural unit may demonstrate negative refraction in the band.

According to Eqs. (6, 7), the effective refraction index $n$, and the impedance Z of the proposed model are calculated as shown in Figs. 5 and 6. In Fig. 5, the real part of effective refractive index Re($n$) is negative in the range of 350–450 MHz, which means the negative refraction of the proposed material in this band. As shown in Fig. 6, it indicates that the real part of impedance Re(Z) is greater than zero and the imaginary part Im(Z) is close to zero in range of 350–450 MHz, which is the band that real part of effective refractive index Re($n$) less than zero.

According to Eq. (8), the effective permittivity $\varepsilon$ and permeability $\mu$ of the novel structural are calculated as shown in Figs. 7 and 8. It can be found that the permeability in the band range of 310–320 MHz and 440–450 MHz is negative. Meanwhile, the permittivity is negative in the band range of 280–320 MHz and 360–450 MHz.

From the above results, it indicates the proposed model has double negative properties in a wide band range which it is flexible in certain applications such as in wireless power transfer system.

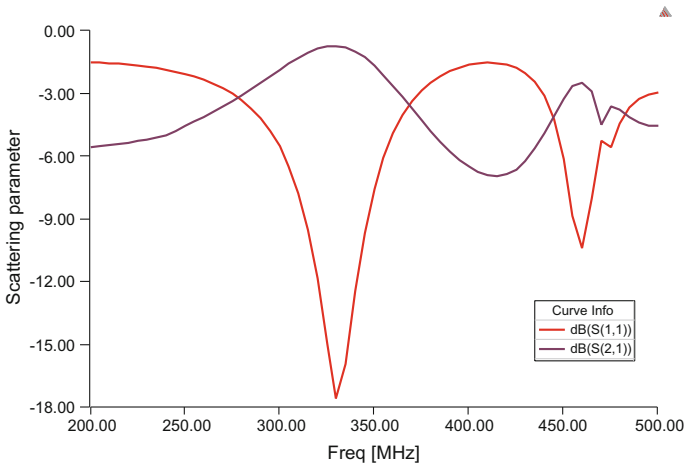



**Fig. 4** Scattering parameter of the proposed model

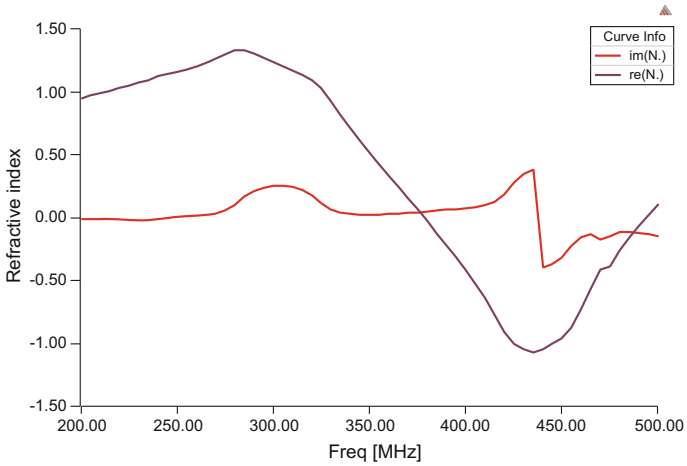



**Fig. 5** Refractive index of proposed model

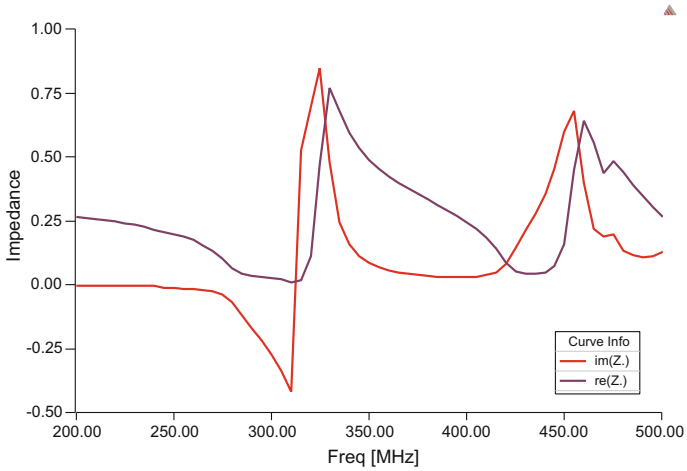



**Fig. 6** Impedance of proposed model

## 5 Conclusion

In this paper, after the introduction of foreign researching situation, we proposed a novel structure of left-handed material based on the theory of negative refraction material, and took relevant in-depth researches. In order to obtain the effective parameters of the structure, the classical S-parameter retrieval method is used to design the material. According to the simulation results using commercial electromagnetic simulation software HFSS, the effective permittivity and permeability

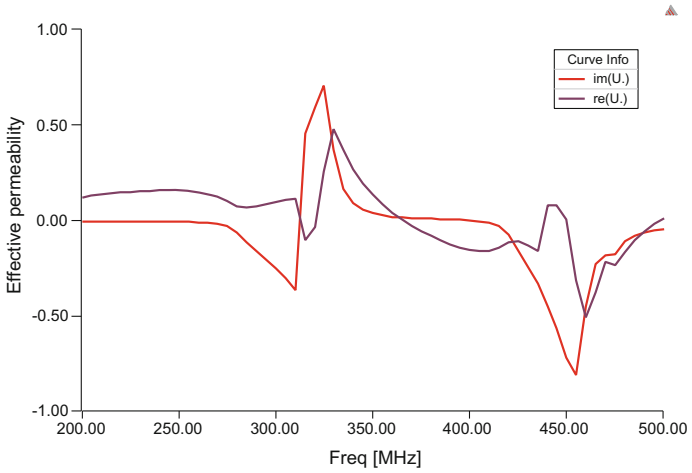



**Fig. 7** Effective permeability of proposed model

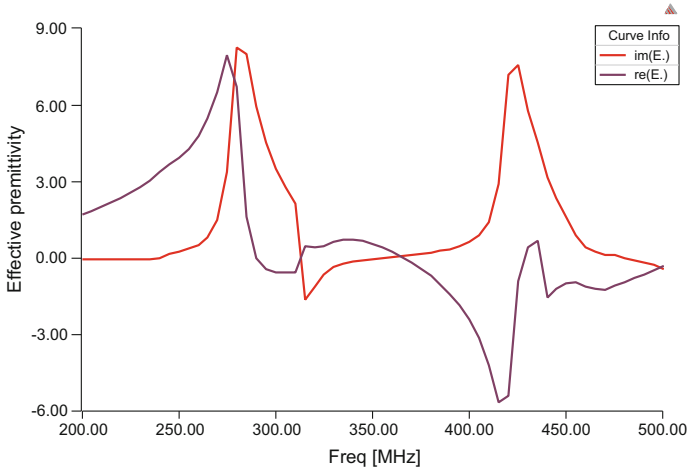



**Fig. 8** Effective permittivity of proposed model

of the proposed model is extracted and the proposed model shows double negative properties in band of 360–450 MHz.

The novel left-handed material unit has the advantages of simple structure, easy fabrication, and it has a relative wide band range with negative permittivity and permeability. Considering these advantages, the proposed model will have greater flexibility applied in the wireless power transfer system in future works.

**Acknowledgements** This research was supported by the Natural Science Foundation of China (61601329, 61603275) and Applied Basic Research Program of Tianjin (15JCYBJC52300, 15JCYBJC51500) and the Tianjin Thousand Youth Talents Plan Project of Tianjin Normal University (ZX110023, ZX0471401505).

## References

1. V.G. Veselago, The electrodynamics of substances with simultaneously negative values of permittivity and permeability. Soviet Physics USPEKI **10**(4), 509–514 (1968)
2. J.B. Pendry, A.J. Holden, W.J. Stewart et al., Extremely low frequency plasmons in metallic mesostructures. Phys. Rev. Lett. **76**(25), 4773–4776 (1996)
3. J.B. Pendry, A.J. Holden, D.J. Robbins et al., Magnetism from conductors and enhanced nonlinear phenomena. IEEE Trans. on Microwave Theor. Tech. **47**(11), 2075–2084 (1999)
4. U.L. Rohde, A.K. Poddar, Metamaterial resonators: theory and applications. Microwave J. **57** (12), 74–76 (2014)
5. D. Schurig, J. Mock, B. Justice, S.A. Cummer, J.B. Pendry, A. Starr et al., Metamaterial electromagnetic cloak at microwave frequencies. Science **2006**(314), 977–980 (2006)
6. W. Cao, B. Zhang, A. Liu, T. Yu, D. Guo, Y. Wei, Gain enhancement for broadband periodic endfire antenna by using split-ring resonator structures. IEEE Trans. Antennas Propag. **60**, 3513–3516 (2012)
7. T. Taubner, D. Korobkin, Y. Urzhumov, G. Shvets, R. Hillenbrand, Near-field microscopy through a SiC superlens. Science **313**, 1595–1595 (2006)
8. X. Zhang, Z. Liu, Superlenses to overcome the diffraction limit. Nat. Mater. **7**, 435–441 (2008)
9. J.A. Schuller, E.S. Barnard, W. Cai, Y.C. Jun, J.S. White, M.L. Brongersma, Plasmonics for extreme light concentration and manipulation. Nat. Mater. **9**, 193–204 (2010)
10. N. Fang, H. Lee, C. Sun, X. Zhang, Sub–diffraction-limited optical imaging with a silver superlens. Science **308**, 534–537 (2005)
11. J. Choi, C. Seo, High-efficiency wireless energy transmission using magnetic resonance based on metamaterial with relative permeability equal to −1. Prog. Electromagn. Res. **106**, 33–47 (2010)
12. A.L.A.K. Ranaweera, T.P Duong, B. Lee, J. Lee, Experimental investigation of 3D metamaterial for mid-range wireless power transfer, in *Wireless Power Transfer Conference (WPTC)*, (2014, IEEE), pp. 92–95, 8–9 May 2014

# Design of a Soft Pack Battery Tab Height Detection Device Based on a Cartesian Robot

**Ming-Shuai Bi, Jia-Song Mu and Yu-Yin Wang**

**Abstract** Equipment is designed for detecting soft battery modules and is used for detecting whether the battery tab reaches its designated position. The system configuration requires visual positioning, a contact probe, a Cartesian robot connected to a host PC, a PLC controller, and C# software for language design. The system has characteristics including strong customization and simple operation.

**Keywords** Visual orientation · Cartesian robots · C#

## 1 Introduction

Environmental pollution and energy shortages prompted States and businesses to begin to vigorously develop the environmental efficiency of power generated by lithium-ion batteries. Over recent years, energy requirements have been increasing in the auto market as it expanded in 2014. New energy car production rose by almost 4 times and 2015 first-quarter output increased 2.5 times. The main influencing factor of electrical vehicle development is battery power. In recent years, the Dongfeng electric vehicle, BYD, and other mainstream companies began using soft power batteries (housings of flexible packaging materials for lithium-ion batteries). Soft power batteries with their high safety, high energy density, low cost, long life, and other advantages have become more and more popular [1]. The quality of the lug of a power lithium-ion battery has a big impact on its performance and lifecycle. Traditional manual testing is inefficient and prone to errors, problems which have seriously hindered the automated manufacturing of lithium-ion batteries [2].

Overseas, European countries and the United States, due to their natural advantages in industrial development, began the study of visual systems very early on. Carnegie Mellon University represents one of a large number of prominent (in terms of machine vision) research universities and research institutions [3].

M.-S. Bi (✉) · J.-S. Mu · Y.-Y. Wang
Tianjin Normal University, Tianjin, China
e-mail: zjdsh@outlook.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_109

Institutions abroad have been leading China, and in industries of good products detection, and products assembled in the achieved has stable of application, while abroad the manufacturers with its industrial automation of height developed, has will Visual system achieved has mature of industrial application, and in world around siege swept to, constantly advance visual system in modern automation of application, achieved modern factory of no of, and intelligent of, and in all aspects of application in the get has stable development [4].

Soft battery power structures use aluminum–plastic packaging. The positive and negative poles have a tab structure, making contact with the conductive row-series connection across different modules. If the conductive row is not in contact with the tab, the battery is defective. Battery tabs and conductive rows have gaps within the millimeter ranges of one another. Therefore, manual measurement is not only error-prone, but is inefficient, and can easily lead to associated dangers. In order to improve the efficiency of detection, and generate accurate measurements, very high efficiency tab-detection equipment is needed. Experiments show that a robot can obtain tab height data quickly and accurately, with a stable working performance.

## 2 The Design of System Structure

Figure 1 shows the tab structure. In this figure A and B represent the tab of the battery and C is the conductive bar which is exposed to the tabs. The product is substandard if the gap between A and C or B and C is more than 0.8 mm.

Figure 2 shows the three-axis robot (x, y, z axes). Axis z is equipped with cameras to gain position coordinates with the three-contact probes used for detecting the tab height [5].

The robot system sends instructions to the PLC through the host computer, and the PLC converts the three-axis drive instructions into three-axis servos according to certain instructions. The three-axis servo controls the axis according to its received commands in order to carry out its related action. The control software works through the Windows 7 operating system. You have to enter the battery parameters into the control software in order to determine the robot's path. Robots can walk precisely to a location by visually locating and adjusting. Figure 3 shows the system's framework.

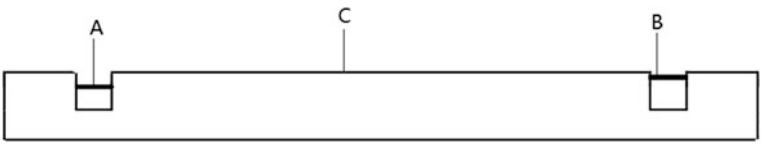



**Fig. 1** Tab structure

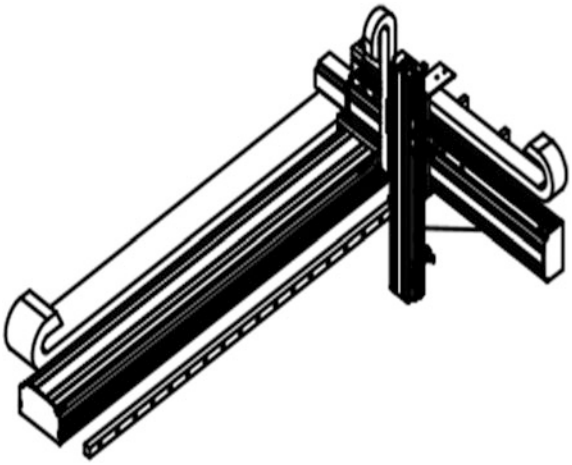


**Fig. 2** Robot's structure

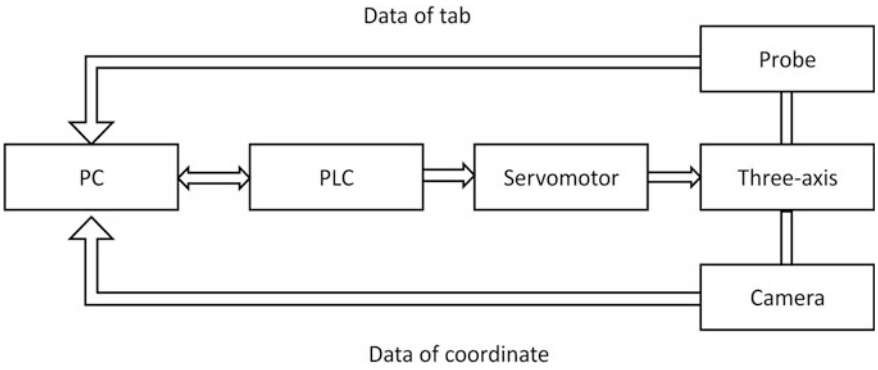



**Fig. 3** Robot system framework

## 3 The Design of the Hardware

Figure 4 shows the robot.

### *3.1 Keyence Probes*

During our research we chose to use the GT-70A Keyence probe which has an accuracy of microns. The traditional sensor probe uses a method of pulse counting. Keyence probes use double pressure which can track the absolute position of an object, not lose any data following high-speed movement, and also will not demonstrate any errors in measurement.

Figure 5 shows that the probe touches the tabs accurately and returns the data.

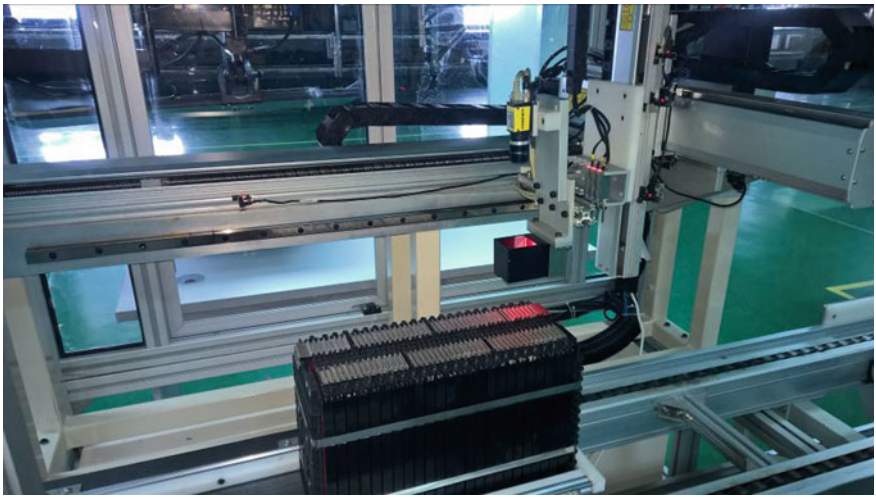


**Fig. 4** The robot

**Fig. 5** Testing

## *3.2 Cognex Cameras*

Super small sized, In-Sight, 5-million-pixel (model 8405), compact-design cameras may be upright or deployed at right angles. The system is compact, equipped with high-performance visual tools, capable of fast communication speeds and high resolutions, and is easily integrated into machines, thus it represents the ideal choice for space-constrained robots.

### 3.3 PLC

The Delta DVP-ES2 has a built-in, 12-bit 4AI/2AO. It can also can be used with a 14-bit AIO extension module with a built-in PID auto-tuning function, providing a complete analog control solution with up to 100-kHz pulse control. It can be used with a variety of control commands (such as masking, benchmarking, immediately change frequency) for each axis.

## 4 Design of Software System

### 4.1 Main Program

The control software works in the Windows 7 operating system, using C# language for development in Visual Studio 2010. The system is programed via manual and automatic procedures. Manual procedures are mainly used for system debugging and system failures. Automatic procedures are mainly used in daily production plants and recycling operations. First, the program begins with the initialization function. The axes must be configured in the initialization function to ensure that they return to the zero position before they initialize. Manual programing includes the setting of manual parameters for servo motors. The automatic program is shown in Fig. 5. When the work piece is in place, it is adjusted before the position coordinates are acquired. At that point, the three axes will be in standard position, thus providing the positioning coordinates. After reaching the first detection position, according to positioning coordinates, adjust the three-axis position, the workpiece is not in place to compensate for the situation. One must then start the probe, which sends data to the PC which is archived. As shown in Fig. 7b, after the detection of a conductive row, the tab will be adjusted to ensure that the probe can detect the tab (Fig. 6).

### 4.2 Adjustment

Due to the fact that the conductive bar is installed manually by a worker, one cannot ensure location accuracy. However, there is a circle on the conductive bar in order to distinguish between positive and negative poles of the battery. The camera can get the coordinates and therefore help the robot to adjust its position (as shown in Fig. 7).

Robots will be adjusted for an error after the conductive bar is detected and then will be allowed to start the next one.

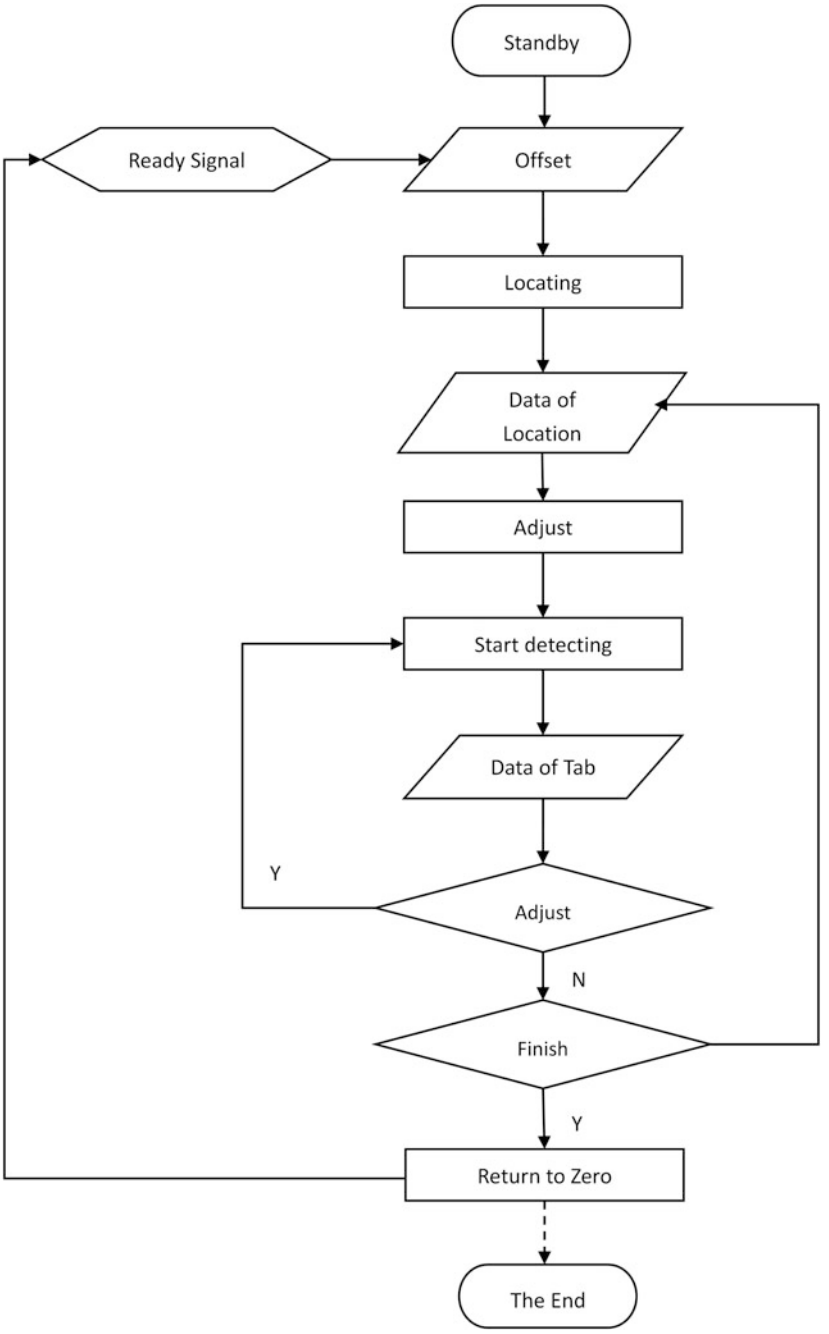



**Fig. 6** System flow chart

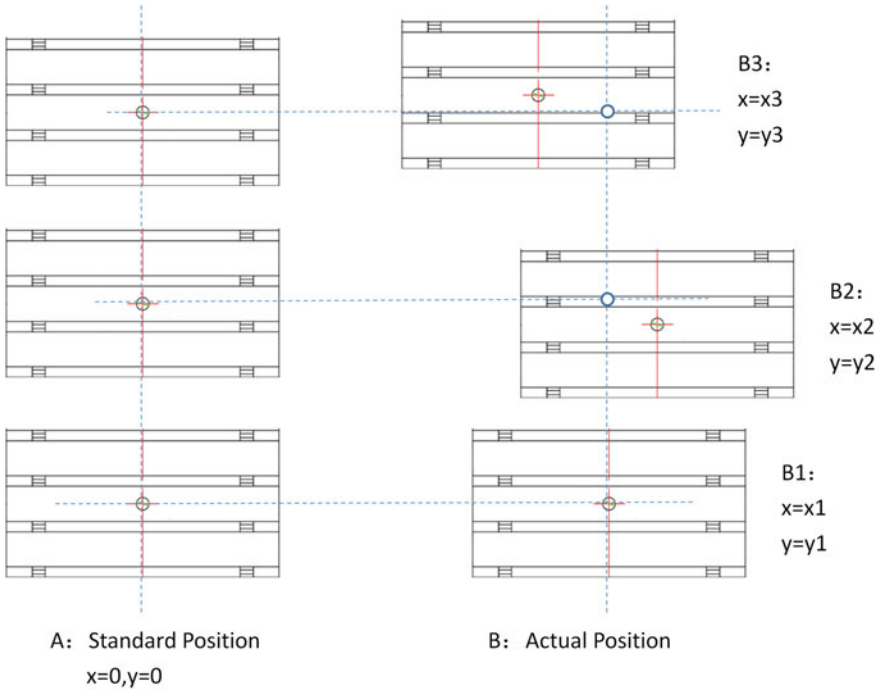



**Fig. 7** Battery module structure

The distance for adjust is thus:

$$\Delta x = x_i - x_{i-1}$$
$$\Delta y = y_i - y_{i-1}$$

## *4.3 Interface Design*

The software interface is divided into two parts, data for detected results on the left-hand side and battery type option on the right.

The software interface uses XAML language developed in Visual Studio 2010 (Fig. 8).


**Fig. 8** Interface design

## 5 Conclusion

In this chapter, we have outlined the design of a robot based on Cartesian geometry to detect battery tabs. The robot was equipped with visual orientation to ensure the accuracy of location measurements. Use of contact detection has advantages such as smaller amounts of data and low system requirements. After field testing, the robot detected a tab in 3 s with testing of the complete battery module taking only 8 min. Compared with traditional manual detection, speeds have been greatly improved. After repeated testing of the 8S9A-type model some 144 tabs were identified. The end of the 8-pole tab has no positioning hole, therefore the robot may miss this case. The rest of the tabs were accurately detected, hence the data is considered accurate. From the overall structure of the system: the hardware and software design has some reference value.

## References

1. Y. Peng, Design of a three-axis pickup robot control device based on a PLC control system. J. Hubei Univ. Technol. **31**(2), 39–41 (2016)
2. M. Luo, Q. Cao, G. Zhang, General controller software design of a three-axis desktop industrial robot. Modern Electron. Tech. **36**(23), 133–138 (2013)
3. J.Y. Pu, Y.J. Zhang, Gap detection by non-contact method. Equip. Electron. Prod. Manuf. **237**, 33–35 (2014)

4. B. Li, Y. Lin, Z. Tan, Automatic packing device of seal based on a Cartesian robot. Light Ind. Mach. **32**(4), 77–79 (2014)
5. Q. Liu, A. Yin, Realization of motion control of a mobile robot based on PLC. Electro-Mech. Eng. **1**, 60–62 (2004)

# Garage Security System Based on Gyroscope

**Hong Yimin and Xu Lei**

**Abstract** With the development of economy and technology, the types and quantity of vehicles are increasing which lead to security problems in storage. To solve this problem, a remote intelligent monitoring system for vehicle storage condition is designed. It has the advantage of real-time monitoring, and sending alarm information once detecting abnormal status of vehicles. Based on the characteristics analyzed by gyroscope sensing data, a high-level precision of security decision function is proposed. The function can further enhance the system's ability of early warning and increase the security level for garaging traffic tools.

**Keywords** Gyroscope · Intelligent monitoring · Security · Similarity · Quaternions

## 1 Introduction

Traditional vehicle security systems are run independently and lack digitized monitoring network. With strengthen of stealing skills, the thieves have a full understanding of the conventional anti-theft system and can easily destroy early warning system. Especially, when the vehicle is moved away from original position, the GPS will lose its effects due to the interaction latency, and will lead to an enormous input of manpower and material resources. In order to reduce the unnecessary cost, we need to strengthen early warning ability and realize the interconnection of alarm information for protecting the target in the first time and reducing the risk of expensive vehicles stolen and destroyed.

The system designed in this paper is an intelligent garage security system based on gyroscope sensor. It can process real-time signal of gyro sensor locally and analyze the safety factor, so that respond to different dangerous levels. The system

H. Yimin (✉) · X. Lei
School of Electronic Information Engineering, Anhui University,
Hefei 230601, China
e-mail: hongyimin22@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_110

combines Wi-Fi with GSM communication technology; it realizes a remote security network of multi-level alarm system which greatly improves anti-theft efficiency and security level of security objectives.

Meanwhile, the system can select the type of anti-theft vehicle through interactive buttons; it makes security system more convenient, diverse, and is widely used.

## 2 System Overview

The designed function is mainly used in the parking lot at commercial and residential area, and the security goal is not just limited to cars, but also includes electric cars, motorcycles, bicycles, and other small vehicles. The system is mainly composed of two parts: one is portable monitoring equipment placed on the vehicle, and another is a monitoring alarm system at guard department. The main work of portable monitoring equipment is to carry on stealing behavior analysis. With their stealing behavior, theft will cause a slight shock or relative displacement of the security target. These signals are real-time collected by gyroscope and analyzed for alarming. When the alarm system of guard department receives signal, it will respond a treatment. When the system does not receive the alarm signal, it will be always on a defensive state, so an abnormal state, the input signal, and low power state will trigger an exception. Users can terminate the process after alarm goes off. If the alarm system receives no response from guard in alarm status, the system will send SMS and voice alarm information through the GSM module to target user for processing.

Portable device is a core part of monitoring system and signals are collected by a data collection front end. Then data are sent to the security monitoring system for real-time monitoring. The whole system block diagram is shown in Fig. 1.

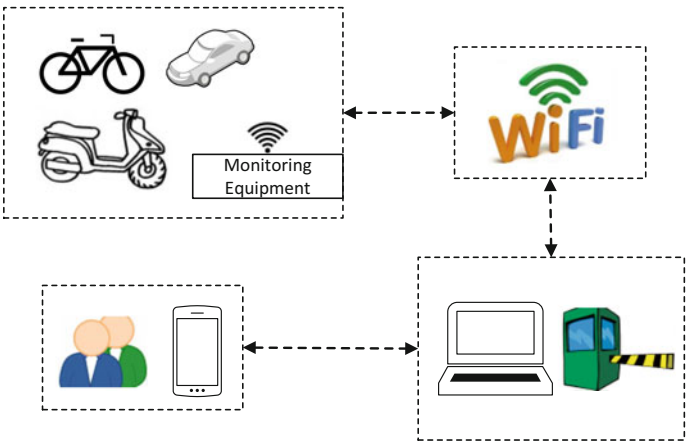



**Fig. 1** The chart of smart garage security system

## 3 Design of the System

The hardware part of the system mainly includes portable monitoring equipment, wireless close-start switch, and GSM SMS sending module; the software part contains a portable monitoring device software systems and visual monitoring system in guard room.

### *3.1 Portable Monitoring Equipment*

This section describes implementation processes from hardware and software aspects, to achieve the following functions: detection of target object's posture; wireless transmission of data; the portable installation; intelligent control; and target multiple types of security options. It is determined that the device has the following components; the block diagram is shown in Fig. 2.

Among them, the power module provides rechargeable power system for devices; Wi-Fi module periodically pushes detected data into monitoring host server. Gyro sensor module is the core components for the device. The wireless control module is used to realize the close range control through the non-contact switch device. MCU module is the intelligent control unit of the system; it can make each module function realized, and have algorithm processing of gyro sensor data, and so on. The speaker module is used to defend the alarm and the working state of the equipment. The interactive module is used for setting the function of the power switch and the mode. The temperature sensor is used for detecting the temperature of the environment of the security target.

For designing software system, the key part is based on the acquired angular velocity and acceleration analysis of the current state of the target. The main program of the system flow chart is shown in Fig. 3.

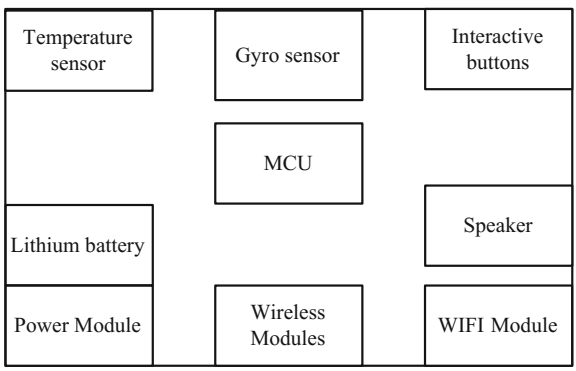



**Fig. 2** The system block diagram of portable monitoring equipment

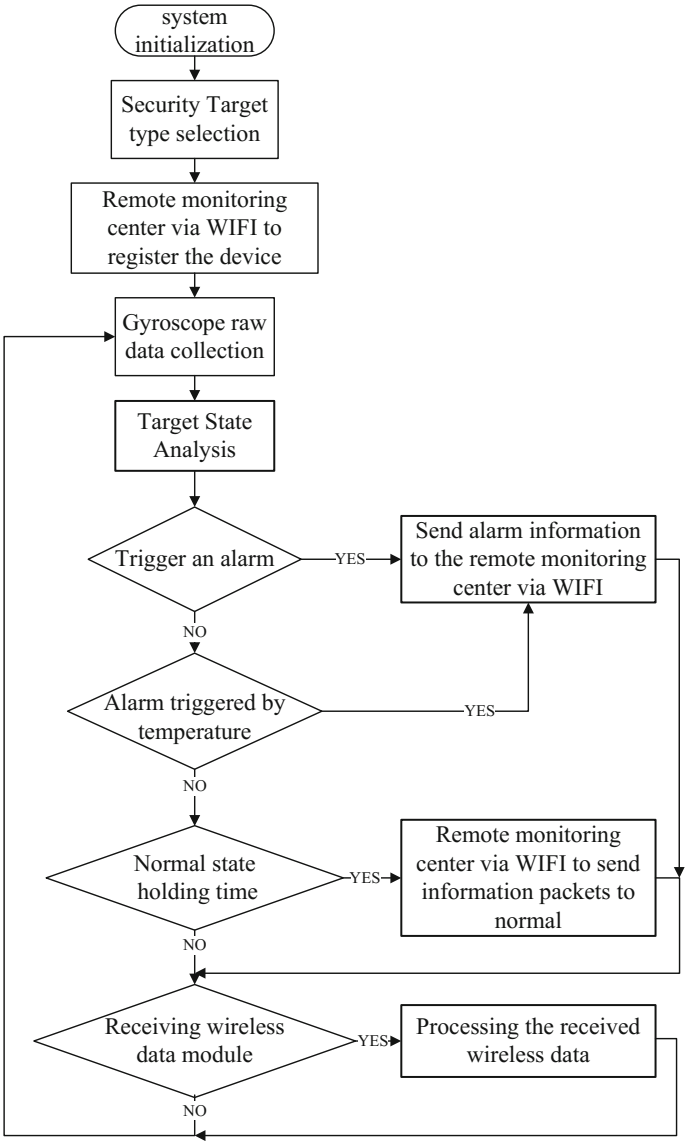



**Fig. 3** Block diagram of software system

Usually, we use the Euler angle (azimuth, elevation angle, roll angle) to determine the posture of moving objects, but the scope of the Euler angle is limited; the dynamic changes of objects cannot be characterized. The computation is also complex; values usually converse from quaternions. For getting sensitive and dynamic changes of object, quaternion is used as the reference vector. In practical

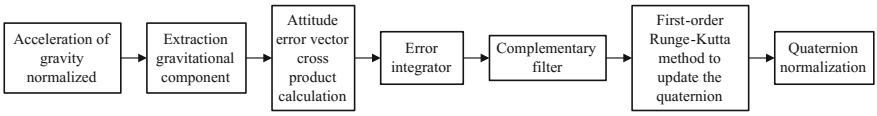



**Fig. 4** Block diagram of data processing

engineering, there are usually the following processes to get the value. The actual algorithm processing process is shown in Fig. 4.

Here we update quaternions mainly through a law-order Runge–Kutta method, the values of quaternions respectively are $q_0$, $q_1$, $q_2$, and $q_3$. Gyroscopes angular velocity of each axis is $w_x$, $w_y$, and $w_z$. The formula is shown below:

$$\begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}_{t+\Delta t} = \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}_{t} + \frac{\Delta t}{2} \begin{pmatrix} 0 & -w_x & -w_y & -w_z \\ w_x & 0 & w_z & -w_y \\ w_y & -w_z & 0 & w_x \\ w_z & w_y & -w_x & 0 \end{pmatrix} \cdot \begin{pmatrix} q_0 \\ q_1 \\ q_2 \\ q_3 \end{pmatrix}_{t}. \quad (1)$$

Then, $t$ is the variable of time and $\Delta t$ is the time of state update.

## 3.2 Alarm Mechanism

The alarm mechanism is set mainly based on the characteristics of security goals. The paper studies three types of transport which is divided by inertial mass grade. Automobiles can be used as a type hierarchy, denoted as class I; the vehicle with two wheels or three wheels can be used as a type hierarchy, denoted as class II; and bicycle can be used as a type hierarchy, denoted as class III. These vehicles' main external force is exerted by human. To achieve an early security alarm in the first time, we need to identify the response characteristics of the inertial mass of these security targets under manpower.

It is obvious that large inertial mass target has relatively small response to external force, and vice versa, so small inertial mass target is vulnerable to noise force of surrounding environment, such as environmental vibration, etc. In this paper, we select appropriate parameters and change threshold value to assess the level of alarm based on the objective of different inertial quality researches.

In this paper, we use quaternions $\vec{Q} = (q_0, q_1, q_2, q_3)$ and acceleration vectors of three-axle direction $\vec{A} = (a_x, a_y, a_z)$ as the reference data. Based on the analysis of collected data and actual situations, target state warning means to have a metric with the changes between current state and original state. Here we use similarity to evaluate the current security level which shows that the smaller the similarity is, the higher the warning state is. Similarity measurement method generally has the

distance measure method and the general similarity function method. Here we use the distance measure method. The commonly used distance measure methods are Minkowski distance (Ming's distance), Euclidean distance (Euclidean distance), Manhattan distance, Mahalanobis distance (the distance of the distance), and so on. In these methods, the Manhattan distance method has low computation, and the error of each element in the vector can be treated equally, so the environmental impacts can be removed to a certain extent.

The mean vector of the four elements for the initial security state is

$$\overrightarrow{Q(t_0)} = (q_0(t_0), q_1(t_0), q_2(t_0), q_3(t_0)). \tag{2}$$

The four element vectors for the current state are

$$\overrightarrow{Q(t)} = (q_0(t), q_1(t), q_2(t), q_3(t)). \tag{3}$$

Manhattan distance formula can get the similarity of quaternions at this time:

$$D_Q = \sum_{i=0}^{3} |(q_i(t) - q_i(t_0))|, \tag{4}$$

where $D_Q$ is the similarity value of quaternions at $t$ time.

In the same way, the mean vector of three-axis acceleration of the initial safety state is

$$\overrightarrow{A(t_0)} = (a_x(t_0), a_y(t_0), a_z(t_0)). \tag{5}$$

The current three-axis acceleration vector state is

$$\overrightarrow{A(t)} = (a_x(t), a_y(t), a_z(t)). \tag{6}$$

Manhattan distance formula can obtain the similarity of three-axis acceleration vector

$$D_A = |(\mathrm{a}_x(t) - a_x(t_0))| + \left|(\mathrm{a}_y(t) - a_x(t_0))\right| + |(\mathrm{a}_z(t) - a_z(t_0))|, \tag{7}$$

where $D_A$ is the similarity value of three-axis acceleration vector at $t$ time.

Common vectors $D_Q, D_A$ will measure the safety factor of the entire system, the final safety factor ($S$) is

$$S = \omega_1 * D_A + \omega_2 * D_Q. \tag{8}$$

Among them $\omega_1 + \omega_2 = 1$, where $\omega_1$ and $\omega_2$ are two reference weighting coefficient parameters of $D_A$ and $D_Q$.

**Table 1** Test environment simulation parameters

| Target type | $\omega_1$ | $\omega_2$ | S (Triggering alarm coefficient) | Triggering times |
|---|---|---|---|---|
| Class I | 0.8 | 0.2 | >0.1 | 3 |
| Class II | 0.4 | 0.6 | >0.05 | 5 |
| Class III | 0.2 | 0.8 | >0.04 | 6 |

For integrating reference parameter of an overall assessment on target safety factor, the suitable coefficient values are obtained from the different types of monitoring targets combined with practical experience in testing. We mainly depend on the acceleration and rotation quaternion sensitivity of inertial mass target to set the coefficient values, and the greater the inertial mass acceleration sensitive, while the rotation quaternion vector parameter insensitive, and are reached from a numerical data analysis and evaluation practice results, the following parameter data through the practice of data, as shown in Table 1.

In Table 1, the number of trigger times displayed in the abnormal state of the target is greater than the number of trigger threshold times. The value is the state value of alarm trigger.

When the state of the security objectives is secure, the data acquisition is shown in Fig. 5.

Under simulated theft state, the acceleration data and the safety factor data are collected and shown in Figs. 6 and 7.

Figure 5 shows, when the system in monitoring process, the mean safety factor is around 0.002. In practice, to prevent the security of the state vector drift, the initial security vector will be updated after the effective time, where effective safety time is $T_s = 300$. If the continued safe state time of security target is greater than $T_s$, the initial state vector data $(\vec{Q}(t_0), \vec{A}(t_0))$ will be updated.

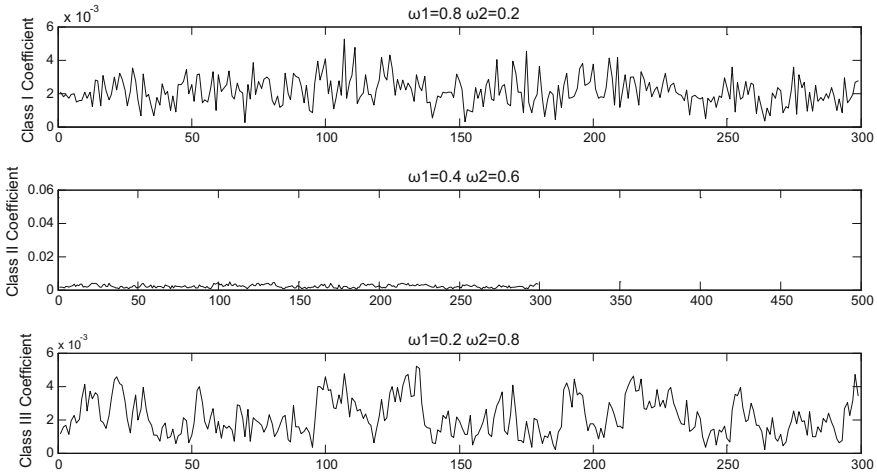



**Fig. 5** Target data at rest security status graph

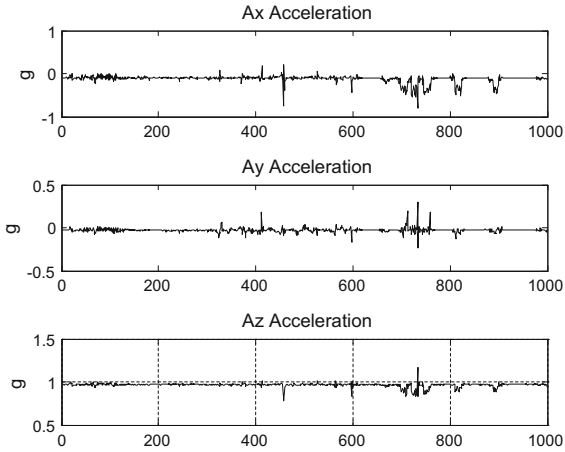



**Fig. 6** Data curve of triaxle accelerometer

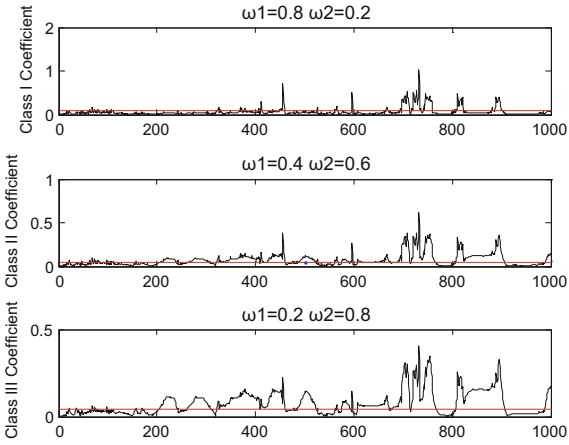



**Fig. 7** Data curve of security factor under different weighting factors

In Fig. 6, the vertical coordinate is the acceleration value, the unit is g, and the horizontal coordinate is the data points collected. It can be seen that the three-axis acceleration fluctuations are very small, but when the detection process of the larger vibration caused by a little obvious fluctuations, the accuracy of the gyroscope is very high.

In Fig. 7, the vertical coordinate represents the safety factor, and the horizontal coordinate represents the data points collected. As can be seen in the detection of the same circumstances, the change of different weight coefficients and numerical safety coefficients is different. It is obvious that the numerical curve of class I distribution at the top, and the numerical curve of class III points at the bottom, which is consistent with our actual needs. For monitoring data of class I, we need

more sensitive data to capture abnormal states. For III class goals, we need non-sensitive monitoring data to remove the interference noise of external environment.

For class III vehicle, since the weight is light and is easy to be touched, the safety factor could exceed the threshold and raise unnecessary alarm. We should set the trigger time as a constant value instead of the instantaneous value (here we first set to 5 s). If you exceed this duration, then you will enter an alarm. If an alarm occurs six times in one minute, then enter the secondary alert. In this case, a simple or a few times touch will not cause an alarm, so as to reduce unnecessary manpower.

### *3.3 Deployment and Withdrawal of Security System Solutions*

When users enter the parking garage, they can collect portable security devices from guard at the entrance. After the completion of the vehicle parking, the security equipment will be placed in a hidden location. When leaving the security vehicle, through the wireless remote control to open the security state, you need withdraw settings; only turn off the device by remote wireless control. When the first-level warning of security equipment is triggered, it will send a warning by the speaker. When the second-level warning is triggered, the guard will receive the alarm for inspection based on the alarm area and the target information. If the second-level alarm is triggered, the guard monitoring personnel in the specified time is not processed; the monitor server will be the target of the head of the household by SMS and telephone voice alarm.

## 4 Software Monitoring System

To reduce the amount of processing data from multi-targets, the remote monitoring server only receives security status of each monitoring node corresponding to an exception handling; the following map is main function interface of the monitor system.

The main parameters of configure system are shown in Fig. 8, mainly for the IP address settings of network server, listen port settings, interface settings of GSM module, and text template design. In the message template, the system uses @ as the beginning of a string variable, and automatically replace the user information from triggered alarming information before sending. Figure 9 shows the status information of all online nodes. Figure 10 shows a user's information list of current abnormal alarm. Monitoring objectives must be handled within the specified time and exceptions must be dealt by monitoring personnel; this time is usually set to minutes, otherwise the abnormal information will be sent to the user's mobile phone for an alarm processing.

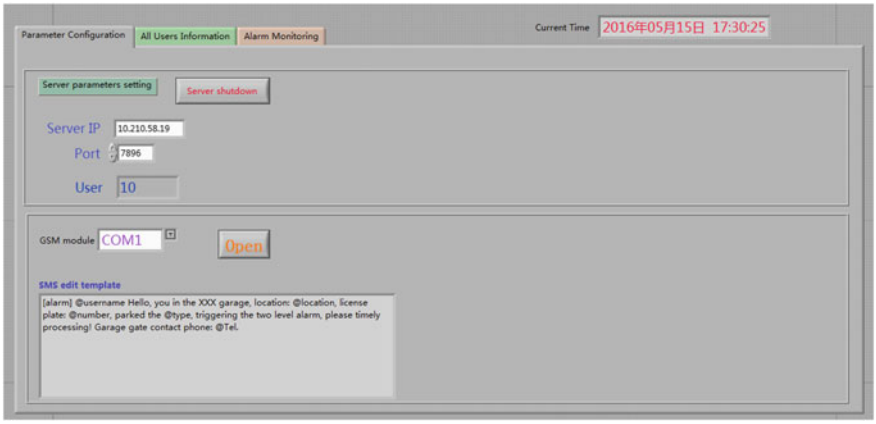
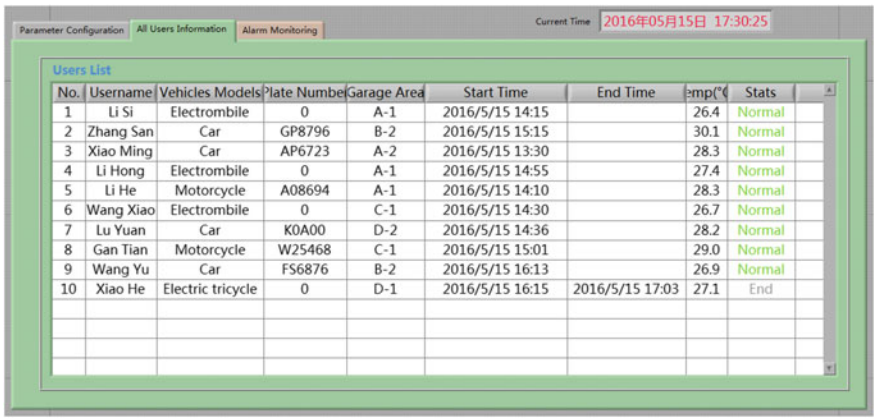
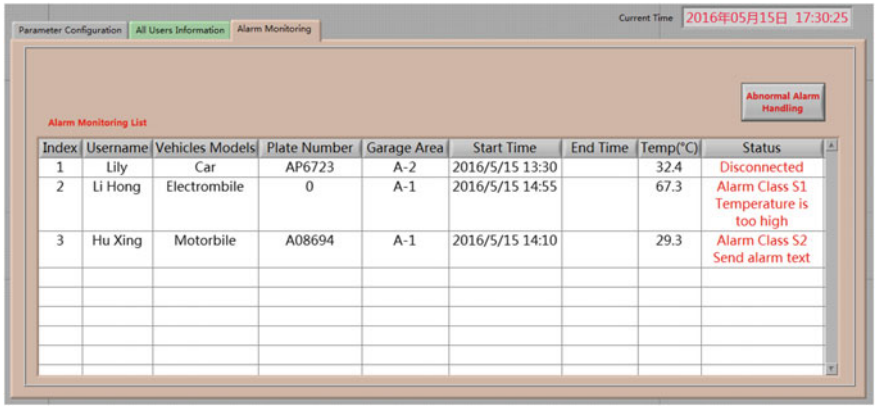



**Fig. 8** The configuration interface of software system




| No. | Username | Vehicles Models | Plate Number | Garage Area | Start Time | End Time | Temp(°C) | Stats |
|---|---|---|---|---|---|---|---|---|
| 1 | Li Si | Electrombile | 0 | A-1 | 2016/5/15 14:15 | | 26.4 | Normal |
| 2 | Zhang San | Car | GP8796 | B-2 | 2016/5/15 15:15 | | 30.1 | Normal |
| 3 | Xiao Ming | Car | AP6723 | A-2 | 2016/5/15 13:30 | | 28.3 | Normal |
| 4 | Li Hong | Electrombile | 0 | A-1 | 2016/5/15 14:55 | | 27.4 | Normal |
| 5 | Li He | Motorcycle | A08694 | A-1 | 2016/5/15 14:10 | | 28.3 | Normal |
| 6 | Wang Xiao | Electrombile | 0 | C-1 | 2016/5/15 14:30 | | 26.7 | Normal |
| 7 | Lu Yuan | Car | K0A00 | D-2 | 2016/5/15 14:36 | | 28.2 | Normal |
| 8 | Gan Tian | Motorcycle | W25468 | C-1 | 2016/5/15 15:01 | | 29.0 | Normal |
| 9 | Wang Yu | Car | FS6876 | B-2 | 2016/5/15 16:13 | | 26.9 | Normal |
| 10 | Xiao He | Electric tricycle | 0 | D-1 | 2016/5/15 16:15 | 2016/5/15 17:03 | 27.1 | End |

**Fig. 9** Monitoring interface of all current normal user information




| Index | Username | Vehicles Models | Plate Number | Garage Area | Start Time | End Time | Temp(°C) | Status |
|---|---|---|---|---|---|---|---|---|
| 1 | Lily | Car | AP6723 | A-2 | 2016/5/15 13:30 | | 32.4 | Disconnected |
| 2 | Li Hong | Electrombile | 0 | A-1 | 2016/5/15 14:55 | | 67.3 | Alarm Class S1 Temperature is too high |
| 3 | Hu Xing | Motorbile | A08694 | A-1 | 2016/5/15 14:10 | | 29.3 | Alarm Class S2 Send alarm text |

**Fig. 10** User information interface with abnormal state

## 5 Summary

In this paper, the design of the security system of the vehicle detection system based on gyro sensor is realized; it could be used in the field of security areas. Portable security devices are convenient and extensible; they could be applicable to a wide range of transport. Through the full research of the characteristics of sensor data in the alarm mechanism, the high accuracy of security level decision function is designed. It has the advantages of easy implementation and small computation. Each sensor node forms a network, and the server can be used as a terminal monitoring node to receive a plurality of data and detect abnormal situation and send out alarm and SMS alarm information, saving manpower and material resources. Multi-level alarm system can maximize the avoidance of abnormal circumstances that no one accepted. The utility model has the advantages of portability, convenient installation, flexible use, high safety factors, etc., and has wide application prospect.

## References

1. X. Yinghui, O. Yangjun, Research and design intelligent vehicle antitheft alarm system based on more integration of information technology. Comput. Digit. Eng. **03**, 114–116 (2009)
2. Z. Ting, Design of smart vehicle antitheft system based on vehicle networking. Appl. Electron. Tech. **7**, 21–22 (2014)
3. P. Singh, T. Sethi, B.B. Biswal et al., A smart anti-theft system for vehicle security (2015)
4. K. Alli, O. Adedokun, Development of a vehicle anti-theft system via GSM network (2015)
5. N. Yamamoto, H. Arai, Ltd. H M C. Antitheft system for vehicle (2012)
6. X. Juanjuan, Design of vehicle remote anti-theft system based on STM32F103. Microcontrollers Embed. Syst. **12**(11), 66–69 (2012)
7. Z. Ronghui, J. Hongguang, C. Tao, Attitude solution for strapdown intertial navigation system based on quaternion algorithm. Opt. Precis. Eng. **16**(10), 1963–1970 (2008)
8. P.G. Laborde, Security device for garage doors: US, US 20120019011 A1 (2012)
9. A.U. Ahmed, T.M. Masum, M.M. Rahman, Design of an automated secure garage system using license plate recognition technique. Int. J. Intell. Syst. Appl. **02**(2), 22–28 (2014)
10. E.D. Williams, Garage door opener security system (2016)
11. S.J. Kurth, J.J. Kurth, Security automatic garage door closer: US, US8561348 (2013)
12. T. Shimomura, Vehicle security apparatus and system: US, US20070222292 (2007)
13. K.E. Flick, Vehicle security system for a vehicle having a data communications bus and related methods: WO, US 6011460 A (2000)
14. H. Brinkmeyer, M. Daiss, G. Schwegler et al., Vehicle security device with electronic use authorization coding: US, US 5708712 A (1998)

# Design and Implementation of "Medicine Chest Butler" Mobile App

**Hengxin Liu, Jin Chen, Lujia Wang, Maolin Ji, Zeng Liu and Hankun Zhang**

**Abstract** Most family does not implement intelligent management of the most commonly used medicine. Little research teams designed the family medicine chest intelligent management. This article is based on this background, committed to the development of smart phone App managing medicine and cloud management platform software system. The mobile terminal will also develop Android and Apple's dual-platform software. The medicine information is obtained by scanning the bar code. The backstage medicine information database maintenance and pharmaceutical science information releasing based on cloud-based platform architecture with more prominent security and scalability. The results show that the software can automatically remind medication and warn expired medicines. It can also well extend family health science knowledge.

**Keywords** Mobile app · Family medicine management · Health science

## 1 Introduction

With the rapid popularization of mobile Internet, the pharmaceutical industry has the potential and practical needs of rapid development. The mobile Internet and cloud hybrid software system developed based on cloud-based platform and a mobile phone belonging to technology-intensive high-input, high-yield industry. However, due to the relatively strong professional medical industry and the difficult software development, there are few companies involved in this area. The overall goal of our software is to achieve intelligent management for the family medicine and science system with independent intellectual property rights, mainly including medicine information entry, medication reminders, alarm expired medicine, and medical knowledge popularization. The completion of the project can form

H. Liu · J. Chen (✉) · L. Wang · M. Ji · Z. Liu · H. Zhang
Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
e-mail: cjwoods@163.com



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_111

industrialization demonstration and marketing capabilities. In view of the field of medicine market prospects, there are expected to benefit for one million people in the long-term with good social benefits and the formation of independent intellectual property rights systematic development of pharmaceutical science system technology.

## 2 Components and Functions and the Software

The project developing family medicine management software is divided into two parts: the cloud and mobile terminal. The mobile terminal provide human interface performing the following functions: family medicine management, medication reminders, warning expired medicine, medication and health science knowledge, providing perfect medicine use and management system for families. The functions of the cloud including backup users' data, analyze each family's medicine usage, provide reasonable suggestions and push health information regularly. Popularize the medicine and health knowledge at the same time of facilitating medication convenience for each family. Overall contents are shown in Fig. 1.

Standing medicine input module: This software provides function of standing medicine entry. The medicine information is from the authoritative medical institutions and manufacturers. The medicine information is obtained by scanning the bar code and interacting to the database. After that the information will be saved to

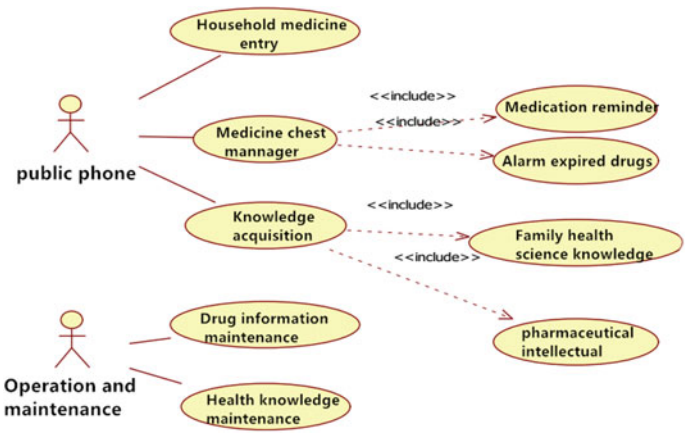



**Fig. 1** Overall research

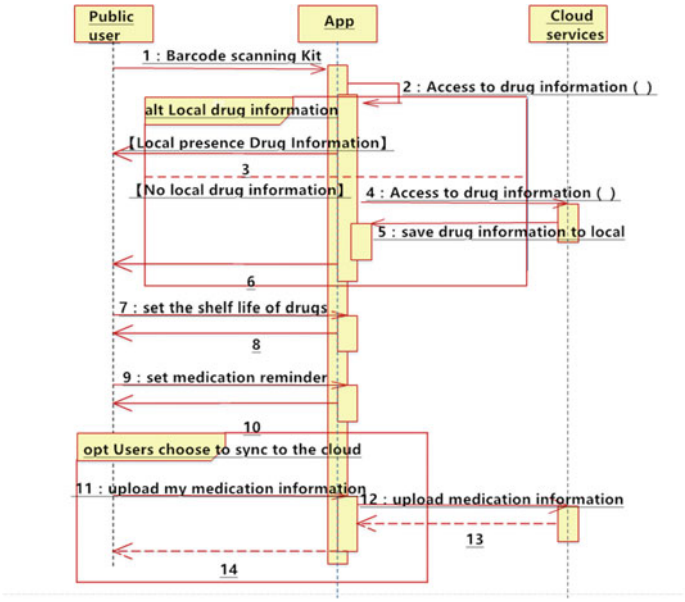



**Fig. 2** Standing medicine input module

the local. The user can manually set the shelf life of medicine、medication reminders and other information. Users can also choose to upload medicine information and sync to the cloud (Fig. 2).

Pharmaceutical science knowledge module: This software providing function of popularizing pharmaceutical science knowledge. The medicine-related knowledge are from authoritative medical institutions and manufacturers. Server-side will update real-time data, ensure the authority and real-time data. The server will establish pharmaceutical community. Users can send the experiences on the processing of medicine usage. The community will invite experts to identity experiences recognized by a majority of users on a regular basis. The recognized information will be added into the database for the user reference forming the atmosphere of information sharing, interactive, promotion of the closed-loop effect.

Family health science knowledge: The software provides home health science knowledge. The functional block diagram is shown in Fig. 4. Help each family member to cultivate eating habits and patterns of health life. By recording the family's diet, it will indicate the shortage of nutrition at this stage, give dietary advice to help family members customized healthy eating plan (Fig. 3).

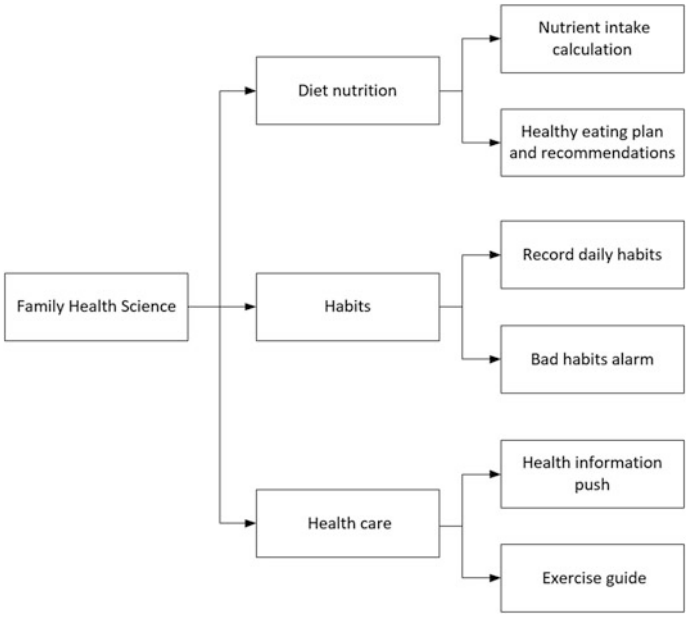



**Fig. 3** Family health science knowledge function block diagram

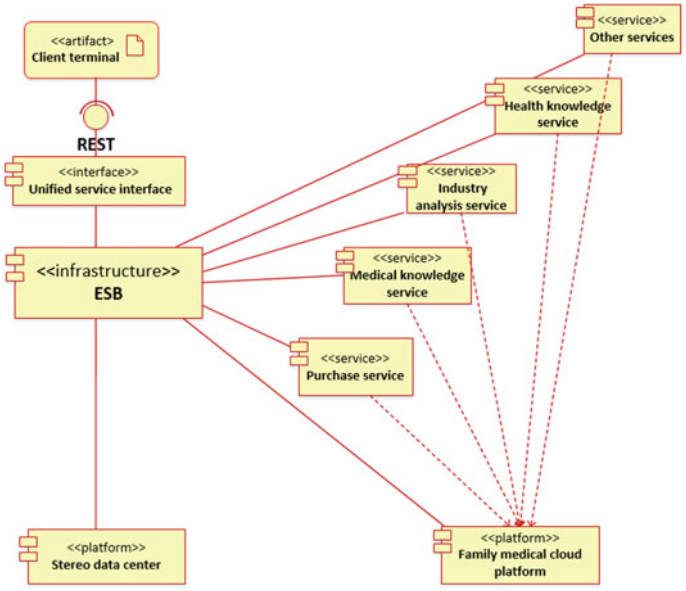



**Fig. 4** SOA-based cloud services platform

## 3 The Key Technological Innovation

(1) The integration of family medicine service based on SOA: SOA (Service-Oriented Architecture) is an important strategy to solve the heterogeneous systems. As the industry grows, more and more services will be integrated into the system [1–3]. In the field of family medicine, how to integrate new services have emerged rapidly、efficiently, seamlessly and smoothly is a key technical problem to be solved in the research project. The technology can provide a viable product evolution mode, thereby ensuring that industry has always been open, advanced position.

(2) Big Data technology in the home medical industry: the following three functions in this project are based on big data analysis: ① Analyzing the region's frequently occurring disease information based on about the area; ② According to the family medicine usage response public medicine using habits; ③ Pushing relevant medical knowledge by the family consumption on the medicine [4–7]. The project can generate a lot of valuable data. To carry the big data applications of this project, we propose the concept of "family medicine perspective Data Center", which is based on the idea to use different data processing techniques to achieve targeted management of the data at all stages of the life cycle and supports processing techniques which is compliance with the respective data features. Its architecture is shown in Fig. 5.

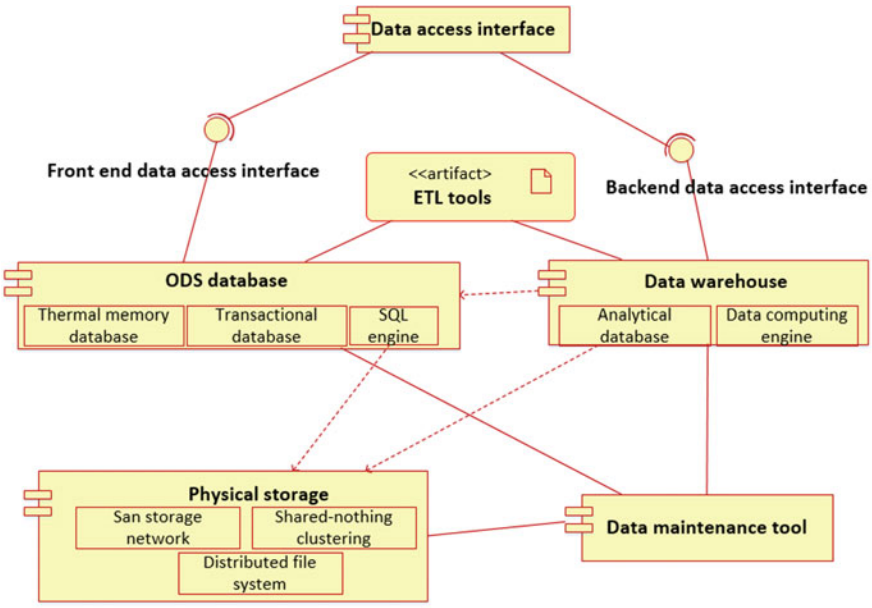



**Fig. 5** "family medicine perspective data center" architecture

(3) REST service interface and HTTPS secure communication: the whole system of the project include the client and server in two parts (C/S), the former showed the phone App program, the latter manifested as cloud services system. Communication between the two through REST interface. The mobile phone network (GPRS/3G/WLAN) is as the link layer protocol and the TCP/IP is as the network transport protocol. HTTP protocol is as an application layer protocol. Application Programming Interface employ REST (Representational State Transfer, characterizing the state transitions) [8–10]. Compared to other implementations, REST Web services model is clearly easier in contrast to the complexity of SOAP and XML-RPC.

## 4 Implementation of the Functions

(1) Choose your medicine

Users can click the different categories of medicine to select needed medicine on the home page. You can also use the search box to do some accurate research for the desired medicine (Figs. 6 and 7).

(2) Standing medicine input module

The medicine information obtained by scanning the bar code the information will be saved to the local.

(3) Medication reminders

Users can manually set the timing. It will automatically remind users to take medicine on the time of medication (Figs. 8 and 9).

(4) Pharmaceutical and family health science knowledge

Users can learn pharmaceutical and health knowledge provided by the software.


**Fig. 6** Home page


**Fig. 7** Standing medicine input module


**Fig. 8** Medication reminders


**Fig. 9** Science knowledge

## 5 Conclusion

With rising living standards, the importance of health care is also rising. The software uses the rapid development of Internet technology meeting the demand for medicine administration to develop the smart home medicine management system. The software aims to achieve family medicine management and medication regularly reminding, to establish a home-based pharmaceutical industry, scientific knowledge publishing system mobile Internet based on the cloud and big data framework; It will initially establish the big data analytics architecture of family medicine industry; It will achieve the integration of family medicine science and professional management by combining mobile phone App and cloud technology.

The software is designed to facilitate the management convenience of family medicine. With the rapid development of mobile Internet, the field of family medicine will have a broader market. This software fit the needs of the public and will have a good social and scientific applications.

**Acknowledgements** Thanks for the support of Google's 2015 college students innovation and entrepreneurship training project in China.

## References

1. C. Song, E. Cho, An integrated design method for SOA-based business modeling and software modeling. Int. J. Softw. Eng. Knowl. Eng. **26**(02), 347–377 (2016)
2. I.R. Chen, J. Guo, F. Bao, Trust management for SOA-based IoT and its application to service composition. IEEE Trans. Serv. Comput. 1–1 (2014)
3. R. Welke, R. Hirschheim, A. Schwarz, Service-oriented architecture maturity. Computer **44** (2), 61–67 (2011)
4. D. Gil, I.Y. Song, Modeling and management of big data: challenges and opportunities. Future Gener. Comput. Syst. **63**, 96–99 (2016)
5. W. Lin, W. Dou, Z. Zhou et al., A cloud-based framework for home-diagnosis service over big medical data. J. Syst. Softw. **102**(C), 192–206 (2014)
6. T. Shah, F. Rabhi, P. Ray, Investigating an ontology-based approach for Big Data analysis of inter-dependent medical and oral health conditions. Cluster Comput. **18**(1), 351–367 (2014)
7. Q. Yao, Y. Tian, P.F. Li et al., Design and development of a medical big data processing system based on Hadoop. J. Med. Syst. **39**(3), 1–11 (2015)
8. S.P. Ong, S. Cholia, A. Jain et al., The materials application programming interface (API): a simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. Comput. Mater. Sci. **97**(5), 209–215 (2015)
9. M. Athanasopoulos, K. Kontogiannis, Extracting REST resource models from procedure-oriented service interfaces. J. Syst. Softw. **100**, 149–166 (2014)
10. P. Selonen, P. Belimpasakis, Y. You et al., Mixed reality web service platform. Multimedia Syst. **18**(3), 215–230 (2012)

# The Analysis of Infrared Dim Target Detection Experiment Based on the Human Eye Time-Limited Model

**Huang Qian, Guo Qin and Feng Liang**

**Abstract** Infrared dim target detection has been one of the key technologies in the Infrared Search and Track (IRST) System field. According to the phenomenon that the dim target could extend into a fake crossed target when the IRST system filtering with the front wire grid, this paper developed a dim target detection experiment software based on the human eye time-limited model. With this software, we could do various dim target detection experiments by changing the scene background, the shape of the dim target with equal energy, the SNR and Gaussian noise variance, etc. Through the detection experiment with six kinds of target, we could verify that in a certain condition, the dim target detection probability would be better when filtering with front wire grid.

**Keywords** Dim target detection · Front wire grid · Time-limited model · Dynamic simulation

## 1 Introduction

Infrared dim target detection is always one key technique of IRST system research. Because this type of targets occupies too few pixels, and there has detector's random noise impacted, the targets' detectable distance and probability is subjected to great limit.

Combination of time filter method and space filter method is generally employed to point target detection. The space filter processes the single image to remove the background interference and extract suspicious targets, the time filter chooses the true target by revelation of consecutive image frames. Target detection and track algorithms can be divided into two types, detect before track (DBT) and track

H. Qian (✉) · G. Qin
Beijing Electromechanical Engineering Research Institute, Beijing, China
e-mail: flownsnow@163.com

F. Liang
China North Vehicle Research Institute, Beijing, China



Q. Liang et al. (eds.), *Communications, Signal Processing, and Systems*,
Lecture Notes in Electrical Engineering 423,
https://doi.org/10.1007/978-981-10-3229-5_112

before detect (TBD). DBT method is very simple and easy to implement, but only adapted to the condition of having a high signal-to-noise ratio. The typical methods include Kalman filtering (KF), multiple hypothesis tracking (MHT), probability data association (PDA), and so on. TBD methods do not care about targets' existence, but first track the possible traces in multi-frames simultaneously and use some strategies to judge them softly, which can avoid traces to left out because of low signal-to-noise ratio and improve detection probability.

Furthermore, methods which change the target's shape to improve detection probability by other optical systems are proposed, for example, point target is shaped into circle target whose pixel grayscales satisfy Gaussian distribution. Recently, we have researched a point detect model based on front wire grid filtering. Our group set up specific optical grid filter in front of the infrared system, which transforms the point target into extendable target having specific shapes, such as crosses. In this condition, system random noise holds the same so that the point target detect probability can be improved. Actually, as the point target is extended to other shapes, increasing pixels the target occupies and dispersing the energy, we have to research further on the impacts through theory and experiments. This article designs detection experiments on human eyes' sensitiveness to point targets and specific shaped targets and derives human eyes' detection feature to dim targets.

## 2 Overview of Time-Limited Model

Researchers have proposed many detection probability models for dim target search in large fields. The typical model [4, 5] is as follows:

$$P(t) = p_\infty (1 - e^{-t/\tau}) \tag{1}$$

where $P(t)$ is the target detection probability in time period $t$, $p_\infty$ is the probability and $\tau$ represents the average detection period when time period is unlimited (long time view, also be called static feature), $\tau$ is relevant to the field status, IRST system and detect task.

Consider the experimental results, this typical model is needed to modify by introducing the delay factor,

$$P(t) = p_\infty [1 - e^{-\max(t - t_0, 0)/\tau}] \tag{2}$$

where $t_0$ is the response delay time. In this article, $t_0$ is about 600 ms, including vision response time and mouse clicked response time.

The typical model supposes the event that experimenters search targets is composed of multiple independent and equiprobable glimpse events, and the probability to find the target is equal for each glimpse. In 2005, Dr. David Wilson supposed that each glimpse is independent but the detection probability varied with time, and the average time is,

$$\tau_{\mathrm{L}} = \tau - \frac{L - t_0 e^{(L-t_0)/\tau}}{e^{(L-t_0)/\tau} - 1} \tag{3}$$

where L is the time to be set in time-limited model. By the experimental results, L is 5 s in this article which means detection probability is becoming more and more stabilized after 5 s.

## 3 Experimental Parameters Analysis and Decision

### *3.1 Decision of Target Gray Value and SNR*

The experiment's objective is to analyze how point targets having the same energy impact human eyes' detection ability because of shape variance in the detectors. Therefore, this experiment supposes two conditions, one is that the background grayscale value and noise variance are not changed when the target shape and energy distribution changes, the other is that targets with different shapes have equal energy.

### *3.2 Analysis of Targets with Different Shapes*

To make sure targets have equal energy, we choose single pixel target, circle target satisfying Gaussian distribution which simulates optical aberration and out of focus, rectangle target satisfying uniform distribution which is met very little in practice but valuable as comparison, continuous cross target satisfying uniform distribution, discontinuous cross target satisfying uniform distribution, discontinuous cross target with diffraction effects and so on, as illustrated in Fig. 1. The cross target is from new front wire grid.

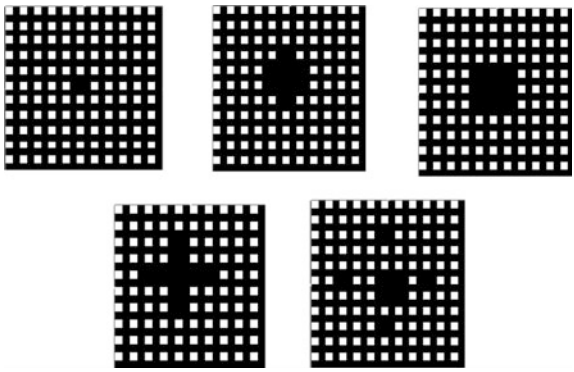


**Fig. 1** Targets with different shapes (point, circle, rectangle, continuous cross, discontinuous cross)

**Fig. 2** The energy distribution rate of six aims

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
Point

$$\begin{bmatrix} 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \\ 1/9 & 1/9 & 1/9 \end{bmatrix}$$
rectangle

$$\begin{bmatrix} 0 & 0.03 & 0 \\ 0.03 & 0.41 & 0.03 \\ 0.03 & 0.03 & 0.03 \\ 0 & 0.03 & 0 \end{bmatrix}$$
circle

$$\begin{bmatrix} 0 & 0 & 0 & 0.1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 0.3 & 0.1 & 0 & 0.1 \\ 0 & 0 & 0 & 0.1 & 0.1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.1 & 0 & 0 & 0 \end{bmatrix}$$
Discontinuous cross 1

$$\begin{bmatrix} 0 & 0 & 0 & 1/8 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1/8 & 0 & 0 & 1/8 & 1/8 & 0 & 1/8 \\ 0 & 0 & 0 & 1/8 & 1/8 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1/8 & 0 & 0 & 0 \end{bmatrix}$$
Discontinuous cross 2

$$\begin{bmatrix} 0 & 0 & 1/9 & 0 & 0 \\ 0 & 0 & 1/9 & 0 & 0 \\ 1/9 & 1/9 & 1/9 & 1/9 & 1/9 \\ 0 & 0 & 1/9 & 0 & 0 \\ 0 & 0 & 1/9 & 0 & 0 \end{bmatrix}$$
Continuous cross

Actually, by the target matrices, the energy can be reallocated when only all the nonnegative factors' sum is one. In this experiment, the six types of targets adopt factors $[a_{i,j}]$ are shown in Fig. 2.

## 3.3 *Decision of Background Brightness and Noise Variance*

As shown in Fig. 3, a nonuniform background image is simulated and given. Suppose the background grayscale is $\bar{B}_g$, $S_0$ is the point target's grayscale value, the

**Fig. 3** Simulation nonuniform background image

noise variance is σ and signal-to-noise ratio is SNR, we conducted analysis toward different targets' detection probability. The energy that each pixel possesses is decided by following method,

$$SNR = \frac{S_0 - \bar{B}_g}{\sigma} \tag{4}$$

Therefore,

$$S_0 = SNR \times \sigma + \bar{B}_g \tag{5}$$

When the point target extends to other shaped targets, the corresponding $[a_{i,j}]$ grayscale is,

$$[S_{i,j}] = SNR \cdot \sigma \cdot [a_{i,j}] + \bar{B}_g \tag{6}$$

Usually, human eye's sensitiveness is determined by eyes' contrast and angle of resolution. In this experiment, the computer background is black, screen illuminance is about 20 lx, and the eye's angle of resolution is 0.85′ [7]. The distance between human eye and screen is 400 mm, and for a computer having 1280 × 800 resolutions, the pixel scale is about 0.25 mm × 0.25 mm, so the experimenter's resolution angle is

$$\theta \approx \tan\theta = 0.25/400 = 2.14^{'} \tag{7}$$

which means human eye's resolution is far beyond pixel angle.

Considered that human eye's contrast is about 0.05, suppose SNR is 9, and the experimenter can distinguish a rectangle having 9 pixels, so

$$\frac{SNR \times \sigma/9 + \bar{B}_g - \bar{B}_g}{\bar{B}_g} > 0.05 \tag{8}$$

and we can obtain $\bar{B}_g < 20\sigma$.

## 4 Experiment Procedures and Results

### *4.1 Experiment Design*

Figure 4 shows UI of the software we developed with C# in VS2008, and the window left is infrared image with noise, the right watch window shows corresponding information, including target existence in experiments, target positions,


**Fig. 4** Experimental interface

mouse clicked positions, total experiment counts, target selected counts, periods between target appearing and mouse clicking, and so on. In the right-down corner of the main window, there are two checkboxes, labeled "turn down" and "display watch window". If "turn down" checkbox is checked, the background will become black.

Experiment procedures:

a. The trained experimenters sit in front of the computer screen, and set the options;
b. Click mouse and some type of target will be displayed in some position of the screen by the set average grayscale, noise, and signal-to-noise ratio within a set time period;
c. When experimenters have detected the target and click mouse immediately, the display screen changed to uniform background;
d. Experimenters move the mouse to the position which they recalled and click the sure button;
e. The software records the relevant data, and continues the next test.

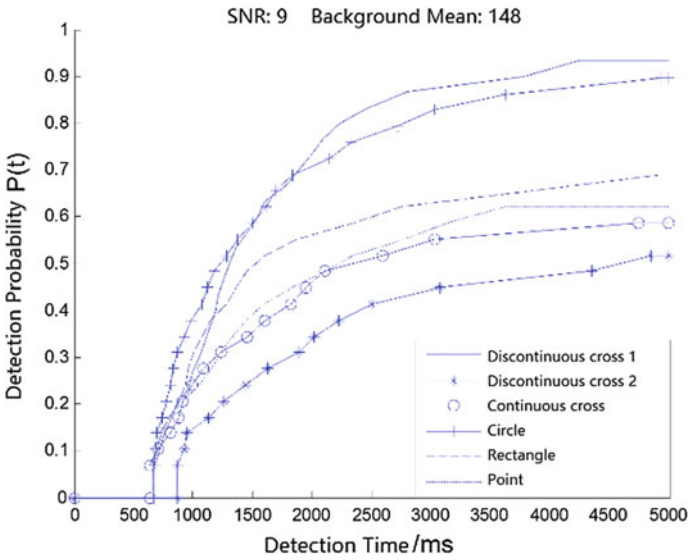



**Fig. 5** Comparison chart of six kinds of target detection results

## *4.2 Experiment Results Analysis*

We chose 10 experimenters, each of whom conducted 24 experiments.

(1) Detection probability analysis of six types of targets, where SNR = 9, σ = 22, $\bar{B}_g = 148$ for single point target.

We chose the results when signal-to-noise ratio is 9 and image average value is 148 after noise is added, as shown in Fig. 5. We can conclude that the continuous cross target from front wire grid has the highest detection probability, and the discontinuous target with uniform distribution is the most difficult to detect, because of big area brings serious energy dispersion.

(2) Proficiency impact experiments.

Choose one same experimenter, complete 20 experiments with the same target and the same background every 10 days, and analyze the detection probability curve.

We considered that experimenters may alter their search mode with more proficiency in tests so that the actual results may be influenced, we designed this experiment, as shown in Fig. 6. It is concluded that detection probability is increasing with time elapsing, which means that the experimenters are still not used to this test in the first time and the results are more reliable and stable when experimenters have fully known the experiment principal and operation method.

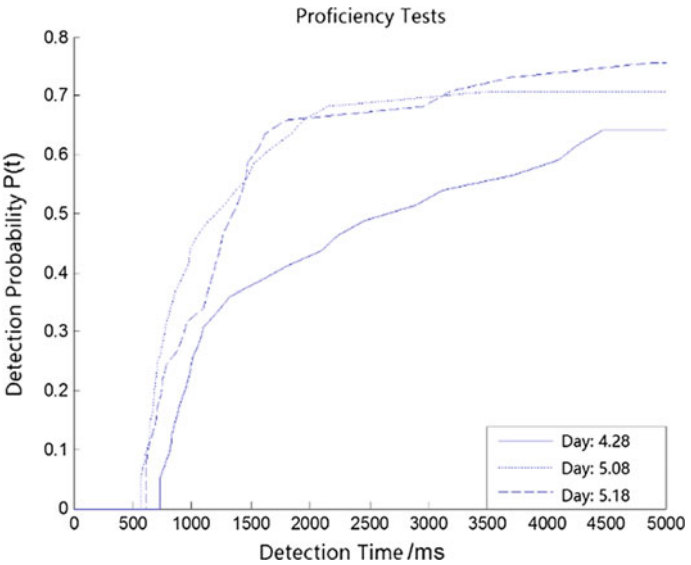



**Fig. 6** Probability curve of the same experimenters do the same experiment detection in every 10 days

## 5 Conclusion

This article designs simulation experiment for verifying the privilege of front wire grid to detect targets, composes simulation software based on human eye time-limited model, and obtains detection time and detection probability feature. The results proved that front wire grid system behaves more benefits in certain conditions.

## References

1. H. Yongsheng, Research of infrared dim small target detection method in complex background. Nanjing: Doctoral Dissertation of Nanjing University of Science and Technology, 2008, 7–13
2. L. Yang, Research on infrared small target detection and tracking algorithm under complex background. Shanghai: Doctoral Dissertation of Shanghai Jiao Tong University, 2006, 1–12
3. Liu Bin, Analysis of diffraction effect of front grid structure in thermal imaging system. J. Phys. **57**(9), 5578–5583 (2008)
4. M. Dudzik, Electro-optical systems design, analysis, and testing. Electro Opt. Syst. Des. (1993)
5. D. Wilson, N. Devitt, T. Maurer, Search times and probability of detection in time-limited search, in *Proceedings. of SPIE* (2005), 5784, 224
6. W. Minghao, Measurement and analysis of image signal to noise ratio of bundled telescope. Opt. Precis. Eng. **17**(1), 92–97 (2009)
7. T. Bai, *Principle and technology of photoelectric imaging* (Beijing Institute of Technology press, Beijing, 2006), pp. 33–35